

Conférence internationale de Géomatique et Analyse Spatiale

*Outils,
méthodes et modèles
en géomatique
pour la production
de connaissances
sur les territoires
et le paysage*

Actes

*Spatial Analysis and GEOmatics
Toulouse, 17, 18 et 19 novembre 2010*

<http://sageo10.univ-toulouse.fr>



magis



Spatial
Simulation
for the
Social
Sciences



Dynafor



Table des matières

BIENVENUE À SAGEO'10 !

COMITÉS

LISTE DES AUTEURS

LISTE DES ORGANISMES PAR PAYS

PROGRAMME

CONFÉRENCIERS INVITÉS / KEYNOTE SPEAKERS

HUI LIN (*CHINESE UNIVERSITY OF HONG KONG*) – WHAT COULD BE AFTER GEOGRAPHIC INFORMATION SYSTEMS?

PAUL BOLSTAD (*UNIVERSITY OF MINNESOTA, USA*) – DOES UNCERTAIN DATA PREVENT COMPLEX MODELS, AND CAN THE WEB AND MOTES SET US FREE?

ROBERT G. PONTIUS JR. (*CLARK UNIVERSITY, WORCESTER, USA*) – THE IMPORTANCE OF ACKNOWLEDGING ERRORS

SESSION 1 : ANALYSE SPATIALE 2D ET 3D

CARTOLIS : VERS UN OUTIL GÉOMATIQUE POUR IDENTIFIER ET CARACTÉRISER LES SEGMENTS DE LISIÈRES FORESTIÈRES (28-41)

Alignier Audrey **, ** — *Espy Philippe* **, ** — *Deconchat Marc* *, ** — *Ladet Sylvie* **, **

PRESENCE-ONLY MODELLING FOR INDICATOR SPECIES DISTRIBUTION:

BIODIVERSITY MONITORING IN THE FRENCH ALPS (42-55)

*Mathilde Redon** — *Sandra Luque**

OPÉRATIONS ET COHÉRENCE POUR L'EXPLORATION DES PROPRIÉTÉS STRUCTURELLES D'UN RÉSEAU DE TRANSPORT (56-72)

Eric Mermet **, ** — *Jean-François Gleyze**

UN SYSTÈME D'INFORMATION GÉOGRAPHIQUE 3D POUR L'EXPLORATION DES RÈGLES D'URBANISME (73-88)

*Mickaël Brasebin** — *Julien Perret** — *Cécile Haëck***

OPTIMISER L'EFFICACITÉ ÉNERGÉTIQUE DU BÂTI ET DE LA VILLE GRÂCE À LA MODÉLISATION 3D. ENJEUX ET PERSPECTIVES D'APPLICATION. (89-105)

*Aurélie Prévost** — *Diana Rodríguez** — *Nathalie Molines**, ** — *Benoit Beckers** — *Isabelle Caudron***

SESSION 2 : SIMULATION SPATIO-TEMPORELLE

SIMULER LA DENSIFICATION DU TISSU URBAIN AU MOYEN D'UN PROCESSUS DE PEUPEMENT (108-123)

*Florence Curie** — *Annabelle Mas*** — *Julien Perret**** — *Anne Puissant*****
— *Anne Ruas****

FUTURE URBAN GROWTH SCENARIOS IN THE MADRID REGION, SPAIN (124-139)

*Plata Rocha, W.** — *Gómez Delgado, M.*** — *Bosque Sendra, J.***

ACQUISITION DE CONNAISSANCES SUR LES DÉPLACEMENTS QUOTIDIENS DES INDIVIDUS DANS UN CONTEXTE DE RISQUES NATURELS. PROTOCOLES D'ENQUÊTE À L'AIDE DE TECHNOLOGIES MOBILES (140-154)

*Sidonie Christophe** — *Paule-Annick Davoine** — *Francis Jambon** — *Isabelle André-Poyaud*** — *S. Chardonnel***

MESURES DE SIMILARITÉ DE TRAJECTOIRES SUIVANT LE MÊME ITINÉRAIRE (155-169)

*Thomas Devogele** — *Laurent Etienne***

SESSION 3 : FOUILLE DE DONNÉES ET CLASSIFICATION AUTOMATIQUE

FOUILLE DE DONNÉES COMPLEXES : DES RELEVÉS TERRAIN AUX DONNÉES SATELLITAIRES POUR LA CARTOGRAPHIE DE PAYSAGES AGRICOLES (172-185)

*Fadi Badra** — *Elodie Vintrou*** — *Agnès Begué*** — *Maguelonne Teisseire**

EXTRACTION DE MOTIFS D'ÉVOLUTION DANS LES SÉRIES TEMPORELLES D'IMAGES SATELLITES (186-197)

*François Petitjean** — *Pierre Gançarski** — *Florent Maseglia***

CLASSIFICATION DES TISSUS URBAINS À PARTIR DE DONNÉES VECTORIELLES — APPLICATION À STRASBOURG (198-211)

*Anne Puissant** — *Nicolas Lachiche*** — *Agnès Braud*** — *Grzegorz Skupinski** — *Julien Perret****

IDENTIFICATION AUTOMATIQUE DE RÉSEAUX DE VOIRIE URBAINE AVEC UN ALGORITHME D'OPTIMISATION PAR COLONIES DE FOURMIS (212-226)

*Xavier Marsault**

SESSION 4 : GEOWEB ET CARTOGRAPHIE COLLABORATIVE

CARTOGRAPHIE ET SIG À L'ÈRE DU WEB 2.0. VERS UNE NOUVELLE GÉNÉRATION DE SIG PARTICIPATIFS (228-242)

*Boris Mericskay** — *Stéphane Roche***

INVASIVES SPECIES TRACKING IN ONTARIO. USING GEOMATICS FOR COLLABORATIVE INNOVATION (243-257)

*Fred McGarry** — *Donald Cowan*** — *Paulo Alencar***

CO-VISUALISATION DE DONNÉES SUR UNE CARTE. AIDE À LA CARTOGRAPHIE DE DONNÉES MULTIPLES ET HÉTÉROGÈNES SUR UN PORTAIL DE DONNÉES GÉOGRAPHIQUES (258-274)

Firas Bessadok — *Sébastien Mustière* — *Anne Ruas*

LES RANDONNEURS DÉFINISSENT LEURS CARTES. EXPLOITATION D'UNE ENQUÊTE SEMI-DIRECTIVE À QUESTIONS OUVERTES AVEC DES OUTILS STATISTIQUES ET LINGUISTIQUES (275-289)

*Catherine Domingues** — *Priscille Baldit-Schneller***

VERS LA CONCEPTION D'UN SERVICE WEB DE CONTEXTE SPATIAL DÉDIÉ AUX TÉLÉPHONES INTELLIGENTS DANS LE CADRE DE JEUX ÉDUCATIFS (290-303)

Elodie Edoh-Alove,*** — *Frédéric Hubert** — *Thierry Badard**

SESSION 5 : QUALITÉ, MÉTADONNÉES ET ONTOLOGIES

DL BASED AUTOMATED CONSISTENCY CHECKING OF SPATIAL RELATIONSHIPS (306-320)

Catherine Roussey — François Pinet***

VERS UN SYSTÈME D'AIDE À L'ÉVALUATION DE LA QUALITÉ DE DONNÉES SPATIO-TEMPORELLES (321-335)

Dounia Azzi — Christine Plumejeaud* — Marlène Villanova-Oliver* — Jérôme Gensel**

INTEGRATING MANY LEGENDS THROUGH ONTOLOGY FOR CARTOGRAPHIC SYMBOLS (336-348)

Roula Karam,*** — Franck Favetta* — Robert Laurini* — Rima Chamoun***

EVALUER, AMÉLIORER LA VISUALISATION DES DONNÉES SPATIALES, LE CAS DES DONNÉES URBAINES. CONSTRUCTION D'UN TEST PSYCHO-COGNITIF SUR LA REPRÉSENTATION VISUELLE DES INDICATEURS DE FLUX (349-363)

Robert Laurini — Rosa Marina Donolo* — Franck Favetta**

SESSION POSTERS

MÉTHODES INTERACTIVES DE DISCRÉTISATION DE DISTRIBUTIONS STATISTIQUES ET/OU/VERSUS SPATIALES (366-369)

Didier Josselin — Benoît Prache*⁺*

ENTRE LISSER ET RABOTER. APPLICATION DU FILTRE AUTO-ADAPTATIF MÉDIEN SUR IMAGE SATELLITES (370-375)

Didier Josselin — Alexandre Ulmer*⁺*

NONLINEAR SOURCE SEPARATION. CONTRIBUTION IN MULTISPECTRAL IMAGE CLASSIFICATION (376-378)

Hela Elmennai,** — Mohamed Anis Loughmari* — Mohamed Saber Naceur**

LES ESPACES D'UNE CONSOMMATION ENGAGÉE :

UNE GRILLE D'ANALYSE DES AMAP EN BOUCHES DU RHÔNE (379-381)

Noé Guiraud

LES CENTRALITÉS COMMERCIALES DE RÉSEAU DANS L'UNITÉ URBAINE DE PARIS : HIÉRARCHIE ET POLARISATIONS (382-384)

*Matthieu Delage**

FOUILLE DE DONNÉES SPATIALES POUR LA CARACTÉRISATION DE PAYSAGES EN LIEN AVEC DES FONCTIONNALITÉS AGRO-ÉCOLOGIQUES (385-387)

Thomas Guyet

INFRASTRUCTURE DE DONNÉES SPATIALES ET INGÉNIERIE TERRITORIALE : UNE APPROCHE EMPIRIQUE À PARTIR DU SIG PYRÉNÉES (388-392)

Pauline Lenormand,** — Laurent Plainecassagne**

LES TRAJECTOIRES DES VÉHICULES D'URGENCES COMME OUTIL DE CAPITALISATION D'ITINÉRAIRES (393-395)

Bilel Soussi — Jean Michel Follin**

VERS UN PROFIL UML POUR LES ENTREPÔTS DE DONNÉES INTÉGRANT LES RÉSEAUX SPATIAUX (396-398)

Mouna Snoussi — Sandro Bimonte* — Myoung-Ah Kang***

ETUDE DE LA REDONDANCE DES RÉSEAUX POUR ÉVALUER LEUR CAPACITÉ À FONCTIONNER EN MODE DÉGRADÉ (399-401)

*Serge Lhomme*** — Damien Serre* — Youssef Diab* — Richard Laganier***

EVALUATION DE LA VULNÉRABILITÉ DES RÉSEAUX FACE AUX RISQUES INONDATION (402-404)

Valère Griot — Serge Lhomme*** — Frédéric Bourgade* — Damien Serre* — Youssef Diab**

**TOXICITY: UN MODÈLE MULTI-AGENT POUR SIMULER LES COMPORTEMENTS DES POPULATIONS
LORS D'UN ACCIDENT INDUSTRIEL TOXIQUE (405-407)**

*Emmanuel Bonnet** — *Élise Beck*** — *David Sheeren**** — *Johnny
Douvinet***** — *Eric Daudé#* — *Marion Amalric##* — *Françoise Duraffour###* —
Paul Salze####

**SPATIAL INFORMATION FUSION. APPLICATION TO EXPERTISE AND MANAGEMENT OF NATURAL
RISKS IN MOUNTAINS (408-412)**

*Jean-Marc Tacnet** — *Martin Rosalie**** — *Jean Dezert**** — *Eric
Travaglini*,*****

**GRAVITATION ET ÉTALEMENT URBAIN ; UN ESSAI DE MÉCANIQUE DU COUPLE VITESSE DE
CIRCULATION ROUTIÈRE ET DENSITÉ DE POPULATION (413-415)**

Cyril Enault

CARTOMOUV' : INFORMER ET PARTAGER SUR LA CARTOGRAPHIE ANIMÉE (P. 416-420)

Groupe Cartomouv

SESSION R & D

**VISUALISATION INTERACTIVE DE RÉSULTATS DE SIMULATIONS NUMÉRIQUES SOUS GOOGLE
EARTH**

Nicolas Merlet,*** — *Marie-Madeleine Maubourguet** — *Denis Dartus**

**LIVEFIRE, DISPOSITIF DE GÉOLOCALISATION AUTOMATISÉE ET TEMPS-RÉEL DE PRISES DE VUES
AÉROPORTÉES DANS LE CADRE DE LA LUTTE CONTRE LES FEUX DE FORÊT**

*Pauline Crombette** — *Christophe Landrieu*** — *Françis Mateu*** — *Jean-
François Galtié**** — *Nicolas Merlet***,*****

MARIAGE DU SIG ET DE LA 3D AU SERVICE DU GRAND PAYSAGE

*Hélène Durand**

APPORTS D'UN GAZETIER POUR L'EXPLORATION DE CONNAISSANCES HÉTÉROGÈNES

*Olivier Bedel** — *Erwan Quesseveur*** — *Ludovic Dupuis****

**EXEMPLE D'UTILISATION DANS L'IDS SIG-PYRÉNÉES D'EASYSOI : SOLUTION OPEN-SOURCE
D'INFRASTRUCTURE DE DONNÉES SPATIALES CONFIGURABLE PAR JOOMLA!**

*Thomas Portier** — *Xavier Merour*** — *Bruno Magoni****

ATELIER ONTOLOGIES GÉOGRAPHIQUES

**CONFÉRENCE INVITÉE : ONTOLOGY-BASED DATA INTEGRATION IN THE GEOSPATIAL DOMAIN: A
CASE STUDY IN SPAIN.**

Oscar Corcho, Universidad Politécnica de Madrid

**ARTICLE : RECHERCHE AUTOMATIQUE DE CONCEPTS ET DE RELATIONS SÉMANTIQUES DANS DES
DOCUMENTS GÉOGRAPHIQUES.**

Marion Laignelet, Mouna Kamel, Nathalie Aussenac-Gilles

**ARTICLE : UNE APPROCHE BASÉE SUR LES ONTOLOGIES ET LES AGENTS POUR L'INTEROPÉRABILITÉ
DES SIG**

Alnafie Emdjed, Hamadou Djamila, Libourel Thérèse

**ARTICLE : AOC, UNE ONTOLOGIE FORMELLE POUR LA MODÉLISATION DE SYSTÈMES COMPLEXES
EN GÉOGRAPHIE**

Patrice Langlois, Eric Daudé, Baptiste Blanpain,

**DÉMONSTRATION : TAXOMAP FRAMEWORK APPLIQUÉ À L'ALIGNEMENT D'ONTOLOGIES
GÉOGRAPHIQUES AU SEIN DU PROJET GEONTO**

Fayçal Hamdi, Chantal Reynaud, Brigitte Safar

**DÉMONSTRATION : VERS UN SYSTÈME EN LIGNE D'AIDE À LA DÉCOUVERTE DU CONTENU DES
BASES DE DONNÉES GÉOGRAPHIQUES**

Ammar Mechouche, Nathalie Abadie, Emeric Prouteau, Sébastien Mustière.

Bienvenue à SAGEO'10 !

La conférence Internationale de Géomatique et d'Analyse Spatiale SAGEO (Spatial Analysis and GEomatics), soutenue par le GDR MAGIS (Méthodes et Applications pour la Géomatique et l'Information Spatiale, le réseau français de recherche en géomatique appuyé par le CNRS) et le groupe de recherche européen S4 (Spatial Simulation for the Social Sciences), est devenue progressivement l'événement national géomatique majeur en France.

L'édition toulousaine 2010 « Outils, Méthodes et Modèles en Géomatique pour la production de connaissances sur les territoires et le paysage » témoigne de nouveau, par son programme riche et diversifié, du statut pluridisciplinaire de la conférence faisant appel, notamment, à l'informatique, la géographie, l'aménagement, l'écologie et l'environnement. Les titres des sessions thématiques – « Analyse spatiale 2D et 3D », « Simulation spatio-temporelle », « Fouille de données et classification automatique », « Geoweb et cartographie collaborative », « Qualité, métadonnées, ontologies » – soulignent la dynamique scientifique de ce carrefour des sciences que sont l'analyse spatiale et la géomatique, autant que les fronts de recherche sur lesquels les 3 conférenciers invités de renommée internationale focaliseront leur propos : comment gérer les erreurs et les incertitudes dans les données et les modèles ? Et quel futur au-delà des systèmes d'informations géographiques ?

La diversité de cette manifestation se matérialise aussi dans la forme : aux présentations d'articles scientifiques se joignent l'exposition de posters, des communications d'activités de R&D dans le milieu professionnel ainsi qu'une série d'ateliers (« Logiciels », « Ontologies géographiques » et « SIG mobiles, réseau de capteurs localisés, SIG ubiquitaires ») et une table ronde sur la formation et la recherche en géomatique.

La réalisation d'une telle manifestation scientifique n'a été possible que grâce à un soutien institutionnel et à l'implication personnelle de nombreux collègues. Nous exprimons notre gratitude à l'Université de Toulouse, notamment ses composantes Université de Toulouse II – Le Mirail qui nous accueille dans ses murs, et l'Institut National Polytechnique – Ecole Nationale Supérieure d'Agronomie de Toulouse, le CNRS et l'INRA. Un grand merci à toutes les personnes qui ont contribué à cette édition : les membres des Comité d'Organisation, Scientifique et de Programme sans oublier le soutien logistique du Centre de Promotion de la Recherche Scientifique de l'Université de Toulouse II. Et un merci tout particulier à tous les auteurs qui, en partageant le fruit de leurs analyses et de leurs expériences, font vivre la recherche d'aujourd'hui et préparent celle de demain.

En vous souhaitant trois journées de discussions et d'échanges fructueux et enrichissants,

Claude Monteil et Martin Paegelow
Co-Présidents de SAGEO'10

Comités

Comité de Programme

DONNAY, Jean-Paul (Belgique)	MONTEIL, Claude (France)
JOLIVEAU, Thierry (France)	PAEGELOW, Martin (France)
LE BER, Florence (France)	SEDES, Florence (France)

Comité Scientifique

MONTEIL Claude (France, co-président du comité scientifique)
PAEGELOW Martin (France, co-président du comité scientifique)

BANOS, Arnaud (France)	LARDON, Sylvie (France)
BATTON-HUBERT, Mireille (France)	LIBOUREL, Thérèse (France)
BOLSTAD, Paul (USA)	MAINGUENAUD, Michel (France)
BORD, Jean-Paul (France)	MAS, Jean-François (Mexique)
BOUJU, Alain (France)	MATHIAN, Hélène (France)
BUCHER, Bénédicte (France)	MUSTIERE, Sébastien (France)
CAMACHO OLMEDO, M. Teresa (Espagne)	PASSOUANT, Michel (France)
CHEYLAN, Jean-Paul (France)	PINET, François (France)
COLLET, Claude (Suisse)	PODOBNIKAR, Tomaz (Slovénie)
CHRISMAN, Nicholas (Canada)	PUISSANT, Anne (France)
CLARAMUNT, Christophe (France)	RAY, Cyril (France)
DAVOINE, Paule-Annick (France)	RUAS, Anne (France)
DE RUFFRAY, Sophie (France)	SALLABERRY, Christian (France)
DEVOGELE, Thomas (France)	SANDERS, Lena (France)
DONNAY, Jean-Paul (Belgique)	SEDES, Florence (France)
DUMOLARD, Pierre (France)	SCHNEIDER, Michel (France)
GAIO, Mauro (France)	SERVIGNE, Sylvie (France)
GANCARSKI, Pierre (France)	SHEEREN, David (France)
HOUET, Thomas (France)	TANNIER, Cécile (France)
JOLIVEAU, Thierry (France)	VANGENOT, Christelle (Suisse)
JOSSELIN, Didier (France)	WACHOWICZ Monica (Pays-Bas)
LE BER, Florence (France)	WEBER, Christiane (France)
LE CORRE, Sébastien (France)	WEMMERT, Cédric (France)
LAFFLY, Dominique (France)	ZEITOUNI, Karine (France)

Comité d'Organisation

Le Laboratoire DYNAFOR, UMR 1201 INRA/INP-ENSAT

MONTEIL Claude, LADET Sylvie, SHEEREN David

Le Laboratoire GEODE, UMR 5602 CNRS/Université Toulouse II-Le Mirail

PAEGELOW Martin, HOUET Thomas

Le Département de Géographie et d'Aménagement de l'Université de Toulouse II

LE CORRE Sébastien, JEGOU Laurent

Le GdR MAGIS

CLARAMUNT Christophe

En partenariat avec :

L'Université de Toulouse II-Le Mirail.

DENUC Fabienne

Le Master SIGMA ("Systèmes d'Information Géographique pour la Maîtrise de l'environnement et l'Aménagement des territoires").

Liste des auteurs

- ALENCAR Paulo, David R. Cheriton School of Computer Science, University of Waterloo, Canada - palencar@uwaterloo.ca
- ALIGNIER Audrey, INRA, UMR 1201 DYNAFOR - Audrey.Alignier@toulouse.inra.fr
- ANDRÉ-POYAUD Isabelle, UMR PACTE Territoire CNRS, Université Joseph Fourier - isabelle.andre-poyaud@ujf-grenoble.fr
- ANSELOT Marc, conseil general du val de marne - marc.anselot@cg94.fr
- AZZI Dounia, LIG - Dounia.Azzi@bvra.e.ujf-grenoble.fr
- BADARD Thierry, Département des Sciences Géomatiques, Université Laval, Canada - thierry.badard@scg.ulaval.ca
- BADRA Fadi, TETIS - Cemagref - fadi.badra@teledetection.fr
- BALDIT-SCHNELLER Priscille, TIM/INaLCO - priscille.schneller@yahoo.fr
- BECK Elise, Université Joseph Fourier (Grenoble) / UMR PACTE 5194 CNRS - elise.beck@ujf-grenoble.fr
- BECKERS Benoit, Université de Technologie de Compiègne - benoit.beckers@utc.fr
- BEDEL Olivier, Alkante SAS - o.bedel@alkante.com
- BÉGUÉ Agnès, TETIS - Cirad - agnes.begue@teledetection.fr
- BESSADOK Firas, COGIT/IGN - firmas.bessadok@ign.fr
- BIMONTE Sandro, CEMAGREF - sandro.bimonte@cemagref.fr
- BOLSTAD Paul, University of Minnesota, USA - pbolstad@umn.edu
- BONNET Emmanuel, UMR IDEES 6266 CNRS - GEOSYSCOM - Université de Caen Basse Normandie - emmanuel.bonnet@unicaen.fr
- BOSQUE SENDRA Joaquín, Universidad de Alcalá, Spain - joaquin.bosque@uah.es
- BOURGADE Frédéric, Ecole des Ingénieurs de la Ville de Paris (EIVP)
- BRASEBIN Mickaël, COGIT/IGN - mickael.brasedin@ign.fr
- BRAUD Agnès, LSIT UMR 7005 CNRS - Université de Strasbourg - agne.braud@unistra.fr
- CARTOMOUV Groupe, UMR Géographie-cités - cartomouv@parisgeo.cnrs.fr
- CAUDRON Isabelle, Communauté Urbaine du Grand Toulouse - isabelle.caudron@grandtoulouse.fr
- CHARDONNEL Sonia, UMR PACTE Territoire CNRS, Université Joseph Fourier - sonia.chardonnel@ujf-grenoble.fr

- CHRISTOPHE Sidonie, Laboratoire d'Informatique de Grenoble, Université Joseph Fourier - sidonie.christophe@imag.fr
- COWAN Donald, David R. Cheriton School of Computer Science, University of Waterloo, Canada - dcowan@uwaterloo.ca
- CROMBETTE Pauline, M2 SIGMA (INP-ENSAT / UT2) - pauline@crombette.fr
- CURIE Florence, COGIT, IGN - florence.curie@ign.fr
- DARTUS Denis, Institut de Mécanique des Fluides de Toulouse - dartus@imft.fr
- DAUDÉ Eric, Université de Rouen / UMR IDEES 6266 CNRS - eric.daude@univ-rouen.fr
- DAVOINE Paule-Annick, Laboratoire d'Informatique de Grenoble, Université Joseph Fourier - paule-annick.davoine@imag.fr
- DECONCHAT Marc, INRA, UMR 1201 DYNFOR - Marc.Deconchat@toulouse.inra.fr
- DELAGE Matthieu, Université Paris 1, UMR Géographie-cités, Equipe PARIS - matthieu.delage@parisgeo.cnrs.fr
- DEVOGELE Thomas, Université François Rabelais (Campus de Blois) - thomas.devogele@univ-tours.fr
- DEZERT Jean, ONERA - jean.dezert@onera.fr
- DIAB Youssef, Université Paris Est - Ecole des Ingénieurs de la Ville de Paris (EIVP)
- DOMINGUÈS Catherine, Laboratoire COGIT – Institut Géographique National - catherine.domingues@ign.fr
- DONOLO Rosa Marina, INSA Lyon (and Tor Vergata University of Rome) - rosamarinad@hotmail.com
- DOUVINET Johnny, Université de Avignon, UMR ESPACE 6012 CNRS - johnny.douvinet@univ-avignon.fr
- DUPUIS Ludovic, University of Greenwich, United Kingdom - L.A.Dupuis@greenwich.ac.uk
- DURAND Helene, Alisé géomatique - helene.durand@wanadoo.fr
- EDOH-ALOVE Elodie, Département des Sciences Géomatiques, Université Laval, Canada - djogbenyuie@yahoo.fr
- ELMANNAI Hela, Ecole Nationale d'Ingénieurs de Tunis -Ecole Supérieure des Communications de Tunis, Tunisia - hela.elmannai@gmail.com
- ENault Cyril, Conseil général du val de marne - cyril.enault@cg94.fr
- ESPY Philippe, INRA, UMR 1201 DYNFOR - Philippe.Espy@toulouse.inra.fr
- ETIENNE Laurent, Institut de Recherche de L'Ecole navale - laurent.etienne@ecole-navale.fr
- FAVETTA Franck, INSA de Lyon - franck.favetta@insa-lyon.fr

FOLLIN Jean Michel, ESGT (école supérieure des géomètres et topographes) le Mans - jmfollin@esgt.cnam.fr

GALTIE Jean-François, UMR 5602 (CNRS/UT2) - galtie@univ-tlse2.fr

GANCARSKI Pierre, Laboratoire des Sciences de l'Image, de l'Informatique et de la Télédétection (LSIIT) - gancarski@unistra.fr

GENSEL Jérôme, LIG - Jerome.Gensel@imag.fr

GLEYZE Jean-François, IGN / Laboratoire COGIT - jean-francois.gleyze@ign.fr

GÓMEZ DELGADO Montserrat, Universidad de Alcalá, Spain - montserrat.gomez@uah.es

GRIOT Valère, Ecole des Ingénieurs de la Ville de Paris (EIVP)

GUIRAUD Noé, Université de Provence - noeguiraud@gmail.com

GUYET Thomas, AGROCAMPUS-OUEST - thomas.guyet@agrocampus-ouest.fr

HAËCK Cécile, IGN Conseil - cecile.haeck@ign.fr

HUBERT Frédéric, Département des Sciences Géomatiques, Université Laval, Canada - frederic.hubert@scg.ulaval.ca

JAMBON Francis, Laboratoire d'Informatique de Grenoble, Université Joseph Fourier - francis.jambon@imag.fr

JOSSÉLIN Didier, UMR ESPACE CNRS Université d'Avignon et des Pays de Vaucluse - didier.josselin@univ-avignon.fr

KANG Myoung-Ah, LIMOS - kang@isima.fr

KARAM Roula, INSA de Lyon - roula.karam@insa-lyon.fr

KILANY CHAMOUN Rima, Université Saint Joseph, Lebanon - rima.kilany@usj.edu.lb

LACHICHE Nicolas, LSIIT UMR 7005 CNRS - Université de Strasbourg - nicolas.lachiche@unistra.fr

LADET Sylvie, INRA, UMR 1201 DYNAFOR - Sylvie.Ladet@toulouse.inra.fr

LAGANIER Richard, Université Paris Diderot (Paris7)

LANDRIEAU Christophe, SDIS 66 - christophe.landrieau@sdis66.fr

LAURINI Robert, INSA Lyon - Robert.Laurini@insa-lyon.fr

LENORMAND Pauline, Dynamiques Rurales, TOULOUSE, France - INRA, UMR AGIR, CASTANET-TOLOSAN - APEM, CASTANET TOLOSAN - paulinelenormand@gmail.com

LHOMME Serge, Université Paris Est - Ecole des Ingénieurs de la Ville de Paris (EIVP) - serge.lhomme@eivp-paris.fr

LIN Hui, Institute of Space and Earth Information Science (ISEIS), Université chinoise de Hong-Kong (CUHK) - huilin@cuhk.edu.hk

- LOGHMARI Mohamed Anis, Ecole Nationale d'Ingénieurs de Tunis, Tunisia - MohamedAnis.Loghmari@isi.rnu.tn
- LUQUE Sandra, Cemagref - sandra.luque@cemagref.fr
- MAGONI Bruno, Depth, Switzerland
- MALINVERNI Eva Savina, SI2G "Geographical Intelligent Information System", Italy - info@si2g.it
- MARSAULT Xavier, Laboratoire MAP-ARIA - xavier.marsault@aria.archi.fr
- MAS Annabelle, CEDETE, Université d'Orléans - annabelle.mas@univ-orleans.fr
- MASSEGLIA Florent, INRIA Sophia Antipolis - Florent.Masseglia@sophia.inria.fr
- MATEU Francis, SDIS66 - francis.mateu@sdis66.fr
- MAUBOURGUET Marie-Madeleine, Institut de Mécanique des Fluides de Toulouse - maubourg@imft.fr
- MCGARRY Fred, Centre for Community Mapping, Canada - mcgarry@comap.ca
- MERICSKAY Boris, Université Laval, Canada - boris.mericskay.1@ulaval.ca
- MERLET Nicolas, IMFT (UMR 5502 CNRS/INP-UPS) / GEODE (UMR 5602 CNRS/UT2) - nicolas@merlet.eu
- MERMET Éric, IGN / Laboratoire COGIT - eric.mermet@ign.fr
- MEROUR Xavier, ASIT-VD, Switzerland
- MOLINES Nathalie, Université de Technologie de Compiègne - nathalie.molines@utc.fr
- MUSTIÈRE Sébastien, COGIT/IGN - sebastien.mustiere@ign.fr
- NACEUR Mohamed Saber, Ecole Nationale d'Ingénieurs de Tunis, Tunisia - Saber.naceur@insat.rnu.tn
- PERRET Julien, COGIT, IGN - julien.perret@gmail.com
- PETITJEAN Francois, Laboratoire des Sciences de l'Image, de l'Informatique et de la Télédétection (LSIIT) - fpetitjean@unistra.fr
- PINET François, Cemagref - francois.pinet@cemagref.fr
- PLAINECASSAGNE Laurent, APEM, F-31320 CASTANET TOLOSAN, France - lp@apem.asso.fr
- PLATA ROCHA Wenseslao, Universidad Autonoma de Sinaloa, Mexico - wenseslao.plata@alu.uah.es
- PLUMEJEAUD Christine, LIG - christine.plumejeaud@imag.fr
- PONTIUS Robert Gilmore, School of Geography, Clark University, Massachusetts, USA - rpontius@clarku.edu
- PORTIER Thomas, APEM - portier@apem.asso.fr

PRACHE Benoît, IUT STID / UMR ESPACE CNRS Université d'Avignon et des Pays de Vaucluse - benoit.prache@etd.univ-avignon.fr

PRÉVOST Aurélie, Université de Technologie de Compiègne - aurelie.prevost31@gmail.com

PUISSANT Anne, LIVE ERL 7230 CNRS - Université de Strasbourg - anne.puissant@live-cnrs.unistra.fr

QUESSEVEUR Erwan, ESO Rennes, Université Rennes 2, UEB - erwan.quesseveur@uhb.fr

REDON Mathilde, Cemagref - mathilde.redon@cemagref.fr

ROCHE Stéphane, Université Laval, Canada - stephane.roche@scg.ulaval.ca

RODRÍGUEZ Diana, Université de Technologie de Compiègne - diana.rodriguez-de-santiago@utc.fr

ROSALIE Martin, Cemagref, INSA Rouen - martin.rosalie@insa-rouen.fr

ROUSSEY Catherine, Cemagref - catherine.roussey@cemagref.fr

RUAS Anne, COGIT, IGN - anne.ruas@ign.fr

SERRE Damien, Université Paris Est - Ecole des Ingénieurs de la Ville de Paris (EIVP)

SHEEREN David, Université de Toulouse / INPT-ENSAT, UMR 1201 DYNAFOR - david.sheeren@ensat.fr

SKUPINSKI Grégoire, LIVE ERL 7230 CNRS - Université de Strasbourg - grzegorz.skupinski@live-cnrs.unistra.fr

SNOUSSI Mouna, CEMAGREF - mouna.snoussi@cemagref.fr

SOUSSI Bilel, ESGT (école supérieure des géomètres et topographes) le Mans - mohamed.soussi@esgt.cnam.fr

TACNET Jean-Marc, Cemagref - jean-marc.tacnet@cemagref.fr

TEISSEIRE Maguelonne, TETIS - Cemagref - maguelonne.teisseire@teledetection.fr

TRAVAGLINI Eric, Cemagref, Université de Savoie - eric.travaglini@cemagref.fr

ULMER Alexandre, IUT STID / UMR ESPACE CNRS Université d'Avignon et des Pays de Vaucluse - alex_andre@hotmail.com

VILLANOVA-OLIVER Marlène, LIG - Marlene.Villanova-Oliver@imag.fr

VINTROU Elodie, TETIS - Cirad - elodie.vintrou@teledetection.fr

ZINGARETTI Primo, SI2G "Geographical Intelligent Information System", Italy - info@si2g.it

Liste des organismes par pays

CANADA

Centre for Community Mapping, Waterloo, Ontario
David R. Cheriton School of Computer Science, University of Waterloo, Ontario
Département des Sciences Géomatiques, Université Laval, Québec

CHINA

Université chinoise de Hong-Kong (CUHK), Institute of Space and Earth
Information Science (ISEIS)

ENGLAND

University of Greenwich

FRANCE

AGROCAMPUS-OUEST, Rennes
Alisé géomatique, Montpellier
Alkante SAS, Cesson-Sévigné
APEM, Foix
CEMAGREF, Bordeaux
CEMAGREF, Clermont-Ferrand
CEMAGREF, UMR TETIS, Montpellier
CIRAD, UMR TETIS, Montpellier
CNRS, ERL LIVE, Strasbourg
CNRS, UMR ESPACE, Avignon
CNRS, UMR GEODE, Toulouse
CNRS, UMR Géographie-cités, Paris
CNRS, UMR IDEES-GEOSYSCOM, Caen
CNRS, UMR IMFT, Toulouse
CNRS, UMR LSIIT, Strasbourg
CNRS, UMR MAP-ARIA, Lyon
CNRS, UMR PACTE, Grenoble
Communauté Urbaine du Grand Toulouse
Conseil Général du Val de Marne, Créteil
École des Ingénieurs de la Ville de Paris (EIVP)
ESGT (école supérieure des géomètres et topographes), le Mans
IGN, Laboratoire COGIT, Saint Mandé
IGN Conseil
INaLCO, Département TIM, Paris
INRA, UMR AGIR, Toulouse
INRA, UMR DYNAFOR, Toulouse
INRA, UMR Dynamiques Rurales, Toulouse
INRIA, Sophia Antipolis
INSA, Lyon
INSA, Rouen
Institut de Recherche de L'École Navale, Lanvéoc-Poulmic

ISIMA, UMR LIMOS, Aubières
IUT STID, Avignon
Master Géomatique SIGMA (INP-ENSAT / UT2), Toulouse
ONERA, Chatillon
SDIS 66, Perpignan
Université d'Avignon, UMR ESPACE
Université d'Orléans, EA CEDETE
Université de Caen Basse Normandie, UMR IDEES-GEOSYSCOM
Université de Paris 1 (Panthéon-Sorbonne), UMR Géographie-cités
Université de Paris 7 (Paris Diderot)
Université de Paris Est, Marne-la-Vallée
Université de Provence, Aix-Marseille
Université de Rennes 2, UMR ESO
Université de Rouen, UMR IDEES
Université de Savoie, Chambéry
Université de Strasbourg, ERL LIVE
Université de Strasbourg, UMR LSIIT
Université de Technologie de Compiègne
Université de Toulouse, INPT/ENSAT, UMR DYNAFOR
Université de Toulouse, INPT-UPS, UMR IMFT
Université de Toulouse, UT2, UMR GEODE
Université Européenne de Bretagne (UEB), Rennes
Université François Rabelais, Blois
Université Joseph Fourier, UMR LIG, Grenoble
Université Joseph Fourier, UMR PACTE, Grenoble

ITALY
SI2G "Geographical Intelligent Information System", Ancona
Universita di Roma Tor Vergata

LEBANON
University Saint Joseph, Beyrouth

MEXICO
Universidad Autonoma de Sinaloa

SPAIN
Universidad de Alcalá, Madrid

SWITZERLAND
ASIT-VD, Lausanne
Depth SA, Monthey

TUNISIA
Ecole Nationale d'Ingénieurs de Tunis
Ecole Supérieure des Communications de Tunis

USA
Clark University, Worcester, Massachussets
University of Minnesota, Minneapolis – St Paul

Programme

1er jour (MERCREDI 17 NOV 2010)

- ▶ 8h30-9h00 Accueil, inscriptions
- ▶ 9h00-9h30 Ouverture du colloque
- ▶ 9h30-10h30 Conférencier invité : **Hui Lin (Chinese University of Hong Kong) – *What could be after GIS ?***
- ▶ 10h30-11h00 Pause

- ▶ 11h00-13h00 ***Session 1 : Analyse spatiale 2D et 3D***
 - ALIGNIER Audrey, ESPY Philippe, DECONCHAT Marc, LADET Sylvie : *Cartolis : vers un outil géomatique pour identifier et caractériser les segments de lisières forestières*
 - REDON Mathilde, LUQUE Sandra : *Presence-only modelling for indicator species distribution: Biodiversity monitoring in the French Alps*
 - MERMET Éric, GLEYZE Jean-François : *Opérations et cohérence pour l'exploration des propriétés structurelles d'un réseau de transport*
 - BRASEBIN Mickaël, PERRET Julien, HAËCK Cécile : *Un système d'information géographique 3D pour l'exploration des règles d'urbanisme*
 - PRÉVOST Aurélie, RODRÍGUEZ Diana, MOLINES Nathalie, BECKERS Benoit, CAUDRON Isabelle : *Optimiser l'efficacité énergétique du bâti et de la ville grâce à la modélisation 3D. Enjeux et perspectives d'application*

- ▶ 13h00-14h30 Déjeuner

- ▶ 14h30-16h00 ***Session 2 : Simulation spatio-temporelle***
 - CURIE Florence, MAS Annabelle, PERRET Julien, PUISSANT Anne, RUAS Anne : *Processus de peuplement pour simuler la densification du tissu urbain*
 - PLATA ROCHA Wenseslao, GÓMEZ DELGADO Montserrat, BOSQUE SENDRA Joaquin : *Future urban growth scenarios in Madrid Region, Spain*

- CHRISTOPHE Sidonie, DAVOINE Paule-Annick, JAMBON Francis, ANDRÉ-POYAUD Isabelle, CHARDONNEL Sonia : *Acquisition de connaissances sur les déplacements quotidiens des individus dans un contexte de risques naturels : protocoles d'enquête à l'aide de technologies mobiles.*
- DEVOGELE Thomas, ETIENNE Laurent : *Similarité de trajectoires suivant le même itinéraire*

- ▶ 16h00-16h30 Pause
- ▶ 16h30-16h45 Remise de prix de thèse du GDR
- ▶ 16h45-18h00 Assemblée générale du GDR MAGIS
- ▶ 18h00-19h00 Réception d'accueil

2ème jour (JEUDI 18 NOV 2010)

- ▶ 8h30-9h30 Conférencier invité : **Paul Bolstad (University of Minnesota, USA)** – *Does uncertain data prevent complex models, and can the web and notes set us free ?*
- ▶ 9h30-10h00 Pause

▶ 10h00-11h30 *Session 3 : Fouille de données et classification automatique*

- BADRA Fadi, VINTROU Elodie, BÉGUÉ Agnès, TEISSEIRE Maguelonne :
alimentaire
- PETITJEAN Francois, GANCARSKI Pierre, MASSEGLIA Florent : *Extraction de motifs d'évolution dans les Séries Temporelles d'Images Satellites*
- PUISSANT Anne, LACHICHE Nicolas, BRAUD Agnès, SKUPINSKI Grégoire, PERRET Julien : *Classification automatique d'îlots urbains - application à Strasbourg*
- MARSAULT Xavier : *Identification automatique de réseaux de voirie urbaine avec un algorithme d'optimisation par colonies de fourmis*

- ▶ 11h30-11h45 Pause

► 11h45-1300 Présentation des posters scientifiques en salle

- JOSSELIN Didier, PRACHE Benoît : *Méthodes interactives de discrétisation de distributions statistiques et/ou/versus spatiales*
- JOSSELIN Didier, ULMER Alexandre : *Entre lisser et rabotter. Application du filtre autoadaptatif médien sur image satellites*
- ELMANNAI Hela, LOGHMARI Mohamed Anis, NACEUR Mohamed Saber : *Nonlinear source separation contribution in image classification*
- GUIRAUD Noé : *Une grille d'analyse des AMAP en Bouches du Rhône : les espaces d'une consommation engagée*
- DELAGE Matthieu : *Les centralités commerciales de réseau dans l'unité urbaine de Paris : Hiérarchie et polarisations*
- GUYET Thomas : *Fouille de données spatiales pour la caractérisation de paysages*
- LENORMAND Pauline, PLAINECASSAGNE Laurent : *Comment l'ingénierie territoriale utilise une solution open-source : l'exemple d'une Infrastructure de Données Spatiales basée sur le logiciel EasySDI*
- SOUSSI Bilel, FOLLIN Jean Michel : *Les trajectoires des véhicules d'urgences comme outil de capitalisation d'itinéraires*
- SNOUSSI Mouna, BIMONTE Sandro, KANG Myoung-Ah : *Un Profil UML pour entrepôts de données de réseaux spatiaux*
- LHOMME Serge, SERRE Damien, DIAB Youssef, LAGANIER Richard : *Etude de la redondance des réseaux pour évaluer leur capacité à fonctionner en mode dégradé*
- GRIOT Valère, BOURGADE Frédéric, LHOMME Serge, SERRE Damien, DIAB Youssef : *Evaluation de la vulnérabilité des réseaux face aux risques inondation*
- BONNET Emmanuel, BECK Elise, SHEEREN David, DOUVINET Johnny, DAUDÉ Eric : *TOXICITY: Un modèle multi-agent pour simuler les comportements des populations lors d'un accident industriel toxique*
- TACNET Jean-Marc, ROSALIE Martin, DEZERT Jean, TRAVAGLINI Eric : *Spatial information fusion : application to expertise and management of natural risks in mountains*
- ENAULT Cyril : *Gravitation et étalement urbain ; un essai de mécanique du couple vitesse de circulation routière et densité de population*
- Groupe CARTOMOUV: *Cartomouv' : informer et partager sur la cartographie animée*

► 13h00-14h30 Déjeuner

► 14h30-17h30 Ateliers en parallèle :

- Atelier **Logiciels** du groupe de travail Mutualisations (14h30-17h30)
- Atelier **Ontologies géographiques** (14h30-17h30)
- Atelier **SIG mobiles, réseau de capteurs localisés, SIG ubiquitaires** du Groupe-Projet Mobilité (14h30-17h30)
- Table-ronde : **Formation à la recherche en géomatique** (14h30-16h30)
- Présentations d'activités de **R&D dans le milieu professionnel** (16h45-17h30)
 - MERLET Nicolas, MAUBOURGUET Marie-Madeleine, DARTUS Denis: *Visualisation interactive de résultats de simulations numériques sous Google Earth*
 - CROMBETTE Pauline, LANDRIEU Christophe, MATEU François, GALTIE Jean-François, MERLET Nicolas : *LiveFire, système de géolocalisation automatisée et temps-réel de prises de vues aéroportées*
 - DURAND Helene : *Modélisation en 3D de la Baie du Mont St Michel*
 - BEDEL Olivier, QUESSEVEUR Erwan, DUPUIS Ludovic : *Apports d'un gazetier pour l'exploration de connaissances hétérogènes*
 - PORTIER Thomas, MEROUR Xavier, MAGONI Bruno : *EasySDI, an open-source solution for Spatial Data Infrastructure configurable via Joomla!*

► 17h30-19h30 Visite du centre-ville de Toulouse

► 19h30-23h00 Dîner de gala

3^{ème} jour (VENDREDI 19 NOV 2010)

► 8h30-9h30 Conférencier invité : **R.G. Pontius (Clark University, Worcester, USA) – *The importance of acknowledging errors***

► 9h30-10h00 Pause

► 10h00-12h00 **Session 4 : Geoweb et cartographie collaborative**

- MERICKSKAY Boris, ROCHE Stéphane : *Cartographie et SIG à l'ère du Web 2.0 : vers une nouvelle génération de SIG participatifs*
- MCGARRY Fred, COWAN Donald, ALENCAR Paulo : *Invasive Species Tracking in Ontario, using Geomatics for Collaborative Innovation*
- BESSADOK Firas, MUSTIÈRE Sébastien, RUAS Anne : *Co-visualisation de données sur une carte*
- DOMINGUÈS Catherine, BALDIT-SCHNELLER Priscille : *Les randonneurs définissent leurs cartes*
- EDOH-ALOVE Elodie, HUBERT Frédéric, BADARD Thierry : *Vers la conception d'un service web de contexte spatial dédié aux téléphones intelligents dans le cadre de jeux éducatifs*

► 12h00-13h30 Déjeuner

► 13h30-15h00 **Session 5 : Qualité, métadonnées, ontologies**

- ROUSSEY Catherine, PINET François : *Using DL for spatial incoherence detection*
- AZZI Dounia, PLUMEJEAUD Christine, VILLANOVA-OLIVER Marlène, GENSEL Jérôme : *Vers un système d'aide à l'évaluation de la qualité de données spatio-temporelles*
- KARAM Roula, FAVETTA Franck, KILANY CHAMOUN Rima, LAURINI Robert : *Map Conflation: Towards an ontology for cartographic symbols*
- LAURINI Robert, DONOLO Rosa Marina, FAVETTA Franck : *Améliorer la visualisation des données urbaines, construction d'un test psycho-cognitif sur la représentation visuelle des indicateurs spatiaux en 2D*

► 15h00-15h30 Discussion finale - Clôture du colloque

Extension (SAMEDI 20 NOV 2010)

Workshop *Land Change Modeling Methods: calibration, validation and extrapolation*, animé par **Professor R.G. Pontius (9-17h)**

Conférenciers invités / Keynote speakers



Hui Lin

(Chinese University of Hong Kong)

—
What could be after Geographic Information Systems?

GIS, after continual development in last 40 years, has been widely used in various fields by researchers, governmental officials, businessman, and many professionals and non-professionals. With its root from maps, GIS has more functions including spatial analysis and static spatial modeling. However, many GIS users today are looking for a platform which is of geo-process modeling functions, such as wild fire modeling and air pollution spreading simulation. The framework of GIS with a geo-coded database shows its bottleneck for this kind of dynamic modeling. What should we do for integrating the geo-coded database and the geo-process models? Virtual geographic environments (VGE) could be an answer as a new framework beyond GISystem. At the same time, we may need VGE for various studies on social geographies.

Hui Lin is Professor and Director of Institute of Space and Earth Information Science (ISEIS) of The Chinese University of Hong Kong (CUHK). He is Director of Hong Kong Base of National Remote Sensing Center of China. He graduated from Wuhan Technical University of Surveying and Mapping in 1980, received Master of Science degree from Graduate School of Chinese Academy of Sciences in 1983, and Ph.D. from University at Buffalo in 1992. Hui Lin joined CUHK in 1993 with major research interests in satellite remote sensing, virtual geographic environments (VGE), and urban and historic GIS. He was elected academician of International Eurasia Academy of Sciences in 1995 and member of the Expert Committee of China Lunar Exploring Program in 2008. Hui Lin is the founding president of International Association of Chinese Professionals in Geographic Information Science (CPGIS) and the chief editor of the international journal *Annals of GIS* published by Taylor & Francis. For a list of his publications and more information on the Institute of Space and Earth Information Science, visit: <http://www.iseis.cuhk.edu.hk/eng/>.

**Paul Bolstad***(University of Minnesota, USA)*

—

Does uncertain data prevent complex models, and can the web and motes set us free?

Science is an incremental endeavor, and our understanding becomes more parameter-rich and data demanding as we discover nuances in system response. Many of our spatial models rest on a well-documented, process-rich foundation, a silhouette of the real world developed in controlled studies at points or patches and validated against site-specific observation. These models often perform poorly when applied across space, in part because of an inappropriate model, and in part because data contain errors. Progress requires we parse the causes. Are our models too complex for the available data, or do they ignore important components of spatial interaction? How do we know? If models are to blame, what strategies lead to appropriate spatial models? If data are to blame, can an instrumented universe and self-informing, collaborative communities provide the answer? I'll discuss these questions in the context of landuse and climatic impacts on the water cycle.

Dr. **Paul Bolstad** is a Professor at the University of Minnesota, USA. He earned a PhD in 1990 from the University of Wisconsin – Madison. He has a long standing research experience with a wide range of domains: he has been called an ecosystems ecologist, a geographer, a forester, a geomaticist, and a modeler. His work spans these disciplines, and particularly on carbon and water cycling from leaf to regional scales, mostly in forest ecosystems and has published over 110 peer-reviewed papers in these areas. He serves on the editorial board of four journals, and has been awarded best annual paper in a category by the Society for Landscape Ecology and the American Society of Photogrammetry and Remote Sensing. His teaching focuses on spatial data collection and analysis. He is the author of book entitled “GIS Fundamentals”, used by over 300 universities and colleges worldwide. For a list of his publications and more information, visit: <http://www.paulbolstad.net/>.



Robert G. Pontius Jr.
(Clark University, Worcester, USA)

—
The importance of acknowledging errors

It is essential to acknowledge errors in both data and models, so that we can focus research on the most important challenges. If scientists feel professional pressure to claim consistently that data and models are correct or acceptable, then scientists will not report important weaknesses in the methods, which prevents progress in science. This presentation illustrates a collaborative exercise among land-change modelers to present results in a transparent manner. Examination of 13 applications of land-change models shows that there is more error than correctly predicted change for 12 of the 13 cases at the fine resolution of the raw data, but that most of the errors vanish at coarser resolutions. This exercise has liberated the community of modelers to report results clearly and honestly, and has helped to re-structure the agenda for future research.

R. G. Pontius is an applied statistician with expertise in Geographic Information Science (GIS), ecological modeling, and land change science. He is Associate Professor in the School of Geography at Clark University, Massachusetts, USA. He earned a Bachelor's degree in Mathematics, a Master's degree in Applied Statistics, and a Doctorate in Environmental Science from the State University of New York. A major product of his doctorate was the land change simulation model Geomod. Methods of map comparison for the sake of validation of predictive models are his most recent scholarly products. Several of these methods have been incorporated into the GIS software Idrisi, which has over 100,000 users worldwide. He has used the methods to lead an international exercise that synthesized numerous land change modeling applications with respect to predictive accuracy. He is on the editorial board of four journals and is author of more than 70 scientific articles. For publications and more information, visit: www.clarku.edu/rpontius.

Session 1 : Analyse spatiale 2D et 3D

Cartolis : vers un outil géomatique pour identifier et caractériser les segments de lisières forestières

Alignier Audrey *, ** — **Espy Philippe***, ** — **Deconchat Marc** *, ** — **Ladet Sylvie***, **

* INRA, UMR 1201 DYNAFOR, BP 52627, F-31326 Castanet-Tolosan, France

** Université de Toulouse, UMR 1201 DYNAFOR, F-31326 Castanet-Tolosan, France

RÉSUMÉ. Les lisières forestières sont des composantes clés des paysages ruraux car elles influencent des processus écologiques à l'interface entre les milieux. Les lisières varient selon leur physionomie, leur orientation, leur histoire et leur topographie, mais peu de méthodes sont disponibles pour identifier ces différents types de lisières et pour les cartographier sur une large étendue. Nous proposons d'identifier et de cartographier des segments de lisière basés sur la subdivision morphologique des contours de forêt. Ces segments sont caractérisés par eux mêmes et par requêtes sur d'autres données spatiales. Une chaîne de traitement basée sur des outils d'ArcGis, appliquée sur la sortie des outils GUIDOS, est utilisée pour identifier les segments de lisière. Nous fournissons des exemples de variables descripteurs de segments de lisières forestières obtenus à partir d'un modèle numérique de terrain et d'une carte d'occupation du sol. Les résultats montrent la faisabilité de cartographier automatiquement la variabilité des lisières sur une large étendue. Ces résultats ouvrent de nouvelles perspectives pour construire un nouvel outil d'analyse de la dynamique des paysages et de leurs effets sur la biodiversité.

ABSTRACT. Forest edges are key components of rural landscapes because they influence ecological processes at the interface between habitats. There is a variability of forest edges, according to their physiognomy, orientation, history and topography, but few methods are available to identify these different types of edges and to map them at a large scale. We propose to identify edge segments based on morphological subdivision of forest boundaries. These segments are mapped and characterized by request on other spatial data. A procedure based on ArcGis tools, applied on the output from GUIDOS tools, is used to identify edge segments. We provide examples of edge segments descriptors obtained from a digital elevation model and from a landcover map. Results showed the feasibility of the automatic mapping of edge variability over a large scale. It opens new perspectives to build up a new tool for analyzing landscape dynamics and their effects on biodiversity.

MOTS-CLÉS : Lisière forestière, SIG, métrique paysagère, segment de polyligne, morphologie, généralisation

KEYWORDS: Forest edge, GIS, landscape metric, segment of polyline, morphology, generalization.

1. Introduction

Les lisières forestières sont des éléments clés des paysages ruraux par l'influence qu'elles exercent sur les processus écologiques importants pour la conservation de la biodiversité (Murcia, 1995) aux interfaces entre milieux, en fournissant habitat et ressources à de nombreuses espèces animales et végétales. Pour mieux comprendre leurs effets sur la biodiversité, leurs rôles fonctionnels dans les paysages et les inclure dans les plans de gestion, il est nécessaire de prendre en compte et de caractériser leur diversité. En effet, les lisières d'un paysage ne sont pas toutes équivalentes. Elles se caractérisent par de nombreux attributs reflétant leur origine, leur gestion, leur structure, leur fonction et leur dynamique (Cadenasso *et al.*, 2003; Strayer *et al.*, 2003). Ainsi, l'orientation, l'histoire ou encore la topographie modifient l'effet des lisières sur l'environnement et la végétation (Alignier *et al.*, 2010).

Le rôle reconnu des lisières en écologie du paysage a conduit les scientifiques à rechercher des méthodes pour les caractériser à partir des données spatialisées et des outils de la géomatique. De nombreuses méthodes ont été développées pour la détection, la classification et la quantification des lisières (Fortin *et al.*, 2000; Zheng *et al.*, 2000 ; Fagan *et al.*, 2003). Ces approches se répartissent en trois groupes répondant à des questions et des démarches différentes.

Un premier groupe concerne les méthodes qui visent à détecter et identifier les lisières à partir de données spatialisées (point ou raster). Les filtres de détection de discontinuité (edge detection) disponibles dans les logiciels de traitement de données satellitaires correspondent à ce type de méthode. De leur côté, Fortin *et al.* (2000) ont développé des méthodes pour détecter des discontinuités dans une distribution de semis de points, qui peuvent représenter des arbres par exemple. Ces méthodes s'appliquent particulièrement pour détecter des lisières dans des situations où la répartition des arbres est graduelle, comme dans le cas des transitions entre les savanes arborées et les forêts. L'utilisation d'images aériennes semble être moins fréquente. Esseen *et al.* (2006) proposent cependant un échantillonnage par une grille appliquée à des photos aériennes ; les lisières sont détectées au niveau de leur intersection avec la grille.

Le second groupe rassemble les méthodes qui s'appuient sur des métriques de tache, au sens de McGarigal *et al.* (2009). L'outil le plus connu est FRAGSTAT (McGarigal *et al.*, 2002) qui calcule de très nombreuses métriques paysagères, dont certaines portent spécifiquement sur les lisières de tache dans une matrice homogène. Les données de base sont des cartes binaires raster sur lesquelles des taches de forêt sont identifiées par des ensembles de pixels. Le logiciel extrait des descripteurs des taches par des opérations morphologiques. Plusieurs indices mesurent des caractéristiques relatives aux lisières. Il s'agit notamment de la longueur totale de lisière (Total Edge), de la complexité du contour de la tache (Landscape Shape Index), du ratio longueur de lisières/surface des bois (Perimeter-Area Ratio). Un autre outil GUIDOS (Graphical User Interface for the Description of image Objects and their Shapes), proposé gratuitement par l'Union Européenne, permet de quantifier et de cartographier la fragmentation des milieux forestiers (Vogt *et al.*, 2007). D'une couche binaire bois/non bois, GUIDOS crée un ensemble de classes nouvelles qui correspondent notamment aux cœurs de bois, aux lisières, aux haies isolées et aux haies reliées à des bois (ou branches). Les lisières sont définies dans GUIDOS comme une certaine épaisseur sur le contour des principales taches forestières identifiées par des analyses morphologiques. Dans ces approches, les lisières et leurs caractéristiques sont des attributs des taches. La lisière est considérée comme étant homogène sur l'ensemble du pourtour de la tache. Les métriques de tache permettent donc d'obtenir des informations sur les lisières contrairement à la première démarche.

Le troisième groupe s'appuie sur une conception différente dans laquelle la lisière d'un bois est considérée comme un ensemble de segments de caractéristiques différentes. Ce point de vue assez récent a été formalisé notamment par Zeng *et al.* (2005) qui ont introduit le concept de « segment de lisière ». Ils proposent ainsi de développer des métriques de lisière (ou de ligne) différentes des métriques de tache, pour évaluer la fragmentation forestière. D'un point de vue géomatique, les lisières sont représentées par des lignes (ou polylignes) qu'on peut caractériser par leur morphologie (longueur, orientation) et par des données supplémentaires relatives aux milieux adjacents ou, grâce à un modèle numérique de terrain, relatives à la position dans la pente, l'exposition, la pente par exemple.

Notre approche s'apparente au deuxième groupe pour identifier les lisières individualisées des cœurs de bois et les distinguer des autres éléments boisés (outil GUIDOS) et au dernier groupe pour définir des segments de lisières. Cependant, elle ne se restreint pas à l'application de ces méthodes mais elle apporte des innovations par rapport aux travaux antérieurs en caractérisant de manière plus fine des segments de lisières qui correspondent à des unités spatiales pertinentes pour les écologues. Cette méthode, appelée Cartolis, est une étape vers la construction d'un outil géomatique intégré de mesure des métriques paysagères basé sur les segments de lisières.

2. Mise au point de la chaîne de traitement

2.1. Justification de la plateforme logicielle personnalisée

Afin de construire une maquette de l'outil visé, permettant de tester les différentes étapes de la méthode, nous avons opté pour le logiciel ArcGis 9.3¹ édité par la société ESRI en raison de sa stabilité et de la facilité à y automatiser les traitements. Des extensions complémentaires, payantes (ETGeoWizards², Spatial Analyst³) et gratuites (HawthsTools⁴, Easy Calculate⁵) ont été mobilisées pour certaines phases. Cependant, dans la perspective de construire un outil personnalisé à notre problématique mais ouvert pour des études similaires à travers le monde, nous avons privilégié l'utilisation de scripts libres de droit afin d'avoir accès au code et ainsi pouvoir le modifier ultérieurement pour nos spécifications. S'ajoute à notre configuration le logiciel gratuit GUIDOS élaboré par le JRC (Joint Research Center) de l'Union Européenne évoqué en introduction. L'outil final envisagé devrait s'apparenter à GUIDOS et pourra être employé en complément.

2.2. Choix d'une représentation à base de lignes

Nous avons opté pour le format vectoriel afin de cartographier les lisières forestières. En effet, nous souhaitons définir des objets spatiaux correspondant à ce qui est observé communément par les écologues dans les bois, c'est à dire des portions de lisières relativement rectilignes. Une représentation en mode raster ne permet pas d'identifier ce niveau d'agrégation et cette contrainte morphologique de rectitude. Les bois dans les régions tempérées présentent des formes polygonales qui résultent des contraintes

¹ édité par la société ESRI : <http://www.esrifrance.fr/>

² édité par la société ET SpatialTechniques : <http://www.ian-ko.com/>

³ édité par la société ESRI : http://www.esrifrance.fr/Spatial_Analyst.asp

⁴ édité par Beyer, H. L. 2004. Hawth's Analysis Tools for ArcGIS: <http://www.spatalecolology.com/htools>

⁵ édité par la société ET SpatialTechniques : <http://www.ian-ko.com/>

sociales (cadastres) et techniques, notamment du fait de la linéarité des opérations agricoles. Certaines limites de bois, dans les contacts avec des zones d'enfrichement, sont parfois moins rectilignes et aisées à identifier, mais ces cas restent assez rares dans les paysages d'Europe atlantique pour être négligés. Les bois sont alors représentés par des polygones, la bordure est représentée par la polyligne de contour, elle même composée de ses sous-parties, les segments. Par définition, une polyligne est un ensemble de segments de ligne, tel que le point final d'un segment soit le point de départ du segment suivant (Bolstad, 2008). Pour ces objets vecteur, le fichier informatique est constitué d'une séquence de nombres indiquant : l'identifiant de l'objet, le nombre de points (sommets) définissant cet objet et les coordonnées X et Y de chacun de ces points. Aussi il est aisé de passer d'un type de primitive à un autre par des outils de conversion : Polygone vers Polygones, Polygones vers Segments, etc.

2.3. Application d'une représentation en segments de ligne pour les lisières

La méthode proposée comporte trois grandes phases et démarre d'une carte d'occupation du sol exhaustive en mode raster couvrant une large étendue. Si les données initiales sont en mode vectoriel, la méthode est aisément adaptable mais n'a pas été testée. Dans la première phase, il s'agit de localiser les lisières dans le paysage étudié ; dans la seconde phase, les segments de lisière sont créés et identifiés ; la troisième phase vise à les caractériser par des variables définies par les écologues. Ces phases correspondent à des étapes de travail distinctes de déroulement de la méthode (Figure 1).

Phase 1: localiser les lisières (étapes 1 à 3 de la Figure 1)

Comme tous les éléments boisés d'un paysage ne sont pas nécessairement assez grands ou de conformation assez compacte pour contenir une zone de lisière, nous avons adopté la démarche proposée par GUIDOS pour distinguer dans un paysage plusieurs types de couvert boisé selon leur configuration spatiale. Cette analyse est appelée MPSA⁶ (pour Morphological Pattern Spatial Analysis) et fait intervenir différents filtres morphologiques à paramétrer (traitement des formes) (Ostapowicz *et al.*, 2008 ; Soille *et al.*, 2009). Ils permettent d'identifier des zones de lisières qui correspondent aux bordures des "taches de coeur", qui sont distinctes d'autres taches, plus petites ou moins compactes telles que les haies isolées ou les branches (Vogt *et al.*, 2007). En adoptant cette méthode, Cartolis sera directement compatible avec GUIDOS et s'appuie ainsi sur un outil de référence pour définir les bordures correspondant aux lisières forestières.

⁶ MPSA est une séquence personnalisée d'opérateurs mathématiques morphologiques ciblés sur la description de la géométrie et la connectivité des éléments de l'image. Cette méthode peut être appliquée à n'importe quelle échelle et à tous types d'images numériques dans tous les domaines d'application (Vogt, 2008).

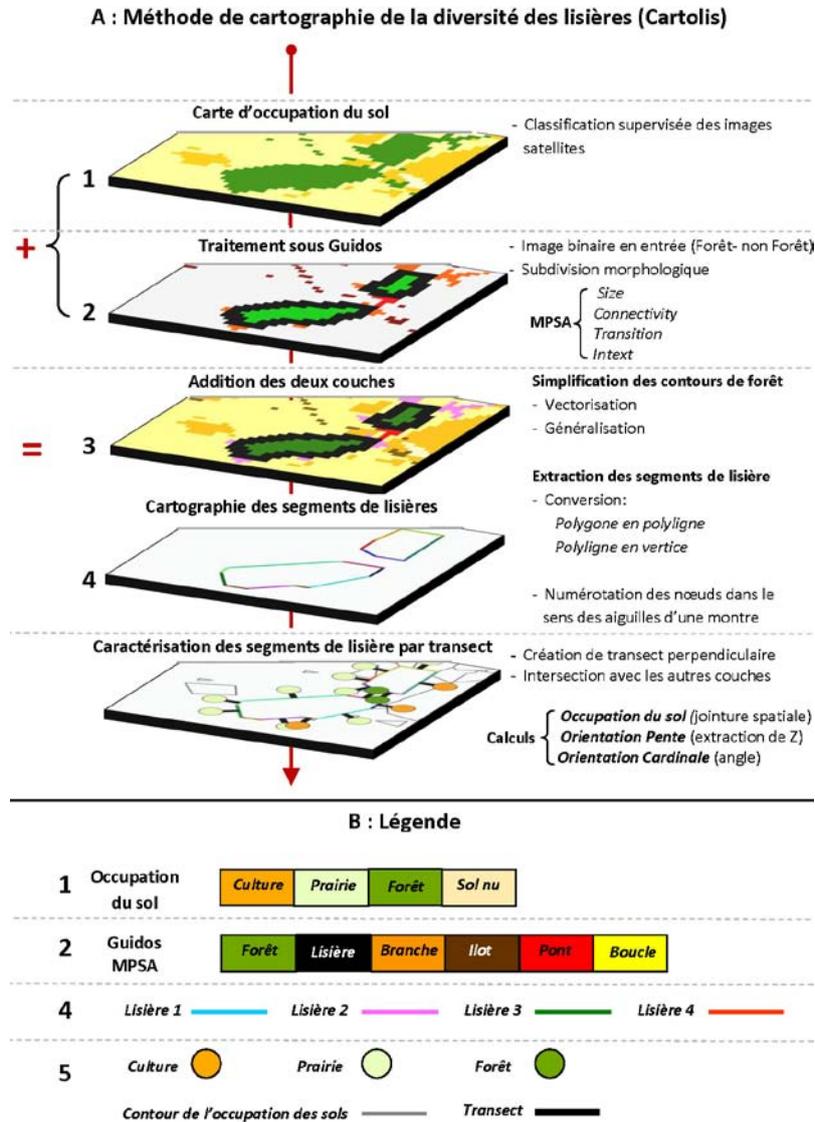


Figure 1 : Etapes de la méthode Cartolis développée en utilisant les fonctionnalités par défaut dans ArcGis et d'autres outils additionnels (extensions et scripts). En partie A, nous décrivons les étapes les plus importantes de la méthode **Cartolis** sur un extrait d'image comportant 2 bois. En partie B, la légende illustre les symboles utilisés.

Dans un premier temps (Figure 1, étape 1 et Tableau 1), il est nécessaire de posséder, en donnée d'entrée, une carte de l'occupation des sols. Dans notre cas, elle est issue d'une classification supervisée d'images SPOT 5 à la résolution de 10 m. Pour des raisons de temps de calcul, nous avons décidé d'extraire une zone de 193 km² de la carte d'entrée. A partir de cette carte, nous conservons uniquement les classes d'éléments boisés pour obtenir une image binaire bois/non bois (Figure 1, étape 2). Cette image est importée sous GUIDOS et traitée à l'aide de la fonction MPSA paramétrée au préalable. En sortie, nous obtenons une nouvelle classification des différents éléments boisés, constituée de sept classes (cœur de bois, lisière, trouées, branches, etc.). Nous avons paramétré la profondeur de lisière à 20 m (EdgeWidth=2) conformément à la littérature scientifique (Gonzalez, 2006).

Le reste de la méthode s'effectue sous ArcGis. L'image raster de sortie de GUIDOS comportant les différentes classes d'éléments boisés est combinée à la carte d'occupation du sol (Figure 1, étape 3). La combinaison des deux données assure le maintien de la topologie des entités boisées et agricoles, règle importante pour la suite des opérations.

Phase 2: identifier les segments de lisières (étape 4 de la Figure 1)

La carte combinée en mode raster est convertie en vecteur par la méthode proposée en standard par ArcGis. A ce stade, les lisières forment des polygones entourant les coeurs de bois, adjacents à d'autres polygones correspondant aux autres occupations du sol, boisées ou non. L'étape suivante de généralisation est une étape fondamentale dans notre démarche qui aura des impacts directs sur la suite du déroulement de la méthode. En effet, nous partons du principe que les bois ont des formes géométriques polygonales relativement simples. Elle consiste donc à simplifier la forme des polygones en réduisant leur nombre de sommets tout en conservant leur topologie, par le biais de l'algorithme de Douglas-Peucker. Un paramètre de tolérance (T) doit être défini, il représente la distance maximale entre la polygone généralisée et celle d'origine. Nous avons opté pour T = 20 m (valeur identique à la profondeur des lisières). Il est important d'effectuer cette généralisation sur l'ensemble du couvert afin de conserver la topologie des bois relativement aux autres occupations du sol, en vue de requêtes spatiales de voisinage⁷. Cette étape fait intervenir l'outil ETGeoWizards qui intègre une fonction « Generalize Polygons ».

Dans l'étape suivante, les bois ayant des lisières (tache de coeur + lisière) sont extraits afin de faciliter la manipulation des fichiers. Les lisières et coeur sont fusionnés et les trouées sont systématiquement rebouchés afin de ne pas induire de lisières intérieures. Les contours de polygones de bois sont convertis en polygones puis en segments de lisière (Figure 1, étape 4). Une attention particulière est portée à la numérotation explicite des lisières, l'objectif étant de retrouver facilement une lisière à l'aide de son identifiant unique et de localiser le bois auquel elle appartient. Cette opération s'est avérée compliquée à mettre en œuvre dans ArcGis, du fait du passage des outils de généralisation de l'image entière qui ne conservent pas une numérotation intra-bois logique. Aussi, nous avons opté pour une procédure de renumérotation, effectuée sous le tableur Excel qui utilise les coordonnées XY des points de départ et d'arrivée des segments.

⁷Cette étape de généralisation est parfois longue à réaliser avec un PC de bureau et sur l'image de 193 km², mais cette durée dépend du nombre de polygones concernés.

Étapes	Processus	Paramètres	Valeurs	
1	Classification	Résolution des images SPOT 5	10 m	
2	Traitement sous Guidos : analyse morphologique MPSA	Foreground Connectivity , ensemble de 3 x 3 pixels. Le pixel central est connecté aux pixels adjacents.	8 pixels	
		Edge Width , définit la largeur ou l'épaisseur en pixel des classes <i>non-cœur</i> . La distance réelle en mètre correspond au nombre de pixels du bord multiplié par la résolution en pixel de l'image.	2 pixels	
		Transition , pixels situés sur une lisière ou une perforation où le cœur est intersecté avec une boucle ou un pont. Si la transition est mise à 0 alors les bords des cœurs seront fermés.	0	
		Intext , permet de distinguer les caractéristiques internes de l'extérieur, où les éléments internes sont définis comme étant délimités par une perforation.	0	
3	Combinaison	Pas de paramètres		
4	Généralisation	Generalize tolerance , réduit le nombre de sommets nécessaires pour représenter un polygone, en utilisant l'algorithme de Douglas-Peucker.	20 m	
5	Création du transect 90°		Script python, qui crée des lignes perpendiculaires. Il utilise une couche vecteur ligne et génère des lignes perpendiculaires à chaque enregistrement avec la longueur indiquée .	40 m
	Calculs	Occupation du sol	Intersection du point situé à l' extrémité du transect à l'occupation du sol.	40 m
		Orientation pente	Les caractéristiques de Z sont obtenues à l'aide du Modèle Numérique de Terrain d'une résolution donnée.	25 m
			Comparer le segment de lisière et le transect avec la même longueur .	40 m
			Résultats : - Dans la pente avec tolérance - Bas de pente - Haut de pente	1 m
Orientation cardinale	Calcul de l'angle de la polyligne en degré.	360 degrés		

Tableau 1 : Description des paramètres utilisés dans les différentes étapes du processus de Cartolis. Les numéros d'étape sont détaillés dans la Figure 1.

Phase 3 : caractériser des segments de lisières (étape 5 de la Figure 1)

Les segments de lisières présentent des caractéristiques propres qui peuvent être obtenues directement, comme la longueur, l'orientation, l'appartenance à un bois donné, la localisation. Conjointement, d'autres caractéristiques peuvent être extraites à partir d'autres couches d'information ou du reste de l'information contenue dans la carte initiale d'occupation du sol. Une fois les segments de lisières détectés, les métriques paysagères de chacune des lisières choisies par les écologues, comme étant majeures pour différencier écologiquement les lisières, sont calculées. Plus précisément, trois variables ont été récupérées dans la table attributaire de la couche vecteur de segments de lisières: l'Orientation Cardinale (OC), l'Orientation par rapport à la Pente (OP), l'Occupation du Sol voisine (OS) (détaillées ultérieurement) (Figure 2).

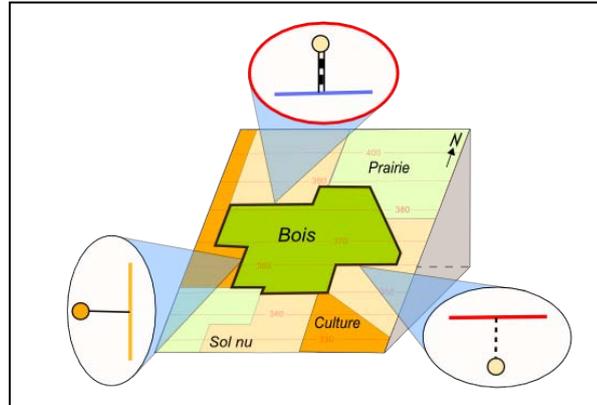


Figure 2 : Bloc diagramme 3D illustrant les trois métriques calculées des segments de lisières à savoir l'orientation, la position dans la pente et la nature du milieu voisin. Les lignes de niveaux aident à interpréter la position dans la pente. Pour la sémiologie, se reporter aux Figures 3 et 4. Par exemple, la lisière encadrée de rouge est une lisière Nord, orientée haut de pente, au voisinage d'un sol nu.

L'orientation cardinale (OC) de la lisière correspond à son exposition en fonction de sa position par rapport au bois auquel elle appartient. Elle est obtenue à partir de l'azimut⁸ du segment. L'OC peut faire l'objet d'une discrétisation pour remplacer les plages de valeurs continues par des classes comme: Nord, Nord Est, Sud, Sud Est, Ouest, etc. comme l'illustre la Figure 4. Nous avons modifié un script de l'extension EasyCalculate pour obtenir ce résultat directement.

La position par rapport à la pente (OP) fait intervenir le Modèle Numérique de Terrain de résolution 25 m⁹. Pour connaître la position des segments de lisières par rapport à la pente, nous choisissons d'utiliser des transects (segments extérieurs au bois, perpendiculaires aux lisières et tenant compte de la pente lorsqu'on se dirige vers l'extérieur du bois) de longueur fixe (40 m, afin qu'ils puissent rencontrer au moins deux cellules du MNT). Ces transects sont réalisés par rotation des segments de lisières à l'aide d'un script en langage Python dans lequel on spécifie la valeur de l'angle (90°) mais également la longueur du transect. Les valeurs d'altitude Z du MNT sur les lisières et les transects sont récupérées en utilisant l'outil ETGeoWizards. Par comparaison de ces valeurs entre lisière et transect, nous identifions si la lisière est orientée dans le sens de la pente, vers le bas de pente ou vers le haut de pente (Figure 2).

Plusieurs possibilités ont été envisagées pour récupérer l'information de l'occupation du sol (OS) voisine de chaque segment. Nous avons choisi d'utiliser les transects réalisés précédemment par souci d'homogénéité de la méthode. L'occupation du sol a été obtenue par jointure spatiale en intersectant le point situé à l'extrémité du transect, c'est-à-dire à 40 m du centre du segment de lisière, avec la couche d'occupation du sol. Le seuil de 40 m correspond au double de la hauteur moyenne d'un houppier d'arbre présent en lisière. Avec cette valeur, nous tenons compte du masque induit par la projection des houppiers sur le polygone adjacent au bois étudié

⁸ L'azimut est l'angle horizontal entre la direction d'un objet et une direction de référence, ici le nord géographique.

⁹ Source : IGN, BD Topo

2.4. Recherche d'une sémiologie graphique personnalisée

Nous avons réfléchi à la définition d'une sémiologie graphique adaptée pour représenter les lisières sur une carte. Il s'agit d'élaborer des règles permettant l'utilisation d'un système graphique de signes assurant une bonne transmission de l'information (Bertin, 1967). L'objectif est que la solution graphique choisie soit efficace pour assurer la correspondance entre les propriétés des données et les propriétés de la variable visuelle qui les représentera. Les contraintes que nous avons à prendre en compte sont les suivantes :

- il faut cartographier plusieurs variables pour chaque segment;
- ces variables sont qualitatives (OP et OS) ou quantitatives (OC);
- la variable visuelle doit différencier les individus sans introduire de hiérarchie entre les descripteurs utilisés.

La grille sémiologique choisie « CartolisSymbol » est basée sur des variations de forme car elle est différenciatrice des variables calculées et elle permet de transcrire des variables qualitatives. Mais il faut que le nombre de formes employées soit limité (5-7 max) et que les formes retenues offrent une forte capacité de séparation (Figure 3).

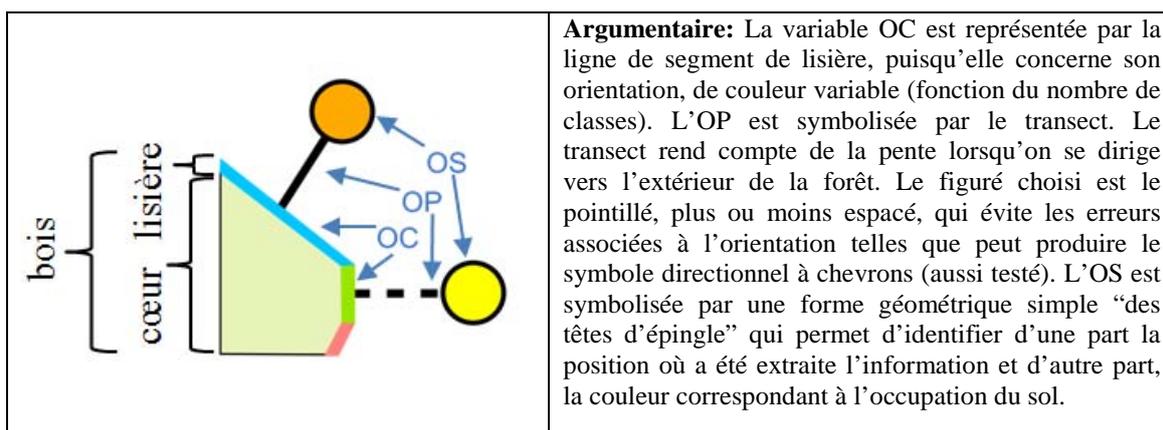


Figure 3 : Explication de la grille sémiologique « CartolisSymbol ».

3. Résultats

3.1. Du point de vue géomatique

Les résultats se présentent d'une part sous forme cartographique synthétique (Figure 4), et d'autre part sous forme d'un tableau des caractéristiques des segments de lisières qui constitue une synthèse statistique descriptive de leur diversité et variabilité (Tableau 2). Ce dernier est soumis à des analyses statistiques sous R (R Development Core Team, 2008) afin d'identifier des patrons particuliers en lien avec les descripteurs de lisières, par des analyses inférentielles et des corrélations non-paramétriques.

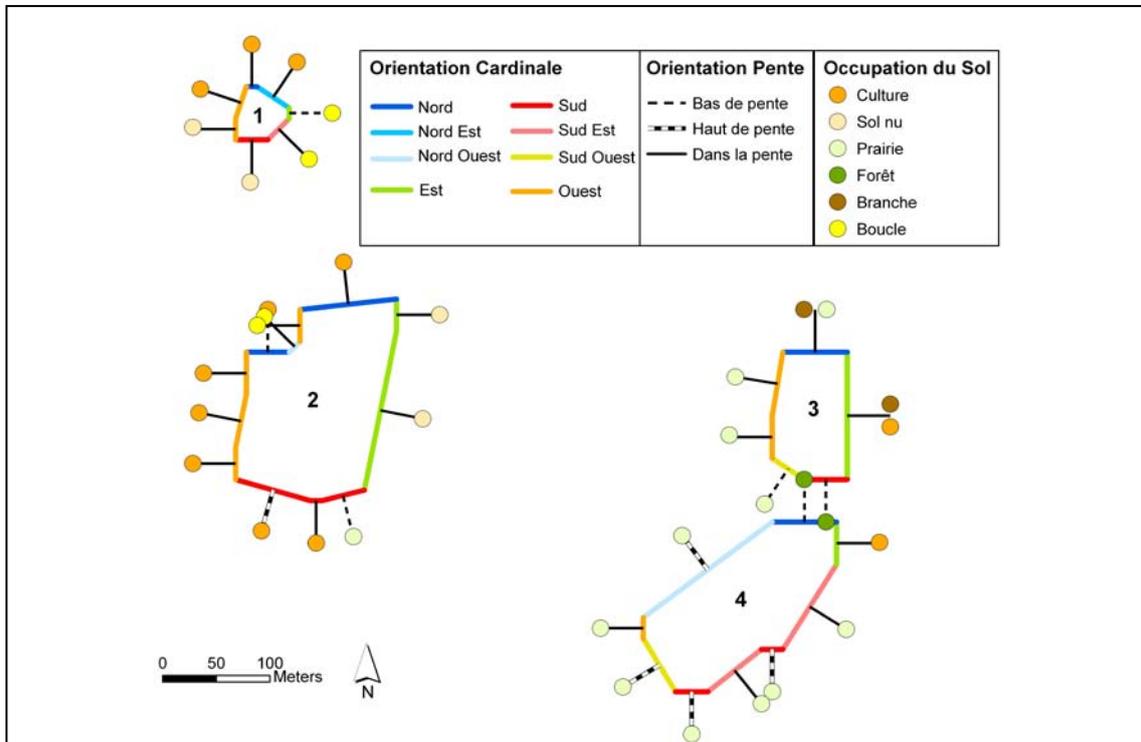


Figure 4 : Carte synthétique des trois variables caractéristiques des segments de lisières, calculées pour une zone test composée de quatre bois et 34 segments de lisière après l'application de la sémiologie CartolisSymbol. L'OC est discrétisée en huit classes, l'OP est représentée selon les trois classes possibles et l'OS est représentée selon un gradient d'extensification croissant (des cultures jusqu'aux éléments boisés). Les branches et boucles sont deux milieux arborés interprétés par GUIDOS. On note qu'un segment de lisière peut être voisin de deux occupations du sol (exemple de la lisière nord du bois n°3).

Etapes	Processus	Statistiques	
1	Classification	Bois = 13,38	193 km ² répartie en %
		Culture = 34,72	
		Prairie = 32,24	
		Sol nu = 10,3	
		Autre = 0,63	
2	Traitement sous Guidos : analyse morphologique MPSA	Cœur = 8,77	
		Lisière = 4,61	
		Eléments boisés = 8,74	
4	Vectorisation	109 bois	
	Généralisation	11 194 lisières	
5	Calcul des variables	101 bois	
		4921 lisières	
		Cf. Figure 5	

Tableau 2 : Statistiques générales sur les entités manipulées au cours des étapes de Cartolis. Les numéros d'étape sont détaillés dans la Figure 1.

A partir d'une carte d'occupation des sols couvrant 193 km², nous obtenons la cartographie et la caractérisation de 4921 lisières contenues dans 101 bois de superficie supérieure à 3 hectares (Tableau 2).

Un atout majeur du projet SIG construit ici est sa capacité de navigation dans la base de données obtenue (variables calculées) et son interrogation à la fois dans sa dimension spatiale (carte) et sa dimension attributaire (table). On peut donc établir des requêtes attributaires, fondées sur des relations logiques et des requêtes spatiales ou sur une combinaison des deux pour sélectionner des objets ayant une caractéristique particulière. Un certain nombre de requêtes ont été définies dont par exemple, une requête attributaire étendue basée sur une analyse multi-critères : il y a 72 lisières (sur 4921) situées au sud, en haut de pente et avec comme voisinage principal une culture.

3.2 Du point de vue écologique

La longueur totale de lisières sur la zone d'étude est de 197 km ce qui correspond à une longueur de lisière de 1km/km² environ. On rencontre majoritairement des lisières orientées au Sud, dans le sens de la pente et adjacentes à des prairies (Figure 5). La longueur d'un segment de lisière est de 40 ± 33 m en moyenne. Il n'y a pas de corrélation significative entre la superficie des bois et la longueur moyenne des lisières (test de corrélation de Spearman $\rho=-0,04$; $p=0,67$).

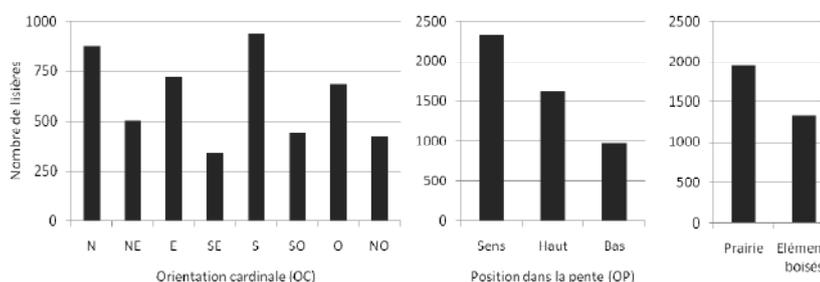


Figure 5 : Nombre de lisières observées en fonction de leurs attributs (orientation, position dans la pente et nature du milieu voisin) sur la zone d'étude de 193 km².

La longueur moyenne de lisière varie significativement en fonction de l'orientation (test de Kruskal Wallis $W=25,62$; $df=7$; $p<0,01$), la topographie ($W=203,39$; $df=3$; $p<0,01$) et la position dans la pente voisin ($W=123,09$; $df=5$; $p<0,01$). Ainsi, une lisière orientée au Sud mesure en moyenne 43 ± 35 m alors qu'une lisière orientée à l'Est mesure en moyenne 35 ± 28 m. Une lisière orientée en sens de pente mesure 46 ± 38 m contre 32 ± 27 m pour une lisière orientée en bas de pente. A titre d'exemple, le bois ID_27 a une superficie de 5,2 ha. Il est composé de 17 segments (soit 991 m) de lisières dont 14 (soit 755 m) sont adjacents à une prairie et 3 (soit 236 m) à une culture. La longueur moyenne de ses segments est de 58 m. Il possède 256 m de lisière orientée au Sud contre 141 m de lisière orientée au Nord.

4. Discussion

La méthode Cartolis répond aux objectifs initiaux et son application sur une zone de large étendue confirme son opérabilité. La méthode combine des outils pré-existants applicables sous le logiciel ArcGis. L'originalité de ce travail réside dans l'enchaînement d'opérations liées sous forme d'une séquence de travail où la compréhension et la maîtrise de chaque étape est essentielle. Avec cette méthode, le concept de segment de lisière devient opérationnel et offre la possibilité d'analyser les paysages selon de nouvelles métriques qui mettent en avant les fonctions d'interfaces entre milieux. Elle produit des résultats cartographiques via l'élaboration d'une grille sémiologique adaptée et tabulaire qui n'étaient pas disponibles aisément jusqu'à maintenant. La méthode Cartolis est par conséquent une méthode d'analyse spatiale en 2D novatrice et efficace pour caractériser la diversité des lisières dans le paysage.

Cependant, la mise en oeuvre complète de la méthode s'est révélée plus compliquée que prévue, nous obligeant pour chaque étape à trouver des solutions cohérentes. La principale difficulté provient du traitement chronologique des trois phases sur lesquelles un travail de calibrage a été réalisé. Une analyse de sensibilité des différents paramètres et l'utilisation de scripts libres de droit contribuent à renforcer la maîtrise des outils employés. Il est important d'être vigilant lors de la phase 2, en particulier lors de la généralisation qui simplifie la forme des bois pour obtenir les segments de lisières recherchés. L'étape de généralisation est l'étape qui prend le plus de temps de calcul et son paramétrage est donc à effectuer en fonction des besoins. Les étapes 1 et 3 sont modulables en fonction des données d'entrée utilisée et du nombre de variables calculées. L'analyse de sensibilité s'insère dans une démarche de construction d'un cadre d'étude rigoureux et extrapolable à des échelles plus larges et à d'autres sites.

L'optimisation de la méthode est envisageable avec l'utilisation d'outils plus intégrés comme COGO d'ET GeoWizards. En mode vecteur, la morphologie des objets peut induire des situations diverses et complexes dont il faut tenir compte. Ainsi, les taches en tores (ou trouées) constituent un cas particulier que nous avons choisi de ne pas prendre en compte en opérant une fermeture des trouées, mais d'autres situations particulières peuvent se présenter. D'autres difficultés sont d'ordre attributaire comme par exemple l'héritage de la numérotation entre les différentes primitives graphiques: du polygone jusqu'au segment de lignes. Cela a été résolu en utilisant des transformations dans un logiciel de calcul externe. A terme, ces transformations devront être directement prises en compte dans la méthode. Enfin, il est prévu d'automatiser la méthode et de proposer une application indépendante et complémentaire de l'outil de caractérisation de la fragmentation forestière, GUIDOS.

Cartolis, par un diagnostic paysager sur de larges emprises géographiques, fournit aux gestionnaires des espaces ruraux de nouveaux éléments d'analyse permettant de rendre compte du devenir des territoires et des usages qui en sont faits. Notre étude montre qu'il est possible d'extraire un certain nombre de données sur les lisières à partir d'image satellitale. Avec cette méthode, nous sommes en mesure de fournir aux gestionnaires des indications sur les attributs des lisières au sens de Strayer et al. (2003) mais également des informations sur le contexte local ou régional (Harper et al. 2005) ou encore sur la fragmentation (McGarigal et al. 2009). Dans un contexte de changement global, la méthode contribue efficacement à la détection de règles de comportement spatio-temporel des lisières pouvant servir de base à l'élaboration de scénarios d'évolution des paysages.

5. Conclusion et perspectives

Aujourd'hui, après plusieurs tests et analyse de sensibilité effectués, nous sommes en mesure de proposer une méthode qui combine plusieurs outils pour traiter des données spatiales et caractériser de façon originale les paysages, en complément d'outils standards comme GUIDOS. Le nombre de paramètres de l'outil a été réduit autant que possible et des règles explicites ont été définies pour choisir leurs valeurs afin de rendre la méthode la plus transparente et explicite pour les utilisateurs.

La méthode proposée permet d'identifier et caractériser des segments de lisières au moyen d'une sémiologie adaptée. Même si la méthode fait appel à de nombreuses étapes qui n'étaient pas prévues initialement, elle a montré son efficacité lors de la confrontation des résultats avec le terrain. La méthode n'est pas tributaire d'un type de données d'entrée. De plus, son découpage en trois phases chronologiques cohérentes nous semble primordial pour moduler cette maquette en fonction de nouvelles données ou de nouvelles variables à calculer en fonction des besoins. Ainsi, Cartolis est généralisable à n'importe quelle zone d'étude sans contrainte d'étendue spatiale (si ce n'est la contrainte associée au temps de calcul). Il est prévu d'appliquer cette méthode sur la couche forestière européenne de 2006 (ForestMap, Joint Research Center) combinée à la carte d'occupation des sols 2006 (Corine Land Cover, IFEN), de résolution identique et toutes deux libres de droit. Ceci nous permettra de tester Cartolis dans un paysage du Nord de l'Europe par exemple et de comparer les tendances observées sur la configuration spatiale des bois et la diversité de leurs lisières.

Remerciements : Ce travail a reçu le soutien financier du département de recherches Sciences pour l'Action et le Développement (SAD) de l'INRA. Il s'insère dans le cadre de la thèse d'Audrey Alignier ayant reçu une bourse du ministère de la Recherche et de l'Enseignement Supérieur.

6. Bibliographie

- Alignier A., Deconchat M., Variability of forest edge effect on vegetation implies to reconsider its assumed hypothetical pattern, *Applied Vegetation Science*, 2010 (in press).
- Bertin J., Sémiologie Graphique. Les diagrammes, les réseaux, les cartes, Paris, La Haye, Mouton, Gauthier-Villars, 1967. 2e édition : 1973, 3e édition : 1999, EHESS, Paris.
- Bolstad P., GIS Fundamentals, A First Text on Geographic Information Systems, 3rd Edition from Atlas Books, 2008.
- Cadenasso M.L., Pickett S.T.A., Weathers K.C., Jones C.G., A framework for a theory of ecological boundaries, *Bioscience*, vol. 53, n°8, 2003, p. 750-758.
- Esseen P.A., Jansson K.U., Nilsson, M., Forest edge quantification by line intersect sampling in aerial photographs, *Forest Ecology and Management*, vol. 230, n°1-3, 2006, p. 32-42.
- Fagan W.F., Fortin M.J., Soykan C., Integrating edge detection and dynamic modeling in quantitative analyses of ecological boundaries, *Bioscience*, vol. 53, n°8, 2003, p. 730-738.

- Fortin M.J., Olson R.J., Ferson S., Iverson L., Hunsaker C., Edwards G., Levine D., Butera K., Klemas V., . Issues related to the detection of boundaries, *Landscape Ecology*, vol.15, n°5, 2000, p. 453-466.
- Gonzalez M., Diversité des plantes ligneuses de fragments forestiers dans les Coteaux de Gascogne: Importance des facteurs locaux et du contexte paysager, Thèse de doctorat, Université Toulouse III, 2006.
- Harper K.A., Macdonald S.E., Burton P.J., Chen J., Brosfokske K.D., Sauders S.C., Euskirchen E.S., Roberts D., Jaiteh M.S., Esseen P.A., Edge influence on forest structure and composition in fragmented landscapes, *Conservation Biology*, vol. 19, n° 3, 2005, p. 768-782.
- McGarigal K., Cushman S.A. Neel M.C., and Ene E., FRAGSTATS: Spatial Pattern Analysis Program for Categorical Maps. Computer software program produced by the authors at the University of Massachusetts, Amherst, 2002. Available at the following web site: www.umass.edu/landeco/research/fragstats/fragstats.html
- McGarigal K., Tagil S., Cushman, S., Surface metrics: an alternative to patch metrics for the quantification of landscape structure, *Landscape Ecology*, vol. 24, 2009, p. 433-450.
- Murcia C., 1995. Edge effects in fragmented forests - implications for conservation, *Trends in Ecology & Evolution*, vol. 10, n°2, 1995, p. 58-62.
- Ostapowicz K., Vogt P., Riitters, K.H., Kozak J., Estreguil, C., Impact of scale on morphological spatial pattern of forest, *Landscape Ecology*, vol. 23, n°9, 2008, p.1107–1117.
- R Development Core Team, R: A language and environment for statistical computing, Vienna, Austria, 2008.
- Soille P., Vogt P., Morphological segmentation of binary patterns, *Pattern Recognition Letters*, vol. 30, 2009, p. 456-459.
- Strayer D.L., Power M.E., Fagan W.F., Pickett S.T.A., Belnap J., A classification of ecological boundaries, *Bioscience*, vol. 53, n°8, 2003, p. 723-729.
- Taylor R.S., Oldland J.M., Clarke M.F., Edge geometry influences patch-level habitat use by an edge specialist in south-eastern Australia, *Landscape Ecology*, vol.23, n°4, 2008, p. 377-389.
- Vogt P., Riitters K.H., Estreguil C., Kozak J., Wade T.G., Mapping spatial patterns with morphological image processing, *Landscape Ecology*, vol. 22, n°2, 2007, p.171-177.
- Vogt P., MSPA GUIDE Release: Version 1.2, European Commission, Joint Research Centre (JRC), December 2008, 6p.
- Zeng H., Wu X.B., Utilities of edge-based metrics for studying landscape fragmentation, *Computers, Environment and Urban Systems*, vol. 29, n°2, 2005, p.159-178.
- Zheng D., Chen J., Edge effects in fragmented landscapes: a generic model for delineating area of edge influences (D-AEI), *Ecological Modelling*, vol.132, n°3, 2000, p.175–190.

Presence-only modelling for indicator species distribution: Biodiversity monitoring in the French Alps

Mathilde Redon ⁽¹⁾ & Sandra Luque ⁽¹⁾

(1) Cemagref, Institute for Agricultural and Environmental Engineering Research, Grenoble, France. mathilde.redon@cemagref.fr

ABSTRACT

The problem in biodiversity monitoring and conservation is that usually exist vast gaps in available information on the spatial distribution of biodiversity that poses a major challenge for the development of biodiversity indicators and regional conservation planning. Within this context, models that establish relationships between environmental variables and species occurrence have been developed to predict species distribution over large areas. We present an example using two indicator bird species, Tengmalm owl (*Aegolius funereus*) and Pygmy owl (*Glaucidium passerinum*). Maximum entropy (Maxent), a presence-only modelling approach, is used to model the distribution of these two species within a large study area in the French Alps. Despite biased sampling design, this method performs very well in predicting spatial distribution of the two owl species and brings useful information to help decision-making concerning the protection of valuable habitats.

RESUME

*Le problème actuel dans les domaines de la gestion et de la conservation de la biodiversité est que les informations disponibles sur la distribution spatiale de la biodiversité sont souvent incomplètes. Dans ce contexte, des modèles qui établissent des relations entre des variables environnementales et des observations d'espèces ont été développés pour prédire la distribution des espèces à grande échelle. Nous présentons un exemple basé sur deux espèces emblématiques : la chouette de Tengmalm (*Aegolius funereus*) et la chevêchette d'Europe (*Glaucidium passerinum*). Le Maximum d'entropie (Maxent), une méthode de modélisation basée uniquement sur des données de présence, a été utilisée pour modéliser la distribution de ces deux espèces de chouettes sur une grande zone d'étude dans les Alpes françaises. Malgré un échantillonnage biaisé, cette méthode donne un très bon résultat pour la répartition potentielle des deux espèces et apporte des informations très utiles pour aider à la prise de décision concernant la protection des habitats à forte valeur écologique.*

KEY WORDS

Maxent, distribution modelling, conservation planning, biodiversity indicators, cavity-nesting owls, French Alps

MOTS CLES

Maxent, modélisation de la répartition d'espèces, planification de la conservation, indicateurs de biodiversité, chouettes cavernicoles, Alpes françaises

Introduction

Improving knowledge on the distribution of indicator and emblematic but locally poorly known species is of great importance for managers as well as for naturalists (Baldwin, 2009). These species can be used as a surrogate for biodiversity monitoring and conservation (Lindenmayer *et al.*, 2000). Still vast gaps in available information on the spatial distribution of biodiversity exist, that poses a major challenge for the development of relevant biodiversity indicators for regional conservation and forest management planning.

In addition, the development of spatial knowledge on the habitat requirements and ecology of these species would facilitate conservation of a great number of related species.

Models that establish relationships between environmental variables and species occurrence have been developed and are widely used with many applications in conservation and management-related fields (Cowley *et al.*, 2000; Elith *et al.*, 2006; Gibson *et al.*, 2004; Pearce and Ferrier, 2000; Stockwell and Peterson, 2002). They can help to guide additional field work, by identifying unknown population locations. It also supports management decisions with regard to biodiversity, to determine suitable sites for reintroductions or to assist selection of protected areas (Baldwin, 2009). First, these models were mainly developed for presence-absence data modelling. However, absence data are often lacking or biased and a new generation of models adapted to presence-only data modelling have been proposed (Baldwin, 2009; Hirzel *et al.*, 2002; Phillips *et al.*, 2006). An important number of such methods exist and differ from their data requirements, statistical models used, output formats, performance in diverse situations (Elith *et al.*, 2006; Guisan and Zimmermann, 2000).

Much of them are based on the ecological niche theory (Hirzel and Le Lay, 2008; Phillips *et al.*, 2006). Ecological-niche based models generally define a function that links the fitness of individuals to their environment (Hirzel and Le Lay, 2008). Thus, theoretically, if we know precisely the habitat characteristics of a species, it is possible to rebuild its ecological niche from the environmental variables describing its habitat.

In this study, we have chosen to use the Maximum Entropy modelling approach, which is a relatively recent method developed by Phillips (2006). Maxent is a presence-only modelling approach with a proved good potential to predict wildlife distribution. Despite biased sampling design, this method performs very well in predicting spatial distribution of species data (Elith *et al.*, 2006; Pearson *et al.*, 2007).

Herein we aim to predict the distribution of two owls' species, Tengmalm owl (*Aegolius funereus*) and Pygmy owl (*Glaucidium passerinum*), in the French Alps. These species have specific habitat needs and their presence reflects those of numerous other forest-dwelling species. They are considered as relicts from Ice Age and need quite cold areas (LPO, 2008), which make them good candidates to develop further studies in relation to global climate changes.

In addition, their distributions are poorly known and their protection statuses are not well defined (Loose *et al.*, 2003). It is also important to denote that Pygmy owl populations in the Vercors Mountains area (Alps range) represent the occidental limit of the European range of the species (Anonyme, 2007). It represents also an additional stake to learn more about distribution and habitat structure of this species at the limit of its range.

Requirements and distribution of Tengmalm owl are less well known because this species is nocturnal and discreet (Hakkarainen *et al.*, 2008); therefore census data are difficult to gather. Modelling its potential distribution will allow to improve knowledge on its habitat and ecological needs. Several local surveys efforts took place in order to develop a census of the populations within the “Vercors” region. But these works are limited to very small areas, and a distribution model which covers the entire mountain region would be very useful to help to define adequate surveys efforts for the future while at the same time will provide an overview of the likely distribution of the two species.

1. Material and methods

1.1 Case study area and species occurrence data

This work was conducted within the Vercors’ Natural Regional Park (VNRP), located at the frontier between northern and southern French Alps (Figure 1; <http://parc-du-vercors.fr>). It covers 206 000 hectares with 139 000 hectares of forests. Approximately a half of these forests are Public (State and municipalities forests) and the rest is in the hands of private stakeholders. The main tree species are Silver Fir (*Abies alba*), Norway Spruce (*Picea abies*) and European Beech (*Fagus sylvatica*).

We used Tengmalm owl (*Aegolius funereus*) and Pygmy owl (*Glaucidium passerinum*) point counts data from several surveys conducted in the ‘Hauts Plateaux du Vercors’ Natural Reserve (HPVNR), which is located within the VNRP. The Reserve is mainly composed of three main forest types: i) mixed uneven-aged beech/spruce/fir forests, ii) pure quite sparse even-aged or uneven-aged spruce forests and at high elevation, iii) pure sparse naturally even-aged Mountain Pine forests. The bigger State forest in the Reserve has recently been classified as an Integral Biological Reserve (IBR).

All local surveys have take place in this State forest and mainly in the IBR. The two owls’ species are from the North European boreal forests and the cold sparse spruce Reserve forests look-like their original habitat. Therefore, local people generally thought that the range of the two species is limited to these particular forests in the Vercors.

In this part of the Alps, Pygmy owl depends on cavities carved by the Great spotted woodpecker (*Dendrocopus major*) and Tengmalm owl by the Black Woodpecker (*Dryocopus martius*) for breeding.

These cavity providers favour respectively spruce and beech trees to breed. It implies that the presence of the woodpeckers and their host trees are likely to be important habitat variables for the two owls.

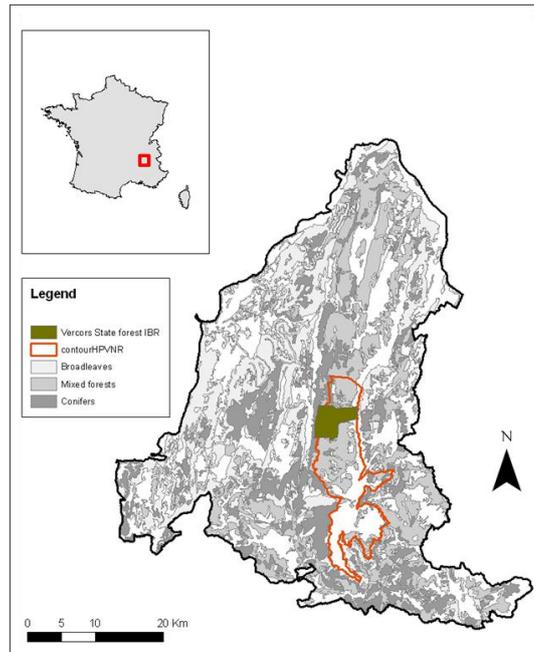


Figure 1. Study area localisation

The point counts data come from the observation network of the National Forest Office and the “Ligue de Protection des Oiseaux” (Bird protection organisation), an NGO which aims to improve knowledge on the local fauna species.

These data are a combination of visual and eared bird contacts in addition to nests locations. Each contact point is located with a Global Positioning System (GPS). The reliability of these data is very heterogeneous because each data source has its own sampling design and its own database system. We therefore harmonize data before integration into a common database.

It is important to denote that despite the low precision of visual and eared occurrence data we include them into our database since these are owl’s activity centres within their territory. Gathering of all available point counts gives 95 points for Pygmy owl and 76 for Tengmalm owl.

The resulted dataset is composed of presence points represented as latitude/longitude coordinates, and then no absence points are considered. This is a common issue when someone works with wildlife surveys data (Anderson *et al.*,

2003; Chefaoui and Lobo, 2008). Therefore the interest of models like Maxent, as aforementioned, is the use of presence data only for the computation of the habitat modelling.

1.2 Modelling algorithm

Maxent was first developed to make predictions or inferences from incomplete information in many different fields (Phillips *et al.*, 2006). Since recently it has been widely used as a general approach for presence-only modelling of species distribution. It estimates the less constrained distribution of training points compare to random background locations with environmental data layers defining constrains (Baldwin, 2009). The results show how well the model fits the location data as compared to a random distribution (Phillips *et al.*, 2006; Phillips *et al.*, 2004).

An increasing number of comparisons with others modelling methods show that Maxent performs better than much of them (Ortega-Huerta and Peterson, 2008; Ward, 2007; Yun-sheng *et al.*, 2007). This performance is due to the numerous advantages of the method. For example, it can be run with very few occurrence data (Hernandez *et al.*, 2006; Pearson *et al.*, 2007), it manages with different kind of environmental data, as continuous and discrete datasets, without any transformation (Ortega-Huerta and Peterson, 2008) and it includes a parameter to avoid over-fitting (Phillips *et al.*, 2006). The strategy to manage with lack of absence data, based on random pseudo-absence, performs better than other ways such as those of ENFA (Hirzel *et al.*, 2002) or BIOCLIM (Busby, 1991) (Wisn and Guisan, 2009). It also shows percentage variable contributions to the final model and response curves for each variable, which allows targeting those which can be deleted to improve model performance and to detect variable correlations (Baldwin, 2009). The output is continuous that allows a great flexibility in threshold choice and results representation. Resulting maps can be very easily exported in a GIS database (Phillips *et al.*, 2006). Outputs are also easy to interpret from an ecological point of view.

Maxent presents also some drawbacks, the main is linked to the exponential model for probability which can give very large predicted values for environmental conditions outside the range present in the study area. The method can have therefore some difficulties to extrapolate to other study area (Peterson *et al.*, 2007; Phillips *et al.*, 2006).

1.3 GIS environmental data and model implementation

We used a set of environmental data based on the knowledge of the species ecology and factors affecting distribution of the species within the entire study area (Array 1).

These data are represented as raster layers with a 50 m resolution, which is those of the most restrictive raster included in the analysis. We used ArcGIS 9.3 to prepare the different data layers.

Implementation of Maxent requires that rasters were perfectly overlapping and have exactly the same number of cells. A single raster mask delimiting the study area was therefore used to assure that all raster layers have the same dimensions.

For the two species, we used 25% randomly selected occurrence data for cross-validation, leaving the remaining 75 % for analysis, as done by (Ward, 2007).

We implemented the model with freeware Maxent developed by (Phillips *et al.*, 2005). It is friendly use, as species occurrence training and test files and environmental data layers are automatically recognize by the application.

We used simultaneously continuous and discrete data. We let almost all default parameters, but we set the regularization value to 1 for the two species. To evaluate the relative contribution of each variable to the model we first chose to see the jackknife test of variable importance. It shows how each variable contributes to the model by testing variation of model gain with and without including each variable separately. We also analyzed the response curves which show the response of each variable to presence probability.

We first include all the environmental variables in the model. We then delete those which did not show any significant contribution to the model.

Six variables were finally selected for Pygmy owl: elevation, topography, land cover, mean annual temperature, forest / non-forest map and presence of Norway Spruce; and five for Tengmalm owl: Land cover, elevation, slope, forest/non-forest map and European Beech presence. Fifty model replicates were run for each species and we select the best among the 50. The contribution of each variable can be visualized through jakknife tests of variable importance and calculation of variables percentage contributions.

GIS layer	Data source	Pygmy owl	Tengmalm owl
Elevation*	French DEM (French National Geographic Institute)	×	×
Aspect		×	×
Slope°		×	×
Topography ⁺		×	×
Forest habitats (Alpine National Botanic Conservatory (ANBC))	Natural habitats map from ANBC	×	×
<i>Dendrocopus major</i> presence	Data from the Ligue de Protection des Oiseaux	×	
<i>Dryocopus martius</i> presence			×
Land cover*	CORINE LAND COVER 2006 (level 3)	×	×
Mean annual temperature ⁺	AURHELY model from Météo France	×	×
Norway Spruce presence ⁺	Database from ANBC	×	
European Beech presence°			
Forest / non forest*	Join Research Center JRC	×	×

Array 1. *Environmental variables used in Maxent modelling, by owl species.*
 (*significant contribution to the two species' models; +significant contribution to Pygmy owl model; °significant contribution to Tengmalm owl model).

Maxent provides three output formats. We select the logistic output as generally recommended. The result is a continuous value between 0 and 100. Each resulting raster pixel contains a value reflecting how well the predictive conditions for each pixel are.

We then export results into ArcGIS 9.3 in order to apply a threshold value to produce the occurrence map. Applying a threshold is the last step of many species modelling approaches. It is necessary to transform the probability map in presence/absence data. Many methods exist to determine the presence threshold. Objective thresholds seem to be more effective than subjective ones (Liu *et al.*, 2005). We used 10 percentile training presence (threshold 0.345 for Pygmy owl and 0.259 for Tengmalm owl) as suggested by (Phillips and Dudík, 2008). This threshold value provides a better ecologically significant result when compared with more restricted thresholds values. Therefore, the use of only one threshold value gives a very narrow overview of the species distribution. In addition, for conservation purpose it is more useful to have a presence gradient which is more realistic and easier to validate with expert knowledge. Hence, we produce output maps with four ranges of presence probability: 0-25% as unsuitable habitat, 25-50% as acceptable habitat, 50-75% as quite suitable and > 75% as suitable habitat.

2. Model evaluation

To evaluate model results, the best method would have been to use an independent data set. However, for the two owl's species, observation data are spatially aggregated and it would have had no sense to use data located in the same place than the training data to evaluate the performance of the model.

We test models performance with several other tools.

Maxent calculates the AUC (Area Under the receiver operating Curve) for each run. It is a standard, threshold-independent method for model evaluation. This method was initially developed for presence-absence data. In Maxent, absence data are replaced by random points (Phillips *et al.*, 2006). AUC tests if a prediction is better than random for any possible presence threshold. It varies between 0.5 when the result is not better than a random selection and 1 when the result is significantly better than random.

Another evaluation method lies in the analysis of the two types of prediction errors provided by all presence/absence models: false negatives (omission error) and false positives (commission error) (Ward, 2007). Maxent algorithm calculates an omission rate for training and test data. Omission rate indicates the percentage of test localities that falls into pixels not predicted as suitable for the species (Phillips *et al.*, 2006). It should be low for a good model performance. With presence-only

modelling, is it very difficult to calculate commission errors as no absence date is available and commission error often increases with omission decreasing (Hernandez *et al.*, 2006). We therefore use different evaluation tools to overcome this difficulty.

AUC and omission rate were nevertheless used to select the best model among fifty replicates for the two species. Models with the lowest training and test omission rates and with the highest AUC were chosen, as generally done by Maxent users. We also verify if all training points were predicted with a high probability.

In addition, expert knowledge was included in the validation process. We compare models results with a mental map drawn by an expert of the two owl species on the Vercors Mountains. We also gather a group of forest managers with good naturalist skills to validate the results.

3. Results and discussion

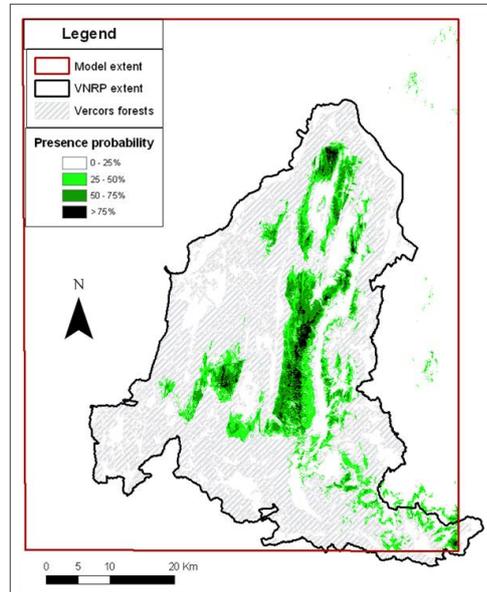
Despite the aggregation patterns of the training data, Maxent modelling method has successfully predicted owl's presence on a large range in the Vercors Mountains (Figure 2). Results show that the approach has a good capability to capture the intrinsic species-habitat relationships and to reconstruct the ecological niche even in sites where any data were available.

Moreover, this method performed very well in predicting potential spatial distribution of the two owl species. Test omission rates are null at minimum training presence threshold for training and test datasets (rate = 0.000 for Pygmy owl and Tengmalm owl) and low at 10 percentile training presence threshold (Tengmalm owl: 0.093 for training data and 0.000 for test data and Pygmy owl: 0.095 for training data and 0.000 for test data).

For the two species, models show an AUC value very close to 1 for test data (0.992 for Tengmalm owl and 0.996 for Pygmy owl), which generally proves good model performance. However, when species have a narrow range (or training data are spatially aggregated), AUC is often overestimated (Phillips and Dudík, 2008), which is certainly the case here. The high AUC values are therefore not sufficient to evaluate model performance in this study.

Mean training data predictive rate is 0.65 (SD = 0.21) for Pygmy owl and 0.63 (SD = 0.20) for Tengmalm owl. The model predictive capacity seems to be good then based on the well predicted calibration points. In addition, for the two species, the resulting presence/absence maps overlap very well with the expert mental map and are in accordance with manager's knowledge.

a)



b)

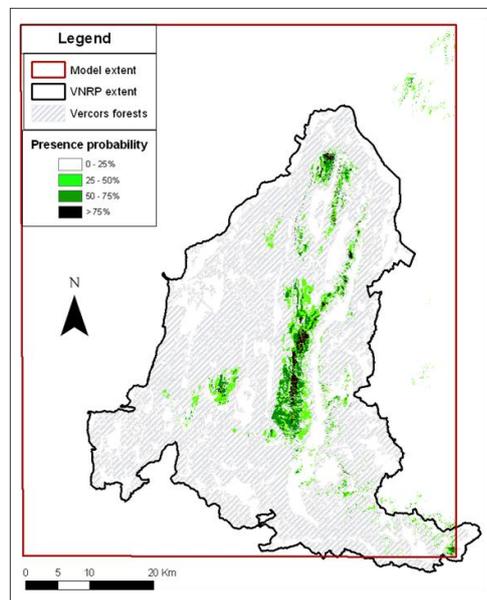


Figure 2 . Results of Maxent model on Vercors Mountains a) Tengmalm owl and b) Pygmy owl, with four ranges of presence probability.

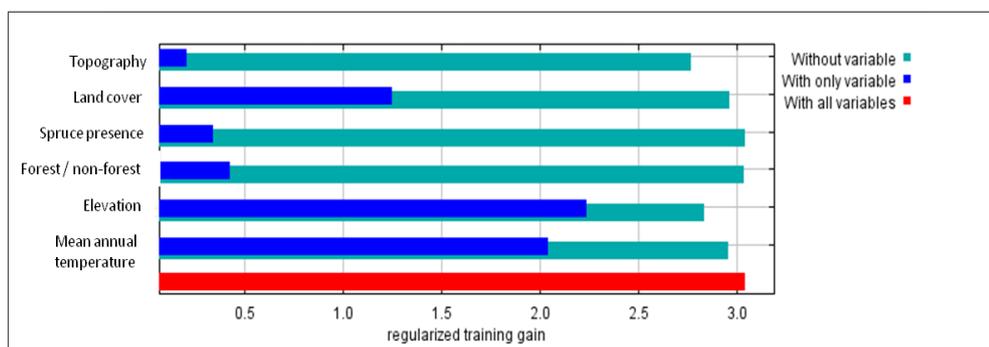
For Pygmy owl, the environmental variable with highest gain when used in isolation is elevation (Figure 3a), which therefore seems to contain the most useful information by itself. It is supported by local expert knowledge and literature where this species is linked with a quite narrow altitudinal range (between 1100 and 2200 meters). Mean annual temperature brings also high gain to the model. These two variables could be considered as very correlated, as temperature generally decreases with elevation, but omission of one of the two variables decreases model performance. This is not the case for Tengmalm owl, where removal of mean annual temperature increases elevation gain as it captures the gain deleted by removal of temperature.

The environmental variable that contains the most information that is not present in the other variable is topography (called “altitopo”), because it decreases the gain the most when is omitted (Figure 3a). As for mean annual temperature, this variable seems to be quite correlated with elevation, but removal of one of them decreases also model performance.

The environmental variable with highest gain for Tengmalm owl is elevation which is also the variable that decreases the most the gain when it is omitted (Figure 3b). Elevation brings therefore the most useful information by itself and contains the most information that is not represented by the other variables.

The linkage of the two owl’s species with elevation can be due to the fact that they are dependent of particular tree species to breed, these trees themselves growing in limited altitudinal ranges. They also need quite cold conditions, that are often linked with high elevation or soil depressions, which are frequent in limestone soils as present in the Vercors Mountains.

a)



b)

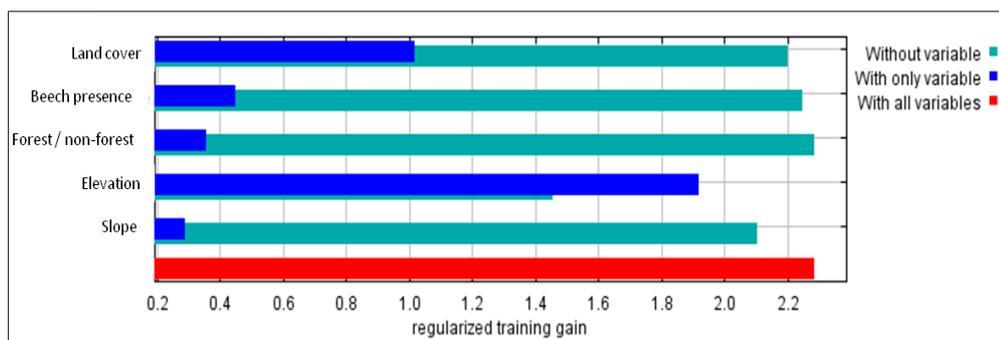


Figure 3. *Jakknife of regularized training gain a) Pygmy owl and b) Tengmalm owl.*

For the two species, the resulting distribution is wider than would be expected by local knowledge (see Figure 2). This result is not surprising because some observations have been done in Vercors forests outside of the HPVNR and in an adjacent mountain area with a different type of forest habitats (i.e. more humid, more productive and with closed canopy conditions).

The distribution maps bring new information on these poorly known species. For example, they are certainly present in some of the most productive Vercors forests, where their conservation will therefore become a new stake for forest management. In addition, the grain and extent of the resulting presence/absence maps allow their use at different scales. They can be used at management units scale (few hectares) to better integrate owl's conservation in forest planning as well at entire forests scale to avoid suitable sites when extension of road network is planned, for example.

However, it is important to note that species could not be present in a site even if they are predicted. Other factors, not taken into account in the analysis, can explain species absence. They can be for example: predator presence (notably Tawny owl

(*Strix aluco*) for the two owls and European pine marten (*Martes martes*) for Tengmalm owl), sites far from existing population and not yet colonized and a lack of prey resources (little mammals, passerine birds, etc.). The integration of a quantitative analysis of landscape heterogeneity, in terms of structure and composition, is underway to better take into account matrix characteristics.

4. Conclusion

Combination of GIS tools with species distribution modelling algorithm shows to have a good potential for species monitoring. In this study GIS was first used as a platform to gather, homogenise and prepare data layers before running the model. These steps are determinant to assure model quality and, in this study, the model algorithm Maxent performed very well in predicting potential spatial distribution of the two owls' species. GIS is then used to facilitate results interpretation and to create maps of presence probability useful for forest managers and naturalists.

This kind of modelling process would be useful to follow the evolution of their spatial distribution in years to come. Furthermore, the tools developed can be applied in assessing biodiversity value of both managed and protected forest areas to help decision-making concerning the protection of valuable habitats. As GIS tools are commonly used in management planning and Maxent algorithm is free and quite simple to implement, this kind of method can be easily adapted and implemented by local managers interested in species monitoring.

Sites of predicted presence would for example guide naturalists' future work in order to identify other suitable areas where the bird distribution is unknown while at the same time facilitate selection of areas with high ecological value. As numerous public forests are managed for wood production in the Vercors, these maps would allow to better integrating biodiversity conservation into management planning. In addition, as these species are linked to cold habitats, they could serve as good indicators of climate change with further work including temporal analysis.

Distribution modelling of these species is among the first attempts to model suitable habitat distribution of cavity-nesting owl species in France. We hope it will launch the use of such methods, which aim to improve species ecological knowledge and facilitate species censuses and conservation.

References

- ANDERSON, R.P., LEW, D. AND PETERSON, A.T., 2003. Evaluating predictive models of species' distributions: criteria for selecting optimal models. *Ecological Modelling*, 162(3): 211-232.
- ANONYME, 2007. La Chevêchette d'Europe, Association Le Petit Entremont, Saint-Pierre d'Entremont
- BALDWIN, R.A., 2009. Use of Maximum Entropy Modeling in Wildlife Research. *Entropy*, 11: 854-866.
- BUSBY, J.R., 1991. BIOCLIM: A bioclimate analysis and prediction system. In: C.R. Margules, Austin, M. P. C. (Editor), Nature Conservation: Cost effective Biological surveys and data analysis. CSIRO, Australia, pp. 64-68.
- CHEFAOUI, R.M. AND LOBO, J.M., 2008. Assessing the effects of pseudo-absences on predictive distribution model performance. *Ecological Modelling*, 210: 478-486.
- COWLEY, M.J.R., WILSON, R.J., LEON-CORTÉS, J.L., GUTIÉRREZ, J.L., BULMAN, C.R. AND THOMAS, C.D., 2000. Habitat-based statistical models for predicting the spatial distribution of butterflies and day-flying moths in a fragmented landscape. *Journal of Applied Ecology*, 37: 60-72.
- ELITH, J. *et al.*, 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 29: 129-151.
- GIBSON, L.A., WILSON, B.A., CAHILL, D.M. AND HILL, J., 2004. Spatial prediction of rufous bristlebird habitat in a coastal heathland: a GIS-based approach. *Journal of Applied Ecology*, 41: 213-223.
- GUISAN, A. AND ZIMMERMANN, N.E., 2000. Predictive habitat distribution models in ecology. *Ecological Modelling*, 135(2-3): 147-186.
- HAKKARAINEN, H., KORPIMÄKI, E., LAAKSONEN, T., NIKULA, A. and SUORSA, P., 2008. Survival of male Tengmalm's owls increases with cover of old forest in their territory. *Oecologia*, 155: 479-486.
- HERNANDEZ, P.A., GRAHAM, C.H., MASTER, L.L. AND ALBERT, D.L., 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography*, 29: 773-785.
- HIRZEL, A.H., HAUSSER, J., CHESSEL, D. AND PERRIN, N., 2002. Ecological-niche Factor Analysis : How to compute habitat-suitability maps without absence data ? *Ecology*, 83(7): 2027-2036.
- HIRZEL, A.H. AND LE LAY, G., 2008. Habitat suitability modelling and niche theory. *Journal of Applied Ecology*, 45: 1372-1381.
- LINDENMAYER, D.B., MARGULES, C.R. AND BOTKIN, D.B., 2000. Indicators of Biodiversity for Ecologically Sustainable Forest Management. *Conservation Biology*, 14(4): 941-950.
- LIU, C., BERRY, P.M., DAWSON, T.P. AND PEARSON, R.G., 2005. Selecting thresholds of occurrence in the prediction of species distributions. *Ecography*, 28: 385-393.

- LOOSE, D., BRUNEAU, G. and ADESTRO, J., 2003. La Chevêchette d'Europe (*Glaucidium passerinum*) sur la Réserve naturelle des Hauts-plateaux du Vercors, Centre Ornithologique Rhône-Alpes Isère, RN Hauts-plateaux Vercors, Grenoble.
- LPO, 2008. Cahier technique Chevêchette et Tengmalm, Mission Rapace.
- ORTEGA-HUERTA, M.A. AND PETERSON, A.T., 2008. Modeling ecological niches and predicting geographic distributions: a test of six presence-only methods. *Revista Mexicana de Biodiversidad*, 79: 205-216.
- PEARCE, J. AND FERRIER, S., 2000. An evaluation of alternative algorithms for fitting species distribution models using logistic regression. *Ecological Modelling*, 128(2-3): 127-147.
- PEARSON, R.G., RAXWORTHY, C.J., NAKAMURA, M. AND PETERSON, A.T., 2007. Predicting species distributions from small numbers of occurrence records : a test case using cryptic geckos in Madagascar. *Journal of Biogeography*, 34: 102-117.
- PETERSON, A.T., PAPES, M. AND EATON, M., 2007. Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. *Ecography*, 30: 550-560.
- PHILLIPS, S.J., ANDERSON, R.P. AND SCHAPIRE, R.E., 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190: 231-259.
- PHILLIPS, S.J., DUDÍK, M. AND SCHAPIRE, R.E., 2004. A Maximum Entropy Approach to Species Distribution Modeling, Proceedings of the 21st International Conference on Machine Learning, Banff, Canada.
- PHILLIPS, S.J. AND DUDÍK, M., 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*, 31: 161-175.
- STOCKWELL, D.R.B. AND PETERSON, A.T., 2002. Effects of sample size on accuracy of species distribution models. *Ecological Modelling*, 148(1): 1-13.
- WARD, D.F., 2007. Modelling the potential geographic distribution of invasive ant species in New Zealand. *Biological Invasions*, 9: 723-735.
- WISZ, M.S. AND GUISAN, A., 2009. Do pseudo-absence selection strategies influence species distribution models and their predictions? An information-theoretic approach based on simulated data. *BMC Ecology*, 9(8): 1-13.
- YUN-SHENG, W., BING-YAN, X., FANG-HAO, W., QI-MING, X. AND LIANG-YING, D., 2007. The Potential Geographic Distribution of *Radopholus similis* in China. *Agricultural Sciences in China*, 6(12): 1444-1449.

Opérations et cohérence pour l'exploration des propriétés structurelles d'un réseau de transport

Éric Mermet
Jean-François Gleyze

Laboratoire COGIT / IGN – Université Paris-Est
73 Avenue de Paris
F-94165 SAINT MANDE CEDEX

eric.mermet@ign.fr
jean-françois.gleyze@ign.fr

RÉSUMÉ. L'analyse des propriétés structurelles d'un réseau de transport permet d'apporter un éclairage nouveau et complémentaire aux analyses de réseau dites thématique. Or, l'analyse structurelle se heurte à des complexités combinatoires, algorithmique et visuelle. Pour cela, nous avons mis au point un modèle théorique permettant de répondre à ces trois complexités. Notre modèle a été mis en oeuvre au travers d'un outil offrant la possibilité d'explorer un réseau à l'aide d'un langage graphique, dans lequel l'utilisateur peut combiner des blocs-cartes et des opérations. Ainsi dans cet article, nous nous intéressons à la définition de ces opérations et aux cohérences qu'elles se doivent d'entretenir entre elles.

ABSTRACT. Analysis of structural properties of a transportation network allows new lighting and complementary to thematics networks analysis. However, structural analysis is hampered by combinatorial, computational and visual complexity. That's why we developed a theoretical model to answer to these three complexities. Our model has been implemented through a tool that provides an opportunity to explore a network using a graphical language, in which the user can combine blocks and cards operations. So in this article, we focus on the definition of these operations and on the consistency they must maintain between them.

MOTS-CLÉS : Réseau de transport, analyse et exploration, propriétés structurelles, complexité

KEYWORDS: Transportation network, analysis and exploration, structural properties, complexity

1. Introduction

L'analyse des réseaux a beaucoup évolué au travers des siècles. De la résolution du problème des sept ponts de Königsberg par Euler dans le premier tiers du 18ème siècle jusqu'au 20ème siècle qui a vu se développer la théorie des graphes par Berge, les réseaux ont beaucoup été analysés d'un point de vue mathématique et cela jusque dans les années 70. On trouve par exemple des études portant sur le gain en accessibilité de villes américaines suite aux différents aménagements autoroutiers durant les années 50 et 60 [MUR 72] ou encore [KAN 63] qui développe des indices structurels tentant de lier géométrie du réseau, emprise spatiale et accessibilité. Ces deux études se basent sur la structure même du réseau. Par la suite et jusqu'aujourd'hui, avec l'avènement de la géographie quantitative et des sciences sociales, les études sont plus anthropisées. L'humain prend sa place comme paramètre du réseau en tant qu'utilisateur. La densification des villes durant les années 70 y joue pour beaucoup. En effet, les moyens de transports ont besoin d'être optimisés en terme de flux, de population, d'horaires, etc., jusqu'au suivi des déplacements par différentes enquêtes ou en Ile-de-France par des cartes électroniques de transport [COU 05]. Il reste cependant des études basées sur la structure des réseaux de transport comme [DUP 91], [CAT 97], [GEN 01] ou encore [GLE 05]. Dans ces derniers travaux, nous définissons que les potentialités relationnelles d'un réseau (c'est à dire en quoi celui-ci, de part sa structure, prédispose à des phénomènes) sont mises en évidence par le biais d'indicateurs relationnels calculés à partir des chemins sur le réseau. Notre travail s'intègre dans ce "mouvement" du tout structurel. Or dans une telle approche, nous rencontrons des difficultés inhérentes aux différentes complexités d'un réseau. Il en existe trois sortes :

- la complexité combinatoire due au grand nombre de chemins existants sur un réseau,
- la complexité des différents algorithmes (calculs de chemins ou d'indicateurs relationnels),
- la complexité dans la mise en valeur visuelle des informations et de leurs appréhensions par l'utilisateur.

Ainsi, nous avons dans un premier temps mis en place un modèle à quatre paramètres (l'espace de relation Origine-Destination (OD), la mesure, la vue et la légende) qui sied à l'analyse des composantes structurelles d'un réseau de transport [MER 09a]. Par la suite et basé sur ce modèle, nous avons construit un outil logiciel d'exploration des composantes structurelles par combinaison de blocs-cartes [MER 09b] [MER 10]. Cet outil nous permet d'agencer les blocs-cartes entre eux dans un langage graphique et de construire une exploration de réseau pas à pas. Il est possible de combiner ces blocs-cartes entre eux par le biais d'opérations sur les quatre paramètres du modèle. Or l'expérience montre qu'il est nécessaire de mettre en place une vérification de la cohérence des opérations possible. Dans ce cadre, cet article s'attache à décrire les opérations possibles selon les différents paramètres et à valider leur cohérence. Nous présenterons donc brièvement dans un premier temps le modèle à quatre paramètres et l'outil d'exploration des propriétés structurelles d'un réseau par combinaison de

blocs-cartes graphiques. Dans un second temps, nous nous attacherons à décrire les opérations du langage graphique selon le paramètre mis en jeu. Enfin, ces dernières opérations seront mises en œuvre dans le cadre de cas d'études.

2. Un langage graphique pour l'exploration des propriétés structurelles d'un réseau de transport

2.1. L'apport de l'analyse structurelle

Il existe actuellement une pléiade d'outils permettant d'étudier un réseau (Arcgis Network Analyst, Network WorkBench) et de le paramétrer selon ses propres données d'entrées (flux, passagers, populations, etc.). Ces logiciels abordent l'analyse de réseau d'un point de vue thématique, *i.e.* les données d'entrées, concrètes et dénombrables, sont attributaires des composantes (nœuds, arcs et chemins) du réseau. Dans une telle approche, le constat est qu'il n'est pas possible de faire la part entre la thématique (exogène au réseau, facteur anthropisé et/ou spatialisé) et le structurel (facteur endogène au réseau). L'approche que nous avons construite se base sur ce postulat : avant d'effectuer une étude thématique, il peut être intéressant de dégager des informations sur la structure du réseau pour en comprendre le fonctionnement.

Un réseau est structuré en un nombre n de nœuds, un nombre m d'arcs et n^2 chemins possibles entre paires de nœuds. L'analyse structurelle (figure 1) repose sur :

- l'implantation des nœuds dans un espace géographique,
- la topologie (connexion des nœuds entre eux par des arêtes),
- la métrique (pondérations et valuations attributaires sur les nœuds et les arcs),
- et enfin les relations Origine-Destination (OD) entre tous les nœuds du réseau.

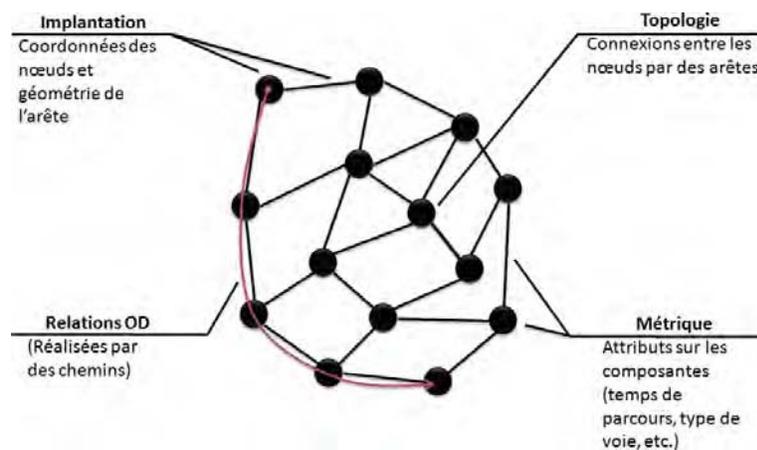


Figure 1. Les quatre composantes structurelles en analyse des réseaux

À partir de ces OD, nous construisons des indicateurs relationnels à partir desquels il est possible d'extraire de l'information par le comptage de chemins passant par une composante ou encore par la longueur de ces chemins. Ainsi, les forces et faiblesses du réseau pourront apparaître plus clairement [GLE 05] sans qu'il ne soit question d'usagers. L'analyse structurelle apporte alors un éclairage nouveau au réseau au préalable d'une analyse classique. En particulier, nous pensons que cette approche est pertinente pour les analyses de vulnérabilité et de résilience.

2.2. Les différentes complexités liées à l'analyse des réseaux

L'analyse visuelle et la géovisualisation demandent de mettre en évidence ce qui est visible et attendu mais aussi d'aider à la découverte de l'inattendu et de ce qui est invisible au premier abord [THO 04]. Dans notre thématique, l'analyse exploratoire des propriétés structurelles des réseaux de transport, ce fait est particulièrement vrai. En effet, notre problématique rejoint celle de la géovisualisation [DYK 05] à savoir :

- une masse de données importante qui proviennent de la combinatoire liée à la nature de graphes souvent denses. Un nombre n important de sommets induit une combinaison quadratique n^2 de relations OD. Les chemins - décrits formellement par une succession ordonnée de nœuds et d'arcs - reliant Origine et Destination deviennent alors volumineux à stocker en mémoire.

- des temps de calcul important émanant d'algorithmes de fortes complexités. À titre d'exemple, calculer tous les plus courts chemins est fait par un algorithme dit All Pairs Shortest Path (APSP) de complexité en $O(n^3)$ [CHA 07]. Ou encore, les complexités des algorithmes d'indicateurs tels que centralité intermédiaire [PAR 72] ou accessibilité moyenne [PIT 79] sont de l'ordre $O(n^2)$.

- des problèmes de visualisation permettant de faire émerger de l'information exploitable. Une fois différents indicateurs calculés, il est bien entendu nécessaire d'en obtenir un rendu visuel par une représentation sur le réseau d'étude. Or, d'une part il existe une infinité de façon de visualiser un phénomène. D'autre part, il faut que la représentation permette à l'utilisateur d'appréhender au mieux le phénomène qu'il étudie.

2.3. Le modèle : un outil de réduction des complexités

Comme bien souvent dans toute résolution de problème et afin d'en diminuer la complexité, il est nécessaire de fragmenter le problème en plusieurs sous problèmes, de les identifier précisément et de trouver des solutions partielles pour en faire une solution globale [REY 04]. Pour répondre précisément aux trois complexités, nous avons construit un modèle théorique. Ce modèle (voir Fig.2), basé sur la définition de carte d'exploration de réseaux, a été mis au point suivant une étude de cas des besoins d'un utilisateur explorant les propriétés structurelles d'un réseau de transport [MER 09a]. Notre modèle est construit autour de quatre paramètres :

– L'**espace** pour la sélection de nœud et donc de chemin d'intérêt au départ de ce nœud. C'est l'espace relationnelle $E [1]$ de la carte d'exploration.

$$E = \{OD_1, OD_2, \dots, OD_k\}$$

$$\text{avec } k = n^2 \text{ si le graphe est dirigé et } OD_n = (n_i, n_j), n_i \neq n_j \quad (1)$$

– La **mesure** permet de valuer quantitativement un comportement du réseaux par une formule mathématique pour laquelle l'espace relationnel est son ensemble de définition.

– La **vue** permet de jouer avec le niveau de granularité visuel par agrégation de sommets et changement d'échelle,

– La **légende** permet de classer les résultats par classes de valeurs et d'y associer des gammes de couleurs afin d'en faire une analyse visuelle plus instinctive.

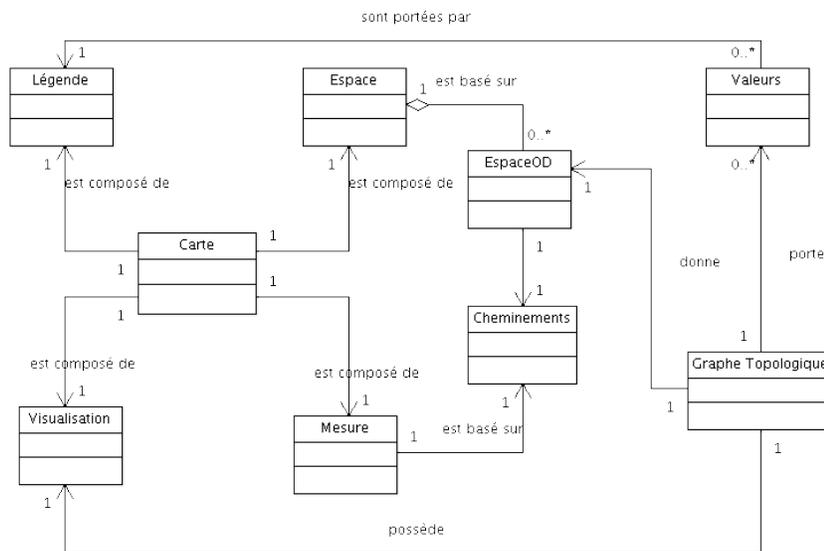


Figure 2. Le modèle UML simplifié regroupant les quatre paramètres autour de la carte d'exploration

Les quatre variables de notre modèle sont modulables selon la volonté de l'utilisateur. Ainsi, il lui est possible par exemple de modifier l'espace de définition, pour n'en sélectionner qu'une portion (*i.e* des chemins d'intérêt). Cette action a pour effet d'une part de diminuer la combinatoire des chemins existants et d'autre part de diminuer les temps de calcul des différents algorithmes de mesures. L'utilisateur peut également effectuer des agrégations de zones du réseau afin de diminuer la complexité visuelle. En fait, et pour généraliser ces exemples, le modèle est construit de façon à ce que

chacune des complexités trouve en regard et pour réponse deux paramètres du modèle (voir Fig. 3).

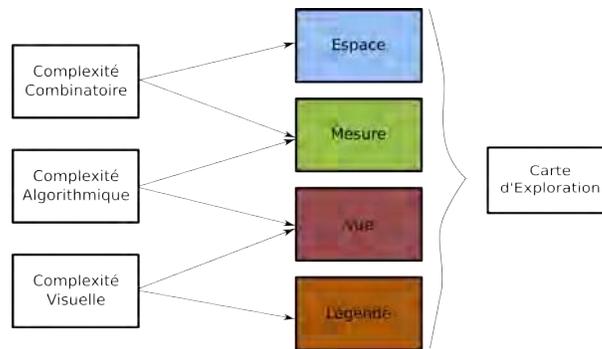


Figure 3. *Trois complexités, quatre variables de réduction des complexités*

Chaque variation d'un des paramètres peut donner lieu (choix utilisateur) à une nouvelle carte. Chaque nouvelle action est enregistrée dans une pile historique, permettant ainsi à l'utilisateur de revenir sur une carte créée préalablement afin de comparer différents résultats.

2.4. GeoGraphLab : un outil d'exploration de réseaux

Comparer visuellement les résultats de deux cartes n'est pas chose facile. Il est plus cohérent de mettre au point une méthode permettant de créer une nouvelle carte résultante des deux cartes à comparer (en effectuant par exemple un ratio). Pour ce faire, nous avons mis au point un environnement graphique permettant la combinaison de cartes par le biais d'un langage graphique. Pour cela, chaque carte (au sens d'une carte d'un réseau) est une représentée par un bloc-carte comprenant les quatre paramètres. Il est dès lors possible à l'utilisateur de croiser les différents paramètres des cartes entre eux par des connecteurs (voir Fig. 4).

De fait, cette méthode ne permet pas seulement de comparer deux cartes de manière quantitative en en créant une nouvelle. Notre outil permet également de créer de nouvelles cartes en s'appuyant sur un ou plusieurs des paramètres des cartes d'exploration déjà stockées en mémoire. Cette approche exploratoire permet à l'analyste de réseau d'avancer pas à pas dans son raisonnement. De plus, ce processus permet de diminuer les temps de calcul et donc les complexités à un niveau méta puisque l'utilisateur ne crée que les cartes dont il a besoin au fur et à mesure de l'exploration plutôt que de cartographier divers indicateurs dont les résultats pourraient ne pas être exploitables.

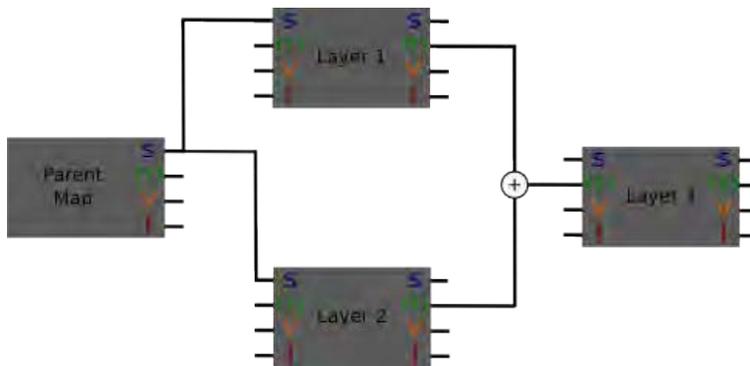


Figure 4. À partir d'une carte parente dont on copie l'espace, on effectue une addition $mesure(Layer1) + mesure(Layer2)$ pour donner une carte résultante Layer3.

3. Combinaison de cartes à l'aide d'opérateurs

Dans cette section, nous allons définir et décrire les opérations possibles, différentes selon le paramètre du modèle mis en jeu. La mise en pratique de ces différentes opérations est illustré dans la section suivante.

3.1. Les opérations sur l'espace

Les indicateurs structurels traduisent des informations observées au niveau relationnel afin de quantifier les propriétés structurelles du réseau. Mathématiquement, ces indicateurs correspondent à une agrégation d'informations quantitatives mesurées sur un ensemble de relations OD. Classiquement, il s'agit de combinaisons linéaires : ainsi, l'éloignement moyen d'un sommet est la moyenne pondérée des longueurs de plus court chemin associées aux relations OD qui émanent de ce sommet, la centralité intermédiaire d'une composante est le poids total des relations OD transitant par cette composante, etc. En tant qu'agrégations, ces indicateurs résumant certains aspects de l'organisation des chemins au sein du réseau. En contrepartie, ils ne permettent pas d'en apprécier toutes les nuances. Ainsi, une même valeur d'éloignement moyen peut résulter de distributions de longueurs d'écart-types très différents ou - à distributions de longueurs similaires - être associées à des ensembles de relations OD spatialement très différents. Pour cette raison, il peut être pertinent de calculer un indicateur en redéfinissant l'ensemble des relations OD à partir desquelles il est calculé par défaut : il s'agit d'une opération " sur l'espace ".

En pratique, l'espace associé au calcul d'un indicateur peut-être défini :

- en appliquant un filtrage à un espace existant à partir d'un critère donné : ce filtrage peut être conditionné par les valeurs d'indicateurs précédemment calculés ou par des informations thématiques. Par exemple en mesurant non plus la moyenne des lon-

guez des plus courts chemins depuis un sommet, mais la moyenne des longueurs des plus courts chemins à des sommets singuliers (mis en valeur par d'autres indicateurs ou décrits par des attributs thématiques - les gares par ex.).

- en opérant une combinaison entre plusieurs espaces existants (voir Fig. 5) : un nouvel espace peut être déduit de plusieurs espaces par les opérations classiques de la théorie des ensembles (réunion, intersection, exclusion, etc.) Par exemple, on veut observer la centralité produite par les relations OD émanant de Saint-Lazare ET dont le sommet destination est situé rive gauche.

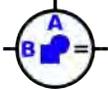
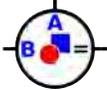
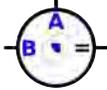
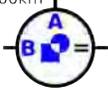
<p>Nous considérons deux espaces relationnels représentés graphiquement par un cercle et un carré. Le premier espace contient un ensemble de relations est-ouest, le second quant à lui représente les relations supérieures à 100 km.</p>	<p>Réunion</p> <p>$eS = A \cup B$ $eS = \text{relations est-ouest} + \text{relations} > 100 \text{ km}$</p> 	<p>Soustraction</p> <p>$eS = A - B$ $eS = \text{relations est-ouest} < 100 \text{ km}$</p> 
	<p>Intersection</p> <p>$eS = A \cap B$ $eS = \text{relations est-ouest} > 100 \text{ km}$</p> 	<p>Exclusion</p> <p>$eS = A \cup B - A \cap B$ $eS = \text{relations est-ouest} < 100 \text{ km}$ et non est-ouest $> 100 \text{ km}$</p> 

Figure 5. Les opérations ensemblistes pour l'espace des OD

- en appliquant une nouvelle pondération à un espace existant : il s'agit ici de moduler un espace existant en jouant sur les pondérations des relations OD qu'il contient.

- en appliquant des pondérations unitaires, surfaciques ou thématiques. Le fait de passer d'une pondération structurelle à une pondération thématique permet notamment de voir l'effet de la structure du réseau sur un phénomène thématique. En particulier, en adoptant des pondérations de population, on entre dans le cadre de tous les indicateurs classiques de potentiel de population [PUM 97] [Esp05]. Le changement de pondération intervient également lorsque les accès au réseau sont restreints. Par exemple, le service du métro parisien en mode dégradé pendant des événements nocturnes (nuit blanche ou fête de la musique). La demande en déplacement se cale sur les stations ouvertes et il faut re-répartir les poids en comparant le Voronoï habituel des stations et celui des stations restées ouvertes.

3.2. Les opérations sur la mesure

Les informations collectées au niveau relationnel sont agrégées sous forme d'indicateurs à des niveaux supérieurs (notamment au niveau des sommets et des arcs) afin d'être cartographiées et analysées. Dans une démarche exploratoire, il peut s'avérer utile de modifier l'expression mathématique d'une mesure ou de combiner mathé-

matiquement plusieurs mesures afin de produire de nouvelles informations et étayer l'analyse d'un phénomène spatialisé : on parle d'opération "sur la mesure".

Les opérations sur la mesure doivent être effectuées et interprétées avec précaution, car elles peuvent poser des problèmes de cohérence :

- lorsque l'on calcule les mesures aux niveaux supérieurs : si l'on applique une opération non linéaire sur les informations relationnelles et si ces informations doivent être moyennées pour produire les indicateurs sur les composantes, l'ordre des opérations importe sur le résultat (typiquement : "le logarithme de la moyenne n'est pas égal à la moyenne des logarithmes");

- lorsque l'on analyse la combinaison mathématique de deux mesures : la mesure résultante doit avoir un sens physique pour pouvoir être interprétée et utilisée dans une étape suivante de l'exploration. Par exemple, le fait de multiplier le degré d'un sommet avec son éloignement moyen n'est pas mathématiquement incohérent, en revanche, le sens et l'utilité de cette opération restent à justifier contextuellement.

3.3. Les opérations sur la vue

Les informations collectées à l'échelle des relations OD sont agrégées aux niveaux supérieurs. En regroupant les relations OD qui émanent d'un sommet donné ou qui transitent par une composante donnée, on construit ainsi des indicateurs de types accessibilité ou centralité. Ce principe d'agrégation peut être étendu à des regroupements de composantes. En regroupant des composantes ou en éclatant des regroupements existants, l'analyste peut apprécier l'impact des propriétés structurelles d'un réseau à différentes échelles pour une même représentation : on parle d'opération "sur la vue".

Lorsque l'on regroupe des composantes, les informations observées à l'échelle de ces composantes nécessitent d'être agrégées à un niveau supérieur. La règle d'agrégation n'est pas universelle : elle dépend naturellement de la nature de l'indicateur et de la règle ayant régi l'agrégation des informations au niveau des composantes. Par exemple, la centralité intermédiaire d'une composante est définie comme le poids total des relations transitant par une composante : pour un groupe de composantes, on transpose cette notion en calculant le poids total des relations qui transitent par au moins l'une de ces composantes. La notion d'éloignement moyen est moins immédiate à transposer. Au niveau méso, elle pose en effet la question de la longueur de plus court chemin d'un sommet S à un groupe de composantes. En fonction du contexte de l'étude, il revient à l'analyste de définir cette longueur : ce peut être la moyenne des longueurs des plus courts chemins de S aux sommets du groupe, le maximum de ces valeurs, la médiane, etc.

Les opérations sur la vue consistent donc :

- à définir les regroupements de composantes qui seront opérés : ceux-ci peuvent être dictés par l'analyste ou régis par un critère donné, mais ils peuvent également ré-

sulter de la combinaison de regroupements existants (par exemple des regroupements étudiés dans des étapes antérieures de l'exploration) grâce aux opérations classiques de la théorie des ensembles (voir Fig. 6) (réunion, intersection, exclusion, etc.) ;

- à spécifier la règle régissant l'agrégation aux niveaux supérieurs.

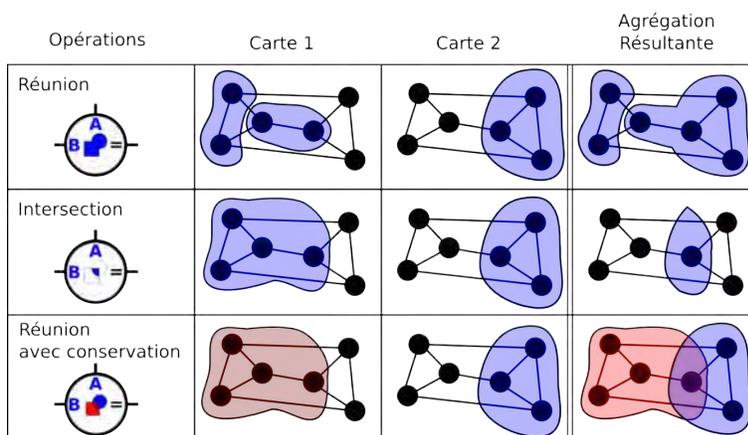


Figure 6. Les opérations possibles pour les différentes agrégations

3.4. Les opérations sur la légende

Pour une vue donnée et des valeurs d'indicateur à cartographier, il s'agit enfin d'adopter une légende, c'est-à-dire de définir :

- les formes et couleurs pour les nœuds et les arêtes,
- les seuils des intervalles dans lesquels seront réparties les valeurs à cartographier,
- les couleurs des intervalles.

Le découpage des intervalles peut être entièrement déterminé par l'analyste, ou bien déduit d'une règle. Les règles les plus classiques consistent à fixer un nombre d'intervalles et à imposer :

- un *découpage en intervalles de longueurs égales* (l'analyste spécifie alors la valeur autour de laquelle les intervalles sont distribués et la longueur de ceux-ci),
- ou un *découpage en intervalles d'effectifs égaux* (le découpage est alors automatiquement réalisé à partir des quantiles de la distribution).

La sémiologie graphique des intervalles dépend quant à elle du type d'indicateur (qualitatif, quantitatif, expression d'un taux, d'une différence, etc.) et des règles de sémiologie cartographique [BER 99]. Cela étant, elle peut être paramétrée par une teinte et un dégradé de saturation, un dégradé entre deux teintes, etc.

Dans une démarche exploratoire, le choix de la légende (au sens que nous venons de définir) prend toute son importance dès lors que l'on souhaite comparer des indicateurs. Pour réaliser cette comparaison, il s'agit alors d'ajuster les seuils ou les effectifs des intervalles et d'adopter des sémiologies graphiques cohérentes : on parle d'opération "sur la légende". En pratique, les opérations sur les légendes consistent à combiner les seuils des intervalles, les règles de découpage des intervalles ou les sémiologies graphiques utilisées antérieurement dans l'exploration.

4. Cas d'applications des différentes opérations

Nous présentons dans cette partie des exemples illustratifs des opérations définies dans la section précédente pour les quatre paramètres.

4.1. *Les opérations ensemblistes sur l'espace*

Les principales opérations relatives à l'espace sont le filtrage des relations OD (ne sélectionner qu'une partie d'intérêt dans l'espace relationnel complet) et la combinaison de deux espaces (eux-mêmes obtenus par filtrage ou par combinaison). Considérons l'exemple du métro parisien (voir Fig. 7.1) évoqué dans la partie précédente dans laquelle on souhaitait observer la centralité produite par les relations OD émanant de Saint-Lazare ET dont le sommet destination est situé rive gauche. L'outil développé nous permet d'effectuer une sélection des relations selon deux méthodes. En premier lieu, nous effectuons une sélection des relations issues d'un nœud sélectionné. Dans la figure 7.1, nous sélectionnons les relations émanant de la station Saint-Lazare. En second lieu, la sélection porte sur les relations dont les origines se situent dans une zone Nord (rive droite) et dont les Destinations se situent dans une zone sud (rive gauche) comme le montre la figure 7.2. Enfin l'étape de combinaison consiste à réaliser une intersection des deux espaces relationnels, opération réalisée dans le langage graphique comme le montre la figure 7.3. On obtient un espace relationnel résultant donnant les relations émanant de la station Origine Saint-Lazare (rive droite) et dont les nœuds de destinations sont situés sur la rive gauche. On peut alors produire une carte de centralité intermédiaire à partir de l'espace relationnel calculé (voir Fig. 7.4).

Notons que pour réaliser cette carte, plutôt que de calculer les 30000 relations Origine-Destination de plus courts chemins, nous n'avons eu à calculer uniquement 121 relations. Ce qui permet, comment nous l'avons vu dans le chapitre 1, de diminuer les complexités combinatoire et donc algorithmique (le calcul est quasiment instantané pour un indicateur de centralité intermédiaire ou d'éloignement moyen) et visuelle : la carte donne précisément les informations que l'utilisateur cherchait. Il peut par la suite raffiner son analyse en jouant sur les classes et couleurs de la légende.

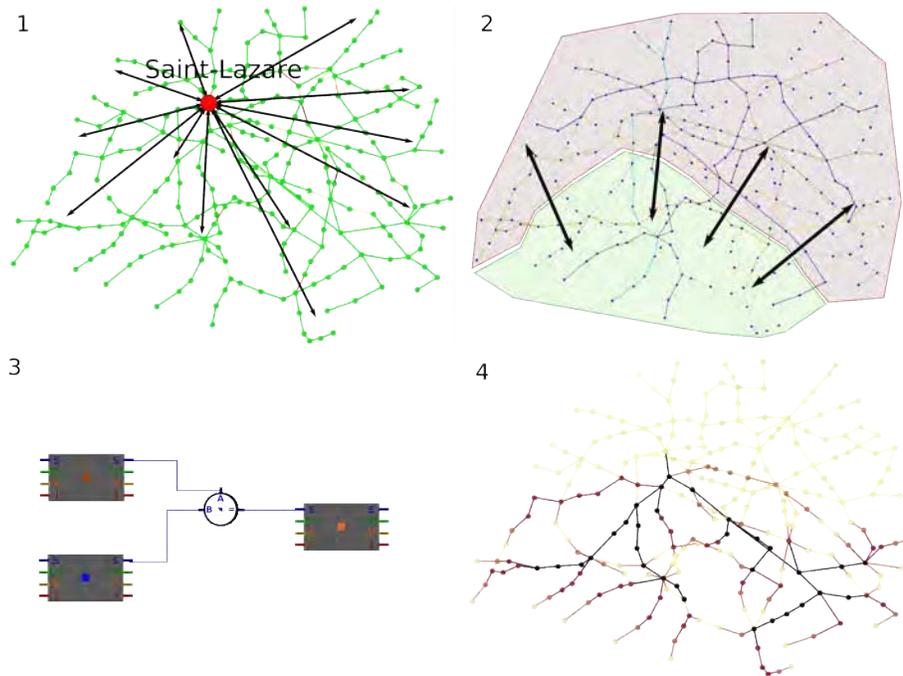


Figure 7. La première carte montre une sélection d'un espace relationnel rive-gauche rive-droite, la seconde montre une sélection des relations émanant de Saint-Lazare, l'opération graphique (3) permet d'effectuer l'intersection des deux espaces définis. On observe en (4) la carte produite par le calcul d'un indicateur de centralité intermédiaire résultante du nouvel espace.

4.2. Les opérations mathématiques sur la mesure

Les opérations sur la mesure permettent de combiner les mesures de deux blocs-cartes (voir Fig. 4) à l'aide de l'outil GeoGraphLab. Nous présentons ici quelques exemples d'opérations sur la mesure.

Modification de l'expression mathématique d'une mesure :

[CAT 97] construit un indicateur de mesure d'efficacité du réseau non pas à partir de l'éloignement moyen mais à partir de la vitesse moyenne. Au niveau d'un sommet O , on collecte les valeurs de vitesse moyenne pour atteindre tous les autres sommets D du réseau (distance à vol d'oiseau / temps de parcours OD sur le réseau). Comme il s'agit d'une vitesse, il faut faire une moyenne harmonique des vitesses observées au niveau relationnel et non pas une moyenne arithmétique.

Comparaison mathématique de mesures de même unité :

Dans des cas de mesure de l'impact de l'endommagement d'un sommet S sur l'éloignement moyen de tous les sommets du réseau, on calcule pour chaque OD l'allongement de la longueur de plus court chemin (comparaison de la longueur de plus court chemin avant et après endommagement de S, par ratio ou par différence [CHA 01]) et au niveau de chaque sommet, on agrège les informations collectées pour les relations émanant du sommet (voir Fig. 8).

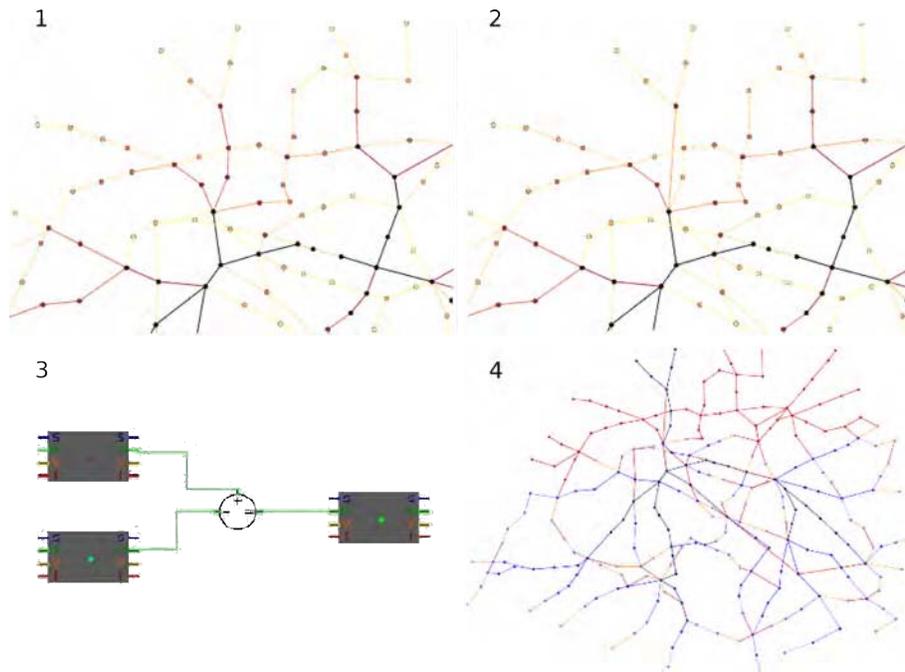


Figure 8. La première carte (1) montre une portion du réseau du métro parisien non modifié, tandis qu'on a ajouté un tronçon sur la seconde (2), l'opération graphique (3) de ratio des deux cartes d'entrées, permet d'obtenir la carte résultante (4) de différence de centralité.

Combinaison mathématique de mesures d'unités différentes :

Lorsque l'on souhaite comparer la sollicitation des sommets d'un réseau (type : centralité intermédiaire) en tenant compte de la richesse des connexions qu'ils offrent (type : centralité de degré), il peut être intéressant d'effectuer un ratio entre centralité intermédiaire et centralité de degré. On effectue également des ratios ou des combinaisons linéaires d'indicateurs hétérogènes lorsque l'on cherche à expliquer la variance d'un indicateur à partir d'autres indicateurs (par exemple lors d'une régression [GLE 05] pour expliquer la vulnérabilité élémentaire des composantes d'un réseau à partir de leur centralité intermédiaire, ou [FOL 07], pour expliquer la fréquence des trajets piétonniers à partir d'indicateurs structurels descriptifs de la géométrie du réseau).

4.3. Cas d'application d'agrégations

Il existe plusieurs méthodes d'agrégations liées à différents besoins d'études d'un réseau de transport. Nous recensons différents regroupements de composantes :

- Un regroupement de composantes dicté par le mode de fonctionnement du réseau : certains réseaux nécessitent d'emblée de modéliser leurs infrastructures nodales par des sous-réseaux (c'est le cas notamment des réseaux de métro pour lesquels une station à correspondance est en fait un sous-réseau où les quais de chaque ligne sont identifiés par des sommets distincts (voir Fig. 9), eux-mêmes reliés par des arêtes matérialisant les couloirs de correspondance). Pour une infrastructure nodale de ce type, les informations relationnelles sont collectées au niveau des sommets et arêtes constitutives du sous-réseau associé, puis elles sont agrégées à l'échelle de l'infrastructure nodale elle-même, représentée par un sommet [GLE 05].

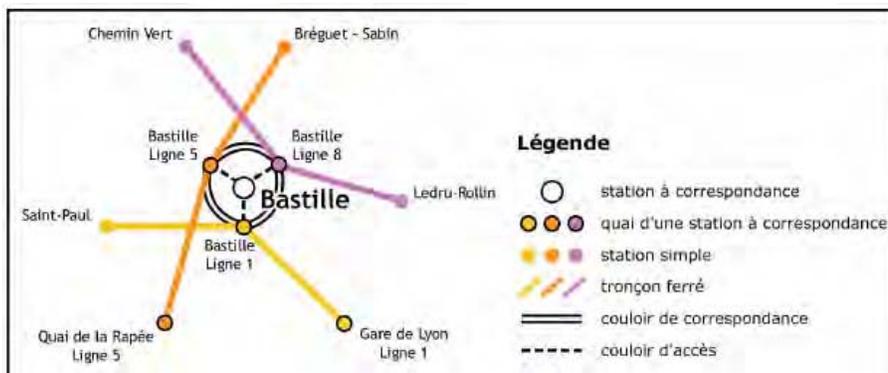


Figure 9. Modélisation des stations à correspondance selon la méthode des zooms nodaux - Exemple de la station Bastille.

- Un regroupement de composantes dicté par le niveau d'analyse souhaité sur les différentes parties du réseau : une des caractéristiques les plus singulières des réseaux réside dans la portée spatiale que peuvent avoir certains phénomènes pourtant très localisés : ainsi peut-on être amené à analyser l'impact d'un aménagement local sur l'organisation globale des chemins dans des zones complètement éloignées du réseau. Pour cela, il s'agit de pouvoir observer les propriétés structurelles du réseau à une échelle fine sur les zones d'intérêt locales, et à une échelle agrégée sur les zones d'intérêt globales. [CHA 97] propose ainsi d'analyser l'impact de l'aménagement du péri-phérique de la ville de Tours à l'échelle de la ville et à l'échelle du territoire français en cartographiant un indicateur de type accessibilité sur des regroupements de sommets de granularité variable.

- Un regroupement de composantes dicté par une analyse préalable (type clustering) : la taille et l'ordre de certains réseaux sont tels qu'il est nécessaire de réaliser au préalable des opérations de clustering pour pouvoir les appréhender et les ana-

lyser (c'est par exemple le cas du réseau des liaisons aériennes entre les aéroports mondiaux, pour lequel [AMI 05] proposent des regroupements de sommets (i.e. d'aéroports) grâce à un indice de clustering local).

Enfin, la cartographie des indicateurs reflète certains aspects des propriétés structurelles du réseau, qu'il convient éventuellement de compléter en éclatant à nouveau certains regroupements pour observer la structure locale des chemins et nuancer les analyses produites dans la vue agrégée. En ce sens, les opérations sur la vue encouragent les allers-retours entre les différentes échelles de représentation pour appréhender la portée spatiale des propriétés structurelles.

4.4. Le cas de la combinaison de légendes

La démarche analytique dicte l'opération à effectuer sur la légende. En particulier, lorsque l'on compare un même indicateur dans deux contextes différents, il est pertinent de garder les mêmes seuils de découpage ou les mêmes palettes de couleurs. Par exemple, si l'on voulait comparer des variations de centralité intermédiaire pour différents scénarios d'endommagement. À l'inverse, on peut choisir des seuils différents lorsque les indices à cartographier sont de nature différentes (voir Fig.10) avec une couleur pour la centralité et en jouant sur la surface du nœud pour la variable d'accessibilité.

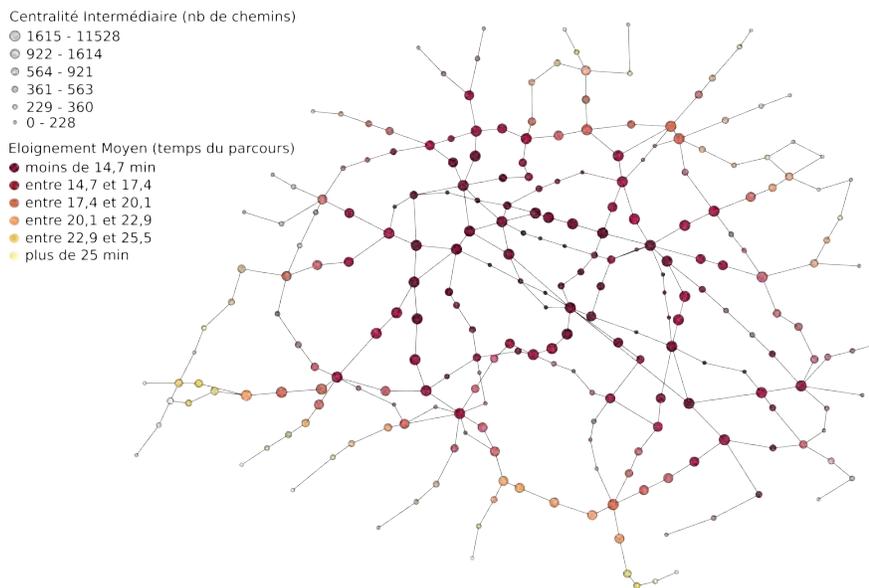


Figure 10. Carte combinée d'un indicateur de centralité au centre des nœuds (légende du haut), et d'un indicateur d'éloignement moyen en bordure des nœuds (légende du bas).

5. Conclusion

Notre approche est basée sur l'analyse exploratoire de données, telle que définie par [CAU 08], dans le but d'extraire de l'information qui n'est pas facilement perceptible et de la visualiser de la façon la plus appropriée et compréhensible. Cet article part du constat qu'un réseau est un objet complexe. Nous dégagons trois complexités liées à l'analyse exploratoire des propriétés structurelles d'un réseau de transport : 1. la complexité combinatoire liée à la masse de données à traiter du fait du nombre de chemins sur le réseau, 2. la complexité algorithmique des différents indicateurs relationnels, et enfin, 3. la complexité visuelle liée à la difficulté pour l'utilisateur d'extraire de l'information.

Ces trois complexités sont liées les unes aux autres. Il est donc nécessaire de mettre en place une stratégie d'exploration reposant sur un grand principe de résolution de problèmes : la séparation des complexités de façon à segmenter le problème.

Le modèle à quatre paramètres que nous avons développé vise à réduire ces complexités de façon à fractionner le l'analyse structurelle du réseau. Il repose sur les notions d'espace relationnel, de mesure dictée par une propriété mathématique liant les relations sur le réseau, de vue permettant l'agrégation de composantes du réseau et enfin de légende facilitant la visualisation d'informations par regroupement en différentes classes et couleurs.

Nous avons développé un outil permettant une exploration arborescente à l'aide d'un langage graphique. L'utilisateur peut alors, à l'aide de blocs-cartes qu'il combine entre eux par des liens simples ou par combinaison d'opérateurs, effectuer une analyse exploratoire. À un niveau méta apparaît ainsi une diminution des complexités puisque :

- l'utilisateur peut sélectionner uniquement l'espace relationnel nécessaire,
- le temps de calcul des des différents indicateurs relationnels se voit amélioré puisque l'espace est restreint,
- les cartes produites sont alors mieux ciblées sur un phénomène et il est donc plus simple d'interpréter ces cartes,
- il est possible de combiner les différentes cartes obtenues par croisement de mesures, de vues et de légendes sans avoir à refaire une série de calcul. Cette méthode de croisements permet d'être encore plus spécifique dans la recherche d'informations.

6. Bibliographie

- [AMI 05] AMIEL M., MÉLANÇON G., ROZENBLAT C., « Réseaux multi-niveaux l'exemple des échanges aériens mondiaux de passagers », *Mappemonde*, vol. 79, 2005.
- [BER 99] BERTIN J., *Sémiologie Graphique. Les diagrammes, les réseaux, les cartes.*, 1999.
- [CAT 97] CATTAN N., GRASLAND C., « Les différentiels d'accessibilité des villes moyennes en France », rapport, 1997, CNRS - Equipe P.A.R.I.S, URA 1243.
- [CAU 08] CAUVIN C., ESCOBAR F., SERRADJ A., *Cartographie thématique*, Hermès Science publications, 2008.

- [CHA 97] CHAPELON L., « Offre de transport et aménagement du territoire : évaluation spatio-temporelle des projets de modification de l'offre par modélisation mutli-échelles des systèmes de transport », PhD thesis, Université de Tours, Laboratoire du CESA, 1997.
- [CHA 01] CHAPELON L., « Analyse de réseau par destruction / hiérarchisation des liens », *Groupe de travail : "Dynamiques spatiales méditerranéennes et systèmes territoriaux émergents, application à l'arc méditerranéen"*, , 2001.
- [CHA 07] CHAN T. M., « More algorithms for all-pairs shortest paths in weighted graphs », , 2007, p. 590–598, ACM.
- [COU 05] COUREL J., MEYERE A., NGUYEN-LUONG D., « Répartition géographique des déplacements : une nouvelle approche », rapport, 2005, Les cahiers de l'Enquête Globale de Transport, IAURIF.
- [DUP 91] DUPUY G., *L'urbanisme des réseaux : théories et méthodes*, 1991.
- [DYK 05] DYKES J., MACEACHREN A.-M., KRAAK M.-J., *Exploring Geovisualization*, Elsevier, 2005.
- [Esp05] *European Territorial Research in Progress, Conference Proceedings of the 1st ESPON Scientific Conference*, 2005.
- [FOL 07] FOLTÊTE J.-C., « Quel rôle du réseau de voirie urbaine sur la distribution des flux piétonniers ? », *Les Cahiers scientifiques du Transport*, vol. N°52, 2007, p. 27–44.
- [GEN 01] GENRE-GRANDPIERRE C., « Contexte réseautique et exposition au risque d'accident », *Colloque Risques Besançon*, , octobre 2001.
- [GLE 05] GLEYZE J.-F., « La vulnérabilité structurelle des réseaux de transport dans un contexte de risques », PhD thesis, Université Paris VII, 2005.
- [KAN 63] KANSKY K. J., « Structure of Transportation Networks : Relationships Between Network Geometry and Regional Characteristics », *Departement of Geography Research*, vol. Paper 84, Chicago : University of Chicago Press, 1963.
- [MER 09a] MERMET E., « Construction d'un modèle pour l'analyse et l'exploration des propriétés structurelles d'un réseau de transport », *Neuvièmes Rencontres ThéoQuant*, , 2009.
- [MER 09b] MERMET E., « Exploration des propriétés structurelles d'un réseau de transport par combinaison interactive de cartes », *Colloque International de Géomatique et d'Analyse Spatial SAGEO'09, 25-27 novembre, Paris (France)*, , 2009, page 5.
- [MER 10] MERMET E., « GeoGraphLab : a tool for exploring structural characteristics of transportation network », *13th AGILE International Conference on Geographic Information Science, Guimaraes, Portugal*, , 2010, page 4.
- [MUR 72] MURACO W. A., « Intraurban Accessibility », *Economic Geography*, vol. 48, 1972, p. 388–405.
- [PAR 72] PARLEBAS P., « Centralité et compacité d'un graphe », *Mathématiques et Sciences Humaines*, vol. 39, 1972, page 22.
- [PIT 79] PITTS F. R., « The Medieval River Trade Network of Russia Revisited », *Social Networks 1978/1979*, , 1979, p. 285–292.
- [PUM 97] PUMAIN D., SAINT-JULIEN T., *L'analyse spatiale, tome 1 : Les localisations dans l'espace*, 1997.
- [REY 04] REY J.-F., *Calculabilité, complexité et approximation*, Vuibert Informatique, 2004.
- [THO 04] THOMAS J., *Illuminating the Path : The Research and Development Agenda for Visual Analytics*, NVAC, 2004.

Un système d'information géographique 3D pour l'exploration des règles d'urbanisme

Mickaël Brasebin* — Julien Perret* — Cécile Haëck**

* *Laboratoire COGIT, IGN*

** *IGN Conseil*

73 avenue de Paris

94160 Saint Mandé

{prenom.nom}@ign.fr

RÉSUMÉ. Cet article présente l'utilisation d'un SIG 3D pour représenter, analyser et explorer les règles les plus courantes trouvées dans les plans locaux d'urbanisme. Un modèle est proposé pour permettre la saisie et la manipulation des règles. Ces dernières s'appliquent sur un espace géographique structuré. Ces éléments sont utilisés au sein d'un SIG3D afin d'explorer les règles d'urbanisme et leurs implications à l'aide d'analyse spatiale 3D et d'analyse visuelle.

ABSTRACT. This article presents the use of a 3D GIS to represent, analyse and explore the most common urban rules found in french local urban planning. A model is proposed in order to allow the capture and the management of urban rules. These rules are applied of a structured geographic environment. Both elements are used through a 3D GIS to explore urban planning regulations and their implications thanks to 3D spatial analysis and visual analytics.

MOTS-CLÉS : SIG 3D, règles d'urbanisme, analyse spatiale, géovisualisation

KEYWORDS: 3D GIS, urban planning rules, spatial analysis, geovisualisation

1. Introduction

Le développement du territoire français est élaboré à travers un ensemble complexe de documents, plans ou schémas à différentes échelles (schémas de cohérence territoriale, plans locaux d'urbanisme, plans de déplacements urbains, etc.). Ces documents reflètent à la fois les différentes stratégies de l'aménagement du territoire et de l'urbanisme, mais aussi la traduction de ces stratégies en contraintes auxquelles la réalisation des projets est soumise. Les logiques sous-jacentes font référence à différents domaines (construction, logement, déplacement, transport, urbanisme, etc.) pour lesquels les niveaux de pertinence sont là aussi différents. Ces difficultés soulèvent ainsi de nombreuses questions et demandent ainsi la conception de nouvelles techniques pour mieux comprendre, représenter, évaluer, gérer et concevoir la ville [BAI 08].

Les bases de données urbaines 3D permettent une meilleure compréhension du territoire grâce à la modélisation, la visualisation et la simulation de phénomènes complexes. Les Systèmes d'Information Géographique (SIG) représentent un outil central pour ces techniques, notamment à travers l'utilisation de telles bases de données. Le projet *TerraMagna*¹ porté par les pôles de compétitivité *Advancity*² et *Cap Digital*³ a ainsi cherché à valoriser les bases de données urbaines 3D et à créer des solutions logicielles innovantes pour la conception, la simulation, mais aussi pour le dialogue sur la ville entre les experts (urbanistes, architectes, etc.), les élus et la population. Le travail présenté dans cet article donne suite à une réflexion commencée dans le cadre de ce projet par le laboratoire *COGIT* de l'IGN en collaboration avec *IGN Conseil* (service de l'IGN) et la société *Bionatics*⁴. Elle a donné lieu à l'expression des besoins d'utilisateurs potentiels (collectivités locales, établissements publics d'aménagement, etc.) par entretiens et attestant de la pertinence d'un outil d'exploration des règles d'urbanisme.

Différents travaux ont cherché à modéliser différentes échelles de l'environnement urbain par les propriétés recherchées lors d'un processus de conception. Une telle modélisation déclarative permet ainsi de fournir des outils pour la conception architecturale ou pour la conception urbaine [DON 92, DES 95], d'évaluer les enveloppes urbaines réglementaires [FAU 01, LE 03], de prendre en compte l'accessibilité visuelle [NIV 99]. [MUR 04] propose l'utilisation d'un SIG 3D pour visualiser l'état d'une ville, visualiser la capacité d'un îlot urbain et évaluer un projet de construction. [FAL 05] propose un outil de visualisation 3D de connaissances urbaines et le prototypage de projets pour améliorer la communication autour d'un projet et impliquer les non professionnels. Les travaux présentés dans cet article visent à proposer des outils permettant d'intégrer la spécificité des connaissances urbaines propres aux règles d'urbanisme françaises. Nous nous intéressons par ailleurs à leur représentation graphique et à la sémiologie adaptée. Ces travaux sont donc aussi à mettre en relation avec un effort similaire effectué pour la représentation des PLU en 2D [CNI06].

1. <http://www.terramagna.org>
2. <http://www.advancity.eu>
3. <http://www.capdigital.com>
4. <http://www.bionatics.com>

2. Un outil pour représenter et explorer les règles d'urbanisme

Pour pouvoir étudier les règles d'urbanisme sur un territoire, il est nécessaire de modéliser 2 types d'informations : les règles (*cf.* section 2.1) et l'espace géographique (*cf.* section 2.2). Les règles intégrées dans l'outil ont permis de proposer un modèle et une interface pour les représenter. L'espace géographique est décrit grâce à un schéma qui nécessite une phase d'intégration lors du chargement des données. Les règles et l'environnement s'inscrivent dans le fonctionnement global de l'outil (*cf.* section 2.3).

2.1. Règles

2.1.1. Sélection des règles

Le PLU (Plan Local d'Urbanisme), amené à terme à couvrir tout le territoire français, est élaboré au niveau de la commune ou de l'intercommunalité. Ces entités administratives sont découpées en zones sur lesquelles s'appliquent 14 articles dont l'intitulé est fixe (l'article 10 concernera par exemple toujours la hauteur des constructions) mais dont la forme reste libre. Dès lors, il existe une grande variété de formulations différentes pour exprimer les règles d'un article.

Afin que l'outil développé puisse être fonctionnel, une sélection a été effectuée sur les règles qu'il intègre. Celles s'appliquant aux éléments n'apparaissant pas dans les jeux de données sont écartées. Par exemple, c'est le cas de celles qui concernent les dépassements (balcons). Elles n'ont pas d'intérêt dans le cadre de ce travail, ces éléments étant absents des données à notre disposition. Il a également fallu ne pas tenir compte des règles ne concernant pas directement le bâtiment (les articles concernant le stationnement et les espaces verts n'entrent pas dans ce cadre) car le but de ce travail est de se focaliser sur les constructions. De plus, comme il n'y a pas de formulation type concernant les règles des différents articles, il a fallu effectuer un tri afin de ne conserver que les formulations les plus couramment utilisées dans les PLU.

Le tableau 1 réalisé dans le cadre d'une synthèse menée en collaboration avec IGN Conseil présente les articles et formulations conservées. La synthèse relève également les informations attributaires ou relationnelles influant sur les paramètres des règles. Par exemple, le recul d'un bâtiment par rapport à une route dépend de la largeur de celle-ci.

2.1.2. Modèle proposé

Un modèle décrivant les règles est proposé à partir des différents éléments issus de la synthèse. Ce modèle retranscrit la logique du plan de zonage : une règle s'applique à une zone composant le plan de zonage. Chaque règle est traduite en termes d'antécédent/conséquence. Ainsi, une règle est composée de zéro ou plusieurs antécédents et d'une ou plusieurs conséquences. Si aucun antécédent d'une règle n'est vérifié alors

N°	Intitulé	Formulations retenues
1	Type d'occupation ou d'utilisation du sol interdit	Aspect des façades à prendre en compte
2	Type d'occupation ou d'utilisation du sol soumise à des conditions particulières	Destination fonctionnelle des bâtiments
6	Implantation des constructions par rapport aux voies et emprises publiques	Distance de la limite d'emprise des voies à la limite d'emprise d'un bâtiment
7	Implantation des constructions par rapport aux limites séparatives	Distance de la limite de parcelle à la limite d'emprise du bâtiment Bande de constructibilité
8	Implantation des constructions les unes par rapport aux autres sur la même propriété	Distance minimale entre deux bâtiments
9	Emprise au sol	Défini par le CES
10	Hauteur maximum des constructions	Hauteur d'un bâtiment limitée Différence de hauteur entre deux bâtiments limitée
11	Aspects extérieurs	Limitation des matériaux utilisés pour le toit et les murs Pente des toits
14	Coefficient d'occupation des sols	Défini par le COS

Tableau 1 – Synthèse des éléments retenus pour l'outil

les conséquences ne sont pas appliquées à la parcelle. Par exemple, une règle indiquant (avec en vert l'antécédent et en rouge la conséquence) :

Si la voie a une largeur inférieure ou égale à 7m Alors un recul de 3m doit être effectué

ne s'applique qu'aux parcelles bordées par des routes de ce gabarit. La figure 1 présente le schéma UML du modèle où sont distinguées les classes représentant les antécédents et les conséquences.

Dans la suite de cet article, nous nous focaliserons sur des règles que nous jugeons particulièrement intéressantes vis-à-vis de la 3D. Nous allons les présenter, ainsi que la formulation retenue et leur représentation.

Hauteur maximale d'un bâtiment :

Il s'agit de limiter la hauteur maximale H des bâtiments d'une parcelle (*cf.* figure 2a). Cela revient à dire que la géométrie des bâtiments doit être incluse dans le solide issu de l'extrusion de valeur H de la parcelle.

Coefficient d'Occupation du Sol (COS) :

Le COS est calculé à partir de la surface hors œuvre nette (superficie de tous les étages - superficie des locaux techniques, 5% d'isolation, etc.) divisée par la surface

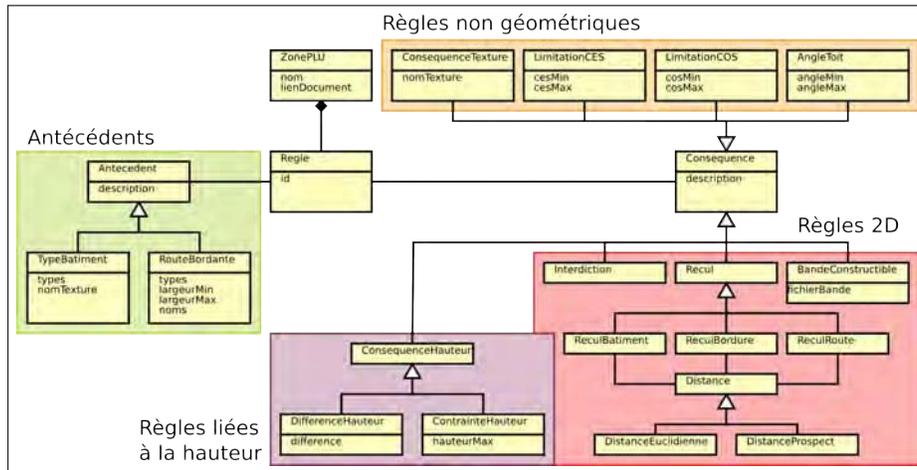


Figure 1 – Schéma des règles implémentées dans l'outil. Dans le cadre vert, les règles concernant les antécédents et dans les autres cadres celles concernant les conséquences.

de la parcelle. Empiriquement, la surface hors œuvre nette est approximée dans cet article en utilisant 80% de la surface hors œuvre brute (ensemble des planchers). De plus, la hauteur d'un étage a été fixée à 3m afin de déterminer la surface des planchers (cf. figure 2b) et ainsi la surface hors œuvre brute.

Distance par rapport à une entité (route bordante, parcelle voisine, bâtiment) :

Il s'agit de limiter la distance minimale entre ces différents éléments. Deux types de distance sont pris en compte (cf. figures 2c et 2d) : la *distance euclidienne* et la *distance dite du prospect*. La distance de prospect s'exprime sous la forme d'une équation affine dépendant de la hauteur : $D = a \times H + H_{ini}$, où a est un coefficient positif (la pente) et H_{ini} la hauteur à respecter en bordure de l'entité.

2.1.3. Interface de création des règles

Une interface graphique a été développée en Java afin de permettre la saisie et l'export de règles dans notre modèle. La bibliothèque JAXB (Java Architecture for XML Binding) [JAX10] est utilisée pour l'implémentation du modèle à partir d'un fichier de schéma (.xsd) et l'export au format XML. L'interface graphique contient une fenêtre permettant de définir les zones d'un PLU et d'ajouter ou de supprimer de nouvelles règle et une autre permettant de définir le contenu d'une règle.

Par ailleurs, pour simplifier sa manipulation, une description textuelle de chaque règle est générée à la saisie. Le fichier XML exporté contient cette description pour permettre une meilleure compréhension à la lecture de celui-ci.

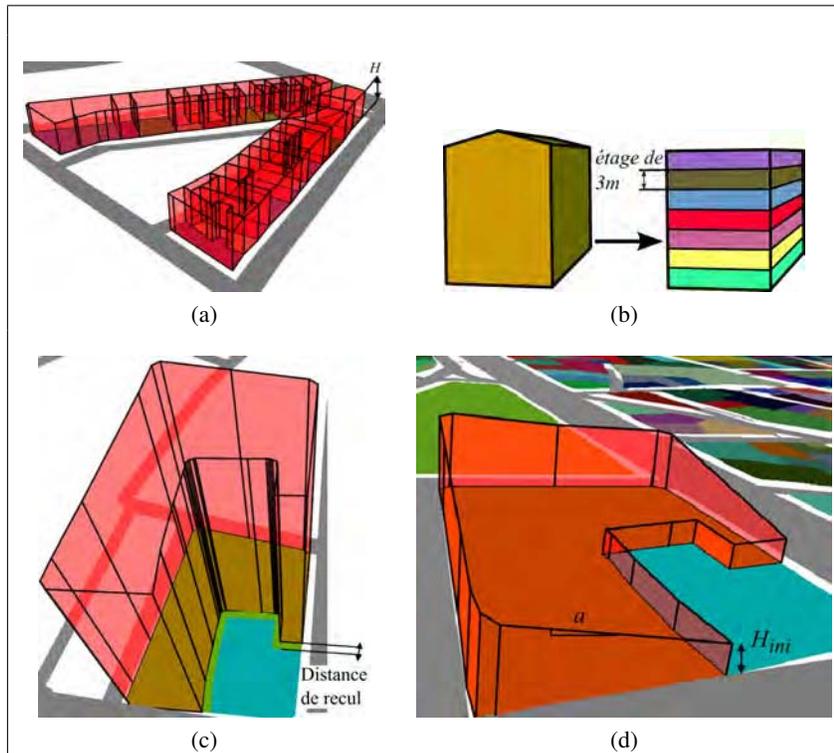


Figure 2 – (a) Limitation d'un ensemble de parcelles à la hauteur H . (b) Découpage d'un bâtiment par étages de $3m$. (c) Recul de $3m$ par rapport à la parcelle rose. (d) Recul par distance de prospect avec $a = 0.5$ et $H_{ini} = 2m$.

2.2. Environnement

2.2.1. Sources des données

L'application développée utilise 3 bases de données de l'IGN différentes (cf. figure 3) :

- **les bâtiments**, issus du Bati3D qui contient des données 3D de bâtiments avec leurs toits au format CityGML (<http://www.citygml.org/>),
- **les parcelles**, importées depuis la BD Parcellaire qui contient les informations cadastrales en 2D,
- **les routes**, chargées à partir de la BD Topo qui décrit en 2D5 le territoire français.

Le format utilisé pour les deux dernières couches appartenant toutes deux au référentiel à grande échelle (RGE) est le format ESRI Shapefile [ESR98]. Il est par ailleurs également possible d'y intégrer un plan de zonage associé aux règles. Dans le cadre de cet exposé, les différentes zones nécessaires ont été saisies manuellement.

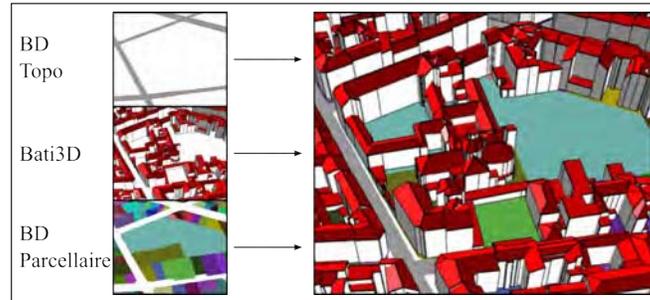


Figure 3 – Les couches composant l'environnement géographique.

2.2.2. Schéma Géographique

Afin de simplifier les raisonnements à mener à partir de règles saisies, les éléments géographiques sont organisés sous forme d'un modèle (*cf.* figure 4). Ce schéma contient également le minimum d'informations mentionnées dans notre modèle de règle. Il fait ressortir un certain nombre d'informations attributaires et de relations entre objets nécessaires au fonctionnement global de cette application. Par exemple, il est indispensable de connaître les bâtiments construits sur une parcelle pour en déterminer le COS.

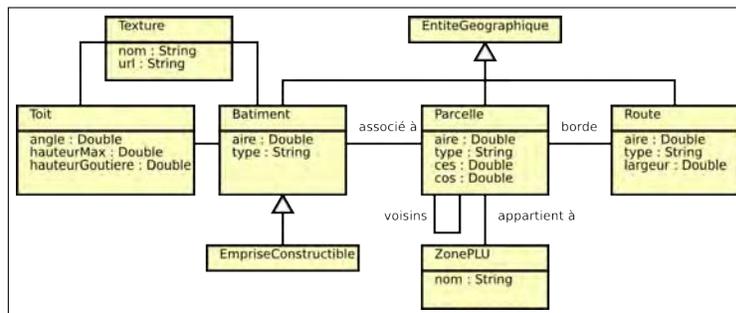


Figure 4 – Schéma de l'environnement géographique.

2.2.3. Intégration et chargement des données

Un prétraitement, effectué lors du chargement des données, permet de créer les relations entre une parcelle et les bâtiments qu'elle contient (un bâtiment n'appartient qu'à une seule parcelle), entre les parcelles voisines, et entre une parcelle et les routes la bordant.

Afin de déterminer le voisinage entre routes et parcelles, il est nécessaire de connaître la géométrie des bordures de route. Néanmoins, dans la BDTopo, les routes sont représentées par des géométries linéaires (par leur axe). Ainsi, la géométrie des bordures

de route est approximée par application d'une zone tampon dont la taille correspond à l'attribut *largeur* de la route. Le lien entre un bâtiment et son toit est quant à lui extrait directement des données importées. L'application permet alors de représenter et d'interroger dans le navigateur 3D les relations afin de vérifier si elles ont été correctement effectuées.

2.3. Architecture du système

Trois utilisations types de l'application ont été imaginées :

- 1) vérification des règles d'urbanisme sur les bâtiments d'un ensemble de parcelle ou sur un solide importé,
- 2) visualisation des contraintes liées aux règles sous forme d'un solide généré automatiquement,
- 3) calcul d'un volume respectant les règles à partir d'emprises au sol ou d'un solide paramètre.

La figure 5 présente l'architecture globale de l'application. Un environnement et un fichier de règles tels que décrits dans la section précédente sont utilisés comme données d'entrée de l'application. Le processus présenté est ensuite appliqué à chaque parcelle importée.

Tous les développements présentés dans cet article ont été développés sur la plateforme SIG Open-Source *GeOxygene* (<http://oxygene-project.sourceforge.net/>). Le dépôt en Open-Source de la composante 3D de la plateforme est imminent [BRA 09]. Ce module s'appuie sur la norme ISO 19107 décrivant un schéma géométrique [ISO03]. Il fournit notamment des outils pour charger des données (format Shapefile, CityGML, etc.), des fonctions géométriques 3D (opérateurs booléens, etc.) et également un visualisateur dont sont issus la plupart des images de cet article.

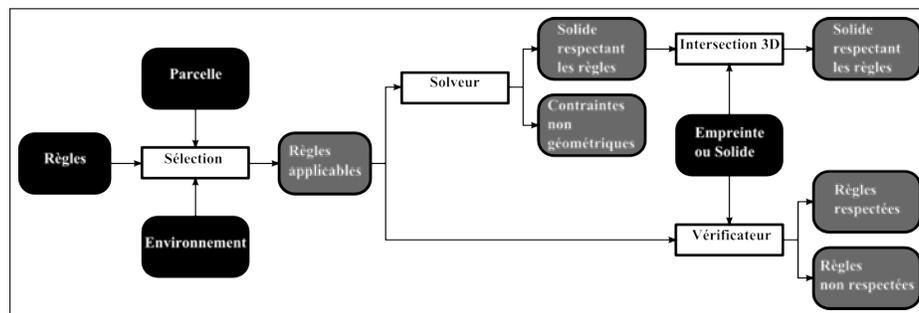


Figure 5 – Architecture globale de l'application.

3. Résultats

Dans cette partie, nous présentons les résultats proposés par les différentes fonctionnalités.

3.1. Vérification du respect du PLU

La première utilisation présentée (cf. figure 6) est la vérification des règles du PLU sur un ensemble de parcelles. L'application vérifie si pour chaque parcelle l'ensemble des règles qui la concernent s'appliquent aux bâtiments existants ou fictifs qu'elle contient. À l'échelle d'une parcelle, cette utilisation permet de vérifier si un nouveau projet respecte les contraintes imposées par le PLU. À l'échelle d'un quartier ou d'une ville, cette fonctionnalité est utile pour visualiser les règles les moins respectées lors de la révision ou de la modification d'un PLU.

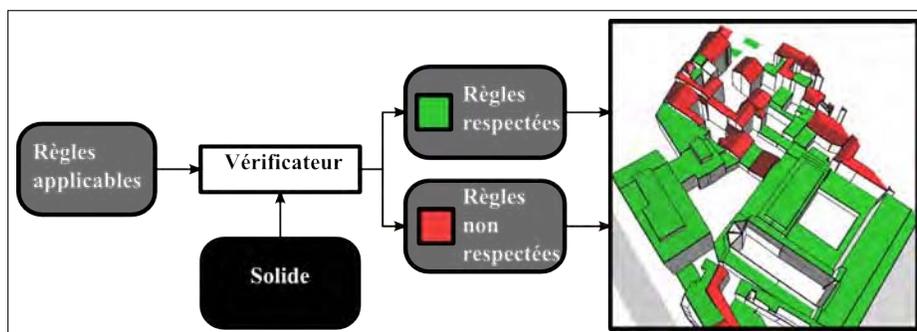


Figure 6 – Vérification d'une règle concernant les angles de toit.

Le résultat de cette utilisation est double. Pour chaque parcelle est associée automatiquement la liste des règles qui sont vérifiées et celles qui ne le sont pas. De plus, une nouvelle couche contenant les incohérences repérées est générée. Cette couche permet à l'utilisateur de zoomer sur une incohérence particulière et de visualiser grâce à une représentation adaptée les règles enfreintes. Pour chaque type de règle, une représentation adaptée est élaborée. Nous présentons ici celles qui concernent les règles présentées dans le chapitre précédent.

Les contraintes liées à la distance de prospect ou à la hauteur maximale d'un bâtiment sont représentées par un plan dont l'altitude représente la hauteur maximale acceptée (cf. figures 7a et 7b). Ce mode de représentation permet de bien distinguer la partie du solide qui respecte la contrainte de celle qui ne la respecte pas.

Concernant le recul par rapport à d'autres entités en distance euclidienne, plusieurs tests ont été effectués (cf. figures 7c et 7d). Le premier consiste à présenter par un trait les distances trop courtes par rapport aux limites des parcelles. Malheureusement, cette représentation ne permet de voir qu'en un seul point la trop grande proximité entre

objets et n'était pas représentable lorsque les deux objets se touchaient. Finalement, le volume non-constructible est utilisé lorsque la règle n'est pas respectée. Le choix d'une représentation volumique et non d'un plan est fait pour pouvoir apprécier le cas où la parcelle serait tellement petite que les solides s'intersecteraient. Un coefficient de transparence est appliqué pour permettre de co-visualiser les parties en accord avec la règle de celles qui ne le sont pas.

Pour le COS, diverses représentations des incohérences sont abordées (*cf.* figures 7e et 7f). Si le COS d'une parcelle est supérieur à la valeur définie par les règles, on peut :

- représenter un volume type cubique atteignant la valeur maximale tolérée,
- extruder les parcelles pour présenter un résultat correspondant au COS maximal,
- afficher les solides de la parcelle transformée de manière à respecter le COS en diminuant l'altitude,
- changer la couleur de la parcelle,
- afficher au dessus de la parcelle de manière textuelle, le COS actuel et le COS toléré.

L'affichage textuel illustré figure 7e est choisi pour signifier si le COS n'est pas respecté (en vert si il est trop faible, en rouge si il est trop grand). En effet, les différentes représentations utilisant des formes géométriques (*cf.* figure 7f) que nous avons testées ne sont pas intuitivement compréhensibles par l'utilisateur et difficiles à distinguer des bâtiments.

3.2. Calcul d'une enveloppe respectant les règles

La seconde fonctionnalité présentée est le calcul automatique, à partir de règles applicables à une parcelle, de formes respectant les contraintes géométriques. Ce que l'on appelle, ici, *contrainte géométrique* est entendue comme un contrainte modifiant de manière immédiate le volume maximal constructible d'une parcelle. Les autres contraintes, dites *contraintes non géométriques*, sont celles que l'on ne peut interpréter qu'une fois le volume maximal obtenu. L'application permet de gérer 4 types de contraintes non géométriques : l'angle des toits, le matériau utilisé (sous forme de texture), le COS et le CES. Cette fonctionnalité décrite par la figure 8 permet de se rendre compte de manière visuelle de l'impact d'un PLU ou de ses variations sur des parcelles non-bâties.

La détermination se fait grâce à un solveur (*cf.* figure 9) qui utilise les fonctions géométriques de GeOxygene. Certaines opérations 3D nécessaires, telles que l'intersection, étant couteuse en termes de temps, une stratégie est utilisée pour limiter la répétition ce type d'opération.

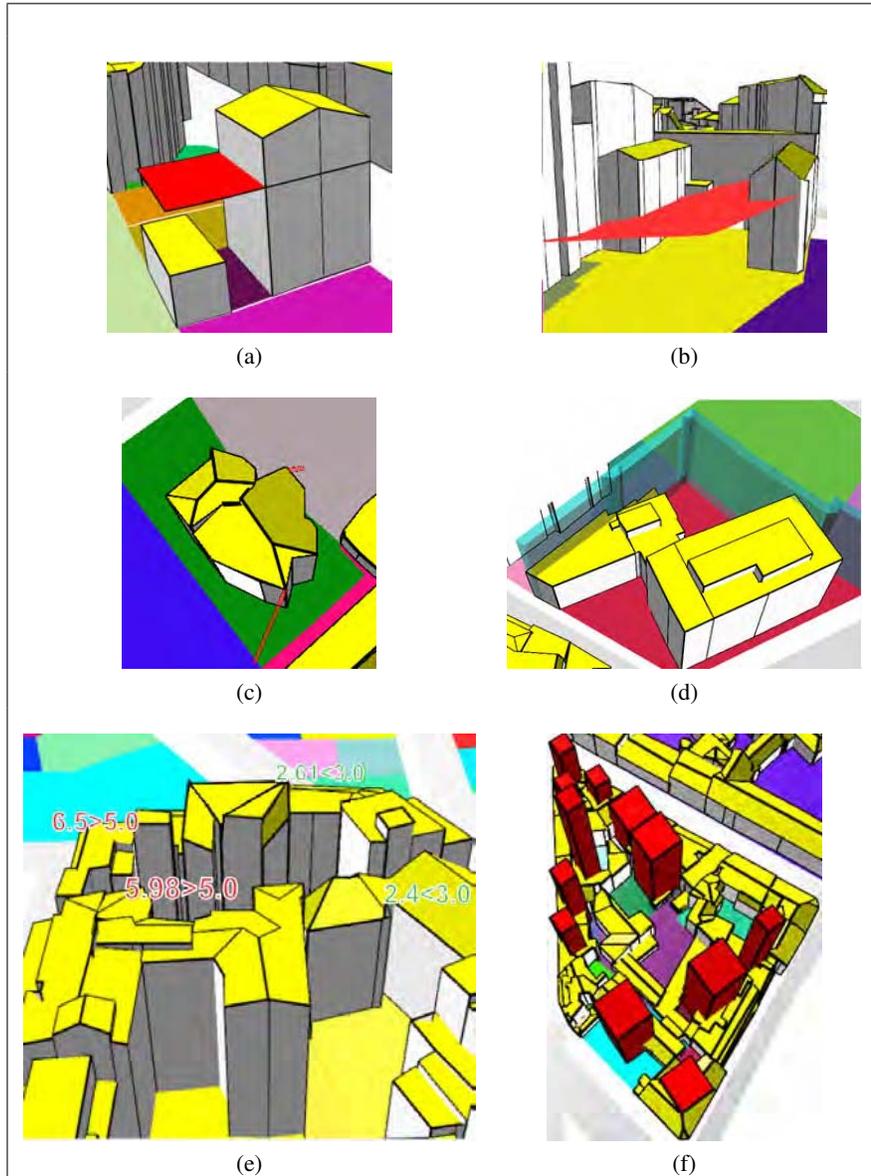


Figure 7 – Non respect de différentes règles. **(a)** Hauteur maximale de 7m. **(b)** Distance de prospect par rapport à la limite de la parcelle. **(c)** et **(d)** « Recul de 2m par rapport aux limites de parcelles ». **(c)** Visualisation par un trait rouge. **(d)** Visualisation par un solide bleu. **(e)** et **(f)** Coefficient d’Occupation des Sols. **(e)** Visualisation par extrusion de bloc. **(f)** Visualisation par affichage des valeurs.

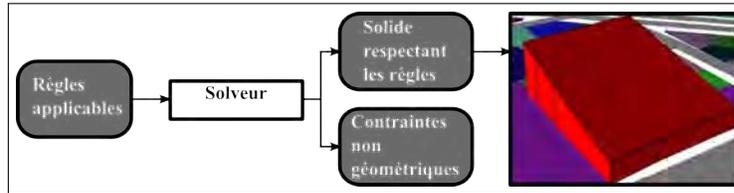


Figure 8 – Création de solide respectant les règles.

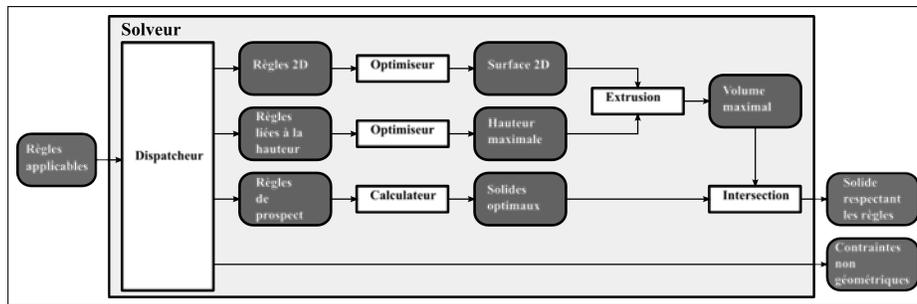


Figure 9 – Organisation du solveur.

La première étape consiste à trier les différentes règles applicables sur la parcelle. Les règles sont classées en 4 groupes qui sont traitées dans cet ordre : les règles 2D, les règles de hauteur, les règles distance de prospect puis, les règles non géométriques.

Les *règles 2D* limitent en 2D la surface de la parcelle (recul en distance euclidienne par rapport à d'autres entités, zone de constructibilité). Elles sont traitées entre elles pour délimiter la surface constructible de la parcelle. Même si aucune règle 2D n'est définie, on retire de la surface constructible de la parcelle les bâtiments qu'elle contient. Si le résultat est une multi-géométrie, chaque élément est traité séparément.

Les *règles de hauteur* limitent la hauteur maximale des bâtiments et les différences de hauteur entre bâtiments. Une hauteur maximale de la parcelle est définie à l'aide des règles de hauteur. Si une règle de type «différence de hauteur entre bâtiments» est applicable, la hauteur minimale acceptable est ajoutée aux règles non géométriques.

Les *règles distance de prospect* font appel à une distance de prospect par rapport à d'autres entités. Le traitement de chaque règle engendre un solide correspondant à la définition de la règle (cf. section 2.1.2).

Les *règles non géométriques* n'interviennent pas directement dans le résultat final dans le cadre du travail présenté ici. L'utilisateur peut également utiliser ce volume final pour vérifier s'il respecte en l'état les paramètres de COS et de CES.

Pour pouvoir générer un volume constructible maximal final, un premier volume est créé issu de l'extrusion de la zone constructible de la parcelle suivant la hauteur

maximale définie pour celle-ci. Ce volume intersecté par l'ensemble des solides définis par les distances de prospect donne le volume constructible final. La figure 10 présente les différentes étapes de générations du solide final à l'aide des règles suivantes : hauteur maximale de 15 m, distance de prospect par rapport à la route de 5m sur la route et un coefficient de pente (a) de 2, recul par rapport aux bordures de parcelle de 4m, recul par rapport à la route de 2m.

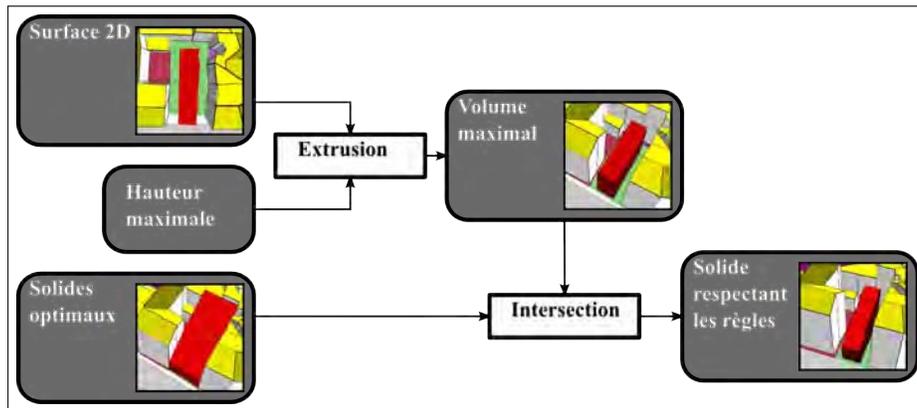


Figure 10 – Etapes pour la génération d'un volume constructible maximal pour une parcelle.

3.3. Utilisation du volume maximal dans le cadre d'un avant projet

Le résultat du calcul précédent reste un peu brut, il propose des résultats qui sont encore loin dans leur forme de résultats plausibles. Pour tenter de les intégrer dans une aide d'avant projet, la possibilité de croiser cette information de volume maximum avec des propositions faites par l'utilisateur est développée. Actuellement, 2 options sont offertes : ajout d'un polygone saisi dans l'application et extrudé, ajout de solides modélisés à l'aide d'un autre outil (CAO/DAO par exemple) et chargés dans l'application. Ainsi, le logiciel propose au final un solide respectant au maximum la forme initiale et les règles géométriques appliquées à la parcelle (cf. figure 11).

4. Conclusion

Cet article propose un outil permettant d'intégrer une partie des règles d'un plan local d'urbanisme (PLU) et de les exploiter à l'aide d'analyse spatiale 3D. L'application permet la visualisation de l'impact de règles en 3D, de vérifier leur respect et de proposer des solutions quant à l'intégration d'un nouveau projet. Le SIG3D intervient ainsi comme un outil judicieux pour mieux appréhender l'espace (notamment vis-à-vis des règles de hauteur) et les contraintes réglementaires. Il s'agit ainsi de fournir

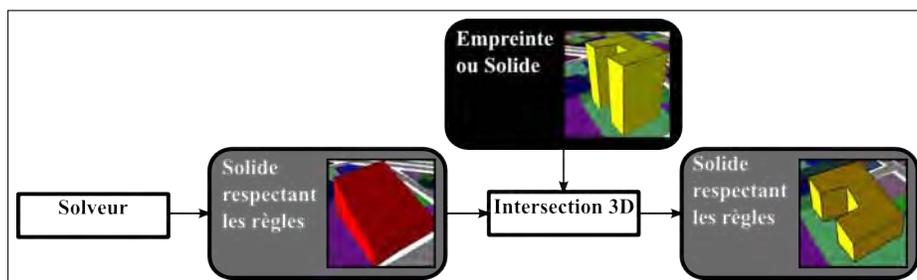


Figure 11 – Calcul d'un volume optimal à partir d'une empreinte.

une méthodologie pouvant aider à la conception, à la concertation et à la décision dans le cadre de projets d'urbanisme. Le PLU reste néanmoins le document de référence, opposable au tiers, et des telles méthodologies ne sauraient se substituer aux méthodes administratives en place. Par ailleurs, le processus mis œuvre requiert la formalisation des règles des plans locaux d'urbanisme et offre ainsi un retour intéressant quant à leur formulation permettant éventuellement d'en souligner les ambiguïtés. À terme, l'intégration de l'évolution de telles règles au cours du temps permettra d'étudier de façon systémique leur impact sur l'évolution du territoire au sein d'une plate-forme de simulation des dynamiques urbaines [PER 09, CUR 10, PER 10].

Actuellement, ces travaux en sont au stade de prototype. À ce stade, les informations concernant les contraintes non-géométriques ne sont pas encore utilisées. Il est possible d'utiliser le vérificateur de règles pour voir si le résultat produit les respecte. De nombreuses perspectives concernant ces règles sont envisageables. La texture pourrait en effet être utilisée par l'application d'une texture non-photoréaliste adaptée et choisie dans une bibliothèque. En ce qui concerne le COS, il pourrait être proposé à l'utilisateur d'indiquer la SHON, un nombre d'étages ou de réduire le bâtiment existant. Pour finir, l'angle de toit pourrait être utilisé pour proposer à l'utilisateur un toit calculé à partir du squelette droit du bâtiment.

L'approche et l'outil proposé ont été testés à travers l'implémentation de quelques PLU, mais une validation auprès de spécialistes du domaine (urbanistes, décideurs, géographes, etc.) reste à effectuer. Les tests ont notamment fait ressortir l'absence de certaines règles relativement fréquentes telles que celles concernant l'alignement des bâtiments qui pourraient être ultérieurement ajoutées. L'ajout de nouvelles règles peut également être pertinent dans le cas où l'on disposerait de données plus précises géométriquement ou sémantiquement (renseignement des balcons par exemple). Les images aériennes et terrestres acquises par un véhicule [DEV 09] et les informations pouvant en être extraites (comme les ouvertures sur les façades [HAM 10] les marquages ou sol ou les panneaux de signalisation [SOH 10]) ouvrent en effet de nouvelles possibilités pour l'utilisation d'un SIG3D pour la conception, la concertation et la prise de décision autour d'un projet urbain.

Pour finir, un travail d'évaluation de l'impact des simplifications effectuées sur la qualité des résultats devra être mené. Il portera en particulier sur l'impact du seuil de 3m pour la hauteur des bâtiments (pour l'évaluation du nombre d'étages pour le calcul du COS), mais aussi sur le choix des données utilisées pour les bâtiments (Bati3D, BDTopo, etc.) et les routes (notamment pour leur largeur). Les problèmes de compréhension des représentation 3D des volumétries urbaines seront par ailleurs traitées dans le cadre d'un travail de recherche plus global sur la sémiologie 3D.

5. Bibliographie

- [BAI 08] BAIN P., MAUJEAN S., THEYS J., « Agora 2020, Vivre, habiter, se déplacer en 2020 : quelles priorités de recherche ? », rapport, 2008, MEEDDAT/DRAST.
- [BRA 09] BRASEBIN M., « GeOxygene : An Open 3D Framework for the Development of Geographic Applications », *12th International Conference on Geographic Information Science (AGILE'09)*, Hanovre, Germany, juin2009.
- [CNI06] « Proposition de sémiologie pour l'édition des PLU à partir d'un SIG », rapport, octobre2006, Groupe de travail "informatisation des PLU", CNIG, pdf.
- [CUR 10] CURIE F., PERRET J., RUAS A., « Simulation of urban blocks densification », *13th AGILE International Conference on Geographic Information Science*, May 2010.
- [DES 95] DESMONTILS E., « Les modeleurs déclaratifs », rapport de recherche n° RR-IRIN-95, septembre1995, Institut de Recherche en Informatique de Nantes, Nantes.
- [DEV 09] DEVAUX A., PAPANODITIS N., PRECIOSO F., CANNELLE B., « Face Blurring for Privacy in Street-level Geoviewers Combining Face, Body and Skin Detectors », *IAPR Machine Vision Applications*, May 2009.
- [DON 92] DONIKIAN S., « Une approche déclarative pour la création de scènes tridimensionnelles : application à la conception architecturale », Thèse de doctorat, Université de Rennes 1, 1992.
- [ESR98] ESRI, « Document de spécification du format ESRI Shapefile », 1998, pdf.
- [FAL 05] FALQUET G., MÉTRAL C., « Integrating Urban Knowledge into 3D City Models », *1st International Workshop on Next Generation 3D City Models*, Bonn, juin2005, ISPRS WG III/4, EuroSDR, DGPF, and University of Bonn, pdf.
- [FAU 01] FAUCHER D., « UrbanLab modélisation déclarative des enveloppes urbaines réglementaires », Thèse de doctorat, Université de Nantes, 2001.
- [HAM 10] HAMMOUDI K., DORNAIKA F., SOHEILIAN B., PAPANODITIS N., « Extracting Outlined Planar Clusters of Street Facades from 3D Point Clouds », *IEEE/CRV - Seventh Canadian Conference on Computer and Robot Vision*, 2010.
- [ISO03] « Geographic information - Spatial Schema ISO Draft International Standard 19107 :2003 », 2003.
- [JAX10] « Site internet du projet JAXB », 2010, website.
- [LE 03] LE V.-T., « Modélisation des volumétries autorisées de construction du bâtiment basées sur les règlements d'urbanisme », Rapport de DEA, École d'architecture de Nancy, 2003.

- [MUR 04] MURATA M., « 3D-GIS Application for Urban Planning based on 3D City Model », *24th Annual ESRI International User Conference*, août2004, p. 9-13, pdf.
- [NIV 99] NIVET M.-L., « De Visu : un logiciel pour la prise en compte de l'accessibilité visuelle dans le projet architectural, urbain et paysager », Thèse de doctorat, Université de Nantes. École d'architecture de Nantes, 1999.
- [PER 09] PERRET J., BOFFET MAS A., RUAS A., « Understanding Urban Dynamics : the use of vector topographic databases and the creation of spatio-temporal databases », *24th International Cartography Conference (ICC'09)*, November 2009.
- [PER 10] PERRET J., CURIE F., GAFFURI J., RUAS A., « A Multi-Agent System for the simulation of urban dynamics », *10th European Conference on Complex Systems (ECCS'2010)*, September 2010, to appear.
- [SOH 10] SOHEILIAN B., PAPARODITIS N., BOLDO D., « 3D road marking reconstruction from street-level calibrated stereo pairs », *ISPRS Journal of Photogrammetry and Remote Sensing*, , 2010.

Optimiser l'efficacité énergétique du bâti et de la ville grâce à la modélisation 3D. Enjeux et perspectives d'application.

Aurélie Prévost* — **Diana Rodríguez *** — **Nathalie Molines*, **** — **Benoit Beckers*** — **Isabelle Caudron*****

* *Equipe de Recherche AVENUES - GSU, Université de Technologie de Compiègne - 1 rue du Docteur Schweitzer, BP 20529, F-60205 Compiègne cedex*

** *CRENAM – (CNRS -UMR 5600), Université de Saint Etienne - 6 rue Basse des Rives, F-42023 Saint Etienne cedex 02*

*** *Règlementation et droit des sols, DGDUD - Communauté Urbaine du Grand Toulouse - 1 place de la légion d'honneur BP 35821, F-31505 Toulouse cedex 5*

RÉSUMÉ. Les villes et plus particulièrement le bâti consomment énormément d'énergie. Si en France la réglementation thermique et les principes d'urbanisme destinés à améliorer l'efficacité énergétique existent, on montrera les limites des méthodes de calculs des apports solaires actuelles et la nécessité de réaliser des simulations d'apports solaires sur un contexte urbain modélisé et réaliste. L'article présentera les limites actuelles des modèles 3D urbains pour la simulation solaire, puis un modèle 3D urbain de la ville de Compiègne exploitable pour les simulations physiques. On conclura en montrant les différentes perspectives d'application de ce type d'outil au bâti, et à la ville.

ABSTRACT. Built-up areas and more particularly buildings are important energy-consumers. If, in France, energetic regulation and principles of town planning intended to improve the energy efficiency do exist, we will show the limits of the methods of calculation for solar radiation and the necessity of making solar simulations on a realistic urban model. The paper will explain the current lack of 3D urban models for solar simulation, and then will present a 3D urban model of the city of Compiègne, adapted for simulations. The conclusion will show the different perspectives of application of such a tool, on the building, and for urban areas.

MOTS-CLÉS : modèle 3D urbain, simulation solaire à l'échelle urbaine, réglementation thermique, principes d'urbanisme, formes urbaines et ensoleillement.

KEYWORDS: 3D urban model, solar simulation in the urban scale, energetic regulation, principles of town planning, urban forms and solar energy.

1.Introduction

Comme le souligne Grenier, le secteur résidentiel et tertiaire français consomme plus de 43% de l'énergie finale (Données françaises de 2006, dans [GRE 07]). Au

sein même du parc bâti, ces consommations varient selon l'âge, le type de bâti, de chauffage et d'énergie utilisée [BRO 07].

Au-delà des consommations du bâtiment en tant que tel, la forme urbaine est également mise en cause : on connaît l'attrait des français pour la maison individuelle, qui « est un rêve pour 82% des français » [DJE 04] et les problèmes liés au développement du périurbain et à la mobilité.

La prise de conscience de cette consommation énergétique excessive a amené à la définition de principes pour une conception du bâti moins gourmande en énergie. Nous connaissons les solutions techniques sur les équipements performants du bâti (techniques d'isolation, matériaux, matériel de ventilation, de chauffage, de refroidissement...) qui permettent de limiter les consommations d'énergie. Ces solutions sont nombreuses et trouvent leur place dans la conception des bâtiments neufs, voire dans la réhabilitation énergétique de l'existant.

Cet article traite principalement du parc bâti à venir et de la réflexion sur l'efficacité énergétique la plus en amont possible qui l'accompagne : la conception doit désormais (ré)intégrer des principes logiques sur la localisation, l'orientation et la disposition des pièces de la construction avant même de penser à l'optimisation par les équipements techniques.

Par ces différents biais, la construction peut devenir « passive », « positive » ou encore « bioclimatique ». Derrière ces appellations se cachent des objectifs de consommations énergétiques chiffrés, encadrés par la réglementation énergétique à l'échelle du bâti.

Or l'article montrera aussi que la question énergétique et sa réglementation sont en fait dépendantes de la question de l'imbrication des échelles à savoir celle du bâti, du quartier et de la ville. Pour cela, il expliquera les limites de calculs thermiques sur un bâtiment « isolé », et présentera les outils à l'échelle de la ville qui peuvent être mobilisés, notamment les documents d'urbanisme qui peuvent conditionner les formes urbaines et architecturales.

A partir de ces éléments de contexte, il abordera la question d'une évaluation énergétique qui prend en compte l'échelle supérieure au bâti, c'est-à-dire l'îlot ou le quartier. La méthode d'évaluation développée par notre laboratoire met en œuvre la création d'une maquette 3D de l'environnement urbain, et un outil de simulation des apports solaires.

Nous verrons ainsi dans une première partie quelles règles cadrent ces problématiques énergétiques. Dans une deuxième partie, nous soulignerons les pré-requis d'une simulation solaire et les limites des modèles 3D actuels. Nous terminerons cet article en présentant le modèle 3D GISOL et ses perspectives d'application.

2. La réglementation au service de l'efficacité énergétique

Il existe deux moyens réglementaires de traiter la question énergétique : la première est la Réglementation Thermique à l'échelle du bâti et la seconde passe par

l'urbanisme réglementaire, c'est-à-dire les règlements des plans locaux d'urbanisme (PLU) qui conditionnent les évolutions de l'urbanisation et plus particulièrement la forme et l'implantation des constructions.

2.1. La réglementation à l'échelle du bâti

Les évolutions de la Réglementation Thermique (actuellement, RT2005 et bientôt RT2012) vont dans le sens d'une très forte incitation à l'efficacité énergétique. Issue du protocole de Kyoto visant à réduire les émissions de gaz à effet de serre en provenance des bâtiments, les Réglementations Thermiques ont pour but de fixer une limite maximale à la consommation énergétique des bâtiments neufs pour le chauffage, la ventilation, la climatisation, la production d'eau chaude sanitaire et l'éclairage.

Le maître d'ouvrage s'engage lors de sa demande de permis de construire à appliquer les règles de construction et donc la réglementation thermique.

Il doit pouvoir justifier du respect de la RT 2005 :

- soit au moyen d'une étude thermique. Le maître d'ouvrage doit dans ce cas pouvoir fournir une synthèse d'étude thermique.
- soit sans calcul à l'aide de solutions techniques agréées par le ministre chargé de la construction et de l'habitation.

Cependant, les modes de calcul associés à ces RT se focalisent sur le bâtiment en occultant son environnement. Ils ne permettent pas d'évaluer correctement un « simple » ilot urbain composé de parties construites à différentes époques et avec des techniques variées. Plus qu'un problème de calcul, on peut penser qu'il s'agit d'une limitation intrinsèque de techniques nécessairement basées sur des valeurs forfaitaires et des calculs simplifiés. La thermique du bâtiment devrait être complétée par une thermique urbaine qui suive ses propres raisonnements et génère une réglementation originale pour le contrôle de l'efficacité thermique à l'échelle urbaine (problème multiéchelle, voir [BEC 09]).

Les plans locaux d'urbanisme définissent des règles applicables simultanément à deux échelles : ils contiennent des règles d'urbanisme générales à toute une zone urbaine, dont les principes concrets s'appliquent à toute nouvelle construction comprise dans cette zone. Ainsi, de cette manière, s'ils peuvent recommander voire inciter (sans pour autant obliger), à l'intégration des dispositifs « écologiques » intégrés au bâti tels que toits végétalisés, panneaux photovoltaïques, etc., les documents d'urbanisme permettent également d'intervenir sur la question de l'ensoleillement du bâti. Ils peuvent alors, par leurs règles, conditionner deux éléments primordiaux : l'orientation des constructions et la gestion des masques solaires créés par l'environnement urbain.

En premier lieu, la question de l'orientation du bâti est souvent un élément négligé dans les plans locaux d'urbanisme qui se contentent souvent « en milieu traditionnel, à reproduire la règle quasi générale de la ville ancienne qui est l'implantation en bordure de voie. Or cette règle n'est pas la seule. On doit la combiner avec d'autres facteurs locaux tels que l'orientation par rapport au soleil, aux vents dominants, aux pentes... » [LAS 02]. D'une manière générale, on

s'accordera à dire que l'orientation Nord-Sud est favorable pour l'ensoleillement et la ventilation traversante.

Concrètement, de manière réglementaire, seulement deux articles des PLU sur les 14 mobilisables, sont obligatoires (les articles 6 et 7). Le premier est relatif aux implantations des constructions par rapport aux voies et emprises publiques, et le second à l'implantation des constructions par rapport aux limites séparatives. L'article 6 permet deux modalités d'écriture : « soit une implantation en limite de la voie ou de l'emprise publique (on utilise alors l'expression, parfois au prix d'un abus de langage, « à l'alignement » et l'objectif est de créer un front bâti sur voie), soit à une implantation en retrait de cette limite. » [PIT 08]. Quant à l'article 7, il oblige à une implantation du bâti soit sur une ou deux des limites de l'unité foncière, soit à une distance définie de ces limites. L'orientation est contrainte par la structure viaire lorsqu'on est « à l'alignement », et est en revanche libre lorsque la construction est « en retrait » (s'il s'agit d'un retrait minimal et non imposé) et sous couvert d'une liberté d'implantation par rapport aux limites séparatives. Cependant, même lorsque le demandeur du permis de construire a le loisir de choisir l'orientation qui convient le mieux, on observe de manière générale que les constructions s'alignent les unes par rapport aux autres, selon une orientation certes en retrait mais en « façade sur rue », même s'il n'y a pas de règle qui y oblige. Cela peut s'expliquer par le fait que les demandeurs des permis de construire s'alignent sur la forme du tissu existant, ce qui est parfois le cas des zones pavillonnaires. Une telle modification des habitudes ne peut se faire que par un rôle de conseil : les instructeurs des permis de construire devraient informer les demandeurs sur ces questions. Le rôle des architectes serait aussi primordial, mais la situation française par rapport à la maison individuelle ne favorise pas ces innovations : seulement 5% des maisons individuelles sont dessinées par des architectes [MAS 08].

Il va de soi que l'on ne peut pas imposer une orientation Nord-Sud sur tout un territoire communal. Premièrement, le choix d'une orientation N-S est à corréluer à la limitation des masques solaires induits par l'environnement de la construction. En effet, privilégier une orientation optimale des façades Nord-Sud n'a plus vraiment d'utilité si le rayonnement solaire est masqué par les constructions voisines. De plus, certains secteurs d'intérêt patrimonial et historique fort ne peuvent pas voir leur paysage urbain évoluer radicalement. En revanche, il s'agit de dépasser un mode de gestion du droit des sols conservateur, sur les secteurs où cela ne se justifie plus. Enfin, notons aussi que les règlements graphiques des PLU permettent d'assouplir le système en offrant la possibilité d'appliquer des règles sur les seules parcelles désignées par les documents graphiques [ADE 06] ce qui permet de travailler à une échelle plus fine que celle du zonage courant.

Si l'ensoleillement du bâti est optimisé par des principes d'orientation adéquats, le PLU devra également permettre l'intégration de dispositifs architecturaux destinés aux protections solaires tels que les brise-soleil (par le biais de l'article 11 relatif à l'aspect extérieur des constructions et l'aménagement des abords).

Ensuite, le deuxième enjeu, celui de la limitation des masques solaires sur les constructions, peut être appréhendé par l'urbanisme réglementaire par trois biais : l'alignement des constructions entre elles, la distance les séparant, et leurs hauteurs.

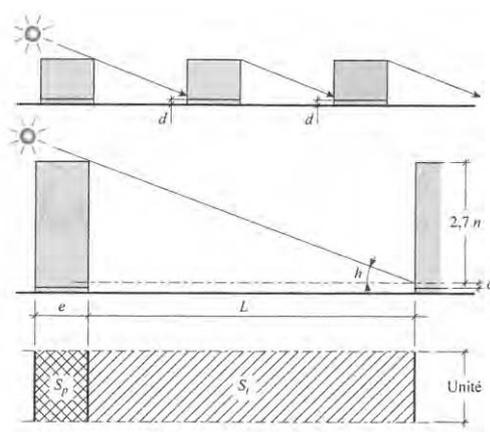
Les principes dépendent évidemment des caractéristiques des zones urbaines concernées, notamment du point de vue de leur densité bâtie.

Dans les zones où la densité urbaine est importante et où il est difficile de compenser l'effet des masques par une distance suffisante entre les constructions, il peut être judicieux de garantir un alignement relatif entre les constructions afin de limiter l'occultation de l'une par l'autre (l'article 6 permet de fixer un recul imposé par rapport à la voie).

La question de la conservation de la lumière en cœur d'îlot en quantité suffisante peut être déterminée en garantissant une constructibilité moindre à l'arrière des unités foncières. Enfin, dans certaines configurations, il peut être intéressant de limiter la hauteur des constructions voisines (notamment au sud) pour limiter les masques et pour optimiser le potentiel d'implantation de panneaux solaires sur les toits [ADE 06].

Dans les zones moins denses, où il n'y a pas de nécessité (ou de possibilité) de densifier, il s'agira de limiter les masques en garantissant une distance minimale entre les constructions. Les PLU ont d'ailleurs les moyens de définir la distance minimale entre construction en fonction de leur hauteur [ADE 06]. Pour les constructions en vis-à-vis, séparées par la voirie, les règlements peuvent soit définir la hauteur des constructions en fonction de la largeur de la voie, soit définir un recul par rapport à la voie selon la hauteur des constructions.

L'intégration d'un paramètre solaire à un calcul simple de distance entre bâtiments selon leur hauteur peut aussi être théoriquement envisagée, en s'inspirant du principe énoncé par Fernandez et Lavigne [FRE 09]. Ces auteurs ont montré, sur un cas simplifié de succession de bâtiments parallélépipédiques de même gabarit, une relation entre le Coefficient d'occupation des sols (COS), h : la hauteur du soleil au pire moment de l'année, n : le nombre de niveaux des immeubles, L : la distance entre deux immeubles et e : leur largeur.



Les relations déterminées sont les suivantes :

$$\tan(h) = (2,7n - a) / L$$

D'où

$$L = (2,7n - a) / \tan(h)$$

Et le COS,
$$\frac{S_p}{S_t} = \frac{k}{\frac{1}{h} + \frac{2,7 - a/n}{e \times \tan(h)}}$$

Avec d le soubassement, a : la hauteur des allèges des fenêtres, k ($k < 1$) qui caractérise la proportion de surface de plancher habitable sur la surface.

Figure 1. Schéma représentant un ensoleillement de façade, premier niveau compris. D'après (Fernandez et al., 2009), p.291

2.2. La gestion de la question de l'énergie par l'urbanisme réglementaire à l'échelle de la ville

A l'échelle urbaine, la loi de programmation fixant les objectifs de la politique énergétique de 2005, dite loi POPE, prévoit de réduire l'intensité énergétique de 2% par an d'ici 2015, puis de 2,5% par an d'ici 2030 et atteindre 10% de la production d'énergie par les énergies renouvelables en 2010. Pour ce qui est de ses orientations effectives, la loi POPE implique directement les collectivités pour agir sur la forme urbaine par des principes d'urbanisation plus « raisonnés ». Les orientations de la politique énergétique définies dans l'annexe de la loi du 13 juillet 2005 précisent que « Les collectivités compétentes définissent des politiques d'urbanisme visant, par les documents d'urbanisme ou la fiscalité locale, à une implantation relativement dense des logements et des activités à proximité des TC et à éviter un étalement urbain non maîtrisé » [BRO 07]. En définitive, la loi POPE mobilise la planification urbaine et ses leviers que sont l'urbanisme réglementaire et ses documents-outils : le SCOT (Schéma de Cohérence Territoriale) à l'échelle supra-communale et le PLU (Plan Local d'Urbanisme) à l'échelle communale.

Concrètement, les orientations de planification contenues dans le PLU doivent être compatibles avec celles du SCOT, et ces orientations sont traduites de manière concrète dans le règlement du PLU. Ce dernier conditionne les constructions futures, que ce soit pour l'occupation du sol, la volumétrie, ou encore la disposition. Un zonage du territoire communal permet également l'adaptation des règles dictées selon la morphologie urbaine générale et les caractéristiques des espaces.

Ainsi, le règlement du PLU permet d'agir sur la question de l'énergie, du point de vue de la forme urbaine et des déplacements, par le biais de son zonage réglementaire. Il favorisera, par exemple, la mixité fonctionnelle pour éviter de tomber dans l'écueil du « zoning » cher à Le Corbusier [COR 71].

Pour augmenter le nombre de logements en ville afin de répondre à la forte attractivité des agglomérations, il peut aussi encourager la densification de certains secteurs géographiques et, pourquoi pas, apporter dans les tissus périurbains de nouveaux types d'habitats dits « intermédiaires » ou « alternatifs ». Cela peut avoir un effet sur l'efficacité énergétique de la ville, car les bâtiments collectifs, même d'échelle raisonnable, ont globalement une conductance par unité de volume meilleure que la maison individuelle [FER 09].

Le PLU peut finalement s'attacher à la mise en œuvre d'orientations inter-sectorielles, telles que la cohérence entre transports urbains et urbanisme. Il s'agit de jouer sur des modulations de la densité bâtie à proximité de ces infrastructures. L'idée de construire avec une densité suffisante au voisinage des stations de transport en commun pourrait renforcer l'attractivité de ces modes en facilitant les déplacements des habitants [JOU 07].

Nous avons dans cette première partie mis en évidence les limites des réglementations thermiques pour la prise en compte de l'énergie, et montré les leviers mobilisables d'un point de vue de l'urbanisme réglementaire. Or, si ces règles influent sur la forme des constructions et plus largement de l'urbanisation, il semble nécessaire de pouvoir en quantifier les effets. Disposer d'un modèle 3D urbain est un pré requis pour ce type d'études. Nous verrons au chapitre suivant que

les modèles 3D actuels ne sont pourtant pas adaptés et qu'il est nécessaire d'en créer un nouveau type, répondant à des critères spécifiques.

3. Les modèles 3D urbains actuels et leurs limites pour les études solaires

3.1. Les besoins autour de la simulation solaire

Les enjeux autour du gisement solaire sont, essentiellement, au nombre de quatre. Le premier d'entre eux est celui du potentiel solaire, offert par les toits et le haut de certaines façades, pour l'installation de panneaux solaires thermiques et photovoltaïques. La demande actuelle impose à la fois une quantification globale de ce potentiel, y-compris à l'échelle territoriale, mais aussi une connaissance très fine des meilleurs emplacements à l'échelle locale, pour un usage individuel ou mutualisé. Dans le cas du photovoltaïque, la forte sensibilité aux ombrages, même partiels et temporaires [WOY 03], exige une modélisation prenant en compte tous les saillants du toit (par exemple, les cages d'ascenseur et les cheminées).

Le second enjeu est celui des apports solaires, qui participent du bilan énergétique des bâtiments, celui-ci étant fortement modifié dans un environnement dense, où les édifices se font de l'ombre les uns aux autres. Il faut distinguer les apports solaires sur les surfaces vitrées de ceux qui atteignent les façades opaques. Comme nous l'avons mentionné plus haut, il est malaisé d'utiliser ces apports solaires corrigés dans le cadre de la réglementation thermique actuelle. Un point d'entrée à l'échelle urbaine est offert par les études d'impact, qui consistent à évaluer l'impact d'une nouvelle construction sur l'ensoleillement des bâtiments existants. Ce problème devient directement quantifiable lorsque le rendement d'installations thermiques ou photovoltaïques est affecté.

Le troisième problème est celui de l'illumination naturelle, et du « droit au ciel », tel qu'il se manifeste dans les quartiers les plus denses où des réglementations spécifiques peuvent s'imposer (voir le cas extrême de Hongkong, [NG 03]). Pour l'éclairage, la couleur des façades joue certainement un rôle important, mais qu'il reste encore très difficile de quantifier à l'échelle urbaine, à cause des temps de calcul qui deviennent considérables. Toutefois, des progrès importants sont probables à moyen terme, et des études originales viendront sans doute bientôt enrichir la réflexion réglementaire à ce sujet.

Enfin, l'ensoleillement et l'ouverture des rues sur le ciel jouent un rôle très important dans la constitution des microclimats urbains [MAR 03]. A cause de la dimension multi-physique du problème (échanges radiatifs, aéraulique, évapotranspiration,...), les simulations ne sont pas encore assez précises pour pouvoir guider la réglementation, et des progrès marquants ne peuvent raisonnablement pas être attendus avant plusieurs années.

Nous avons montré précédemment que les évolutions récentes de l'urbanisme réglementaire produisent une demande de simulation numérique, touchant en particulier aux différents aspects du gisement solaire. Dans le paragraphe suivant, nous montrerons les progrès récents dans la modélisation 3D des ensembles urbains ainsi que leurs limites, puis nous décrirons un modèle 3D réalisé pour permettre des

simulations de l'ensoleillement relatives aux deux premiers enjeux rappelés ci-dessus : le potentiel solaire et les apports solaires.

3.2. Limites des modèles 3D actuels

Depuis quelques années, de nombreuses collectivités ont lancé des projets liés à la réalisation de maquettes numériques 3D d'ensembles urbains. Les applications des modèles géométriques sont très variées, par exemple le développement urbain, l'aménagement du territoire, la concertation publique et l'accompagnement de projets architecturaux. On peut aussi les utiliser comme support pour différentes simulations physiques (radiation solaire, illumination, thermique urbaine, etc.) et simulations d'inondations.

Les données SIG (Système d'Information Géographique) sont des données géographiques géo-référencées, qui contiennent entre autres les contours au sol, et la hauteur des bâtiments. De telles bases de données permettent de construire une maquette numérique 3D, laquelle est constituée d'un socle élaboré à partir d'un Modèle Numérique de Terrain (MNT), représentant le relief, sur lequel on peut ajouter des informations diverses (photographies aériennes, cartes d'occupation du sol, bâtiments, etc.).

En 2007, la norme CityGML (OGC, 2008) a été validée sur le plan international par l'OpenGeospatial Consortium, afin de fournir une structuration standard des bâtiments dans un environnement urbain [KOE 08]. Elle définit cinq niveaux de détails consécutifs (Level of Detail, LoD), étiquetés de 0 à 4.

Le *LoD 0* est constitué du modèle numérique de terrain (MNT), sur lequel une image aérienne ou une carte peut être drapée. Le *LoD 1* est le modèle de blocs, les bâtiments sont représentés par des polygones extrudés avec des toits plats. Le *LoD 2* contient des bâtiments avec une typologie de toits et les surfaces thématiquement différenciées. La végétation peut également être représentée. Le *LoD 3* indique un modèle avec des bâtiments détaillés, comportant des façades (avec ouvertures) et des toits complexes. Des textures de haute résolution peuvent être drapées sur ces éléments. Il contient aussi de la végétation détaillée. Le *LoD 4* enrichit le *LoD 3* en ajoutant une modélisation intérieure des bâtiments (pièces, portes intérieures, escaliers et mobilier) (OGC, 2008).

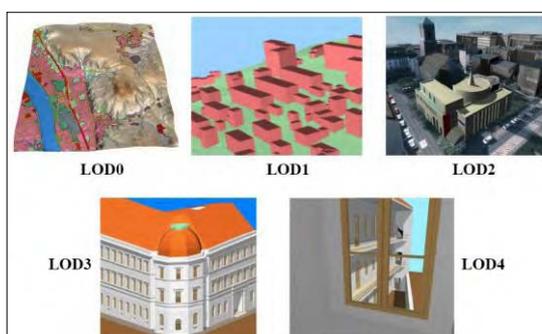


Figure 2. Les 5 niveaux de détail (LoD) définis par CityGML (Source : IGG Uni Bonn).

Néanmoins, ces modèles ne sont pas forcément adaptés pour les simulations physiques à l'échelle urbaine. Les modèles LOD1 sont le résultat d'une simplification très forte, les bâtiments sont représentés sous forme de polyèdres simples, les fenêtres, les portes et les pentes de toits sont négligées, même les hauteurs des bâtiments sont le résultat d'approximations et de généralisations. Les modèles LOD2 sont insuffisants parce que les toits sont issus d'une approximation typologique et les fenêtrages apparaissent comme des textures et non comme des objets. Les modèles LOD3 et LOD4 sont très détaillés et par conséquent, très lourds à manipuler pour les simulations. Le niveau de détail est très précis, car le but principal est d'obtenir des rendus hyperréalistes.

La simulation à l'échelle urbaine de phénomènes physiques, tels que la lumière naturelle et les échanges radiatifs, est généralement très lourde. La géométrie doit donc être très simplifiée. Les bâtiments sont modélisés de manière à maintenir les grandes surfaces (façades et toits) correctement orientées et sans solution de continuité. Il est également nécessaire de modéliser les fenêtrages, ainsi que, éventuellement, les protections solaires et le terrain. Actuellement, on ne peut obtenir de tels modèles que par l'emploi de méthodes manuelles : il est nécessaire de prendre des mesures *in situ*, des photographies et de dessiner le modèle dans un logiciel de CAO.

La reconstitution volumétrique d'ensembles urbains pour les simulations physiques est une question difficile. Comme le souligne Siret [SIR O6] « le nombre, l'hétérogénéité et la complexité des formes bâties sont autant de facteurs rendant généralement laborieuse et coûteuse la réalisation de maquettes 3D de larges zones urbaines ». L'étude à l'échelle urbaine implique une simplification drastique de la géométrie qui permette toutefois de préserver la fiabilité des résultats. Voici donc le problème fondamental : trouver et justifier une simplification appropriée de la ville quand on étudie l'impact de la radiation solaire à l'échelle urbaine. Il faut aussi déterminer la fiabilité de ces modèles et le degré de simplification adéquat.

4. Le modèle GISOL 3 et la simulation des échanges radiatifs sur la ville de Compiègne

4.1. Le modèle GISOL 3 : un modèle adapté aux simulations solaires

Afin de dépasser les limitations des modèles 3D urbains disponibles et de réaliser à l'échelle urbaine des simulations de phénomènes physiques tel que le potentiel solaire, un modèle géométrique 3D du centre-ville de Compiègne a été élaboré. Ce modèle, simplifié mais rigoureux, a été obtenu par l'emploi de méthodes manuelles. Il se compose de 21.735 triangles sous format STL, séparés en trois couches: façades 3759, fenêtres 7.402 et toits 10.574. (Voir Figure 4).

Les façades et les toits sont réduits aux grandes surfaces, correctement orientées et jointes sans solution de continuité. Les toits ont été modélisés avec les chiens-assis, mais les éléments décoratifs, les cheminées, l'épaisseur du mur et le terrain ont été négligés. Les fenêtres sont des surfaces planes situées à un centimètre au dehors de la paroi extérieure. Ce modèle est donc beaucoup plus élaboré que le niveau LOD01 généralement utilisé dans les simulations. Les nouvelles techniques

automatisées (type LIDAR) donnent des modèles en nuage de points que l'on ne sait pas encore adapter correctement pour la simulation. Il s'agit donc ici d'un modèle intermédiaire, spécialement conçu pour les études d'ensoleillement.



Figure 3. Image satellite de la zone modélisée
Surface modélisée : 61900 m²,
périmètre : 1190 m



Figure 4. Vue sud-ouest du Modèle GISOL 3 dans Heliodon 2.

Le modèle 3D a été testé avec le logiciel Heliodon 2, un outil de simulation interactive qui permet de quantifier l'impact énergétique et visuel de la radiation solaire dans les projets architecturaux et urbains¹. Ici, nous n'utilisons qu'un seul paramètre, le plus simple : l'irradiation solaire directe évaluée à partir de la constante solaire en tenant compte de l'absorption atmosphérique par temps clair. La composante diffuse (lumière du ciel), le couvert nuageux et les réflexions ne sont donc pas prises en compte [BEC 07]

4.2. Validation de GISOL 3 : Effet de l'environnement et de la forme des toits sur les résultats de simulation.

Le but principal de l'étude réalisée est de déterminer la fiabilité des résultats du modèle et d'évaluer si le niveau de détail issu de la simplification est adéquat. Deux critères de validations sont présentés dans ce chapitre : l'environnement et la modélisation plus précise des toits inclinés.

4.2.1. Effet de la modélisation de l'environnement sur les résultats de simulation solaire

Pour réaliser le bilan énergétique d'un bâtiment, la réglementation thermique française actuelle (RT2005) impose de quantifier les apports solaires. Néanmoins, cette quantification est réalisée sans prendre en compte l'effet de masque produit par l'environnement immédiat du bâtiment. Cette simplification nuit à la précision des

¹ Le logiciel Heliodon a été développé par Benoit Beckers et Luc Masset en 2003. La version actuelle Heliodon 2 utilisée ici date de 2009 (www.heliodon.net)

calculs. Le but de ce premier test est de quantifier la marge d'erreur issue de ce type de simplification et de mettre en évidence l'importance de modéliser les alentours du bâtiment.

Il s'agit, dans ce cas, de quantifier les apports solaires d'un bâtiment du modèle, et de comparer les différences entre deux configurations : le bâtiment pris dans le contexte urbain, et le bâtiment pris seul. Les paramètres qui influent sur les résultats sont la saison, l'élément qui reçoit l'énergie, l'orientation ainsi que la hauteur des bâtiments environnants. Voir la figure 5.

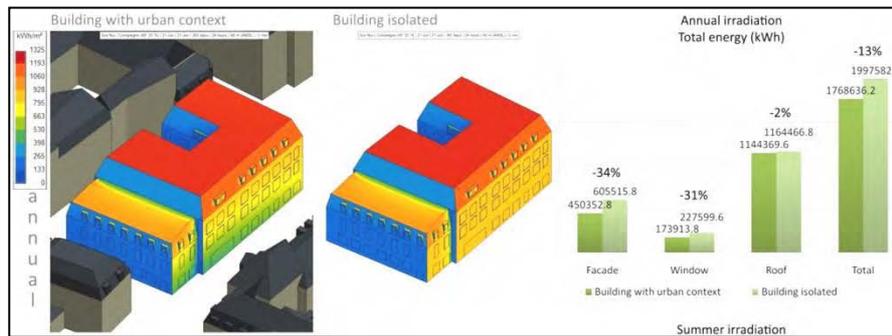


Figure 5. Comparaison des apports solaires dans un bâtiment sans et avec contexte urbain.

Le tableau suivant montre en pourcentage la réduction des apports solaires comme conséquence de la présence du contexte urbain. Les résultats de cette simulation montrent que le bâti environnant réduit considérablement les apports solaires du bâtiment, en particulier sur les façades et les fenêtres. De manière générale, les toits présentent une diminution tout-a-fait négligeable, parce que les bâtiments environnants sont de hauteur similaire au bâtiment étudié.

Période de calcul	Façades	Fenêtres	Toits	Total
Calcul annuel	-34	-31	-2	-13
Calcul du solstice d'été	-29	-21	-1	-9
Calcul du solstice d'hiver	-24	-32	-3	-16

Figure 6. Réduction en pourcentage des apports solaires comme conséquence de la présence du contexte urbain par surface et période de calcul.

Il faut remarquer que pour les calculs d'apports solaires, pour l'évaluation de l'illumination naturelle à l'intérieur du bâtiment et pour les études d'impact, les fenêtres sont les surfaces les plus sensibles.

4.2.2. Effet de la modélisation des toits sur les résultats de simulation solaire.

Dans le nord de la France, on peut trouver des toits très pentus et géométriquement complexes. Néanmoins, la plupart des simulations numériques des phénomènes physiques à l'échelle urbaine est effectuée sur des modèles d'un niveau de détail *LoD1*, c'est-à-dire, que les bâtiments sont représentés avec des toits plats. Cette simplification de la géométrie introduit généralement une erreur supplémentaire sur la précision quant à la hauteur des bâtiments, car dans les bases de données géographiques (type BD TOPO®,) cette précision dépend essentiellement de la source de la donnée, mais aussi de la classe de l'objet [IGN 09]. Une étude de potentiel solaire a été réalisée afin de mesurer la marge d'erreur de modèles *LoD1* et d'évaluer s'il est nécessaire de modéliser les toits avec un niveau de détail plus précis.

Ce test a consisté à comparer les différences d'irradiation sur les toits entre deux configurations : une partie du modèle avec des toits plats et la même partie avec les toits dessinés de manière plus précise. Les calculs d'irradiation ont été faits sur trois périodes : les solstices et le total sur l'année. Le tableau suivant montre les valeurs d'énergie dans les deux modèles par période de calcul, ainsi que les différences en pourcentage. Les résultats de la simulation ont montré que les différences en valeurs dans le calcul annuel sont négligeables (+1%), parce que l'accroissement en été (+6%) est presque entièrement compensé par les fortes pertes hivernales (-33%). D'autres tests montrent que la variation annuelle reste généralement en dessous de 5%.

Période de calcul	Modèles avec des toits plus précis	Modèle avec des toits plats	Pourcentage
Calcul annuel	3075374,9	3121109,3	+1%
Calcul du solstice d'été	16655,3	17711,3	+6%
Calcul du solstice d'hiver	1380,4	930,6	-33%

Figure 7. Valeurs d'énergie dans les deux modèles par période de calcul, ainsi que les différences en pourcentage.

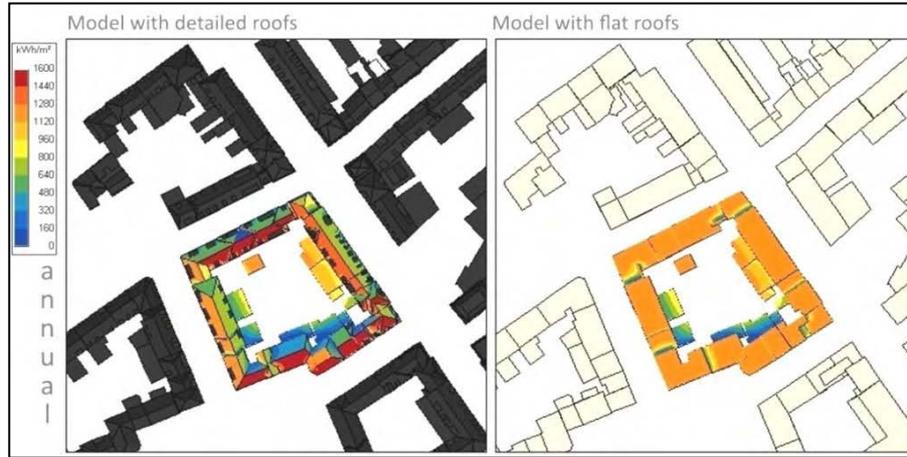


Figure 8. Irradiation annuel solaire sur les toits, à gauche le modèle avec des toits détaillés et à droite le modèle avec des toits plats. (Surface des toits détaillés: 3554,20 m², surface des toits plats: 2830,12 m²)

Les résultats obtenus dans le calcul annuel montrent que si l'objectif de la simulation est d'évaluer le potentiel solaire global, il est possible d'utiliser un modèle LOD01. Néanmoins, ce modèle ne permet pas de déterminer l'emplacement idéal des panneaux solaires, partie cruciale car l'effet d'ombre affecte sérieusement la performance des systèmes solaires actifs (voir Figure 8). On peut constater que seul le modèle avec des toits plus détaillés donne plus de précision sur l'emplacement idéal, ainsi que les valeurs d'énergie et leur distribution. Le modèle avec des toits inclinés devient nécessaire et trouve alors tout son intérêt.

Par ailleurs, le temps de calcul est le double pour les calculs journaliers (solstices) et 2,5 fois plus importants pour le calcul annuel, ce qui reste très acceptable.

Nous avons vu que les modèles géométriques urbains 3D existants n'ont pas un niveau de détail adapté aux critères de modélisation nécessaires pour les simulations solaires, soit par manque de précision (LED01, LED01 et LED02), soit par excès de détail (LED03 et LED04).

Le modèle GISOL 3 nous a permis d'évaluer et de valider le niveau de détail et la simplification nécessaire des modèles urbains pour ces simulations. Pour cela, plusieurs études sur GISOL 3 ont été réalisées, (dont deux présentées ici), la première destinée à valider l'intérêt de la modélisation de l'épaisseur des murs, la seconde concernant la pente des toits, et la troisième à valider l'intérêt de la prise en compte du contexte urbain.

On conclura cette première étude de sensibilité du modèle en affirmant que le niveau de détail du modèle dépend de la finalité de l'étude envisagée et de la précision des résultats attendue. Par exemple, pour l'étude des apports solaires, et sans doute pour les études d'impact comme de l'évaluation de l'illumination

naturelle, il est nécessaire d'intégrer l'épaisseur du mur, car les différences de valeurs sont assez importantes. Néanmoins, il faut prendre en compte les temps de calcul, car le nombre de triangles dans les façades augmentera ce qui rendra le modèle sensiblement plus difficile à manipuler.

Le niveau de détail de la modélisation des toits dépend aussi de l'étude envisagée : pour les calculs de potentiel solaire, elle est évidemment indispensable lorsque l'on s'intéresse à l'emplacement idéal de panneaux solaires. De plus, la prise en compte de toits inclinés sur l'empreinte solaire au sol (quantification des différences de valeurs d'énergie au niveau du sol) devient importante pour les études d'aménagement des espaces publics et pour les études du climat urbain (en particulier pour les rues des quartiers denses des villes du Sud où il peut être intéressant de limiter les apports solaires en été pour conserver la fraîcheur).

Si ces premiers résultats ouvrent un vaste champ de perspectives d'application, l'outil n'est pas encore complètement abouti et cette étude de sensibilité du modèle nécessitera d'être complétée par des tests plus nombreux, afin de définir des indicateurs à portée générale.

Nous avons présenté dans ce chapitre, par l'intermédiaire de ces deux tests, la nécessité de prendre en compte la ville dans son ensemble pour quantifier de manière réaliste les apports solaires. Si nous avons montré des perspectives d'applications à court terme de cet outil, nous verrons en conclusion de l'article d'autres objectifs prospectifs à plus long terme, en amont de la définition des principes d'évolution des villes.

5. Perspectives

Nous avons vu que les méthodes de calcul des apports solaires imposés par la réglementation thermique française en vigueur ne garantissaient pas des résultats proches de la réalité. Les calculs solaires, pour être valables, doivent intégrer le tissu urbain environnant. La méthode présentée ci-dessus, qui associe un modèle 3D urbain adapté aux simulations physiques à un logiciel de simulation solaire permet de fournir de meilleurs résultats. L'objectif n'est pas tant aujourd'hui de vouloir bouleverser les méthodes de réglementation thermique. Nous pensons cependant qu'à l'avenir, lorsque les modèles de villes 3D adéquats se seront généralisés, ce type d'outil pourra lui-même être généralisé pour les études de conception du bâti, pour les études d'impact d'une nouvelle construction sur les autres, et comme outil de labellisation énergétique spécifiquement urbain.

De plus, l'évolution de l'urbanisation est contrainte par l'urbanisme réglementaire. Nous avons vu à ce propos que les plans locaux d'urbanisme français disposent d'un certain nombre de « moyens » réglementaires pour améliorer la situation urbaine et bâtie par rapport à l'énergie. S'il y a des règlements de qualité, avec des règles a priori bien pensées, il n'y a pas pour autant de moyen de « vérifier » quantitativement l'impact de ces règles.

De ce fait, si on projette les développements possibles d'un tel outil, ce type de méthode pourrait servir à la définition en amont de certaines règles. Ainsi, par exemple, si la pente des toits est couramment définie dans les règlements

d'urbanisme, une pente différente mais optimisée pour l'implantation de panneaux solaires pourra être préconisée.

Pour ce qui est de l'implantation des constructions sur l'unité foncière, les règlements d'urbanisme définissent des règles propres à un type particulier de zone. Ces règles (hauteur, emprise au sol, alignement ou retrait par rapport aux voies, distance entre bâtiments...) définissent ce que l'on appelle une « enveloppe » pour les évolutions de l'urbanisation. A l'avenir, on pourra imaginer l'utilisation de ce type de méthode dans la définition de ces « enveloppes ». En effet, par rapport à des orientations d'évolution des tissus urbains, transcrites réglementairement, il serait intéressant de comparer, quantitativement, des scénarios réglementaires par type de zone, par rapport à la question de l'énergie. Si des orientations de développement urbain sont à priori difficilement conciliables (par exemple : densifier, mais limiter les masques solaires), un tel outil pourrait permettre de trouver le meilleur compromis et de définir l'enveloppe réglementaire en conséquence. Ainsi, pour revenir à l'exemple de la pente des toits, si une commune désire optimiser l'utilisation des panneaux solaires, tout en conservant un style de toit proche de l'existant, l'outil pourra déterminer l'ordre d'idée qui permettra ce compromis et quantifier cet apport.

Notons enfin que l'outil, utilisé comme aide à la décision et de communication auprès des élus des villes pourrait favoriser une prise de conscience sur l'impact de règles trop conservatrices ou subjectives définies sur des secteurs géographiques sans réel intérêt patrimonial, et à terme, pourrait favoriser la modification de ces règles.

Que ce soit pour la thématique énergétique ou toute autre thématique de développement durable, nous pensons qu'il est indispensable de s'intéresser aux effets et aux possibilités d'optimisation des règlements d'urbanisme. D'une part, l'urbanisme réglementaire, qui joue pourtant un rôle important dans la « fabrique » de la ville de tous les jours ne fait l'objet d'aucune réelle valorisation. De plus, ces questions d'analyse des effets des documents d'urbanisme se justifient par les éléments d'évaluation environnementale qu'ils doivent comporter.

En effet, depuis la loi Solidarité et renouvellement urbain de 2000, tous les documents doivent comporter un « profil environnemental », une analyse des incidences notables prévisibles du plan, ainsi que la présentation des mesures de compensation prévue. Si la précision scientifique et donc la quantification n'est pas exigée pour ce type d'étude environnementale de droit commun puisqu'il s'agit d'une obligation de moyens et non de résultat (Jegouzo, 2009), on peut supposer que la demande face aux préoccupations environnementales justifiera ce type de simulation.

Mais depuis la directive plans et programmes de 2001, traduite en droit français en 2004, certains documents d'urbanisme doivent fournir une évaluation environnementale renforcée, qui prévoit en plus de l'étude de droit commun une analyse des effets du plan à l'expiration d'un délai de 10 ans après son approbation. Face à cette obligation, on assiste actuellement à une effervescence d'études et de travaux d'Évaluation au sein des collectivités². Ce renforcement de l'exigence

² Voir la journée *Journée Réseau PLU 2008* – Evaluation environnementale des documents d'urbanisme organisée par le CERTU et le CGDD ou ont été présentées diverses méthodes.

environnementale ajoutée aux premières obligations relatives à la loi SRU semble conforter le besoin d'outils nouveaux pour la simulation de phénomènes en milieu urbain.

Quant au problème physique et numérique, pour terminer, on a montré qu'un modèle géométrique non standard est nécessaire pour une bonne qualité de simulation; le modèle ici proposé n'est pas directement opératoire, à cause du travail considérable que son élaboration requiert, mais il pourra servir comme objectif pour les chercheurs en reconnaissance de formes qui travaillent actuellement à l'élaboration semi-automatique de modèles 3D urbains. Des développements ultérieurs de l'étude de sensibilité permettront d'optimiser le niveau de détail requis, mais on peut déjà conclure que le modèle d'un quartier urbain comprendra forcément quelques dizaines de milliers de triangles. C'est un ordre de grandeur atteignable pour les simulations numériques actuelles. Il devrait donc être bientôt possible de résoudre complètement le problème radiatif (ici, nous nous sommes limités à la composante solaire directe) à l'échelle du quartier étendu, voire à l'échelle urbaine. Cependant, les développements numériques complexes impliqués ne seront réellement utiles que s'ils sont réalisés dans le but de mieux quantifier les conséquences physiques de la réglementation urbaine (en termes d'économies d'énergies ou de bilan CO₂, par exemple), de manière à pouvoir guider son évolution. Ce dernier point justifie un travail pluridisciplinaire approfondi dont cette communication présente les tous premiers résultats.

Bibliographie

- [ADE 06] ADEME (2006). Réussir un projet d'urbanisme durable. Méthode en 100 fiches pour une Approche Environnementale de l'Urbanisme (AEU).
- [BEC 07] Beckers, B., Masset L. & P. Beckers (2007). « Una proyección sintética para el diseño arquitectónico con la luz del sol. » 8th Ibero-american Conference of Mechanical Engineering CIBIM08, October 23-25, Cusco, Perú, Libro virtual: ISBN 978-9972-2885-3-1 (see Conference Proceedings, Aspectos medioambientales)
- [BEC 09] Beckers, B. and D. Rodriguez (2009). "Helping architects to design their personal daylight." WSEAS Transactions on Environment and Development Volume 5(Issue 7): pp 467-477.
- [BROU 07] Brouant, J.-P. (2007) PLU et énergie. Fiche 1 Introduction. Séminaire écriture des PLU, GRIDAUH
- [COR 71]Corbusier, L. (1971). La Charte d'Athènes. Paris, Editions du Seuil.
- [DJE 04] Djefal, S. and S. Eugene (2004). "Etre propriétaire de sa maison, un rêve largement partagé, quelques risques ressentis." revue Consommation et modes de vie, CREDOC N°177.
- [FER 09] Fernandez, P. and P. Lavigne (2009). Concevoir des bâtiments bioclimatiques Fondements & méthodes. Paris, Le Moniteur.
- [GRE 07] Grenier, A. (2007). Ville et énergie. Spécificité et complexité de la question en France. Les Annales de la recherche urbaine, © MEDAD, PUCA. n°103: pp.128-136

- [JEG 09] Jegouzo, Y. (2009) Écriture du PLU. L'évaluation environnementale des plans locaux d'urbanisme. Séminaire écriture des PLU, GRIDAUH
- [JPU 07] Jourdan, G. (2007). Débat public sur le projet de grand contournement autoroutier de Toulouse – contribution de Gabriel Jourdan, maître de conférences en urbanisme, université Pierre Mendès France Institut d'urbanisme de Grenoble. Toulouse.
- [IGN 09] Institut Géographique National. (2009).« BD TOPO® Version 2 – Descriptif de contenu ».p8,76-112.
- [KOE 08] Koehl, M., E. Meyer, C. Koussa and C. Lott (2008). SIG 3D ET 3D dans les SIG : Application aux modèles patrimoniaux. GéoEvénement 2008, France (2008).
- [LAS 02] Lassus, P. (2002). Harmonie et règle urbaine, *Economica*, 369p..
- [MAR 03] Martilli, A., Y.-A. Roulet, M. Junier, F.Kirchner, M. Rotach and A. Clappier (2003). "On the impact of urban surface exchange parameterizations on air quality simulations : the Athens case " *Atmospheric Environment*, 37, 2003, pp. 4217–4231.
- [MAS 08] Masbouni, A. (2008). « Dix leçons européennes sur le projet urbain ». dans Cuillier, F (sous la dir. de), *Fabriquer la ville aujourd'hui*, Les débats sur la ville 7, a'urba, édition Confluences.
- [NG 03] Ng, E. (2003). Applying computational simulation results to the development of a design method for daylighting design and regulation in highdensity cities, Eighth Intern. BPSA Conf. Eindhoven, Netherlands.
- [OGC 08] OGC (2008). OpenGIS® City Geography Markup Language (CityGML) Encoding Standard. T. H. K. Gerhard Gröger, Angela Czerwinski, Claus Nagel, Open Geospatial Consortium Inc.
- [PIT 08] Pittard, Y. (2008) L'ÉCRITURE DES ARTICLES 6, 7 ET 8 DU RÈGLEMENT DES PLU. Fiche 3 Les spécificités de l'écriture de l'article 6. Séminaire écriture des PLU, GRIDAUH
- [SIR 06] Siret, D., M. Musy, F. Ramos, D. Groleau and P. Joanne (2006). "Développement et mise en oeuvre d'un SIG 3D en environnement urbain." *RIG 16/2006*: pp 71-91.
- [WOY 03]Woyte, A., J. Nijs and R. Belmans (2003). "Partial shadowing of photovoltaic arrays with different system configurations: Literature review and field test results." *Solar Energy* 74(3), 217-233, 2003

Session 2 : simulation spatio-temporelle

Simuler la densification du tissu urbain au moyen d'un processus de peuplement

Florence Curie* — Annabelle Mas** — Julien Perret*** — Anne Puissant**** — Anne Ruas****

* UMR ISTO, Université de Tours

Parc de Grandmont, 37200 Tours, {prenom.nom}@univ-tours.fr

** Laboratoire CEDETE, Université d'Orléans

10 rue de tours, BP 46527, 45065 Orléans cedex 2, {prenom.nom}@univ-orleans.fr

*** Laboratoire COGIT, IGN

73 avenue de Paris, 94160 Saint Mandé, {prenom.nom}@ign.fr

**** Laboratoire Image, Ville, Environnement, Université de Strasbourg

3 rue de l'Argonne, 67000 Strasbourg, {prenom.nom}@live-cnrs.unistra.fr

RÉSUMÉ. L'approche présentée dans cet article utilise un système multi-agent pour modéliser du point de vue morphologique les propriétés auto-organisatrices des systèmes complexes urbains et vise à fournir un outil permettant d'étudier la croissance urbaine. Le système multi-agent proposé comporte une hiérarchie d'agents topographiques (e.g. les bâtiments, les routes, les cours d'eau, les îlots, la ville) qui peuvent être construits, modifiés, fusionnés, découpés, restructurés et détruits au cours du temps. Le comportement de chaque agent est contrôlé par un ensemble de règles d'évolution, de contraintes et d'actions associées. La question particulière traitée dans cet article concerne les méthodes de peuplement qui permettent la création de nouveaux agents tout en respectant une certaine organisation spatiale et un certain processus de construction. En effet ces méthodes nécessitent de trouver un compromis entre les besoins de la simulation et les connaissances géographiques. Dans cet article, l'accent a été mis sur les zones industrielles et les zones d'habitat de type individuel et collectif.

ABSTRACT. The approach presented in this paper uses a multi-agent system to model the self-organising properties of urban systems from a morphological point of view and aims at providing a tool to study urban growth. The proposed multi-agent system is a hierarchy of topographic agents (e.g. buildings, roads, rivers or urban blocks) that are built, modified, merged, split, restructured and destroyed over time. Each agent's behaviour is controlled by a set of evolution rules, constraints and associated actions. This paper deals especially with populating functions which allow the creation of new agents while respecting a certain spatial organisation and a certain construction process.

MOTS-CLÉS : tissu urbain, simulation urbaine, systèmes multi-agents, croissance urbaine, peuplement

KEYWORDS: urban fabric, urban simulation, multi-agent systems, urban growth, populating

1. Introduction

La simulation des dynamiques urbaines est un domaine de recherche très actif. Parmi les approches existantes, l'approche morphologique présente l'intérêt de limiter le nombre de variables manipulées et de simplifier la mise en place et la maintenance de bases de données historiques permettant l'étude de telles dynamiques puisque les informations morphologiques sont plus facilement mesurables que les informations socio-démographiques par exemple [BAT 86]. De nombreuses méthodes existent pour représenter et étudier les dynamiques spatiales. [ROD 07, LEF 08] proposent notamment un modèle pour étudier les dynamiques urbaines dans la longue durée.

En termes de simulation, les automates cellulaires ont été très utilisés pour la simulation urbaine et ont permis la simulation de nombreux phénomènes complexes tels que la croissance et la ségrégation [COU 97, WHI 97, BAT 04, BAT 08, WER 09]. Néanmoins, comme le font remarquer de nombreux auteurs, les automates cellulaires supposent une certaine homogénéité de l'espace et du voisinage des cellules de l'automate. Ainsi, plusieurs modèles ont été proposés afin d'intégrer l'influence du réseau urbain et de contraintes spatiales sur l'accessibilité des cellules de tels automates [BAT 97]. D'autres auteurs ont par ailleurs proposé d'étendre le formalisme en y intégrant la notion de graphe [O'S 00, O'S 01, BAD 07], qui permet de mieux prendre en compte les voisinages des objets géographiques et les irrégularités de la structure urbaine.

Le projet GeOpenSim vise au développement d'une plate-forme Open Source pour l'analyse et la simulation des dynamiques urbaines à l'aide de bases de données topographiques historiques [PER 09]. Ces dernières sont construites à l'aide de bases de données topographiques récentes ainsi que de photographies aériennes et/ou de cartes historiques. Ces bases de données sont analysées afin d'identifier des règles d'évolution à différents niveaux meso-géographiques (e.g. la ville ou l'îlot), ainsi que des statistiques sur des objets spécifiques comme les bâtiments. Chaque simulation peut ainsi être paramétrée à l'aide de règles d'évolution provenant de l'analyse automatique des données historiques, mais aussi de règles énoncées par des experts. Ces règles sont en principe similaires aux règles de transition des automates cellulaires et peuvent prendre en compte le voisinage des agents ainsi que leurs différents attributs (pour les îlots : la densité, le type d'îlot, la localisation de l'îlot par rapport à la ville, etc.). Ces règles peuvent par ailleurs changer en fonction de la période à laquelle l'évolution simulée a lieu, permettant ainsi d'implémenter des évolutions spécifiques à une période donnée. Ces évolutions peuvent différer en vitesse (la croissance d'une ville n'est pas un processus régulier) et en nature.

Les méthodes de peuplement sont définies afin de modéliser les différentes stratégies que les agents peuvent utiliser pendant leur évolution. La notion de peuplement utilisée ici désigne le processus de remplissage de l'espace, en particulier des îlots, par des objets géographiques tels que les bâtiments ou les routes. Une méthode de peuplement spécifie ainsi la façon dont ce processus de remplissage se déroule. De telles méthodes peuvent, par exemple, spécifier la façon dont les îlots se densi-

fient en reproduisant leur structure interne, en se restructurant ou en diversifiant leur structure [CUR 10, PER 10a, PER 10b]. L'article détaille certaines méthodes de peuplement utilisées dans GeOpenSim pour la densification d'îlots et le paramétrage de telles méthodes.

Dans la première partie de l'article, les méthodes de peuplement sont présentées et détaillées. Les sections 3, 4 et 5 présentent les méthodes de peuplement spécifiques proposées pour les zones industrielles, l'habitat collectif et l'habitat individuel respectivement ainsi que les résultats de simulation obtenus sur chacun de ces types de zone. La section 6 conclut cet article, discute les résultats et présente les perspectives ouvertes par ces travaux.

2. Méthodes de peuplement

La dynamique globale du système est contrôlée par les règles d'évolution. Ces dernières sont appliquées à toutes les échelles meso géographiques, de la ville à l'îlot. En plus de ces règles, la simulation requiert d'autres règles, les méthodes de peuplement, plus spécifiques et locales, afin de peupler la ville avec de nouveaux éléments micro (bâtiments, routes, etc) tout en respectant certaines contraintes quant à l'organisation spatiale souhaitée.

2.1. Présentation

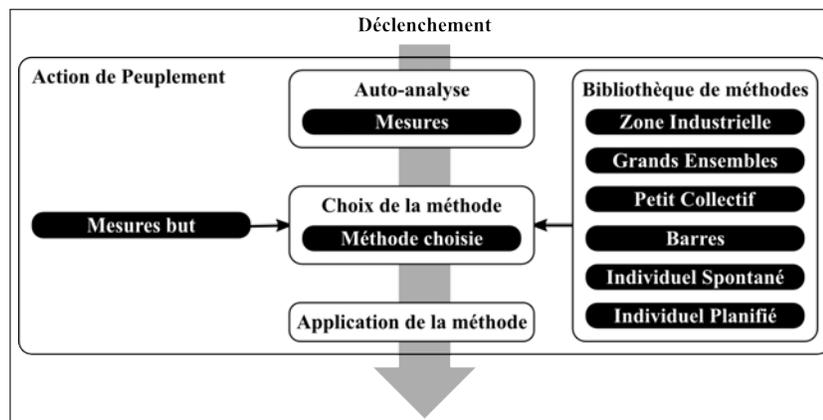


Figure 1 – Choix d'une méthode de peuplement.

Les méthodes de peuplement définissent précisément les processus dont la dynamique est contrôlée par les règles d'évolution. Pour ce faire, les méthodes de peuplement guident la création de nouveaux objets géographiques et sont paramétrées afin de simuler les spécificités de l'évolution d'un élément meso géographique particulier. Un îlot peut, par exemple, se densifier ou se restructurer. Le processus s'effectue en

trois étapes. Dans un premier temps, l'agent (ici l'îlot) analyse son état à l'aide d'un ensemble de mesures (densité, type de zone...). Dans un second temps, il compare ses mesures avec les mesures but fixées par les règles d'évolution et choisit, parmi la liste des méthodes disponibles dans la bibliothèque, celle qui est la plus appropriée (cf. figure 1). Dans la dernière étape, l'agent applique la méthode choisie. Dans la suite de l'article, nous présentons les spécialisations proposées pour les zones industrielles, l'habitat collectif et l'habitat individuel.

2.2. Mesures sélectionnées a priori

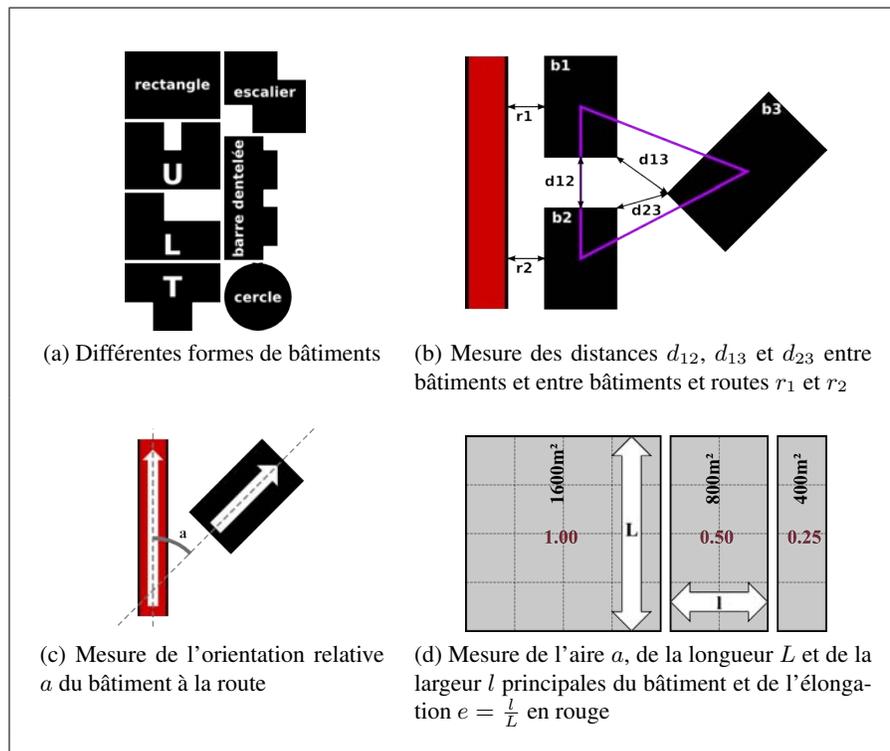


Figure 2 – Mesures sélectionnées.

Afin de définir des méthodes de peuplement génériques pour la construction de bâtiments au sein d'un îlot urbain, un ensemble de mesures morphologiques sur l'îlot et les bâtiments a été défini (cf. figure 2). Ces mesures concernent d'une part l'ensemble des paramètres nécessaires à créer la géométrie des bâtiments parmi lesquels on peut citer la forme (rectangulaire, en L, en U, en T, en forme d'escalier, etc.), la taille (aire, longueur et largeur) et l'élongation, et d'autre part des paramètres de positionnement du bâtiment au sein de l'îlot. Il s'agit ainsi, pour ces dernières mesures, de savoir où et comment placer le nouveau bâtiment dans l'îlot. Les mesures retenues concernent

l'orientation du bâtiment par rapport à la route (la plus proche, celle qui donne accès au bâtiment), la distance du bâtiment à cette route, mais aussi la distance entre ce bâtiment et les autres bâtiments proches de l'îlot (les voisins).

2.3. Problèmes rencontrés par les géographes

La détermination des méthodes de peuplement telles qu'elles sont implémentées dans le système nécessite de décrire l'organisation spatiale des objets à l'échelle de l'îlot urbain. Cette échelle d'analyse est assez inhabituelle pour un géographe. Si le commentaire de carte est un exercice largement pratiqué par le géographe [TIF 00], l'analyse qui en est traditionnellement faite ne correspond pas au niveau d'analyse qui est requis par le système. À la lecture d'une carte topographique [COS 08], un géographe est capable de retracer l'histoire de la ville [WAC 00]. Mais à l'échelle de l'îlot, déterminer comment s'est structuré cet îlot est une toute autre problématique. Le niveau d'analyse requis ici correspond à l'analyse morphologique décrite par [PAN 01] et [ARN 08] entre géographie et architecture : "Il s'agit tout d'abord de reconnaître des différences dans l'organisation matérielle de l'espace ou bien dans la manière dont il est occupé, sur la base de représentations cartographiques. [...] Autrement dit, l'analyse découpe l'espace considéré en parties qui présentent des similitudes internes ou qui se distinguent de leurs environs. Derrière ce découpage, on fait l'hypothèse générale que chaque partie correspond à un mode spécifique de production.[...] L'analyse morphologique traite de l'organisation matérielle, celle d'un réseau de rues par exemple ou bien celle du mode de distribution des constructions sur une parcelle" [ARN 08]. Dans le projet GeOpenSim, l'analyse morphologique est réalisée sur les seuls critères topographiques issus de cartes et bases de données topographiques et ne considère que les formes des objets topographiques. L'objectif final est de concevoir un outil de simulation de l'îlot urbain basé sur des règles de peuplement suffisamment précises pour créer des routes et placer des bâtiments dont l'agencement fera émerger des structures connues de villes industrialisées sans pour autant se soucier de la position et de la taille exactes des bâtiments.

La détermination des méthodes de peuplement est basée sur une analyse visuelle réalisée par des experts géographes. L'objectif de cet article est de présenter le raisonnement mis en place permettant de déduire les paramètres caractéristiques pour construire des méthodes de peuplement. Le passage de l'analyse géographique aux méthodes de peuplement fixant les règles de structuration à l'échelle de l'îlot urbain est loin d'être immédiat et nécessite de nombreux ajustements. Des règles qualitatives, a priori simples pour le géographe, ne sont, en effet, pas forcément simple à transcrire en règles quantitatives utilisables pour la simulation. Cette difficulté, qui constitue tout l'intérêt de ce travail et la difficulté majeure rencontrée, a deux principales explications : (1) les objets rencontrés dans la réalité présentent une grande diversité qu'il est très difficile de formaliser (en formes paramétrées par exemple) et (2) de ce fait, il est très difficile de décrire complètement (dans toute sa complexité) un objet géogra-

phique à l'aide de mesures (ou paramètres) de sorte qu'il soit possible de construire un objet semblable lors d'une simulation.

Dans le cadre de cette étude, trois types de peuplement ont été identifiés et caractérisés : (1) zone à caractère industriel ou commercial (2) habitat collectif et (3) habitat individuel (les zones de centre-ville ont été écartées de cette analyse). À partir de ces trois types de peuplement de base, des peuplements mixtes pourront être ultérieurement dérivés par combinaison. Pour chacun de ces types de base, les experts ont identifiés des zones types sur les agglomérations d'Orléans et de Strasbourg (les deux zones tests du projet GeOpenSim). Une caractérisation période par période de chaque zone est réalisée afin d'obtenir une description des changements récurrents intervenus pour chaque période. Les périodes sont définies entre chaque date ou document de référence. Pour les deux zones d'intérêt (Orléans et Strasbourg), les dates de références sont respectivement 1956/1957, 1976/1978, 1989, 1999 pour les cartes papier au 1/25000 (Orléans) ou les photographies aériennes (Strasbourg), et 2002/2007 pour la BD TOPO®. La comparaison entre zones des changements récurrents nous permet d'établir des comportements types (comportements proches ou identiques) qui sont transcrits en méthodes de peuplement dans le système.

3. Zones industrielles

3.1. *Analyse géographique du processus de peuplement*

Les zones industrielles sont issues de programmes d'aménagement qui déterminent précisément la localisation et l'emprise de ces zones. Les zones industrielles sont morphologiquement très différentes du reste de la ville. La concentration de bâtiments de très grande taille et de formes compactes et variées dans des secteurs bien déterminés de la ville (proche des voies ferrées, des échangeurs, et en périphérie) sont les critères spécifiques de ces zones.

L'analyse visuelle du processus de peuplement des zones industrielles a permis de dégager un comportement assez typique de remplissage (*cf.* figure 3) en trois phases : (1) initialisation : création des premiers bâtiments aux extrémités de la zone avec un accès direct au réseau de communication existant (10% de la zone est exploitée) (2) complément : construction des bâtiments et réseaux complémentaires (environ 60% de la zone est occupée) (3) finalisation : dernières constructions de bâtiments qui permettent d'atteindre la saturation estimée de la zone. Ces trois phases correspondent aux états d'avancement des zones industrielles, qui semblent naturellement liés à la position des zones industrielles dans l'agglomération. Les plus anciennes se situent naturellement au centre et sont saturées (phase 3 atteinte) alors que, parmi les plus récentes, développées en périphérie, certaines permettent encore une densification (phase 2 ou 3 en cours). Pour les plus anciennes, on note parfois une phase de restructuration interne alors que la saturation a déjà été atteinte.

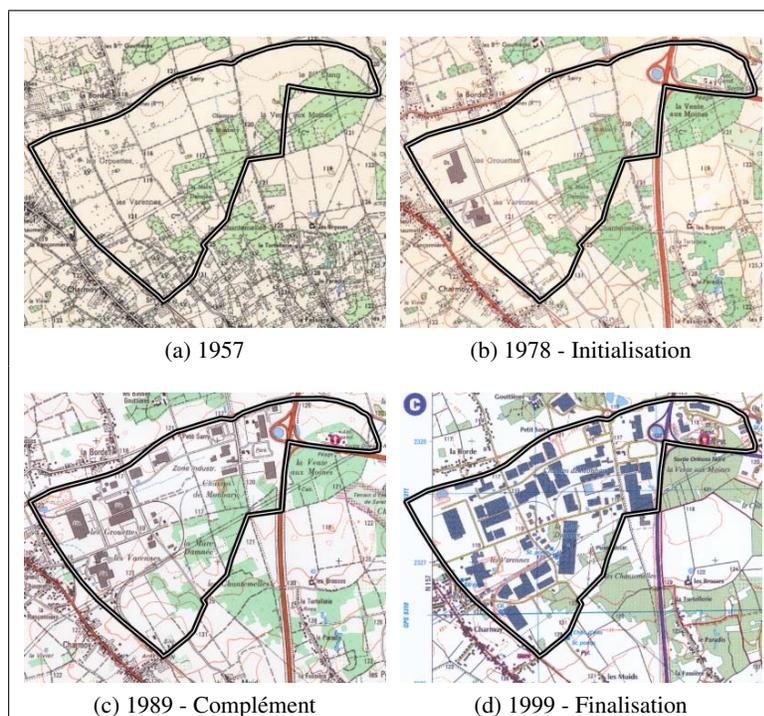


Figure 3 – Exemple d'évolution de zones industrielles - Orléans.

3.2. Ajustement des mesures aux critères géographiques

L'analyse effectuée a ainsi permis de mettre en évidence trois grandes phases de remplissage des zones industrielles. Ces trois phases (initialisation, complément et finalisation) sont fortement dépendantes de l'espace encore disponible dans l'îlot pour la construction de nouveaux bâtiments. Ainsi logiquement, la taille des bâtiments va décroître de la phase d'initialisation à la phase de finalisation, l'espace constructible se réduisant progressivement au fur et à mesure du remplissage de l'îlot.

Afin de reproduire cette densification des zones industrielles, plusieurs stratégies peuvent être envisagées. La première consiste à créer une méthode de peuplement pour chacune des phases de remplissage identifiées. Cette stratégie présente l'inconvénient de nécessiter énormément de seuils notamment concernant la taille des bâtiments qui est un paramètre susceptible de varier d'une zone industrielle à l'autre (en fonction de la place disponible). Une autre stratégie, celle qui a été adoptée ici, consiste à ne créer qu'une seule méthode de peuplement et à choisir la taille du bâtiment à construire en fonction de l'espace libre dans l'îlot. Le choix de la taille du bâtiment se fait par tirage aléatoire entre un minimum (de $200m^2$) et un maximum dépendant de l'espace disponible pour construire le bâtiment. Les principaux critères de cette méthode de

peuplement dérivée de l'analyse géographique des zones industrielles sont présentés dans le tableau 1 (page 120).

3.3. Résultats

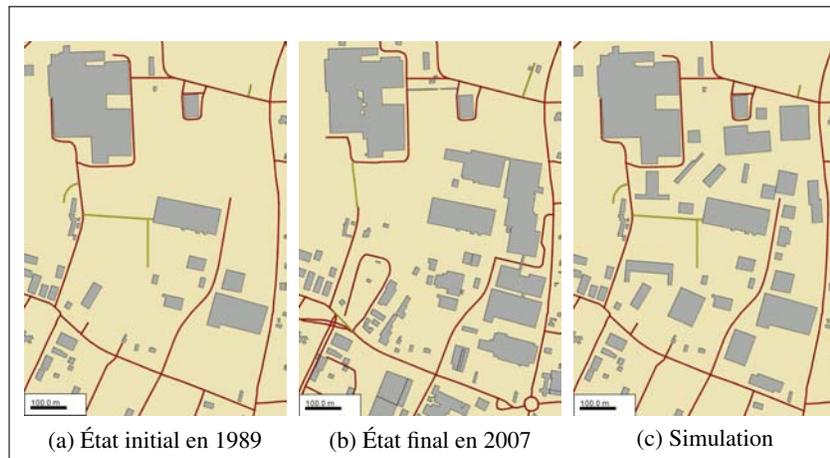


Figure 4 – Simulation d'une densification avec la méthode de peuplement zone industrielle.

Le résultat obtenu pour la densification d'une zone industrielle via la méthode de peuplement définie précédemment (*cf.* figure 4c) peut être comparé à la réalité (*cf.* figure 4b). La distribution des tailles de bâtiments ainsi que le positionnement des bâtiments (par rapport à la route et les uns par rapport aux autres) correspondent en effet assez bien aux observations. Cependant, plusieurs améliorations peuvent être apportées afin d'améliorer le réalisme du peuplement d'une telle zone :

- La forme des bâtiments de ces zones industrielles est extrêmement complexe et variée dans la réalité. Les bâtiments présentent souvent de petits décrochements non reproduits dans cette simulation mais qui pourraient faire l'objet de futurs développements.

- L'orientation des bâtiments peut également être améliorée. Par défaut, les bâtiments sont positionnés parallèlement à la route la plus proche. Ce positionnement n'est pas forcément adapté au niveau des virages.

- Enfin, l'observation de différentes zones industrielles semble montrer qu'il existe une organisation de l'hétérogénéité de la taille des bâtiments présents. Les petits bâtiments semblent se concentrer sur certaines parties de l'îlot. Si cette analyse se vérifie, la taille du nouveau bâtiment créé devra prendre en compte les caractéristiques de ses plus proches voisins afin de leur ressembler.

4. Habitat collectif

4.1. Analyse géographique du processus de peuplement

Trois grands types d'habitat collectif qui correspondent à une évolution chronologique ont été identifiés : les débuts de l'habitat collectif avec les «barres» des années 50-60, les «grands ensembles» des années 70 et le «petit collectif» qui fait office de compromis depuis le début des années 80.

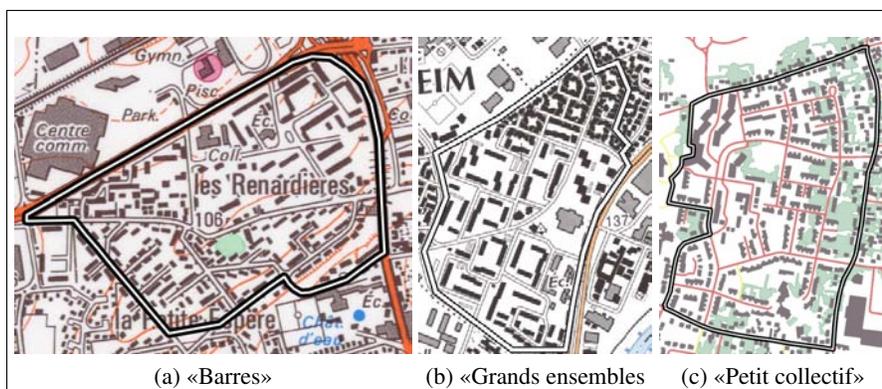


Figure 5 – Différentes zones d'habitat collectif.

«Les barres» correspondent aux premiers immeubles en forme de parallépipèdes étroits et allongés au sol, constitués d'appartements traversants [GAU 03]. De 300 à $900m^2$, de forme rectangulaire, l'élongation est évaluée à $0,1 - 0,3$, la largeur des murs est stable, autour de $10m$. Ces petites barres se situent à $10 - 15m$ de la route (afin de limiter les raccords aux divers réseaux), parallèle, en épis ou perpendiculaire à la route et parallèles entre elles avec des décalages. On peut ainsi obtenir des formes de râteau ou d'arêtes de poisson (cf. figure 5a). Ce type d'habitat correspond aux premières formes d'habitat collectif, il doit donc être simulé pour des dates antérieures à 1960 lors du stade d'initialisation des zones d'habitat collectif.

«Les grands ensembles» correspondent à de grandes opérations immobilières qui se caractérisent [LAC 63] par le nombre de logements qu'ils rassemblent (> 1000 logements) en réponse à la crise du logement des années 60-70. Cet aménagement urbain comporte le plus souvent plusieurs bâtiments en forme de barres ou de tours (au volume allongé vers le haut) construits suivant une architecture et un plan de masse constituant une unité de conception [GAU 03]. Sur les zones de ce type, les tours font 300 à $400m^2$, elle sont parfois doubles (jumelées par l'angle) plutôt groupées en milieu d'îlot (grappe de tours disposée en quinconce par exemple [ALL 04]) et les immeubles peuvent atteindre 800 à $1000m^2$ voire $2000m^2$. Ils sont de formes carrées à rectangulaires, très allongés, positionnés parallèlement ou perpendiculairement les uns aux autres, formant par association des géométries qui construisent des U, L ou S (cf. figure 5b).

«**Le petit collectif**» apparaît ensuite dans les années 80 et continue de se développer à l'heure actuelle. Il s'agit des nouvelles formes d'habitat collectif en réaction au «drame» des grands ensembles. Cet aménagement propose des bâtiments où la verticalité n'est plus de rigueur (3 à 4 étages maximums) avec une emprise au sol plus faible. Il présente 95% de formes très découpées, dentelées (avancées de garage) de 300 à 1000m², parallèles à la route. Des structures complémentaires (hors habitat) sont parfois associées, il s'agit de formes complexes originales, très imposantes (800 à 1200m²) (cf. figure 5c).

4.2. *Ajustement des mesures aux critères géographiques*

Une méthode de peuplement a été créée pour chacun des types d'habitat collectif déterminés lors de l'analyse géographique (cf. tableau 1 page 120). Différentes adaptations ont été nécessaires afin de reproduire au mieux ces zones d'habitat collectif. La principale adaptation concerne sans doute l'épaisseur des bâtiments. Lors de la sélection des paramètres a priori nécessaires à la création de nouveaux bâtiments dans un îlot urbain, nous avons sélectionné : la forme, l'aire et l'élongation. Les premiers tests ont montré que ces 3 paramètres, qui nous paraissaient suffisants pour contraindre la forme de nouveaux bâtiments, ne permettaient pas de respecter certaines contraintes architecturales. Les barres d'habitat collectif sont constituées, la plupart du temps, d'appartements traversants. L'épaisseur de ces bâtiments est donc souvent d'environ 10m, ce que les seuls paramètres d'aire et d'élongation ne suffisent pas à fixer avec précision. L'épaisseur du bâtiment a donc été ajoutée comme nouveau paramètre nécessaire à la création d'un bâtiment. Une nouvelle forme de bâtiment a également été créée pour le petit collectif. Cette forme de barre dentelée assez typique de ce type de peuplement permet d'améliorer le réalisme de notre méthode.

4.3. *Résultats*

Les deux premières méthodes de peuplement d'habitat collectif correspondant aux «barres» et aux «grands ensembles» ont été appliquées sur une zone test de Strasbourg. Les simulations (cf. figures 6c et 6f) montrent des résultats satisfaisants concernant la forme et le positionnement des bâtiments simulés. Comme pour les zones industrielles, l'orientation des bâtiments peut être améliorée. Les arrangements de bâtiments en forme d'épis, de râteau ou d'arêtes de poisson pourront également être intégrés afin d'améliorer le réalisme de ce type de peuplement. La troisième méthode correspondant au «petit collectif» a, quant à elle, été appliquée sur une zone d'Orléans (cf. figure 6i) Le travail réalisé sur la forme des bâtiments pour ce type de peuplement doit encore être poursuivi, les caractéristiques de forme et de position doivent être affinées afin d'orienter vers la route les avancées correspondant aux garages.

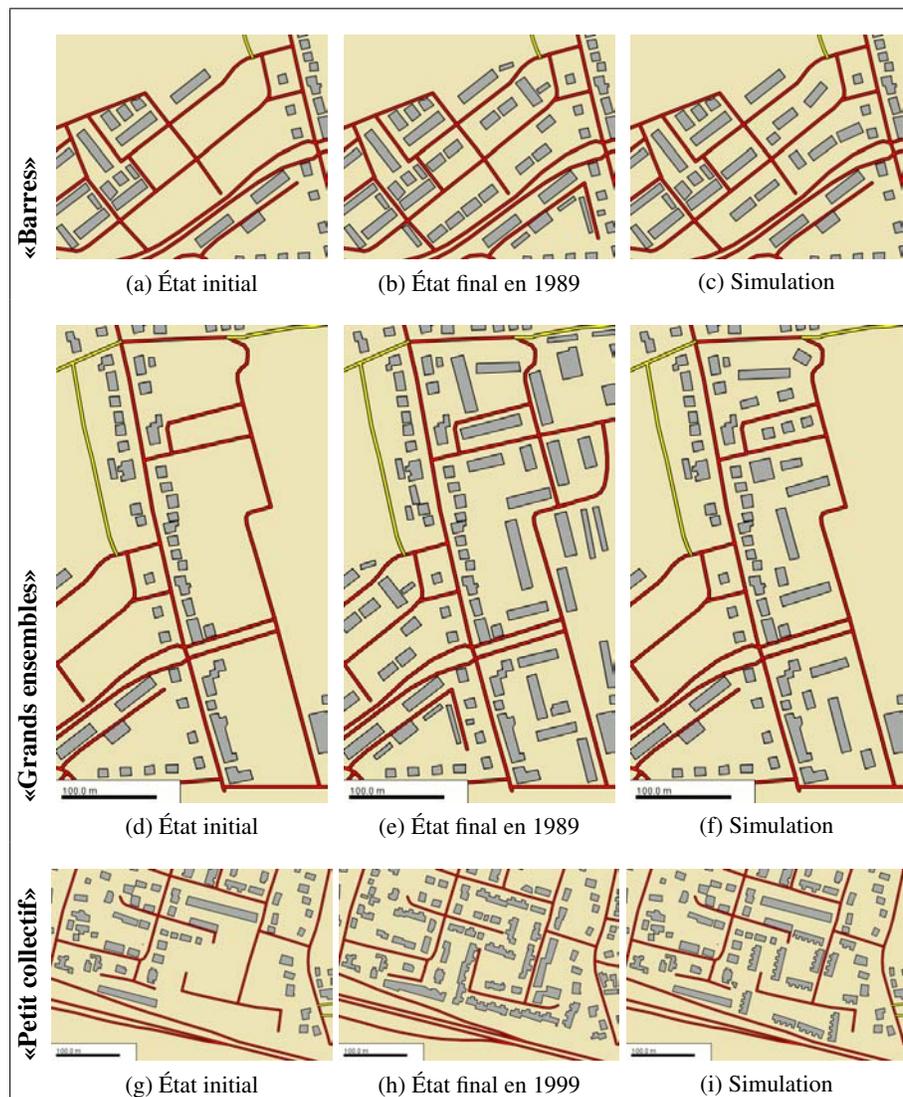


Figure 6 – Simulation d’une densification avec les méthodes de peuplement.

5. Habitat individuel

5.1. Analyse géographique du processus de peuplement

Contrairement à l’habitat collectif, il est difficile d’associer des formes spécifiques d’habitat individuel à une époque. L’étude des cartes et des données topographiques permettent de différencier deux grands types d’habitat individuel (pavillonnaire), hor-

mis l'habitat individuel hérité (présent sur les cartes avant 1957). Il s'agit de l'habitat individuel «spontané» et de l'habitat individuel «planifié».

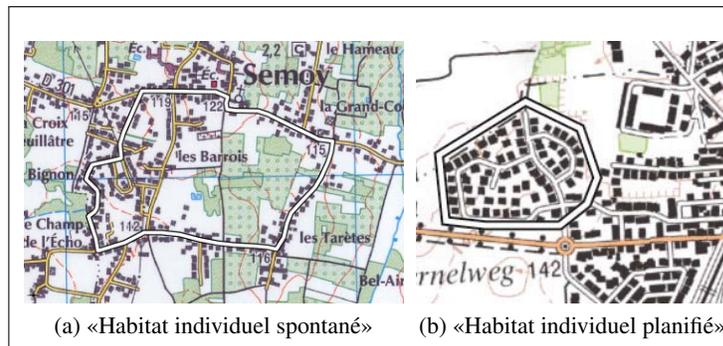


Figure 7 – Zones d'habitat individuel.

«L'habitat individuel spontané» correspond à des pavillons/habitations construites, en complément de l'habitat hérité dans les espaces interstitiels, en priorité le long des axes de communication existants. Ce type d'habitat dit «spontané» n'est pas exempt de règles mais il répond à une logique de continuité linéaire du bâti (cf. figure 7a).

«L'habitat individuel planifié» correspond à des bâtiments construits selon un plan d'aménagement issu du (re)découpage des îlots et de la construction de dessertes sans issue. Ce type d'habitat génère un tissu très différencié. Les nouveaux îlots créés pour cet habitat peuvent former soit des boucles à l'intérieur d'îlots plus grands permettant de relier des routes existantes, soit des lobes (ou nodules) à l'extérieur, en bout de réseau existant (cf. figure 7b). Le découpage est optimisé pour placer et desservir le maximum de bâtiments pavillonnaires. Les pavillons sont principalement de forme carrée à rectangulaire, parallèle à la route.

Sur la forme, la distance à la route et l'orientation des bâtiments, la très grande régularité observée dans le cadre de l'habitat planifié ne se vérifie pas toujours pour l'habitat spontané. En termes de dynamique, les densifications partent des zones construites vers les zones peu ou pas construites (avancée du front) de manière à créer des zones bâties compactes ou établir des liaisons entre zones bâties antérieures (2 centres bourg qui s'agglutinent). On peut noter également une multiplication de l'habitat individuel planifié à partir des années 80.

5.2. Ajustement des mesures aux critères géographiques

Ce type de peuplement a été traité en dernier et après de nombreux ajustements concernant les mesures et les algorithmes mis en place. C'est probablement pour cette raison, mais aussi, peut-être, parce que ces types de peuplement sont plus simples, qu'aucun ajustement n'a été nécessaire. Les paramètres utilisés pour les méthodes de peuplement de type habitat individuel sont listées dans le tableau 1 (page 120).

Nom de la méthode	«Zone Industrielle»	«Barre»	«Grand Ensemble»	«Petit Collectif»	«Individuel spontané»	«Individuel planifié»
Période	1950-actuel	1945-1960	1960-1980	1980-actuel	1945-actuel	1945-actuel
Distance à la route - distribution - min à max	normale 10 à 20m	normale 5 à 15m	normale 5 à 15m	normale 4 à 10m	aléatoire 4 à 15m	normale 8 à 12m
Distance interbâtiment - distribution - min à max	normale 20 à 40m	normale 0 à 10m	normale 5 à 15m	normale 0 à 10m	aléatoire 2 à 15m	normale 8 à 12m
Orientation du bât. - / à la route - / aux bâtiments	non renseigné non renseigné	parallèle parallèle	parallèle parallèle	parallèle parallèle	parallèle non renseigné	parallèle non renseigné
Type de bâtiment 1 : - forme - aire - distribution - min à max - élongation - épaisseur - fréquence	rectangle, L, U, T, escalier aléatoire 200 à 10000m ² non renseigné non renseigné 100%	rectangle normale 300 à 700m ² non renseigné 10m 100%	carré normale 250 à 450m ² non renseigné non renseigné 15%	barre dentelée normale 200 à 1000m ² non renseigné 10m 95%	rectangle aléatoire 60 à 300m ² 0.6 à 1 non renseigné 80%	rectangle normale 100 à 180m ² 0.6 à 1 non renseigné 80%
Type de bâtiment 2 : - forme - aire - distribution - min à max - élongation - épaisseur - fréquence			rectangle, L normale 600 à 1400m ² non renseigné 10m 85%	escalier normale 1500 à 2500m ² 0.6 non renseigné 5%	L, T aléatoire 60 à 300m ² 0.6 à 1 non renseigné 20%	L, T normale 100 à 180m ² 0.6 à 1 non renseigné 20%

Tableau 1 – Principales caractéristiques des méthodes de peuplement.

5.3. Résultats

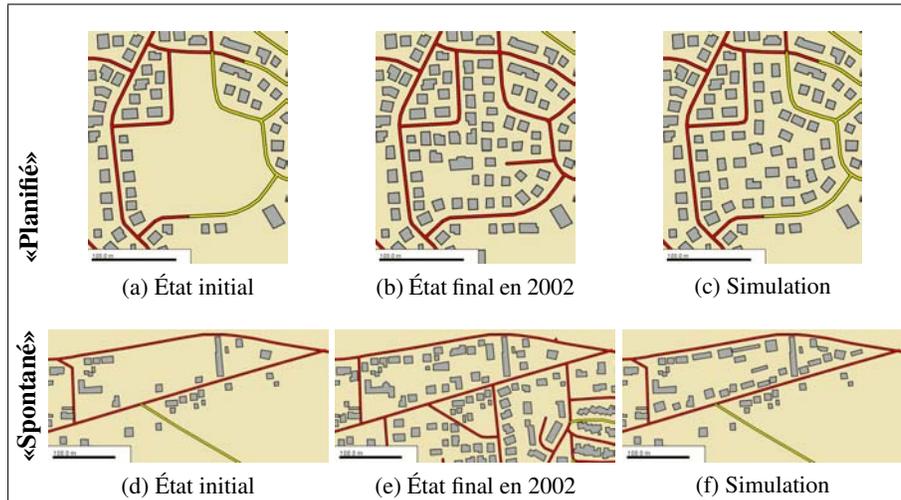


Figure 8 – Simulation d’une densification avec les méthodes de peuplement «individuel planifié» et «individuel spontané».

Les résultats produits par la simulation de la densification de zones d’habitat individuel sont satisfaisants, notamment pour l’habitat individuel planifié. Néanmoins, certaines propriétés des tissus urbains simulés sont à améliorer. En effet, malgré la ressemblance entre les figures 8b et 8c, on remarque l’absence d’alignements entre les bâtiments dans les résultats de simulation. Ces alignements, produits à la fois par le parcellaire et les règles d’urbanisme, peuvent être obtenus lors de la simulation en ajoutant des contraintes spécifiques.

6. Conclusion et perspectives

Dans cet article, une formalisation des processus de peuplement d’îlots (de types industriel, habitat collectif et individuel) a été proposée pour simuler la densification des tissus urbains. Cette formalisation s’appuie sur l’analyse visuelle de cartes topographiques et de photographies aériennes entre 1950 et 2010. Les processus identifiés ont été implémentés sur la plate-forme GeOpenSim dans des méthodes de peuplement. Ces dernières ont été testées dans le cadre de la simulation d’une densification urbaine sur différentes zones d’Orléans et de Strasbourg à l’aide d’un système multi-agents.

Comme le montrent les résultats de simulation présentés, les méthodes de peuplement proposées produisent de très bons résultats, notamment pour l’habitat individuel et collectif (barres et grands ensembles). Lors du travail de formalisation et d’évaluation des résultats, une différence importante a été identifiée entre les mesures (quantitatives) a priori importantes pour la création et le positionnement des bâtiments (par

le simulateur) et les connaissances géographiques (qualitatives) des experts. D'une part, certaines mesures se sont avérées difficiles à utiliser (l'élongation par exemple) et d'autre part, certaines mesures se sont révélées importantes pour le réalisme des résultats de simulation et ont été ajoutées (l'épaisseur par exemple). Ainsi d'autres mesures peuvent être ajoutées afin d'améliorer encore les résultats. Citons, par exemple, la ressemblance entre bâtiments voisins et l'accessibilité au réseau de transport. Par ailleurs, de nouveaux comportements doivent être intégrés à la plate-forme, notamment concernant la construction de nouvelles routes. Les méthodes de peuplement d'îlots définies pourront, à terme, être combinées afin de proposer des peuplements mixtes par association de types différents (ex : îlot qui se densifie à 60% d'individuel et à 40% de collectif). L'intégration de types complémentaires pourra aussi être envisagée si la typologie "individuel, collectif, industriel" devenait trop restrictive. En outre, le caractère générique des règles de peuplement proposées (sur la base de Strasbourg et Orléans) devrait permettre de les utiliser sur d'autres villes. Cependant des adaptations locales seront probablement nécessaires notamment lors de la simulation sur des tissus urbains non français.

7. Bibliographie

- [ALL 04] ALLAIN R., *Morphologie urbaine - Géographie, aménagement et architecture de la ville*, Armand Colin, 2004.
- [ARN 08] ARNAUD J.-L., *Analyse spatiale, cartographie et histoire urbaine*, Parenthèses, 2008.
- [BAD 07] BADARIOTTI A., BANOS A., MORENO D., « Conception d'un automate cellulaire non stationnaire à base de graphe pour modéliser la structure spatiale urbaine : le modèle Remus », *Cybergeo, Sélection des meilleurs articles de SAGEO 2006*, , 2007.
- [BAT 86] BATTY M., LONGLEY P., « The fractal simulation of urban structure », *Environment and Planning A*, vol. 18, 1986, p. 1143-1179.
- [BAT 97] BATTY M., XIE Y., « Possible urban automata », *Environment and Planning B : Planning and Design*, vol. 2, 1997, p. 175-192.
- [BAT 04] BATTY M., BARROS J., ALVES S., « Cities : continuity, transformation and emergence », *Working paper: CASA Working Papers (72)*, Centre for Advanced Spatial Analysis (UCL), London, UK, 2004.
- [BAT 08] BATTY M., « Cities as complex systems : scaling, interactions, networks, dynamics and urban morphologies », *Working paper: CASA Working Papers (131)*, Centre for Advanced Spatial Analysis (UCL), London, UK, 2008.
- [COS 08] COSTA L., ROBERT S., *Guide de lecture des cartes anciennes*, Errance, 2008.
- [COU 97] COUCLELIS H., « From cellular automata to urban models : new principles for model development and implementation », *Environment and Planning B : Planning and Design*, vol. 2, 1997, p. 165-174.
- [CUR 10] CURIE F., PERRET J., RUAS A., « Simulation of urban blocks densification », 13th *AGILE International Conference on Geographic Information Science*, May 2010.
- [GAU 03] GAUTHIEZ B., *Espace urbain : vocabulaire et morphologie*, Editions du patrimoine, 2003.

- [LAC 63] LACOSTE Y., « Un problème complexe et débattu : les grands ensembles », *Bulletin de l'association des géographes français*, n° 318-319, 1963.
- [LEF 08] LEFEBVRE B., RODIER X., SALIGNY L., « Understanding urban fabric with the OH_FET model based on social use, space and time », *Archeologia e Calcolatori*, vol. 19, 2008, p. 195-214.
- [O'S 00] O'SULLIVAN D., « Graph-based cellular automata models of urban spatial processes », PhD thesis, University College London, London, 2000.
- [O'S 01] O'SULLIVAN D., « Exploring spatial process dynamics using irregular cellular automaton models », *Environment and Planning B : Planning and Design*, vol. 28, 2001, p. 687-705.
- [PAN 01] PANERAI P., CASTEX J., DEPAULE J., *Formes urbaines : de l'îlot à la barre*, Parenthèses, 2001.
- [PER 09] PERRET J., BOFFET MAS A., RUAS A., « Understanding Urban Dynamics : the use of vector topographic databases and the creation of spatio-temporal databases », *24th International Cartography Conference (ICC'09)*, November 2009.
- [PER 10a] PERRET J., CURIE F., GAFFURI J., RUAS A., « A Multi-Agent System for the simulation of urban dynamics », *10th European Conference on Complex Systems (ECCS'10)*, Lisbon, Portugal, september 2010.
- [PER 10b] PERRET J., CURIE F., GAFFURI J., RUAS A., « Un système multi-agents pour la simulation des dynamiques urbaines », *actes des 18èmes Journées Francophones sur les Systèmes Multi-Agents (JFSMA'10)*, october 2010.
- [ROD 07] RODIER X., SALIGNY L., « Modélisation des objets urbains pour l'étude des dynamiques urbaines dans la longue durée », BATTON-HUBERT M., JOLIVEAU T., LARDON S., Eds., *SAGEO 2007, Rencontres internationales Géomatique et territoire*, 2007.
- [TIF 00] TIFFOU J., *Commenter la carte topographique aux examens et concours*, Armand Colin, 2000.
- [WAC 00] WACKERMANN G., *Géographie urbaine*, Ellipses, 2000.
- [WER 09] WERNER P., « Application of Cellular Automata and Map Algebra in Studies of Land Use Changes. The Neighborhood Coefficients Method », *Geoinformatica Polonica*, vol. 9, 2009, Polska Akademia Umiejetności - Wydawnictwo, Kraków.
- [WHI 97] WHITE R., ENGELEN G., « Cellular automata as the basis of integrated dynamic regional modelling », *Environment and Planning B : Planning and Design*, vol. 2, 1997, p. 235-246.

Future urban growth scenarios in the Madrid Region, Spain

Plata Rocha, W.* - Gómez Delgado, M.** - Bosque Sendra, J.**

(*) *Escuela de Ciencias de la Tierra, Ciudad Universitaria*
Av. De Las Américas y Blvd. Universitarios 80000. Culiacán, Sinaloa, México.
wenseslao.plata@alu.uah.es

(**) *Departamento de Geografía, Universidad de Alcalá*
C/ Colegios. 28801 Alcalá de Henares, Madrid (SPAIN)
montserrat.gomez@uah.es
joaquin.bosque@uah.es

ABSTRACT: During the last few years urban expansion around the world has occurred without much control or urban planning, causing both regional and global impacts on natural ecosystems. This situation has spurred reflection on current planning policies and has also stimulated research into new tools and methodologies for the simulation of alternative territorial scenarios for the future.

In this respect, the use of MultiCriteria Evaluation (MCE) techniques within a Geographical Information System (GIS) environment has proven to be an efficient option for an optimal allocation of land uses. However, those techniques have not yet been thoroughly explored in the field of scenario simulation. For this reason, the objective of the research presented here has been the generation of two scenarios of urban growth (business as usual, and innovation and sustainability) using MCE techniques. Both scenarios discriminate between three land uses (residential, commercial and industrial), and they have been developed for the Madrid Region. These scenarios simulate the land use structure for 2020 on the basis of the situation in 2000. In that process, fourteen spatial factors have been considered, which include environmental, economic and social aspects.

The results obtained show that some discrepancies exist between the business-as-usual scenario (S1) and the European policies for transport and for the environment. This situation has an impact on the territory, and it causes irreversible damage to the environment. Meanwhile, innovation has been considered a driver of economic development and good planning practice, the positive effects of which have been highlighted in the second scenario (S2).

KEY WORDS: Simulation of urban growth, Geographical Information Systems, MultiCriteria Evaluation, Madrid Region.

1. INTRODUCTION

Very little control has been exercised over urban expansion in recent decades, and little is known about its regional and global impact on different ecosystems. The principle reason for this may be the inefficient regulatory power of the authorities to combat acute land pressure in the region; consequently, it has become necessary to reconsider current planning policies. In this respect, the use of new methodologies and tools for creating spatial simulations of a series of potential alternatives for future land use, within a more participative framework and with a view to providing planners with ample quality information on which to base more efficient planning decisions, is crucial. In this context, the simulation of land-use change scenarios is rich in potential for anticipating the consequences of specific land intervention decisions in the future, thus assisting in the process of urban planning (Schwarz et al., 2010). In fact, several European projects already focus on the development of crucial new planning strategies and tools for maintaining a sustainable relationship between rural and urban environments (PRELUDE, PRospective Environmental analysis of Land Use Development in Europe, PLUREL, Peri-urban Land Use Relationships – Strategies and Sustainability Assessment Tools for Urban – Rural Linkages or TiGrESS, Time-Geographical approaches to Emergence and Sustainable Societies).

Similarly, in a special edition of the *Journal of Agriculture, Ecosystems & Environment* (2006, Vol. 114, Issue 1) and in other scientific journals, a series of studies of land-use scenario simulations in the specific context of Europe have been published, including quantitative assessments of the impact of future land-use change (de Nijs et al., 2004, Verburg et al., 2006), the potential effects of different environmental policies and the influence of climate change on agriculture and biodiversity (Reidsma et al., 2006, Westhoek et al., 2006), or the vulnerability of ecosystems and rural areas to land-use change (van Meijl et al., 2006).

The research carried out in this field has applied a generic methodology, adapted according to the land-use simulation techniques employed. This methodology basically comprises three steps: 1) a qualitative description of the scenarios, 2) a calculation of land demand for each scenario, and 3) a spatial simulation of land uses.

The literature review carried out for the present study confirmed that the special report of the Intergovernmental Panel on Climate Change (IPCC) about emission scenarios (Nakicenovic and Swart, 2000) has been widely used as a starting point for qualitative land-use scenario descriptions on a global or continental scale (Barredo and Gómez Delgado, 2008, de Nijs *et al.*, 2004, Reginster and Rounsevell, 2006, Verburg *et al.*, 2006). Other researchers, however, have conceptualized scenario evolution hypotheses on the basis of consensus among expert groups involved (Hoogeveen and Ribeiro, 2005; Winder 2006; Aguilera Benavente 2009).

With respect to the spatial simulation of land uses, different techniques have been employed for performing quantitative future scenario simulations, such as Cellular Automata (CA) models, Spatial Regression (SR) methods, MultiCriteria Evaluation Techniques (MCE), among others.

In recent years widespread use has been made of models based on CA, which generally employ a probabilistic system determined by factors such as accessibility to urban areas and roads, land-use suitability (usually related to the slope), zoning status and the neighbourhood effect measured for urban land use (Aguilera Benavente 2009; Barredo and Gómez Delgado, 2008, Barredo *et al.*, 2004, de Nijs *et al.*, 2004, Hoogeveen and Ribeiro, 2005, Verburg *et al.*, 2006). Simulations based on SR (Hu and Lo, 2007) have also generally employed these types of factors.

MCE models have been applied, although not exploited to their maximum potential, through the use of additional compensatory techniques for generating future urban growth scenarios (Pettit and Pullar, 2004). It should be mentioned that in these cases, only one scenario is generated rather than

a series of alternative scenarios. The aim of this work is to show and to reinforce the potential of MCE techniques in the simulation of prospective models as opposed to more frequently used models, such as those based on CA.

Another key aspect in the simulation of models is deciding which spatial factors or variables to model. In this connection, useful tools for discriminating the most significant factors in spatial models are, on the one hand, sensitivity analysis (SA) and, on the other, the analysis of explanatory factors by means of statistical regression techniques. Such information may be of great use for guaranteeing the soundness of results or, at least, ensuring the inclusion of the minimum elements necessary for the results to be consistent with expectations and to confer credibility.

Given the present situation in the region of Madrid, we believe that the simulation of future urban growth scenarios (residential, commercial and industrial land use) is of interest for the Community of Madrid and with this in mind, we have explored and increased our understanding of the application of MCE techniques. These techniques enable the land-use suitability of different scenarios to be determined by integrating a combination of different factors with differing restrictions from an environmental and land-planning perspective, thus rectifying some of the deficiencies of the other studies previously mentioned. The scenarios are based on formulations which integrate past urban growth trends and environmental concerns expressed by environmental protection groups. As a result, images of the future are obtained which facilitate the visualization, exploration, assessment and understanding of the driving forces behind two different scenarios, and an assessment of the consequences these could have for the environment in the future.

2. STUDY AREA AND DATA

The region of Madrid has undergone a rapid process of urbanisation in the last 10-15 years (Figure 1), which has principally translated into land occupation in the form of on widespread, dispersed residential housing zones (EEA, 2006).

This process is reflected in the fact that approximately 50,000 ha were transformed into non-natural surfaces between 1990 and 2000, almost another 50% of the surface area in existence in 1990. This urban expansion, however, was not accompanied by any notable population increase with the population only rising over the period by 5.16% (Plata Rocha et al., 2009).

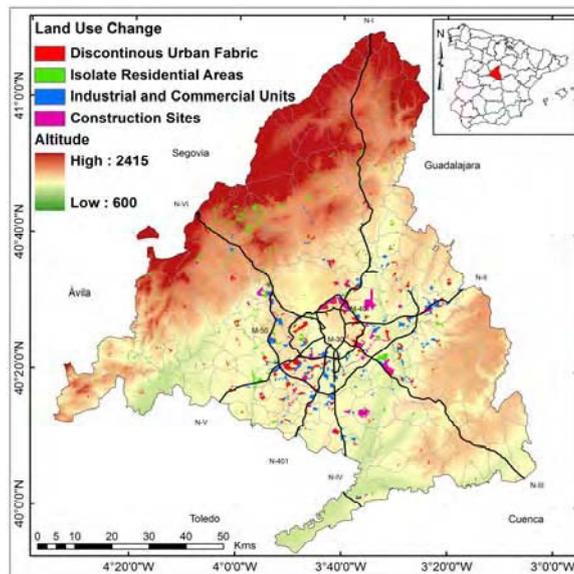


Figure 1. Urban growth in Madrid region between 1990-2000
Source. Corine Land Cover 1990 and 2000

Recent studies have shown that this dynamic is to a great extent the consequence of demand for first and second homes, together with the increase in, and improvement of, transport infrastructures and the favorable socioeconomic situation in Spain during the period (López de Lucio 2003).

Consequently, the urban area of Madrid is changing rapidly, and a new land-use model is emerging – one which suffers from a series of problems of sustainability within a context of weak spatial planning, a common issue in large European conurbations where the regulatory powers of the authorities are insufficient to control the enormous pressures brought to bear by new land uses (López de Lucio 2003).

In order to construct the models, a series of spatial variables were compiled for the base year (2000), using the existing literature on this type of study as a reference, but above all, bearing in mind the availability of information concerning the study area. In the end, the data compiled related to land use (Corine Land Cover), soil type, hydrography, roads, protected natural areas (Red Natura), vulnerability of aquifers to contamination, geotechnical map, undesirable installations, social facilities (medical and educational facilities), topography (altitude, slope, aspect), population and administrative boundaries. This information was collected from various official Spanish sources.

3. METHODOLOGY

There is always an element of uncertainty involved in scenario simulation, which increases the further into the future the simulation is projected. For this reason, we proposed generating medium-term simulations, taking the year 2000 as a starting point (the last year for which *Corine Land Cover* data were available at the beginning of our research) and the year 2020 as the time horizon, in line with the Spanish Ministry of Public Works and Transport's Strategic Plan for Infrastructures and Transport (2004) and the Spanish Ministry of the Environment's Strategy for Climate Change and Clean Energy (2007), both of which present a 2020 time horizon.

Among the existing MCE techniques, weighted linear combination (WLC) is one of the most widely used because of its simplicity. Basically, this compensatory methodology adds the products of n parameters (weights, w_i)

by n variables (maps, X_i) to obtain in this case a map showing the suitability of the territory for a given future land use (M suitability). This is expressed as the following equation:

$$M_{\text{suitability}} = w_1 \cdot X_1 + w_2 \cdot X_3 + \dots + w_n \cdot X_n$$

Non-compensatory techniques are less widely used, but we think their application together with compensatory techniques could enhance the performance of the spatial simulation of each scenario. In this context, the Conjunctive/Disjunctive model enabled a threshold to be fixed for a series of factors, eliminating all the alternatives (areas of land or pixels) which fell above or below the threshold established (Malczewski, 1999, Gómez Delgado and Barredo, 2005).

Finally, the multiobjective assignment of future urban growth (residential, commercial and industrial) to each scenario was performed using the BuildSceneries tool designed specifically for allocation/simulating land use¹ (Rodríguez Álvarez et al., 2009). This tool was used in Idrisi software in order to optimize the time needed to simulate scenarios.

3.1. Qualitative interpretation and Urban land demand of each scenario

Two future urban growth scenarios based on the evolution hypothesis proposed by the group of experts involved in the SIMURBAN project (<http://www.geogra.uah.es/simurban>) were developed.

The first scenario proposed could be called “business as usual” (**S1**), as it assumes the continuation and consolidation of the growth models and intensities witnessed for the period 1990-2000.

The “innovation and sustainability” scenario (**S2**) tries to strike a balance between social, economic and environmental aspects. This scenario proposes more sustainable patterns of urban growth, with a diversification of

¹ Tool available at: <http://www.geogra.uah.es/simurban/publicaciones.php> (last viewed October 2010).

productive systems thanks to the adoption of innovation as an economic driving force.

The future demand for urban land (residential, commercial and industrial) was calculated using a System Dynamics-based model which generated demand for each scenario according to variations in rates of growth or decline for factors such as population and immigration, income, housing, homes, etc. for the years 1990 to 2000. The procedure is explained in detail in Aguilera Benavente et al., 2009. The demand obtained in this study for each scenario, for the year 2020, is shown in Table 1.

Table 1. *Urban land demand (ha) for each of the scenarios in the year 2020*

No.	Scenario	Demand for Residential use	Demand for Productive
<i>S1</i>	Business as usual	28.414	17.356
<i>S2</i>	Innovation and Sustainability	17.618	10.836

Source: Aguilera Benavente et al.,2009

3.2. *Factors selection via sensitivity analysis and factor explaining*

Subsequently, a series of factors were chosen which previous studies had shown to bear a close relationship to urban growth, such as accessibility to roads and urban zones, proximity to areas of environmental interest, etc.

The selection followed the procedure of Plata Rocha (2010) in so far as it explored and determined the influence of certain spatial variables on the emergence and location of new residential and productive zones. Factors such as accessibility to roads, urban zones, industry, commerce and hospitals, and topography exercised a strong stimulus on the growth of new urban zones between 1990 and 2000.

The study carried out by Plata Rocha (2010) in his doctoral research work was also taken into account. This study performs a sensitivity analysis of a simulation model of future urban growth in the Madrid region. To be more precise, we applied a global sensitivity technique (E-FAST, Extended Fourier Amplitude Sensitivity Test) using the SIMLAB software from the Institute for Systems, Informatics, and Safety at the Joint Research Centre of the

European Union. The results show that the most important, influential and/or significant factors in the models are: land use; accessibility to roads, urban areas and commercial zones; the protection of land with greater degrees of productivity; accessibility to hospitals; and the weighted distance as per the most vulnerable and the most affluent populations.

Thus, after considering the results of those studies, we selected a group of variables in order to create the two scenarios of future urban growth (Tables 2 and 3).

3.3. Scenario implementation and spatial assignment rules

In order to model each of the scenarios, the most appropriate factors were selected which would enable the essential and differentiating characteristics of each to be reproduced (Table 2 and Table 3). The factor weights for each land use and scenario were calculated through Saaty's pairwise comparison matrix, in accordance with the consensus of the researchers of SIMURBAN project.

All artificial surfaces, water bodies and courses and protected non-urban zones in existence in 2000 were classified as restricted areas when simulating the two scenarios, in line with current planning legislation in force in all municipal councils in the region.

Once the decision rules for the spatial representation were assigned, the Conjunctive/Disjunctive and WLC techniques were used to perform the simulation in S1 and WLC for S2.

3.3.1. Business as usual scenario (S1)

Here, the assignment of future urban growth was determined by past trends, more precisely, those produced between 1990 and 2000. Before going on to describe the different scenarios, we would like to reiterate the fact that the starting point was the year 2000, a time when the property boom was at its peak in the Community of Madrid.

In an attempt to reproduce the reasons behind the past growth, the factors shown in Table 2 were considered. In this case, the factors of accessibility to roads and urban zones were combined through the use of the conjunctive/disjunctive method, while the eliminatory distance thresholds used were those indicated by an analysis of growth probability for the simulated categories (residential and productive) (Plata Rocha, 2010). We consider that this non-compensatory method is better at reproducing past trends than compensatory techniques like WLC.

Table 2. *Business as usual Scenario: factors used for the simulation*

MCE method	Factors	Residential	Commercial	Industrial
Compensatory	Land use (CLC2000)	<i>Weight = 1.00</i>	<i>Weight = 0.90</i>	<i>Weight = 0.80</i>
	Distance (weighted) to population	-----	<i>Weight = 0.10</i>	-----
	Soil type	-----	-----	<i>Weight = 0.20</i>
Non-compensatory	Distance to roads	< 4000 m.	< 3000 m.	< 3000 m.
	Distance to urban centres	< 3000 m.	< 5000 m.	< 3000 m.
	Transport network nodes	-----	< 3000 m.	-----

3.3.2. Innovation and sustainability scenario (S2)

Up to 12 factors were considered for this scenario with a view to obtaining a more detailed description of land suitability for the allocation of new urban zones. These factors are subsumed within the three basic tenets of sustainable development, uniting economic and social development with environmental protection (Table 3). The aim in this case was to obtain an “optimal” situation, for which reason the use of a compensatory technique was considered to be more appropriate.

Table 3. *Innovation and sustainability scenario: factors and weights used for the simulation*

MCE method	Factors	Residential	Commercial	Industrial
Compensatory	Land use (CLC2000)	0.18	0.15	0.29
	Vulnerability of aquifers to contamination	0.04	0.02	0.06
	Soil type	0.09	0.07	0.12
	Slope	0.03	0.02	0.02
	Distance to Roads	0.16	0.21	0.16
	Distance to urban centres	0.21	0.17	0.12

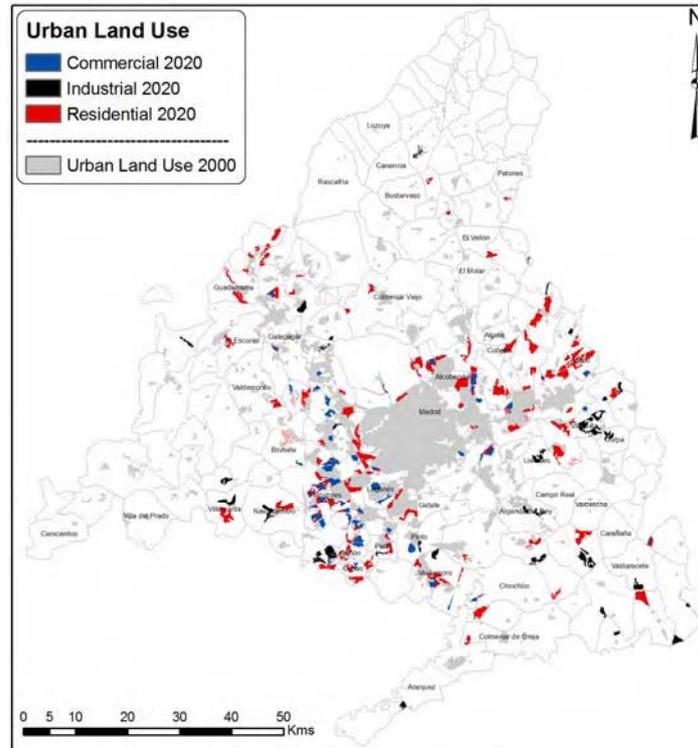
Distance to commercial centres		0.15	
Distance to industry			0.09
Geo-technical-risk	0.08	0.06	0.04
Distance to Hospitals	0.13	0.06	
Distance (weighted) to vulnerable population			0.10
Distance (weighted) to population with spending power		0.09	
Aspect	0.08		

4. RESULTS

On the basis of a first visual analysis of the results, it was possible to confirm that the spatial patterns for land use proposed for 2020 corresponded to the hypotheses assumed for each scenario (Figure 2).

Figure 2a gives the results for the scenario simulating the situation in 2020 if the urban development trends registered for the period 1990 – 2000 continued (business as usual). As can be seen, residential zone growth in this scenario is dispersed throughout the region, primarily consuming land suitable for agricultural and livestock rearing activities. Such dispersal would generate new needs in terms of transport infrastructures with a concomitant increase in the use of private transport, together with further fragmentation of the landscape and an increase in environmental pollution. Commercial and industrial zones proliferated around urban centres, mainly in the south-west and east of the region.

Figure 2b depicts the image of land occupation resulting from a context of innovation and sustainability. The prevailing pattern is one of compact development, where the growth of new residential zones has occurred mainly in proximity to medium and intermediate sized urban centres and where, from a social, economic and environmental point of view, efficient use has been made of land, corresponding to sustainable land planning policies. Commercial zones have been assigned according to the criterion of maximizing access to potential users: located close to residential zones, there is an appreciable distances between them. As for industrial zones,



b) Innovation and sustainability

Figure 2. *Urban growth results for each scenario in the Madrid region (2000-2020)*

5. CONCLUSIONS

With regard to methodology, we can confirm that the results obtained demonstrate that the use of models based on MCE techniques represents an efficient alternative for simulating future urban growth scenarios. In contrast to other prospective models (based on AC, for example), these permit the explicit integration of a wide range of economic, social and environmental factors, thus enabling due consideration of the three fundamental dimensions of sustainable development. Furthermore, we have shown how the combination of compensatory and non-compensatory techniques makes it possible to reproduce more realistic conditions in each of the scenarios.

As for the results obtained for the urban growth scenarios, it was possible to identify the contrasts, imbalances or benefits which might arise in the Madrid region if the evolution hypotheses proposed for each of the future scenarios are accepted or not. These scenarios provide important information concerning the potential environmental, territorial, economic and social impacts which would result if present planning policies, economic growth based on property speculation (S1), and/or sustainable development alternatives continue (S2).

Finally we would like to highlight the fact that scenario simulation represents a useful tool for integrating the opinions of the different agents involved in planning, with a view to formulating more participative and informed territorial strategies.

6. ACKNOWLEDGMENTS

This research was performed in the context of the SIMURBAN project (Analysis and Simulation of Urban Growth using Geographic Information Technologies. Sustainability Evaluation) (SEJ2007-66608-C04-00/GEOG), funded by the Spanish Ministry of Science.

7. REFERENCES

- Aguilera Benavente F., Plata Rocha W., Bosque Sendra J., Gómez Delgado M., "Hacia la simulación de escenarios futuros de demanda de ocupación de suelo en ámbitos metropolitanos", in *Proceedings of II Internacional Conference on Sustainability Measurement and Modelling*, 2009, Terrassa, España, p. 1-21.
- Barredo J. I., Demicheli L., Lavalle C., Kasanko M., McCormick N., "Modelling future urban scenarios in developing countries: an application case study in Lagos, Nigeria", *Environment and Planning B: Planning and Design*, 31, (1), 2004, p. 65-84.
- Barredo Cano J. I., Gómez Delgado M., "Towards a set of IPCC SRES urban land-use scenarios: modelling urban land-use in the Madrid region", in *Modelling Environmental Dynamics, Advances in Geomatic Solutions* Eds. M. Pagelow and M. T. Camacho (Springer), 2008, p. 363- 385.
- de Nijs T. C. M., de Niet R., Crommentuijn L., "Constructing land-use maps of the Netherlands in 2030", *Journal of Environmental Management*, 72, (1-2), 2004, p. 35-42.

- European Environment Agency, *Urban sprawl in Europe, the ignored challenge*, Report No 10. 2006, EEA (Office for Official Publications of the European Communities, Luxembourg).
- Gómez Delgado M., Barredo Cano J. I., *Sistemas de Información Geográfica y Evaluación Multicriterio en la ordenación del territorio*, Paracuellos de Jarama, España, Editorial RA-MA, 2005.
- Hoogeveen Y., Ribeiro T., Land use scenarios for Europe Regional case studies Estonia, The Netherlands, Northern Italy, PRELUDE Report, 2005, European Environment Agency <http://scenarios.ewindows.eu.org>.
- Hu Z., Lo C. P., “Modeling urban growth in Atlanta using logistic regression”, *Computers Environment and Urban Systems*, 31, (6), 2007, p. 667–688.
- López de Lucio R., “Transformaciones territoriales recientes en la región urbana de Madrid”, *Urban*, 8, 2003, p. 124-161.
- Ministerio de Fomento de España, Plan Estratégico de Infraestructuras y Transporte. Definición y evaluación de escenarios alternativos, 2004, http://peit.cedex.es/documentos/escenpeit_final.pdf.
- Ministerio de Medio Ambiente de España, Estrategia Española de Cambio Climático y Energías Limpias horizonte 2007 – 2012 – 2020, 2007, http://www.mma.es/secciones/cambio_climatico/documentacion_cc/estrategia_cc/pdf/est_cc_energ_limp.pdf.
- Nakicenovic N., Swart R., Special Report on Emissions Scenarios, Cambridge University Press, 2000, UK).
- Pettit C., Pullar D., “A way forward for land-use planning to achieve policy goals by using modeling scenarios”, *Environment and Planning B: Planning and Design*, 31, (2), 2004, p. 213-233.
- Plata Rocha, W., Descripción, Análisis y Simulación del Crecimiento Urbano Mediante Tecnologías de la Información Geográfica. El Caso de la Comunidad de Madrid, PhD Thesis, 2010, Universidad de Alcalá de Henares, España.
- Plata Rocha W., Gómez Delgado M., Bosque Sendra J., “Cambios de usos del suelo y expansión urbana en la comunidad de Madrid (1990-2000)”, *Scripta-Nova*, XIII, (293), 2009, <http://www.ub.es/geocrit/sn/sn-293.htm>.
- PLUREL, Peri-urban Land Use Relationships – Strategies and Sustainability Assessment Tools for Urban – Rural Linkages, University of Copenhagen, 2010 <http://www.plurel.net/>.
- Reginster I., Rounsevell M., “Scenarios of future urban land use in Europe”, *Environment and Planning B: Planning and Design*, 33, (4), 2006, p. 619-636.
- Reidsma P., Tekelenburg T., van den Berg M., Alkemade R., “Impacts of land-use change on biodiversity: An assessment of agricultural biodiversity in the European Union”, *Agriculture, Ecosystems and Environment*, 114, (1), 2006, p. 86-102.

- Schwarz N., Haase D., Seppelt R., “Omnipresent sprawl? A review of urban simulation models with respect to urban shrinkage”, *Environment and Planning B: Planning and Design*, 2010, (doi:10.1068/b35087).
- Van Meijl H., van Rheenen T., Tabeau A., Eickhout B., “The impact of different policy environments on agricultural land use in Europe”, *Agriculture, Ecosystems and Environment*, 114, (1), 2006, p. 21-38.
- Verburg P. H., Schulp C. J. E., Witte N., Veldkamp A., “Downscaling of land use change scenarios to assess the dynamics of European landscapes”, *Agriculture, Ecosystems and Environment*, 114, (1), 2006, p. 39–56.
- Westhoek H. J., van den Berg M., Bakkes J. A., “Scenario development to explore the future of Europe’s rural areas”, *Agriculture, Ecosystems and Environment*, 114, (1), 2006, p. 7-20.
- Winder N., Final Report of the project Time-Geographical approaches to Emergence and Sustainable Societies (TiGrESS). School of Historical Studies, University of Newcastle, 2006, United Kingdom. <http://www.tigress.ac/tigshome.htm>.

Acquisition de connaissances sur les déplacements quotidiens des individus dans un contexte de risques naturels

Protocoles d'enquête à l'aide de technologies mobiles

Christophe Sidonie* — **Davoine Paule-Annick*** — **Jambon Francis*** — **André-Poyaud Isabelle**** — **Chardonnel Sonia**** — **Lutoff Céline**** — **Lbath Ahmed***

* *Laboratoire d'Informatique de Grenoble, Université Joseph Fourier*

681 Boulevard de la Passerelle

38402 Saint Martin d'Hères

{sidonie.christophe, paule-annick.davoine, francis.jambon, ahmed.lbath}@imag.fr

** *UMR PACTE Territoire CNRS, Université Joseph Fourier*

14 bis avenue Marie Reynoard

38100 Grenoble

{sonia.chardonnel, celine.lutoff, isabelle.andre-poyaud}@ujf-grenoble.fr

RÉSUMÉ. Nos travaux de recherche s'intéressent à la mise au point de méthodes et d'outils pour l'acquisition et l'analyse d'informations liées à l'étude des mobilités des individus et à leurs adaptations à des contextes variés. Ces travaux sont menés dans le cadre du projet TRACES qui s'intéresse aux comportements humains face à des conditions météorologiques instables, comme les crues rapides du Gard. Notre problématique est de connaître les conditions d'adaptation des mobilités individuelles dans ces situations. Nous avons pour objectif de déterminer les méthodologies utilisant des dispositifs mobiles, les plus adaptées, permettant d'extraire des données géographiques (localisation) et de l'information sémantique (activités, modes, etc.) sur les déplacements des individus en situation réelle.

ABSTRACT. Our research work deals with methodologies and tools to acquire and analyze information about individual mobility and adaptation of mobility to various contexts. Our TRACES project aims at improving knowledge about human behavior during bad weather episodes, such as flash floods episodes in the French region Gard. Our issue is to know conditions of individual mobility adaptation during those episodes. We aim at determining methodologies, based on mobile devices, allowing acquiring both geographical and semantical information about individual movements in real case.

MOTS-CLÉS : géomatique, mobilité, enquête déplacements, motivation, comportement, GPS, sémantique.

KEYWORDS: geomatics, mobility, travel survey, motivation, behavior, GPS, semantics.

1. Introduction

La connaissance des comportements effectifs des individus en périodes d'alertes ou d'épisodes catastrophiques constitue un élément fondamental de la gestion des crises dues à des risques naturels. Aujourd'hui, l'accent est mis sur la prévention, en informant les populations sur les risques encourus et les comportements à adopter en période d'alerte. Les automobilistes ont des difficultés pour reconnaître le risque associé aux routes inondées et beaucoup ignorent les alertes. Il est pourtant difficile de juger de la profondeur et de la vitesse de l'eau traversant une route, surtout pendant la nuit ou pendant des orages. Les automobilistes ne savent pas qu'il suffit de 40 cm de hauteur d'eau pour entraîner un véhicule (Maples *et al.*, 2008). L'évaluation de la capacité des populations à s'approprier les consignes de sécurité et à s'adapter aux situations particulières, voire extrêmes, reste en effet un problème délicat. Les outils et méthodes disponibles pour acquérir ce type d'information relèvent essentiellement d'enquêtes sur le terrain, faites a posteriori, et caractérisées par de nombreuses imprécisions tant sur les déplacements réalisés par les individus que sur les activités et les processus de décision qui les ont déclenchés.

Les travaux de recherche que nous menons s'intéressent ainsi à la mise au point de méthodes et d'outils pour l'acquisition et l'analyse d'informations liées à l'étude des mobilités des individus et à leurs adaptations à des contextes variés. Ces travaux sont menés dans le cadre du projet TRACES¹ qui s'intéresse aux comportements humains, en termes d'adaptation des mobilités individuelles, face à des conditions météorologiques instables pouvant générer un danger mortel : notre terrain d'étude est la région française du Gard et ses crues rapides. Nous avons pour objectif de déterminer les méthodologies les plus adaptées à l'analyse des déplacements des individus en situation réelle. Le projet TRACES regroupe des équipes de recherche en géomatique (LIG-Steamer), en géographie (PACTES/Territoires) et en interaction homme-machine (LIG-MultiCom) (Christophe *et al.*, 2010).

Dans la première partie, nous mettons en évidence les enjeux soulevés par notre projet. Dans la deuxième partie, nous présentons des travaux existants autour des méthodologies d'analyse des mobilités des individus. Dans la troisième partie, nous présentons nos propositions de protocoles d'enquêtes à l'aide de technologies mobiles. Pour finir, nous détaillons les pré-expérimentations mises en place pour les valider, avant d'effectuer l'expérimentation sur le terrain d'étude.

2. Enjeux du projet TRACES

Dans le contexte des crues rapides du Gard, I. Ruin (2007) tente d'analyser les comportements des individus suite à la crue mortelle de 2002, en s'appuyant sur des techniques d'enquêtes traditionnelles : un nombre non négligeable de personnes

¹ <http://traces.ligforge.imag.fr>

poursuit ses activités quotidiennes, d'autres modifient leur emploi du temps pour aller chercher leurs enfants à l'école. Toutefois, de nombreuses incertitudes demeurent : nous ne connaissons ni les déplacements précis (*Où sont les individus à tel moment de la journée ? Par où les individus sont-ils passés ?*), ni les motivations des personnes enquêtées (*Pourquoi emprunter cette route ?*).

L'objectif principal du projet TRACES est de pouvoir proposer et tester de nouveaux protocoles d'enquêtes à l'aide de technologies mobiles, afin d'acquérir des connaissances sur les comportements humains en situation de risque. Plusieurs types d'enjeux se présentent : des enjeux méthodologiques sur l'acquisition d'informations sémantiquement plus riches que la connaissance brute de la localisation à tout moment, des enjeux déontologiques liés au suivi précis et constant des individus, ainsi que des enjeux technologiques sur la faisabilité d'utiliser des technologies mobiles dans notre contexte.

2.1. Enjeux méthodologiques

Nous souhaitons pouvoir connaître les comportements des personnes en termes de mobilité individuelle, dans le contexte de risque de crues rapides. L'objectif est de savoir s'il y a une forme d'adaptation des déplacements quotidiens suite à l'information reçue sur le risque naturel ou à l'observation directe de la dégradation des conditions climatiques : interrompre ou poursuivre les déplacements habituels, changer de mode de transport, changer de trajet, etc. Nous souhaitons aussi déterminer les motivations du changement ou non, de comportement. La question est donc non seulement d'acquérir des informations précises de localisation (les coordonnées géographiques et le temps) mais aussi d'en extraire des informations plus riches d'un point de vue sémantique : activités, modes de transport, facteurs déclenchant la prise de décision, etc. Il s'agit également de savoir comment l'utilisateur est intégré dans la validation et l'enrichissement des informations acquises automatiquement. Pour finir, il est nécessaire de réduire les phases d'interview post-traçage pouvant être perçues comme pénibles. Pour cela, nous souhaitons recourir largement à l'automatisation au moyen de dispositifs mobiles.

2.2. Enjeux déontologiques

Si des méthodes d'enquêtes de déplacement dispensées en face à face ou par téléphone sont utilisées et acceptées depuis longtemps, les méthodes que nous souhaitons expérimenter sont potentiellement intrusives. C'est pourquoi, nous souhaitons évaluer si ces méthodes sont utilisables dans une enquête sur les déplacements d'un individu, en termes de respect de la vie privée. Une première question à se poser est la suivante : est-ce que des personnes vont se porter volontaires pour de telles enquêtes et vont accepter de porter à tout instant un

appareil aussi intrusif enregistrant automatiquement leurs coordonnées, selon des fréquences temporelles définies par le protocole d'enquête ? Si elles sont volontaires pour l'enquête, et donc volontaires pour que leurs positions spatio-temporelles soient connues à tout moment de l'enquête, seront-elles toujours d'accord pour que leurs traces soient exploitées afin d'en extraire d'autres informations personnelles (habitudes, préférences, lieux de passage, etc.) ?

2.3. Enjeux technologiques

Nous faisons l'hypothèse que l'utilisation de dispositifs mobiles de géolocalisation (GPS enregistreurs ou téléphones mobiles équipés de GPS) peut nous aider à acquérir les déplacements des individus pendant la période d'enquête ; ceux-ci pourront être ensuite analysés, de façon automatique ou avec l'aide de l'utilisateur, pour acquérir des informations sémantiques. Nous considérons que les individus peuvent utiliser tout moyen de transport et circuler en intérieur comme en extérieur. L'utilisation de GPS soulève des problèmes techniques liés à la qualité de la réception du signal, notamment en ville, dans les bâtiments ou les véhicules disposant d'habitacles fermés. Même s'il est évident que certains modes de déplacement posent des problèmes insolubles, par exemple le métro souterrain, la sensibilité des nouvelles générations de puces de réception GPS nous permet d'espérer contourner ces problèmes, en proposant notamment des techniques de filtrage et d'interprétation des enregistrements.

La description de ces enjeux nous amène à nous intéresser aux travaux existants proposant des méthodologies d'analyse des mobilités des individus.

3. Etat de l'Art : Méthodologies d'analyse des mobilités des individus

Nous nous inscrivons dans l'approche « activités-centrée » (*activity based approach*), paradigme sur l'étude des mobilités, apparu dans les années 70 (Hägerstrand, 1970, Chapin, 1974, Axhausen *et al.*, 1992) : les comportements de mobilité reposent sur des composantes sociales et des contraintes spatio-temporelles, qu'il convient d'étudier en combinant les trois éléments « temps », « espace » et « activités ». Dans cette partie, nous présentons les méthodologies mises en place pour connaître les déplacements d'individus et pour analyser les comportements humains face aux modifications des conditions météorologiques : les enquêtes type « enquête ménages déplacements » et les enquêtes à l'aide de technologies mobiles.

3.1. Les enquêtes « ménages déplacements » standardisées du CERTU

La connaissance des déplacements des individus est acquise en France, depuis 1973, grâce aux « enquêtes ménages déplacements » (CERTU, 1998) mises en place initialement pour accompagner les politiques de transport dans les villes : ces enquêtes, réalisées en face à face ou par téléphone, questionnent les membres d'un ménage sur leurs déplacements de la veille. Les enquêtés remplissent avec l'enquêteur des renseignements socio-démographiques sur eux-mêmes et sur leur ménage, ainsi que le nombre de véhicules à moteur ou non dans le foyer, le détail de chacun de leurs déplacements (heure et lieu de départ et d'arrivée, mode de transport, personnes accompagnantes, etc.), et les motifs engendrant ces déplacements. Les déplacements sont ensuite décomposés en trajets en fonction du mode de transport mécanisé utilisé. Les lieux sont identifiés par l'indication des zones dans lesquelles ils se trouvent : on parle de zones Origine-Destination, issues d'un découpage spatial dont les mailles dépendent du territoire observé (à l'échelle d'un sous-quartier ou d'une commune selon la densité de personnes au km²). Un déplacement va donc se décrire par exemple de la façon suivante : « *Domicile (zone AB42) à 8h30 -> Travail (zone VG12) à 9h15, composé de deux trajets : le premier en bus incluant 5 minutes de marche à pied pour rejoindre l'arrêt de bus et le second en tramway comprenant 10 minutes de marche pour rejoindre le lieu de travail* ».

Si ces enquêtes permettent de connaître les zones d'origine et de destination pour chacun des trajets effectués par l'individu, elles ne permettent pas de connaître précisément la localisation de l'individu, dans ces zones et tout le long de ses trajets. La connaissance sur les chemins empruntés n'est pas acquise : on ne sait pas par où l'utilisateur est passé, on ne peut donc pas en déduire des informations expliquant son temps de trajet (*s'il a pris le périphérique à 7h, cela expliquerait qu'il ait mis 40 minutes pour relier ces deux zones, plutôt que les 30 minutes requises*), ni des variations sur ces déplacements (*des fois, je passe par cette route, des fois par celle-là*).

Par ailleurs, ces enquêtes reposent sur la mémoire qu'ont les enquêtés de leurs propres déplacements, trajets et motivations. Il n'est pas évident de se souvenir précisément de chacun de ses déplacements, en particulier des petits déplacements (durée et/ou distance courte), de ceux qui sont très habituels (aller chercher du pain), spécifiques à un jour de la semaine (*c'était un jeudi, donc j'ai accompagné ma fille au karaté*), ou encore lorsqu'ils sont nombreux. L'enquêté a également une vision légèrement déformée du temps passé à une activité : les personnes non contraintes par des horaires fixes peuvent rencontrer des difficultés à fournir les heures exactes de leurs trajets. Nous avons-nous-mêmes observé ces imprécisions pendant la phase d'expérimentation décrite en section 5.

Malgré ces inconvénients, ce recueil riche d'information permet de nombreuses exploitations. Ces enquêtes permettent de décrire finement les pratiques de déplacements des ménages des agglomérations (de plus de 100000 habitants), voire

des grandes régions urbaines : l'enquête de 2002 de la grande région urbaine grenobloise a été utilisée par André-Poyaud *et al.* (2008) pour répondre à des questions sur les emplois du temps des individus. D'autres enquêtes portent sur de plus petites agglomérations, mais aucune n'a été réalisée dans le département du Gard.

3.2. Les enquêtes à l'aide de technologies mobiles

Dans le domaine de la géomatique, l'avènement des technologies mobiles a entraîné de profondes évolutions dans le domaine de l'analyse des mobilités. De nombreuses enquêtes se sont intéressées aux déplacements d'objets mobiles : voitures (CMAP, 2007), navires (Noyon *et al.*, 2007, Etienne *et al.*, 2009), etc. Dans cet article, nous nous intéressons spécifiquement aux trajectoires des individus. Sans être exhaustif, nous citons les travaux de (Spek, 2006, Asakura *et al.*, 2007, CMAP, 2007, MIRO, 2007, Millonig *et al.*, 2009, Chaker *et al.*, 2009, Godard *et al.*, 2009, Buard, 2010, Taczanowska, 2010), à partir desquels nous avons tenté de construire une typologie des méthodes d'analyse des mobilités :

– les travaux s'intéressant aux interactions entre la dynamique du milieu et la dynamique des individus l'investissant, comme le travail de Buard (2010) qui consiste à déterminer les origines et destinations des déplacements d'animaux sauvages et à décrire des déplacements-type (lieux attractifs, lieux de pause et couloirs à partir de traces GPS et de comptages). Spek (2006) observe les mouvements et les densités d'individus dans la ville par rapport à l'évolution de l'offre des réseaux de transport en commun, pour améliorer l'aménagement urbain. Taczanowska (2010) utilise des enquêtes GPS dans un parc de loisirs afin de vérifier l'adéquation entre les itinéraires VTT conseillés et ceux réellement effectués.

– les travaux s'intéressant à l'identification de typologies de comportement de déplacement, par exemple Millonig *et al.* (2009) se sont attachés à fournir une typologie des déplacements piétons en zone urbaine, dans le contexte d'une activité de *shopping*. Asakura *et al.* (2007) cherchent à déterminer des caractéristiques topologiques des comportements touristiques à Kobe (Japon).

– les travaux cherchant à simuler les déplacements d'individus, souvent à partir de données acquises dans les deux types de travaux précédemment décrits, en construisant des populations synthétiques dont les comportements peuvent être étudiés dans des plateformes de simulation (MIRO, 2007). Le projet CODIGEOSIM consiste à proposer des plateformes de géo-simulation sur l'expansion de maladies (maladie de Lyme, Virus du Nil) (Chaker *et al.*, 2009, Godard *et al.*, 2009).

En parallèle, le post-traitement des traces GPS brutes acquises par les enquêtes de suivi des individus est un sujet largement traité ces dernières années (Schönfelder *et al.*, 2002, Dykes *et al.*, 2003, Elgethum *et al.*, 2003, Du *et al.*, 2007, Laube *et al.*, 2007) : il consiste en général en cinq étapes -- filtrage, détection de trajets,

accrochage à la carte sous-jacente, détection des modes de transport, détection des activités. Stopher (2008) et Schuessler *et al.* (2009) résument les nombreux travaux de recherche sur ces différentes étapes et proposent eux-mêmes des méthodes de post-traitement. La clé de la réussite du post-traitement réside dans le développement de méthodes appropriées de filtrage des traces GPS. Cependant, ces méthodes dépendent des grandeurs fournies par les appareils GPS, ainsi que de leur qualité : si la plupart des GPS fournissent les coordonnées (latitude, longitude, altitude), la date, et la vitesse, tous ne fournissent pas les informations sur la précision de la trace.

4. Propositions de protocoles d'enquêtes et méthodes d'analyses associées

Pour répondre à notre besoin, nous proposons la démarche d'enquête globale suivante : (1) acquérir les positions (les sujets enregistrent leurs déplacements à l'aide d'un dispositif mobile qu'ils portent sur eux) ; (2) traiter les positions acquises et (3) enrichir la connaissance des déplacements (les sujets répondent à un questionnaire fermé sur leurs déplacements proposé selon différents modes de passation). A partir de cette démarche globale, nous proposons deux protocoles d'enquête. Le premier est basé sur un GPS enregistreur puis un questionnaire en face à face avec un enquêteur ; le second est basé sur un téléphone mobile équipé d'un GPS et disposant d'un questionnaire proactif.

4.1. Protocole A : GPS enregistreur + interview par un enquêteur

Le protocole A est une enquête de type « face-à-face », c'est-à-dire qu'après la phase d'acquisition de la trace, le sujet doit répondre à des questions posées par l'enquêteur. Le post-traitement des traces GPS brutes est réalisé de façon automatique par un enquêteur. La démarche est la suivante :

- (1) Le sujet se déplace avec un GPS pendant la période d'enquête, puis rapporte le GPS au lieu de rendez-vous.
- (2) L'enquêteur décharge les traces GPS brutes et effectue le post-traitement de ces traces (Cf. 4.3). Des arrêts déjà connus peuvent avoir été préalablement mémorisés dans une base de données.
- (3) L'enquêteur et le sujet visualisent les traces GPS brutes et filtrées et s'entretiennent au sujet des déplacements, selon un questionnaire fermé : le sujet valide ou non les arrêts et les trajets détectés puis précise l'activité pour chaque arrêt, ainsi que le mode et les personnes accompagnantes pour chaque trajet. Les nouveaux arrêts identifiés sont mémorisés dans la base de données.

4.2. Protocole B : téléphone mobile avec GPS + questionnaire proactif

Le protocole B fait appel à une enquête automatique réalisée pendant la période d'acquisition de la trace, à l'aide d'un questionnaire proactif prenant lui-même l'initiative de poser des questions à l'utilisateur. Le matériel prévu est un téléphone mobile de type *Smartphone* muni d'un GPS interne et d'un clavier alphanumérique. La démarche est la suivante :

- (1) Le sujet se déplace avec un téléphone pendant la période d'enquête.
- (2) Les traces GPS brutes sont enregistrées en interne dans le téléphone qui effectue lui-même le post-traitement (Cf. 4.3).
- (3) Une fois l'arrêt détecté, le téléphone vérifie si l'arrêt est déjà mémorisé dans la base de données du sujet. Si c'est le cas, l'utilisateur est informé de la nature de l'activité reconnue. En cas d'erreur ou de changement d'activité, l'utilisateur peut désigner une autre activité (« Vous vous êtes arrêté à l'école de vos enfants, sinon quelle est votre activité ? »). Si l'arrêt n'est pas mémorisé dans la base de données, le téléphone demande pro-activement à l'utilisateur la nature du lieu et l'activité réalisée (« Vous vous êtes arrêté, quelle est votre activité ? »).

Les traitements effectués (filtrage, analyse, accès à la base de données des arrêts, déclenchement du questionnaire proactif, enregistrement des réponses) peuvent être effectués au niveau du téléphone ou en partie délégués à un serveur de calcul.

4.3. Des méthodes de post-traitement des traces GPS

Nous proposons des méthodes de filtrage et de détection des arrêts, afin d'une part de faciliter la visualisation et la compréhension des traces par le sujet, d'autre part d'avoir des traces suffisamment précises pour des analyses ultérieures.

4.3.1 Méthodes de filtrage

Les traces GPS brutes acquises possèdent sporadiquement certains points imprécis dont la position est éloignée de plusieurs dizaines de mètres de la position réelle. Ces points imprécis sont obtenus lorsque le nombre de satellites ou leur géométrie n'est pas optimale, par exemple, le long de rues étroites. En outre, lorsque le sujet stationne dans un bâtiment, ces points imprécis deviennent très nombreux et forment des étoiles appelées « positions fantômes » dont certaines branches peuvent être très éloignées de la position réelle.

Afin de minimiser les points imprécis, nous utilisons les grandeurs sur l'estimation de la précision horizontale (HDOP) et sur le nombre de satellites utilisés (SAT) pour calculer la position. Théoriquement, un filtrage sur la seule valeur de la HDOP aurait dû être suffisant. En pratique, nous nous sommes aperçus qu'il n'était pas suffisant et qu'un filtrage sur SAT était plus performant. Plusieurs seuils pour les

valeurs de HDOP et SAT ont été testés. La difficulté du filtrage réside dans le compromis entre la suppression des points imprécis et la complétude de la trace : un filtrage trop restrictif retire l'ensemble des points imprécis mais aussi une partie significative du déplacement de l'individu. Ces seuils doivent être ajustés en fonction des conditions expérimentales. Un GPS utilisé dans un déplacement en terrain dégagé supportera un filtrage plus restrictif qu'un GPS utilisé dans les rues d'une ville entourée de montagnes. Nos tests ont montré que les valeurs de seuil de $HDOP > 2$ et $SAT < 5$ sont un bon compromis pour nos conditions expérimentales.

4.3.2 Méthodes de détection des arrêts

Nous avons utilisé une méthode relativement simple de détection des arrêts : à partir de la trace GPS filtrée, nous extrayons l'ensemble des points consécutifs dont la vitesse est inférieure à 0,5 km/h (~0, 14 m/sec.). Nous avons remarqué dans l'ensemble des traces acquises lors des phases de pré-expérimentation, que peu de points ont des vitesses nulles : les erreurs de positionnement du GPS font que la vitesse calculée est rarement nulle, même en cas d'arrêt prolongé, où beaucoup de points sont acquis à vitesse très faible. Nous choisissons d'améliorer cette présélection de points, en vérifiant le temps écoulé entre deux de ces points. Si le temps écoulé est inférieur à trois minutes, alors nous considérons que le deuxième point correspond au même arrêt que le premier, sinon, c'est un nouvel arrêt détecté.

5. Expérimentation : comparaison d'une enquête EMD et d'une enquête GPS

Préalablement à l'évaluation des protocoles A et B, nous avons effectué deux types de tests techniques. Un premier test avait pour objectif la sélection d'un modèle de GPS enregistreur parmi l'ensemble des GPS présents sur le marché, selon les critères suivants : sensibilité, autonomie, capacité d'enregistrement et facilité de récupération des données. Notamment, l'autonomie et la capacité d'enregistrement des données devaient permettre des expérimentations de 24h au minimum. Au total, 6 GPS ont été sélectionnés sur les données publiées par les constructeurs et testés sur le terrain. Si la sensibilité est à peu près équivalente pour l'ensemble des GPS (utilisant majoritairement des puces basées sur l'architecture SiRFstarIII²), un seul nous permettait d'avoir l'autonomie et la capacité d'enregistrement adéquates. Un second test technique a été réalisé afin de constituer un corpus de déplacements annotés de manière à calibrer les méthodes de filtrage et de détection des arrêts. Ce test a été par la suite complété par une évaluation des meilleurs emplacements de transport du GPS, car nous avons constaté des variations importantes de la qualité des données interindividuelles liée à la position et l'orientation du GPS. Il est ainsi

² <http://www.sirf.com/>

préférable de donner des consignes strictes aux enquêtés afin d'obtenir des données exploitables.

Au moment de la rédaction de cet article, seul le protocole A a été testé. L'objectif de cette expérimentation était dans un premier temps de comparer des enquêtes EMD et GPS, puis dans un deuxième temps de valider notre protocole A d'enquête GPS.

5.1. Déroulement de l'expérimentation

L'expérimentation, réalisée par 13 étudiants, suit les trois phases définies précédemment : (1) acquisition des traces, (2) post-traitement, (3) interviews des sujets. Pour la phase 3, les étudiants sont regroupés en binômes « enquêteur - enquêté » et chacun leur tour, répondent d'abord à l'enquête EMD, puis à l'enquête GPS. La Figure 1 récapitule les trois phases de l'expérimentation et montre un exemple d'une trace brute issue de la phase 1, d'une trace filtrée et d'arrêts détectés issus de la phase 2.

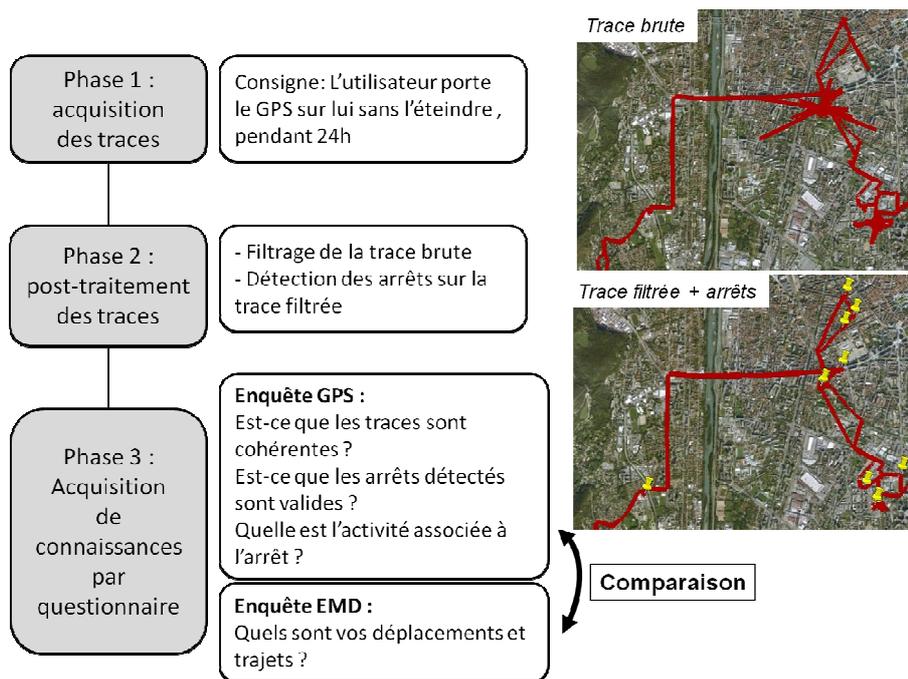


Figure 1. Expérimentation Enquête EMD vs Enquête GPS

L'enquête EMD correspond à une EMD traditionnelle dont nous n'avons conservé que les aspects liés aux déplacements : les sujets doivent décrire minutieusement chacun de leurs déplacements, décomposés en trajets et activités, pendant les dernières 24h. Ils disposent de fiches déplacements et trajets pré-remplies et d'une carte représentant le découpage en zones origine-destination de Grenoble³ permettant de localiser l'origine et la destination de chaque trajet.

L'enquête GPS suit le protocole A : les sujets valident leurs traces brutes et filtrées, les arrêts détectés, et enrichissent leurs activités aux arrêts. Pour la partie visualisation avec le sujet, nous avons choisi d'utiliser dans un premier temps, Google Earth⁴ pour une visualisation ludique des traces brutes et filtrées, dans un deuxième temps, le logiciel SIG QuantumGIS⁵ pour la visualisation et l'enrichissement de la couche des arrêts détectés sur un fond de données topographiques.

Il faut noter que certains arrêts détectés peuvent ne pas être des arrêts significatifs pour l'enquêté, mais seulement des arrêts imposés par le moyen de transport utilisé (par exemple un arrêt au feu). L'enquêté peut également rajouter un arrêt qui n'aurait pas été détecté automatiquement. La Figure 2 montre la table attributaire d'une couche d'arrêts détectés sur un autre cas de trace GPS filtrée : l'utilisateur y précise, pour chaque arrêt détecté, sa validité, et s'il est valide, l'activité associée⁶.

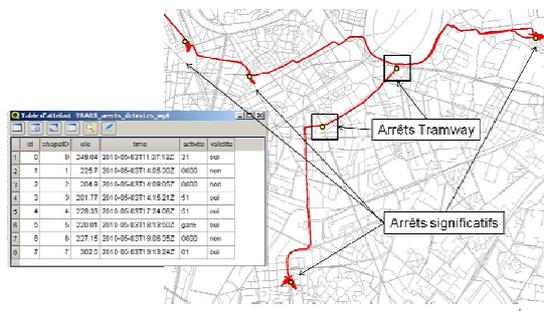


Figure 2. Validation des arrêts et précision de l'activité à l'arrêt sur trace filtrée

³ Le découpage des zones OD a été réalisé par l'Agence d'Urbanisme de Grenoble.

⁴ <http://earth.google.fr/>

⁵ <http://qgis.org/>

⁶ Les activités sont codées selon les catégories fournies par l'EMD (CERTU, 1998)

5.2. Résultats et discussion

Cette expérimentation a permis dans un premier temps de comparer le protocole d'enquête EMD avec notre premier protocole d'enquête GPS. En plus des inconvénients d'une EMD déjà relevés dans la première partie de cet article, nous avons pu identifier les principaux apports d'une enquête GPS par rapport à une EMD en termes :

- d'identification des trajets et des chemins empruntés ;
- de mémorisation des petits arrêts et des petits déplacements dans le temps ou ceux pour lesquels l'activité est jugée « insignifiante » ou « pas importante » ;
- de prise en compte de tous les trajets, même les trajets à pied.

Nous pouvons penser que l'expérimentation proposée possède un biais, car les sujets ont répondu à l'EMD avant de visualiser les résultats de l'enquête GPS, et donc qu'ils ont déjà fait l'effort de mémorisation nécessaire. Malgré cela, en début d'enquête GPS, nous avons observé que les sujets mettent quelques minutes avant de comprendre ce qu'ils visualisent : l'EMD leur a permis de se remémorer les origines-destinations des déplacements, mais pas les chemins empruntés pour les atteindre. L'EMD incite les sujets à penser un déplacement en termes de repères dans le temps et l'espace (début et fin), et non à penser à l'inscription de chaque moment de son trajet dans le temps et dans l'espace. Cette nuance de perception de la notion de trajet occasionne des oublis et des désorientations. Or, une trace GPS permet de réparer cela. Le temps d'observation et de questionnement (« C'est où ? », « Quand est-ce que j'étais là ? ») est nécessaire pour comprendre son empreinte dans l'espace et dans le temps. Ce protocole permet aussi de mettre en évidence des activités « imprévues », i.e. des activités ayant été décidées au dernier moment (acheter du pain, aller poster une lettre, etc.).

L'analyse automatique des traces GPS met en évidence tous les arrêts sur la trace : les sujets peuvent passer du temps à s'interroger sur certains de ces arrêts car ils n'en trouvent pas immédiatement la signification (« je ne me suis pas arrêtée là : pourquoi j'étais là ? ») alors qu'il s'agit d'un arrêt de leur moyen de transport (arrêt du tramway ou de la voiture à un feu rouge, etc.).

Par ailleurs, nous avons observé que l'animation des traces proposée par l'outil utilisé (Google Earth) a présenté ici un intérêt certain. L'animation cartographique a permis aux sujets de réaliser le temps accordé à des arrêts ou à des trajets et a favorisé chez le sujet la reconstruction des cheminements et des arrêts.

Si l'expérimentation a permis de valider certains apports du protocole d'enquête proposé, quelques inconvénients ont toutefois été soulevés, notamment en termes de post-traitement des traces. Dans la partie précédente, nous avons précisé qu'un filtrage trop restrictif retire l'ensemble des points imprécis mais aussi une partie significative du déplacement de l'individu. Or, la validation des traces brutes et filtrées par les étudiants montrent qu'il est nécessaire d'assouplir notre méthode de

filtrage, quitte à conserver des positions fantômes importantes. Elles pourront être filtrées dans un deuxième temps : nous travaillons sur une méthode de recherche de clusters de points sur la trace. En outre, les positions fantômes étant représentatives d'arrêts longs en intérieur, elles correspondent pour la plupart aux arrêts les plus communs (domicile et travail) qui peuvent être connus au préalable.

Cette expérimentation nous a permis de valider le protocole A que nous proposons. Si aucune difficulté majeure n'a été relevée, à part l'apprentissage nécessaire sur les outils Google Earth et QuantumGIS, en revanche le besoin d'un système d'information intégré permettant d'assurer les différentes étapes du protocole a été ressentie, du traitement de la trace, jusqu'à l'annotation sémantique en passage par des fonctionnalités de visualisation spécifiques, tel celui proposé par Guc *et al.* (2009). Les sujets auraient également apprécié qu'un lien soit fait près de leurs arrêts significatifs, avec des activités potentielles (bar avec retransmission d'un match, boutiques, restaurant, etc.), ce qui rejoint nos perspectives de travail.

6. Conclusion

Notre étude préliminaire a montré l'intérêt de l'enquête assistée par un GPS pour l'étude des déplacements. Notamment, le suivi GPS apporte une plus grande précision et fiabilité à la classique EMD. Ce travail se poursuit actuellement pour finaliser le protocole d'enquête basé sur les GPS enregistreurs et afin de développer l'enquête basée sur les téléphones mobiles équipés de GPS.

Dans cet article, nous nous sommes focalisés sur les enjeux méthodologiques et technologiques du projet TRACES afin de proposer une méthode d'acquisition de connaissances sur les déplacements des individus. Nous sommes maintenant à la recherche d'un panel de volontaires dans la région du Gard pour effectuer une expérimentation en situation quotidienne « normale » et en situation de fort risque de crues rapides (automne). La contrainte que nous nous donnons est que les volontaires doivent converger quotidiennement vers un lieu commun qui pourra être le lieu de distributions et retours des appareils, ainsi que de rencontres avec l'enquêteur (lieu de travail, école, etc.). Dans cette recherche, nous sommes confrontés aux enjeux déontologiques décrits dans cet article.

La constitution de ce panel nous permettra alors de vérifier en situation réelle l'acquisition d'informations sémantiques sur les déplacements, afin d'en déduire des connaissances sur les comportements effectifs des individus en période de risque de crues rapides. Ces connaissances permettront d'aider à quantifier la perception du risque par les individus et l'impact des trajets « prévus » et « imprévus » dans la prise de décision, afin d'améliorer les propositions interactives sur les routes à prendre ou à ne pas prendre et les systèmes d'information sur les risques.

7. Bibliographie

- André-Poyaud I., Chardonnel S., Charleux L., Tabaka K., « La mobilité au cœur des emplois du temps des citoyens », in : Chalas Y., Paulhiac F. (Dir), *La mobilité qui fait la ville*, Certu, Lyon, 2008, p. 67-95.
- Axhausen K. W., Gärling T., « Activity-based approaches to travel analysis: conceptual frameworks, models, and research problems », *Transport Reviews*, 12 (4), 1992, p. 323-341.
- Asakura Y., Iyoro T., « Analysis of Tourist Behaviour Based on the Tracking Data Collected Using a Mobile Communication Instrument », *Transportation Research Part A: Policy and Practice*, 2007, vol. 41, issue 7, 2007, p. 684-690.
- Buard E., « Comparaison des échelles spatio-temporelles décrivant les dynamiques de populations animales et les dynamiques d'utilisation du sol », *Actes de Géopoint 2010, 18è biennale de Géographie, Les échelles pour les géographes et les autres*, juin 2010.
- CERTU, L'enquête ménages déplacements « Standard Certu », Guide méthodologique, 1998.
- Chaker W., Proulx M.-J., Moulin B., Bédard Y., « Modélisation, Simulation et Analyse d'Environnements Urbains Peuplés : Approche multi-agent pour l'étude des déplacements multimodaux », *Revue Internationale de Géomatique*, Hermès, vol 19/4, 2009, p. 413-441.
- Chapin F.S., *Human activity patterns in the city: things people do in time and in space*, John Wiley & Sons, New York, 1974, 272 p.
- Christophe S., André-Poyaud I., Chardonnel S., Davoine P.A., Jambon F., Lutoff C., et al., « Human Behaviour Analysis during Remarkable Flood Episodes: Proposition of Automatic Survey Protocols based on Mobile Technologies », *Proceedings of the Workshop on "Movement Research: are you in the flow?"*, AGILE 2010.
- CMAP, Chicago Regional Household Travel Inventory, Draft GPS final report, prepared for Chicago Metropolitan Agency for Planning (CMAP), 2007. http://www.goto2040.org/uploadedFiles/regional_data/TravelTrackerSurvey/TravelTrackerSurveyGPSReport.pdf (2010, June, 17th).
- Du J., Hall L., « Increasing the accuracy of trip rate information from passive multi-day GPS travel datasets: Automatic trip end identification issues », *Transportation Research Part A: Policy and Practice*, 41(3), 2007, p. 220-232.
- Dykes J. A., Mountain D. M., « Seeking structure in records of spatio-temporal behaviour: visualization issues, efforts and applications », *Computational Statistics and Data Analysis*, 43(Data Visualization II Special Edition), 2003, p. 581-603.
- Elgeth K., Fenske R. A., Yost M.G., Palcisko G.J., « Time-location analysis for exposure assessment studies of children using a novel global positioning system instrument », *Environment Health Perspectives*, 111(1), 2003, p. 115-122.
- Etienne L., Devogele T., Bouju A., « Analyse de similarité de trajectoires d'objets mobiles suivant le même itinéraire : Application aux trajectoires de navires », *Revue Ingénierie des Systèmes d'Information (ISI)*, numéro thématique Systèmes d'information et géolocalisation, Servigne, S., Zeitouni, K. (eds.), Hermès, Vol.14 (5), 2009, p. 85-106.
- Godard V., Méha K., Benabderrahmane M.C., « Modeling of human Exposure to Lyme disease Risk in a French forest Landscape », *CODIGEOSIM Workshop on geosimulation*

and mathematical modelling for zoonotic diseases, Toronto, Canada, 19-21 Août 2009, York University.

- Guc B., May M., Saygin Y., Körner C., « Semantic annotation of GPS trajectories », *Proceedings of the 11th AGILE International Conference on Geographic Information Science (AGILE)*, 5-8 Mai 2008.
- Hägerstrand T., « What about people in regional science? », *Papers of the Regional Science Association*, 24, 1970, p.7-21.
- Laube P., Dennis T., Forer P., Walker M., « Movement beyond the snapshot - Dynamic analysis of geospatial lifelines », *Computers, Environment and Urban Systems*, 31, 2007, p. 481-501.
- Maples L.Z., Tiefenbacher J.P., « Landscape, development, technology and drivers: The geography of drowning associated with automobiles in Texas floods, 1950-2004 », *Applied Geography*, 2008.
- Millonig A., Brändle N., Gartner G., « Supporting pedestrian wayfinding by analysing movement patterns », *Proceedings of the XXIVe International Cartographic Conference (ICC'09)*, Santiago de Chile, 15-20 Novembre 2009.
- MIRO, Rapport final du projet MIRO (Modélisation Intra-urbaine des Rythmes quOtidieNs), Septembre 2007.
- Noyon V., Claramunt C., Devogele D., « A Relative Representation of Trajectories in Geographical Spaces », *Geoinformatica*, Springer (Publisher), vol. 11, 4, 2007, p.479-496.
- Ruin I., Conduite à contre-courant. Les pratiques de mobilité dans le Gard : facteur de vulnérabilité aux crues rapides, Thèse de l'Université Joseph Fourier, 2007.
- Schönfelder S., Axhausen K., Antille N., Bierlaire M., « Exploring the potentials of automatically collected GPS data for travel behaviour analysis, A Swedish data source », *GI-Technologien für Verkehr und Logistik*, 13, 2002, pp. 155-179.
- Schüssler N., Axhausen K.W., « Processing GPS raw data without additional information », *Transportation Research Record*, 2105, 2009, p. 28-36.
- Spek, SC van der. « Pedestrian oriented design, designing the sustainable city - a tool to optimize pedestrian routing ». In V Wang, Q Sheng, & C Sezer (Eds.), *Modernization and Regionalism. Re-inventing Urban Identity.*, vol. 2, 2006, p. 693-697.
- Stopher P. R., « Collecting and processing data from mobile technologies », paper presented at 8th International Conference on Survey Methods in Transport, Annecy, Mai 2008.
- Taczanowska K., Tomek H., Brandenburg C., Muhar A. « Investigating the spatial distribution of visitors in recreational areas - Contribution of GPS-based tracking to a better understanding of human spatial behaviour outdoors », *Proceedings of the Workshop on Movement Research: Are you in the flow?*, AGILE 2010, Guimaraes, Portugal.

Mesures de similarité de trajectoires suivant le même itinéraire

Thomas Devogele¹, Laurent Etienne²

1 : Université François Rabelais Tours / LI, 3 place Jean Jaurès, 41000 Blois
e-mail : thomas.devogele@univ-tours.fr

2 : Institut de Recherche de L'Ecole navale, Lanvéoc 29240 Brest
e-mail : laurent.etienne@ecole-navale.fr

RÉSUMÉ. Les systèmes de suivi d'objets mobiles surveillent en temps réel la dernière position de ces objets. Le stockage des données historiques offre des perspectives encourageantes en termes d'analyse de trajectoires et d'enrichissement de ces systèmes par des connaissances sur les déplacements et les comportements habituels liés aux usages. Ainsi, la fouille de données dans les historiques des déplacements d'objets suivant un même itinéraire et de même type permet de définir des patrons spatio-temporels : les routes types. En s'appuyant sur ces patrons, il est alors possible de détecter en temps réel des positions inhabituelles. Or, les positions d'un objet mobile prises individuellement peuvent être normales sans pour autant impliquer que la trajectoire soit habituelle. Il est donc nécessaire de définir des mesures de similarité entre une trajectoire d'un point de vue global et un patron de trajectoires. Cet article définit des mesures de similarité portant sur les aspects spatiaux et temporels. Il propose par la suite de déterminer un indice de similarité basé sur ces mesures et la logique floue.

ABSTRACT. Mobile objects are now equipped with sensors allowing real time monitoring of their movements. Nowadays, the data produced by these sensors is received and stored in spatio-temporal databases in order to visually follow their movements. Data mining on this huge quantity of stored positions allows to extract the behaviour of these mobile objects (spatio-temporal patterns) and to analyze in real time trajectories of mobile objects following the same itinerary. Using these patterns, unusual situations can be detected. Two kinds of unusual situations are distinguished: unusual positions and unusual trajectories. The first one describes position delayed or far away from patterns positions. The second one describes trajectories dissimilar to spatio-temporal patterns. This article defines similarity index based on spatial and temporal measures and fuzzy logic.

MOTS-CLÉS : objet mobile, analyse de trajectoires, fouille de données spatio-temporelles, mesure de similarité, route type, logique floue.

KEYWORDS: mobile object, trajectory analysis, spatio-temporal data mining, similarity measure, main route, fuzzy logic.

1. Introduction

Le suivi d'objets mobiles est utilisé couramment dans de nombreux domaines tels que la migration des animaux (Mefteh et *al.*, 2008), le suivi de phénomènes météorologiques (Lee et *al.*, 2008), les mouvements de foules, de piétons (Knorr et *al.*, 2000), les déplacements de véhicules (automobiles, aéronefs, navires...) (Wang et *al.*, 2006; Kharrat et *al.*, 2008). Or, les systèmes de suivi d'objets mobile, le plus souvent, se contentent de surveiller en temps réel la dernière position de ces objets. Cependant, le stockage des données historiques dans des entrepôts de données offre des perspectives encourageantes en termes d'analyse de trajectoire et d'enrichissement de ces systèmes par des connaissances sur les déplacements et les comportements habituels. Actuellement, le volume de données à traiter en temps réel est de plus en plus important ce qui entraîne une surcharge cognitive des opérateurs. Une aide à la décision est donc indispensable pour faciliter la détection de positions ou de trajectoires inhabituelles. De tels outils d'analyse de trajectoires couplés à une visualisation appropriée offrent, à des opérateurs de surveillance, la possibilité de se focaliser sur certaines trajectoires considérées comme anormales. En fonction de l'application, ces positions peuvent refléter le déplacement d'un animal malade, d'un piéton ayant une attitude suspecte ou un navire s'éloignant de la route pour en éviter un autre. Pour identifier ces dernières, il est nécessaire au préalable de déterminer des patrons spatio-temporels caractérisant les déplacements habituels. Cet article se focalise sur des patrons décrivant des trajectoires d'un même type d'objets se déplaçant dans un espace ouvert, ayant une origine et une destination communes. A l'aide de ces patrons, il doit être possible de définir dans quelle zone est sensée se situer une position d'un objet suivant cet itinéraire en fonction de la durée depuis le départ. Néanmoins, les positions d'un objet mobile prises individuellement peuvent être normales sans pour autant impliquer que la trajectoire soit habituelle. Il est donc nécessaire de définir des mesures de similarité entre une trajectoire d'un point de vue global et un patron de trajectoires. Il est donc indispensable de mesurer la similarité entre une trajectoire et un patron. Ces mesures doivent porter sur les aspects géométriques et temporels. De même, il est souhaitable de disposer d'un indice de similarité global basé sur ces mesures.

L'application de ces travaux au contexte maritime a pour objectif d'augmenter la sécurité maritime dans des zones denses. Ainsi, les comportements inhabituels de navires peuvent être mis en avant. L'entrepôt de données spatio-temporelles étudié dans l'exemple maritime contient 1005 navires ainsi que 4 821 447 positions recueillies sur une période de 30 mois dans la région brestoise. Chaque position est associée à un navire.

Le chapitre suivant de cet article introduit les méthodes employées pour définir les patrons spatio-temporels : les routes types. Les mesures de similarité proposées sont présentées dans le chapitre trois. Enfin, le quatrième chapitre explique comment obtenir un indice de similarité d'une trajectoire et d'un patron à l'aide de ces mesures et de la logique floue. Ces méthodes sont illustrées par des trajectoires de navires.

2. Définition de routes types

Ce chapitre résume succinctement la définition des patrons spatio-temporels de trajectoires dans un espace ouvert. Par espace ouvert, il faut comprendre que les objets ne se déplacent pas sur un réseau contraint tel l'espace ferré et l'espace routier. Un patron caractérise les itinéraires d'objets mobiles de même type et ayant la même origine et la même destination. Une description détaillée est disponible dans (Etienne et *al.* 2009). Ce travail repose sur le postulat suivant : dans un espace ouvert, des objets mobiles de même type ayant le même itinéraire suivent des trajectoires similaires optimisées en termes de temps, de distance et de sécurité. Ces patrons sont appelés **routes types**.

2.1. Extraction des groupes homogènes de trajectoires

Pour chaque type d'objet et chaque itinéraire, la première étape consiste à extraire de l'entrepôt de données le groupe de trajectoires d'un objet de ce type, suivant cet itinéraire. Ce groupe est appelé **groupe homogène de trajectoire (GHT)**. Dans l'exemple de la figure 1, le GHT du type "navires à passagers" suivant l'itinéraire Brest Arsenal - l'Ecole navale est affiché



Figure 1. Groupes des 554 trajectoires de navires à passagers suivant l'itinéraire Brest Arsenal → Lanvéoc Ecole navale

2.2. Détermination de la trajectoire médiane

La deuxième étape consiste à définir une trajectoire médiane. Cette trajectoire est une suite ordonnée des positions médianes. Ces dernières ont pour coordonnées la latitude médiane et la longitude médiane des positions des trajectoires du GHT au même temps relatif normalisé. Ce temps correspond à un pourcentage de la durée pour parcourir cet itinéraire. La figure 2 représente les trajectoires (trait clair fin) du GHT de la figure 1, un nuage de positions à un même temps normalisé (point noir) et la position médiane (point noir cerclé de blanc).

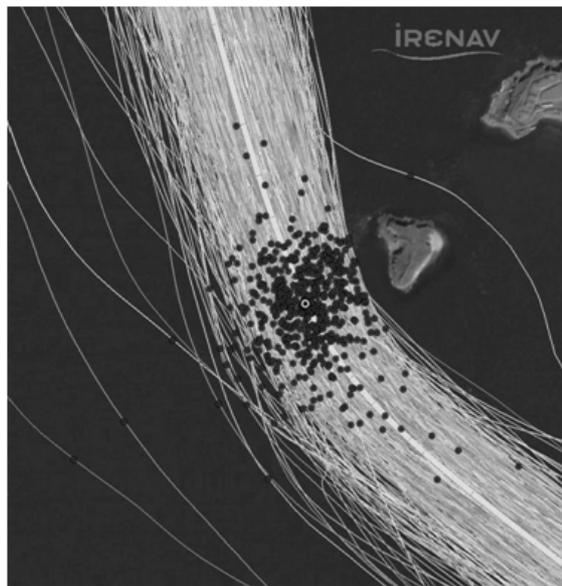


Figure 2. Définition d'une position médiane à partir des positions des objets mobiles au même temps relatif normalisé

L'utilisation des positions médianes permet d'éviter un biais pouvant être lié à des trajectoires anormales ayant un écart très important. La trajectoire médiane est filtrée (Bertrand et *al.* 2007) afin de ne garder que les positions les plus significatives.

2.3. Définition des couloirs spatio-temporels

Une fois cette trajectoire médiane obtenue, la route type doit être complétée afin de connaître l'écartement possible normal d'un point de vue spatial puis d'un point de vue temporel. Dans cet objectif, premièrement, un couloir spatial est défini, il contient P % des trajectoires à droite et P % des trajectoires à gauche. Ce pourcentage P est actuellement fixé empiriquement. Les trajectoires sont ordonnées à l'aide du point d'intersection entre la trajectoire et une **droite perpendiculaire au**

cap (DPC) en chaque position de la trajectoire médiane. Ces points d'intersection étant sur un espace à une dimension, il est possible de les ordonner totalement en fonction de la distance à la position médiane. La figure 3 représente en trait foncé épais, les frontières du couloir spatial autour de la trajectoire médiane pour l'exemple retenu avec un pourcentage de 90%.



Figure 3. Définition du couloir spatial (trait foncé épais) autour de la trajectoire médiane (trait épais clair)

De même, un couloir temporel est calculé. Il regroupe P % des trajectoires en retard et P % des trajectoires en avance à l'intérieur du couloir spatial. Seules les trajectoires à l'intérieur du couloir spatial sont prises en considération pour ce calcul statistique. En effet, les trajectoires à l'extérieur peuvent soit faire un détour ou prendre un raccourci. Il ne faut donc pas les prendre en compte. La DPC est aussi employée, les points d'intersection sont ordonnés en fonction de l'estampille temporelle de temps relatif (durée écoulée depuis le départ).

La route type est donc composée pour chaque GHT, d'une trajectoire médiane et pour chaque position médiane d'un écart spatial à droite toléré (ESDT), d'un écart spatial à gauche toléré (ESGT), d'un écart temporel de retard toléré (ETRT), et d'un écart temporel d'avance toléré (ETAT). A chaque temps relatif de position médiane, il est donc possible de définir cinq zones spatiales :

- dans le couloir à l'heure,
- dans le couloir en retard,
- dans le couloir en avance,
- à droite du couloir,
- à gauche du couloir.

La figure 4, représente les 5 zones obtenues au temps relatif de la figure 3.

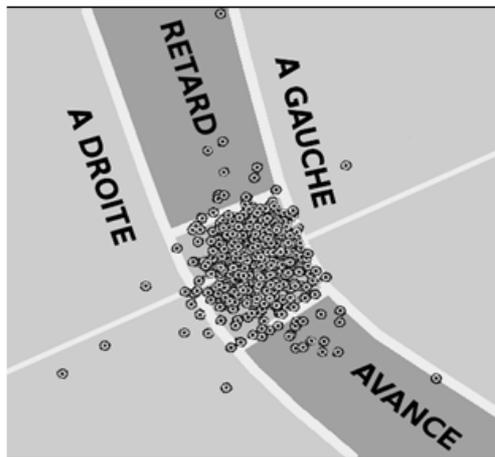


Figure 4. Partitionnement de l'espace en fonction du couloir spatio-temporel à un temps relatif

Il est ainsi possible pour chaque nouvelle position obtenue, de la qualifier en temps réels à l'aide du patron spatio-temporel associé à l'itinéraire suivi par l'objet mobile à cette position.

3. Mesure de similarité entre trajectoires

La qualification des positions est un critère fondamental pour identifier les situations inhabituelles. Néanmoins, cette information n'est pas suffisante. Effectivement, une trajectoire peut être une suite ordonnée de positions qualifiées de "dans le couloir, à l'heure" et être inhabituelle. Par exemple, des suites de changements de direction peuvent entraîner des variations de la trajectoire, la rapprochant de la frontière droite puis de la frontière gauche. Cette trajectoire doit être qualifiée d'inhabituelle. Il est donc indispensable de définir des mesures de

similarité entre trajectoires. (Lee et *al.* 2008) et (Pelekis et *al.* 2007) ont défini des mesures de similarité. Or, ces dernières ne traitent que de l'aspect géographique. Qui plus est, les distances proposées ne tiennent pas compte du contexte. Une distance de 10 m dans un espace de 15 m de large est grande or la même distance dans un espace de 1 000 m de large est relativement faible. De même, il est plus facile de prendre du retard que de l'avance. Nous proposons de définir trois mesures de similarité géographique et trois mesures de similarité temporelle basées sur un appariement entre une trajectoire et une trajectoire médiane.

3.1 Appariement de trajectoires

L'appariement de trajectoires consiste à définir des couples de positions entre des trajectoires partant de la même origine et ayant la même destination. Ce processus définit des couples de positions. Plusieurs options sont possibles pour appairer :

- Appariement des positions au même temps relatif (durée depuis le départ), ce choix donne de mauvais résultats si les vitesses sont dissemblables,
- Appariement des positions au même temps relatif normalisé, ce choix ne permet pas d'appairer des trajectoires partielles, non encore terminées,
- Appariement des positions selon le dynamic time warping (DTW) (Sakoe et Chiba 1978). Ce processus issu du traitement du signal donne de bons résultats, mais ne peut pas être employé pour des trajectoires partielles,
- Appariement des positions selon la distance de Fréchet discrète (Mascret et *al.* 2006). Ce processus d'appariement fournit des couples proches de ceux définis par le DTW et est employable pour des trajectoires partielles.
- Appariement des positions selon la droite perpendiculaire au cap (DPC), consiste à relier une position d'une trajectoire avec la position de l'autre trajectoire qui intersecte la DPC de la première. Cette méthode est rapide et est employable pour des trajectoires partielles, par contre elle n'est pas symétrique. Contrairement aux techniques précédentes, les positions de la trajectoire appariée sont interpolées.

La figure 5 illustre les 5 méthodes possibles d'appariement pour des trajectoires de deux navires. Les couples de positions appariées sont reliées. Une des deux dernières méthodes doit être retenue. Pour l'appariement d'une trajectoire avec une trajectoire médiane, le rôle de ces deux éléments n'étant pas symétrique, l'appariement selon la DPC est retenu. Qui plus est, cette technique (DPC) est déjà employée pour définir les bornes du patron spatio-temporel. Pour cet appariement chaque position appariée n'apparaît que dans un seul couple.

Figure 5. Couples des positions appariées de deux trajectoires selon les 5 méthodes proposées

3.2 Similarité géométrique de trajectoires

Une fois les couples de positions définis par le processus d'appariement, des mesures de similarité tenant compte du contexte peuvent être calculées. Le processus proposé se focalisant sur la comparaison d'une trajectoire d'un GHT avec la route type associée, il est possible de normaliser les distances géométriques en fonction de l'écart spatial toléré (EST_j) à cette position médiane PM_j . De même, l'ESDT et l'ESGT étant différents, l' EST_j est $ESDT_j$ si la position est à droite et $ESGT_j$ sinon. Le processus d'appariement entre la trajectoire T (suite ordonnée de positions P_i) et la trajectoire médiane TM (suite ordonnée de PM_j) renvoie une suite de couple (P_i, PM_j) . Les mesures entre P_i et PM_j sont exprimées relativement par rapport à l'EST en PM_j . La distance géométrique normalisée entre deux positions appariées est donc :

Afin de distinguer la position relative de P_i par rapport à la TM, les distances normalisées sont notées positivement à gauche et négativement à droite. La largeur EST_j est variable. Ainsi, pour la figure 6, deux distances normalisées en fonction de

la largeur du couloir : 110% pour une position légèrement à l'extérieur et 72% pour une position à l'intérieur correspondent à deux distances absolues identiques.

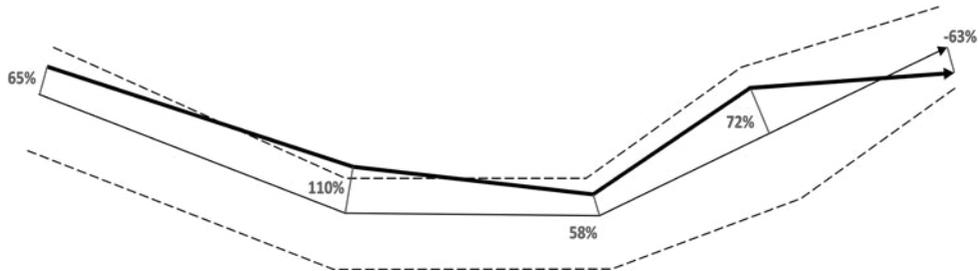


Figure 6. Couples des positions appariées d'une trajectoire avec une trajectoire médiane, les distances sont exprimées en fonction de la largeur du couloir et du côté

Pour la similarité géométrique, trois mesures ont été retenues :

- la distance géométrique maximale : **DGM**
- la distance géométrique moyenne : **DGm**
- la moyenne des deltas géométriques : **ΔGm** ,

DGM représente l'écart géométrique maximal entre des positions appariées, la deuxième l'écart moyen. La troisième mesure représente les variations entre les écarts successifs et renseigne sur un écart de forme entre les deux trajectoires. Pour l'exemple de la figure 6 :

- la DGM vaut 110%, ce qui signifie que la trajectoire est sortie de 10% du couloir au pire.
- la DGm vaut 73,6% $((65\%+110\%+58\%+72\%+63\%) / 5)$, ce qui signifie que la trajectoire est en moyenne à une distance de 73,6% de la trajectoire moyenne par rapport à l'écart spatial toléré.
- la ΔGm vaut 61,5% $((|65\%-110\%| + |110\%-58\%| + |58\%-72\%| + |72\%+63\%|)/4)$, ce qui signifie que l'écart entre la TM et la T vari de manière importante.

Il faut noter que ces mesures sont calculées de manière discrète, uniquement aux positions PM_j . La trajectoire médiane étant filtrée, les positions conservées sont les plus significatives en termes de direction et de changement de vitesse, elle surestime donc les mesures par rapport à une approche continue.

Afin d'illustrer l'intérêt de ses trois mesures, la figure 7 donne 3 exemples de couple de T, TM. Pour le couple (a), la DGM et la DGm sont très proches, mais non

nulles et la ΔG_m est proche de zéro. Ces valeurs décrivent un couple de trajectoires décalées et ayant des formes très similaires. Pour le couple (b), la DGM est assez faible et la ΔG_m est faible. Par contre, la ΔG_m est importante. Ces valeurs décrivent des trajectoires assez proches, mais ayant des formes dissimilaires. Finalement, le couple (c) a une DGM importante, une DGM et ΔG_m faible. Ces trajectoires sont donc la plus part du temps assez proches. Cependant, un écart important sur une courte longueur est présent.

Figure 7. Exemple de trois trajectoires (trait épais) associées à la même trajectoire médiane (trait fin)

Ces 3 mesures spatiales sont complétées par 3 mesures temporelles ayant des significations proches.

3.3 Similarité temporelle de trajectoires

Pour l'aspect temporel, les 3 mesures proposées reflètent de même la distance temporelle maximum (DTM) l'écart moyen (DTm) et la moyenne des deltas temporels (ΔT_m). Ces mesures s'appuient sur la distance temporelle normalisée d_{TN} :

L'ET correspond l'écart en seconde entre les estampilles temporelles relative (durée depuis le départ) entre les positions appariées. Cet écart est normalisé en fonction de l'écart temporel toléré. Ce dernier est égal au retard toléré (ETRT noté négativement) si l'estampille temporelle de P_i est supérieure à celle de PM_j ou à l'avance tolérée (ETAT noté positivement) sinon.

Nous disposons donc de 6 mesures de similarité. Ces mesures sont complémentaires et donc toutes indispensables. Cependant, proposer 6 mesures aux opérateurs n'est pas envisageable, trop d'informations seraient présentes pour chaque objet mobile et donc surchargeraient l'opérateur de surveillance.

4. Indice de similarité et logique floue

Afin de définir un indice unique, deux approches sont possibles. La première possibilité consiste à agréger les 6 mesures à l'aide d'une somme pondérée. Cette solution n'est pas optimum, elle a le défaut de renvoyer un résultat trop lisse et de ne pas correspondre exactement au jugement humain. La deuxième, plus pragmatique et proche du raisonnement humain, est basée sur la logique floue (Zadeth 1965). Elle a été retenue, car elle permet de s'appuyer sur des règles floues exprimées en langage naturel décrivent la similarité. Par exemple, les deux règles floues suivantes sont exprimées pour la partie géométrique :

R1 : Si l'écart géométrique maximum est petit alors la similarité est forte.

R2 : Si la forme de la trajectoire est très différente de celle de la trajectoire médiane, alors la similarité est faible en d'autres termes si la ΔG_m est grande alors la similarité est faible.

D'autres règles doivent être ajoutées pour prendre en compte l'ensemble des critères de similarités spatiales et temporelles. Pour pouvoir raisonner avec ce type de règles floues, la première étape consiste à transformer les valeurs des 6 distances mesurées en valeurs qualitatives linguistiques. Cette étape s'appelle : "**Fuzzification**". Des ensembles flous et des fonctions d'appartenance sont définis. Les trois figures suivantes illustrent cette étape. Pour la DGM, la figure 8 fournit les trois ensembles flous associés à la DGM (petite, moyenne et grande) et les fonctions d'appartenance. Par exemple, si cette distance est de 110%, cette valeur appartient à 80% à l'ensemble DGM petite et à 20% à l'ensemble DGM moyenne.

Figure 8. Les trois ensembles flous et les fonctions d'appartenance pour la Distance Géographique Maximum (DGM)

La figure 9 exprime les trois valeurs associées à la ΔGm .

Figure 9. Les trois ensembles flous et les fonctions d'appartenance pour la moyenne géographique des deltas (ΔGm)

La figure 10 se focalise sur les valeurs associées à la similarité (faible, moyenne et forte) entre des trajectoires. Deux trajectoires identiques ont une similarité égale à 1. Deux trajectoires très différentes ont une similarité proche de 0. Les valeurs et les fonctions associées aux 4 autres mesures ne sont pas détaillées dans cet article.

Figure 10. Les trois ensembles flous et les fonctions d'appartenances associés à l'indice de similarité entre une trajectoire et une trajectoire médiane

La deuxième étape : **l'inférence floue** consiste à raisonner de manière approximative avec les règles floues en partant du principe que : "plus la condition sur les entrées est vraie plus la règle doit être appliquée". Une règle est activée uniquement si ses prémisses ont une valeur de vérité non nulle.

Supposons que nous ayons pour valeur pour la DGM 110% et pour la ΔGm de 35% ce qui équivaut à une trajectoire qui ne s'écarte pas trop de la trajectoire

médiane mais qui n'a pas une forme proche de la trajectoire médiane (telle la figure 7.b). La règle R1 est vraie à 80% (DGM appartenant à 80% à l'ensemble flou "petite"). La similarité selon cette règle est forte à 80 % (voir figure 11).

Figure 11. Application de la règle R1 qui est vraie à 80%

De même pour la règle R2, elle est vraie à 50 % (ΔG_m appartenant à 50% à l'ensemble flou "grande"). Donc la similarité selon cette règle est faible à 50%.

Figure 12. Calcul du centre de gravité

La troisième et dernière étape : la "**défuzzification**" consiste à transformer ces appartenances à des ensembles flous (valeurs qualitatives) à une valeur quantitative interprétable par l'utilisateur, dans notre cas : l'indice de similarité. Pour cela nous avons retenu la méthode du centre de gravité (COG). Elle consiste à calculer le centre de gravité de la surface cumulée des fonctions d'appartenance et à projeter ce centre sur l'axe des abscisses. Pour notre exemple, et nos deux règles, la similarité

est forte à 80 % selon la règle R1 et faible à 50% selon la règle R2. Le centre de gravité projeté renvoie un indice de 58% (voir figure 12).

Il est donc possible à l'aide de ces 6 mesures, de ces règles d'inférence floue et des fonctions d'appartenance d'obtenir un indice de similarité. Plus l'indice est fort, plus les trajectoires sont similaires. Cet indice est plus proche de la perception humaine qu'un indice renvoyé par une somme pondérée. Il est donc fort utile dans le contexte de l'analyse de trajectoires inhabituelles. Il doit être employé de manière complémentaire à l'analyse de positions. Une étude est nécessaire pour optimiser la visualisation de cet indice à l'opérateur afin de ne pas surcharger l'affichage des positions et des trajectoires.

5. Conclusion

Dans cet article, nous avons présenté une technique de fouille de données spatio-temporelles. Elle permet pour chaque itinéraire d'extraire une route type à partir d'un groupe homogène de trajectoires d'objets mobiles de même type. Ce patron est composé d'une trajectoire médiane et de couloirs spatio-temporels. Ce patron est très utile pour qualifier, en temps réel, une nouvelle position d'un objet suivant cet itinéraire. Ainsi, les opérateurs de surveillance du trafic disposent d'un outil d'aide à la surveillance réduisant la charge cognitive en focalisant l'attention de ces derniers sur les objets ayant des trajectoires inhabituelles. Pour aller plus loin, un indice de similarité entre une trajectoire et la route type a été défini. Il permet, à partir de 6 mesures temporelles ou spatiales et la logique floue, de mesurer la similarité entre ces deux éléments.

En ce qui concerne l'analyse de positions, cet outil a été validé à partir d'une base de données volumineuse des déplacements de navires dans la zone de Brest. Pour la comparaison de trajectoires, cette partie demande encore à être affinée puis validée.

Ainsi, à partir d'une simple base de données gérant les positions d'objets mobiles et permettant uniquement de répondre à des questions simples du type "Où est tel objet ?", nous passons progressivement à une base de données contenant ces informations de positionnement, mais aussi des connaissances de plus haut niveau lié au comportement des objets mobiles. Ce type de bases appelées bases de données inductives permettent en manipulant simultanément les données et les connaissances d'analyser le comportement des objets mobiles et de poser des requêtes de plus haut niveau.

Plusieurs travaux futurs dans ce domaine restent à réaliser. Premièrement, les parties définition d'indice de similarité et analyse de trajectoires doivent être consolidées. De plus, pour fournir une interface adaptée aux opérateurs, une étude de sémiologie graphique doit être menée afin d'afficher les informations de similarité de manière optimum. Finalement, ces travaux doivent être testés sur d'autres thématiques telles l'analyse des déplacements des piétons ou les migrations d'animaux.

Bibliographie

- Bertrand, F., Bouju, A., Claramunt, C., Devogele, T. and Ray, C. (2007) Web and Wireless Geographical Information Systems. Lecture Notes in Computer Science, Vol. 4857, Springer Berlin /Heidelberg, chapter Web Architecture for Monitoring and Visualizing Mobile Objects in Maritime Contexts, pp. 94–105.
- Etienne L., Devogele T., Bouju A. (2009) Analyse de similarité de trajectoires d'objets mobiles suivant le même itinéraire : Application aux trajectoires de navires, *Revue Ingénierie des Systèmes d'Information (ISI)*, numéro thématique Systèmes d'information et géolocalisation, Hermès, vol. 14, num.5, pages 85-106
- Lee J., Han J., Li X. (2008) « Trajectory Outlier Detection: A partition-and-Detect Framework », ICDE 2008. IEEE 24th International Conference on Data Engineering.
- Kharrat A., Popa I. S., Zeitouni K., Faiz S. (2008), « Clustering Algorithm for Network Constraint Trajectories », Proc. 13th International Symposium on Spatial Data Handling, Montpellier, France.
- Knorr E. M., Ng R. T., Tucakov V. (2000) « Distance based outliers: algorithms and applications », VLDB Journal, vol. 8, n° 3, pp. 237-253.
- Mascret A., Devogele T., Le Berre I., Hénaff A. (2006) Coastline Matching Process based on the discrete Fréchet distance - In Proceedings of the 12th International Symposium on Spatial Data Handling (SDH), W. Kainz, A. Reidl, and G. Elmes (eds.), Vienna, Austria, Springer-Verlag, pages 383-400
- Mefteh W., Bouju A., Malki J. (2008) « Vers un modèle générique des trajectoires », Proc. Manifestation des Jeunes Chercheurs en Sciences et Technologies de l'Information et de la Communication, Marseille, France.
- Pelekis N., Kopanakis I., Ntoutsis I., Marketos G., Andrienko G. and Theodoridis Y., (2007) "Similarity Search in Trajectory Databases", In the Proceedings of the 14th IEEE International Symposium on Temporal Representation and Reasoning (TIME 2007), Alicante, Spain.
- Sakoe, H. et Chiba, S. (1978) *Dynamic programming algorithm optimization for spoken word recognition*, IEEE Transactions on Acoustics, Speech and Signal Processing, **26**(1) pp. 43-49, 1978, ISSN: 0096-3518
- Wan T., Zeitouni K., Meng X. (2007) « An OLAP System for Network-Constrained Moving Objects », Proc. 22nd Annual ACM Symposium on Applied Computing, Seoul, Korea, pp. 13-18.
- Zadeh, L. A. (1965) "Fuzzy sets". *Information and Control*, vol. 8, num. 3, pages 338–353.

**Session 3 : Fouille de données et
classification automatique**

Fouille de données complexes : des relevés terrain aux données satellitaires pour la cartographie de paysages agricoles

Fadi Badra* — Elodie Vintrou** — Agnès Bégué** —
Maguelonne Teisseire*

* Cemagref, UMR TETIS F-34093 Montpellier, France

** CIRAD, UMR TETIS F-34093 Montpellier, France

{nom}@teledetection.fr

RÉSUMÉ. Dans cet article, nous présentons une approche préliminaire de caractérisation des paysages ruraux et de leurs systèmes de culture à partir de techniques de fouille de données (recherche d'itemsets fréquents). Cette méthode permet de coupler des données de relevé terrain aux indicateurs de texture extraits des images satellites. Sa mise en œuvre sur des données associées au Mali pose les premières bases d'une méthode originale d'extraction de motifs séquentiels à partir de données complexes.

ABSTRACT. Remote sensing mining

MOTS-CLÉS : Fouille de données, itemsets fréquents, image satellite, occupation du sol

KEYWORDS: Data Mining, Frequent Itemsets, Satellite Image, Land Cover

1. Introduction

Motivés par des problèmes d'Aide à la Décision, les chercheurs de différentes communautés (Intelligence Artificielle, Statistiques, Bases de Données ...) se sont intéressés à la conception et au développement d'une nouvelle génération d'outils permettant d'extraire automatiquement de la connaissance de grandes bases de données. Ces outils, techniques et approches sont le sujet d'un thème de recherche connu sous le nom de *Knowledge Discovery in Databases* ou KDD (Extraction de Connaissances dans les Bases de Données) ou Data Mining (Fouille de Données). Elles sont utilisées dans de nombreux domaines d'applications. Les exemples les plus courants sont les compagnies d'assurance, les compagnies bancaires (crédit, prédiction du marché, détection de fraudes), le marketing (comportement des consommateurs, mailing personnalisé), la recherche médicale (aide au diagnostic, au traitement, surveillance de population sensible), les réseaux de communication (détection de situations alarmantes, prédiction d'incidents), l'analyse de données spatiales, etc.

La fouille de données peut être définie par « *Processus non trivial permettant l'extraction automatique de connaissances d'une base de données pour obtenir de nouvelles données, valides, potentiellement utiles et compréhensibles* » [FAY 96]. Bien que le terme de fouille de données représente la découverte de connaissances, il ne constitue en fait qu'une seule des étapes du KDD, qui comprend globalement trois étapes : la préparation des données, l'extraction des données (Data Mining) et leur interprétation.

La première étape consiste à sélectionner uniquement les données potentiellement utiles de la base (opération de filtrage), sur lesquelles on effectue une phase de pré-traitement (gestion des données manquantes ou invalides). Ensuite, les données obtenues passent par une phase de formatage, afin de les préparer au processus de Data Mining. Finalement, la dernière étape est une étape d'analyse et d'interprétation de la connaissance extraite par la fouille de données, pour la rendre lisible et compréhensible par l'utilisateur. Les besoins variés nécessitent des approches différentes telles que la classification, la recherche de corrélations, la segmentation ou encore la détection de déviation.

Notre objectif est de développer et de mettre en œuvre des méthodes de caractérisation des paysages ruraux et de leurs systèmes de culture, en utilisant des techniques de fouille de données permettant de coupler des données hétérogènes comme les relevés terrains à des informations issues des images satellites. Ce projet vise ainsi à proposer une approche alternative pertinente pour définir un mécanisme d'apprentissage basé sur des images satellitaires multi-sources (données spectrales, texturales et temporelles), des données environnementales (climat, relief, type de sol, ...) et des données de terrain, et mettre en évidence des critères d'implication qui n'auraient pas pu être identifiés autrement. Nous procédons suivant deux étapes :

- Adapter un mécanisme de recherche de motifs séquentiels multidimensionnels, comme proposé dans [PLA 10], à la fouille de séries d'images satellitaires et des don-

nées pouvant être mises en corrélation (informations externes) ;

– Proposer un système de catégorisation (classification supervisée) basé sur un mécanisme d'apprentissage afin d'offrir une aide à la cartographie de l'utilisation des sols. Cette solution doit être le moins sensible possible au changement d'échelle. Les informations annexes associées à des relevés terrains ou au croisement de bases de données (modèles numériques de terrain, climat...) seront donc très importantes.

Dans cet article, nous nous concentrons sur la première étape de recherche d'itemsets fréquents à partir de données de télédétection. Un des enjeux de l'application de ces techniques en télédétection réside dans le fait que les données relatives à une même réalité terrain sont par nature multi-source, puisqu'elles résultent du croisement de données issues de différents capteurs, de relevés terrain et bases de données externes. Leur description se fait par ailleurs dans plusieurs dimensions et combine des informations spectrales, spatiales et temporelles. Pour cette raison, nous nous tournons vers des algorithmes d'extraction d'itemsets multidimensionnels [PIN 01], qui peuvent prendre en compte le caractère multidimensionnel des données.

Les résultats présentés dans cet article sont des résultats préliminaires mettant en œuvre un jeu de données restreint en termes d'imagerie satellitaire (images SPOT seulement) et de données exogènes. Cette approche préliminaire simplifiée nous a semblé indispensable pour assurer des bases communes aux deux communautés engagées dans ce travail, les agronomes et les spécialistes de fouille de données.

Dans les sections suivantes, nous décrivons tout d'abord les données étudiées (section 2) pour ensuite présenter les définitions associés aux motifs multidimensionnels (section 3). Puis, nous décrivons en détail le processus d'extraction mis en œuvre ainsi que l'analyse des premiers résultats obtenus (section 5).

2. Description des données

2.1. Contexte et zone d'étude

Le Mali est un pays d'Afrique de l'Ouest, autour de la latitude 14°N. Ce pays possède un gradient climatique Nord-Sud, qui varie de régions subtropicales à semi-arides, et s'étend plus au Nord vers des zones arides et désertiques. Le Mali peut être considéré comme représentatif de la zone soudano-sahélienne, où la forte dépendance à l'agriculture pluviale entraîne une vulnérabilité aux changements climatiques et anthropiques. Par conséquent, afin de mieux connaître le phénomène de mousson en Afrique de l'Ouest et sa variabilité, et d'améliorer les prévisions des impacts de cette variabilité sur l'agriculture et la sécurité alimentaire, une des premières étapes nécessaires est une estimation fiable du domaine cultivé. Une attention particulière a été portée sur 3 zones, le long du gradient climatique malien (tableau 1).

Site d'étude	Précipitations moyennes	Cultures principales	Végétation naturelle (en majorité)	Pourcentage de surface cultivée
Cinzana (zone soudano-sahélienne)	600 mm/an	mil et sorgho	Végétation dégradée et sol nu	43%
Koutiala (zone soudano-sahélienne)	750 mm/an	mil, sorgho et coton	Végétation ouverte et fermée	52%
Sikasso (zone soudanienne)	1000 mm/an	maïs, coton et fruits	Végétation naturelle et dense	40%

Tableau 1. Les principales caractéristiques des trois zones d'étude retenues.

2.2. Les données terrain

Des missions de terrain ont été effectuées au Mali de mai à novembre 2009, dans le but de caractériser les paysages agricoles sahéliens. Au total, 980 points GPS ont été enregistrés, et des paysans de chacune des régions étudiées ont été interrogés. Chaque point relevé a été transformé en un polygone dont il est le centre et auquel a été affecté un type d'occupation du sol.

2.3. Les données image

Pour couvrir de larges zones avec précision, les images SPOT à 10 m de résolution constituent souvent la solution la plus efficace et la plus rentable. Pour cette étude, les images utilisées sont des images SPOT5 orthorectifiées et fusionnées, à 2.5 m en multispectral. Pour des raisons de disponibilité des images sur nos sites d'études, la date d'acquisition est le 14 novembre 2007 pour Cinzana et le 20 novembre 2007 pour Koutiala et Sikasso. Cependant, les classes d'occupation du sol observées on globalement peu évolué en deux ans (données terrain en 2009).

À partir de ces images sont calculés un certain nombre de descripteurs spectraux (indice de végétation) et spatiaux (indices de texture) des états de surface (occupation du sol) :

– *L'indice de végétation NDVI.* Les indices de végétation sont des combinaisons, linéaires ou non, de réflectances dans les bandes spectrales R (rouge), PIR (proche infrarouge) et MIR (moyen infrarouge). Ils permettent de caractériser le couvert végétal en terme de vigueur de la végétation. Ainsi, ces indices permettent de distinguer une surface de végétation verte photosynthétiquement active (indice élevé) avec une sur-

face de sol nu, ou faiblement couverte (indice faible). L'indice le plus souvent utilisé est le « *Normalized Difference Vegetation Index* » [ROU 74] :

$$NDVI = \frac{PIR - R}{PIR + R}$$

L'indice varie de 0,1 (sol nu) à 0,9 (végétation verte et dense).

– *Les indices de texture.* L'hétérogénéité spatiale de la couverture végétale est étudiée au travers de la distribution spatiale des variables radiométriques. La variabilité spatiale d'une image est représentée par le concept de texture. Haralick élargit dans [HAR 79] la définition en décrivant une texture comme un phénomène à deux dimensions : la première concernant la description d'éléments de base ou primitives (le motif) à partir desquels est formée la texture ; la deuxième dimension est relative à la description de l'organisation spatiale de ces primitives.

En télédétection, l'utilisation de variables texturales dans les algorithmes de classification augmente la précision des résultats de classification par rapport à l'utilisation de la seule information spectrale [HAR 73]. Les images utilisées étant des champs continus de variables radiométriques, l'approche microtexturale est plus adaptée. La matrice de co-occurrences (ou matrice de dépendance spatiale) est une des approches les plus connues et les plus utilisées pour extraire des caractéristiques de textures. Elle effectue une analyse statistique de second ordre de la texture, par l'étude des relations spatiales des couples de pixels [HAR 73, HAR 79]. Quatorze indices (définis par Haralick) qui correspondent à des caractères descriptifs des textures peuvent être calculés à partir de cette matrice. Dans cette étude, les indices d'homogénéité, variance, dissimilarité, et contraste ont été calculés sur une fenêtre de 15×15 pixels.

3. Extraction d'itemsets séquentiels multidimensionnels

Le problème de la recherche de motifs séquentiels a été introduit par R. Agrawal dans [AGR 95] et appliqué avec succès dans de nombreux domaines comme la biologie [WAN 04, SAL 09], la fouille d'usage du Web [PEI 00, MAS 08], la détection d'anomalie [RAB 10], la fouille de flux de données [MAR 06] ou la description des comportements au sein d'un groupe [PER 09]. Des approches plus récentes [JUL 08] utilisent les motifs séquentiels pour décrire les évolutions temporelles des pixels au sein des séries d'images satellites. Néanmoins, à notre connaissance, l'étude de la littérature ne fait état d'aucun travaux sur l'application de techniques de recherche de motifs séquentiels couplant données externes et télédétection.

Dans cette section, nous introduisons les définitions relatives à la fouille d'itemsets séquentiels multidimensionnels et décrivons l'algorithme de fouille de données utilisé.

3.1. Itemsets séquentiels multidimensionnels

Soit un ensemble \mathcal{D} de dimensions et $\{\mathcal{D}_R, \mathcal{D}_A, \mathcal{D}_T, \mathcal{D}_I\}$ une partition de \mathcal{D} dans laquelle \mathcal{D}_R désigne les dimensions de référence, qui permettent de déterminer si une séquence est fréquente, \mathcal{D}_A les dimensions d'analyse, sur lesquelles les corrélations sont extraites, et \mathcal{D}_T les dimensions permettant d'introduire une relation d'ordre (généralement le temps). Les dimensions \mathcal{D}_I sont les dimensions ignorées lors de la fouille.

Pour chaque dimension $D_i \in \mathcal{D}$, on note $Dom(D_i)$ son domaine de valeurs. À chaque domaine de valeurs $Dom(D_i)$ est associé une hiérarchie H_i , et l'on suppose que $Dom(D_i)$ contient une valeur particulière notée \top_i (la racine de la hiérarchie). Lorsqu'aucune hiérarchie de valeurs n'est définie sur une dimension D_i , nous considérons H_i comme un arbre de profondeur 1 dont la racine est \top_i et dont les feuilles sont les éléments de $Dom(D_i) \setminus \{\top_i\}$. La figure 1 présente un exemple de hiérarchie de valeurs H_i pour la dimension *site-étude*.

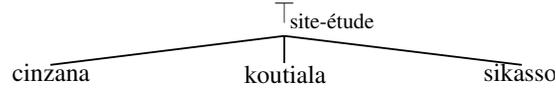


Figure 1. La hiérarchie de valeurs H_i pour la dimension *site-étude*.

Un *item multidimensionnel* $e = (d_1, d_2, \dots, d_m)$ est un m -uplet défini sur les dimensions d'analyse \mathcal{D}_A , c'est-à-dire tel que $\forall i \in [1 \dots m], d_i \in Dom(D_i)$ avec $D_i \in \mathcal{D}_A$ et $\exists d_i \in [1, \dots, m]$ tel que $d_i \neq \top_i$. Par exemple, $e = (\text{sorgho}, \text{cinzana})$ et $e' = (\text{sorgho}, \top_{\text{site-étude}})$ sont des items multidimensionnels qui décrivent des points terrain sur les dimensions d'analyse $\mathcal{D}_A = \{\text{type-de-culture}, \text{site-étude}\}$. On définit une relation d'inclusion \subseteq entre items multidimensionnels : un item multidimensionnel $e = (d_1, d_2, \dots, d_m)$ est inclus dans un item multidimensionnel $e' = (d'_1, d'_2, \dots, d'_m)$ (noté $e \subseteq e'$) si $\forall i \in [1, \dots, m], d_i = d'_i$ ou est une spécialisation de d'_i dans H_i . Dans l'exemple précédent, on a l'inclusion $(\text{sorgho}, \text{cinzana}) \subseteq (\text{sorgho}, \top_{\text{site-étude}})$ car *cinzana* est une spécialisation de $\top_{\text{site-étude}}$ dans la hiérarchie $H_{\text{site-étude}}$.

Un *itemset multidimensionnel* $i = (e_1, e_2, \dots, e_m)$ est un ensemble non vide d'items multidimensionnels non deux à deux comparables par rapport à \subseteq (c.-à-d., $\forall i, j \in [1, \dots, m], e_i \not\subseteq e_j$ et $e_i \not\supseteq e_j$). On définit une relation d'inclusion \subseteq entre itemsets multidimensionnels : un itemset i est inclus dans un itemset i' (noté $i \subseteq i'$) si pour chaque item a de i , il existe un item a' de i' tel que $a \subseteq a'$.

Une *séquence multidimensionnelle* $s = \langle i_1, \dots, i_n \rangle$ est une liste ordonnée non vide d'itemsets multidimensionnels. On définit une relation de généralisation (ou spécialisation) entre séquences multidimensionnelles : une séquence $s = \langle i_1, i_2, \dots, i_n \rangle$

pt-id	date	type-de-culture	site-étude	NDVI-100
1	1	arachide	sikasso	très faible
2	1	mil	koutiala	faible
2	2	mil	koutiala	modéré
3	1	sorgho	koutiala	élevé

Tableau 2. Base de données DB .

pt-id	date	type-de-culture	site-étude	NDVI-100
2	1	mil	koutiala	faible
2	2	mil	koutiala	modéré

Tableau 3. Bloc $B_{(mil, koutiala)}$.

est plus spécifique qu'une séquence $s' = \langle i'_1, i'_2, \dots, i'_m \rangle$ s'il existe des entiers $1 \leq j_1 \leq \dots \leq j_m \leq n$ tels que $s_{j_1} \subseteq s'_1, s_{j_2} \subseteq s'_2, \dots, s_{j_m} \subseteq s'_m$.

Étant donnée une table relationnelle DB , on appelle *bloc* l'ensemble des n -uplets qui ont la même projection sur \mathcal{D}_R . Par exemple, le tableau 3 donne le bloc formé en ne gardant que les n -uplets de la table relationnelle DB donnée au tableau 2 dont la projection sur $\mathcal{D}_R = \{\text{type-de-culture}, \text{site-étude}\}$ est $(\text{mil}, \text{koutiala})$. Le *support* d'une séquence est le nombre de blocs qui contiennent cette séquence.

Étant donné un seuil σ_{min} de support minimum, le but de la recherche d'itemsets séquentiels multidimensionnels est de trouver toutes les séquences dont le support est supérieur ou égal à σ_{min} .

3.2. Méthode

L'algorithme M^3SP [PLA 10] effectue efficacement la recherche d'itemsets séquentiels multidimensionnels. Le choix de cet algorithme est motivé par notre objectif d'intégrer la dimension temporelle de séries d'images dans une prochaine étape.

4. Mise en œuvre

Un processus d'extraction de connaissances a été mis au point (figure 2). Il est constitué de quatre étapes : (1) chargement des données, (2) préparation des données, (3) fouille de données et (4) interprétation et validation des résultats.

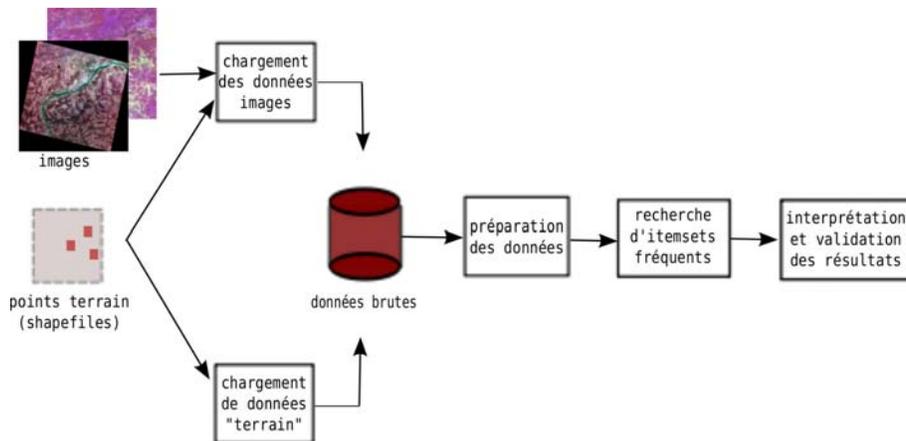


Figure 2. Les différentes étapes du processus d'extraction de connaissances.

Dans l'étape de chargement des données, les données sont collectées et placées dans une base de données (données "brutes"). Cinq indices images sont calculés à partir des images SPOT : l'indice de végétation NDVI et les indices de texture variance, homogénéité, contraste et dissimilarité. À chaque point terrain est associée une valeur de chacun de ces indices, qui est calculée en prenant la moyenne de la valeur de l'indice sur les pixels contenus dans un polygone carré centré sur ce point (figure 3). Deux jeux d'indices sont ainsi calculés avec des tailles de polygones de 100 m et 200 m de côté.

L'étape de préparation des données a pour but de constituer l'ensemble d'apprentissage et de formater les données pour pouvoir être traitées par l'algorithme de fouille. L'ensemble d'apprentissage est constitué en sélectionnant parmi l'ensemble des points terrain les points qui se situent sur une des images SPOT et qui correspondent à des relevés en zone de culture. On obtient 498 points dont la répartition par site d'étude et en culture / non culture est donnée par le tableau 4.

Site d'étude	Culture	Non culture	Total
Cinzana	138	85	223
Koutiala	105	78	183
Sikasso	46	46	92

Tableau 4. Les points terrain de l'ensemble d'apprentissage.

Chaque point de l'ensemble d'apprentissage est décrit par les valeurs qu'il prend dans un ensemble de dimensions D_i :

- id-pt : entier qui identifie chaque point terrain de manière unique

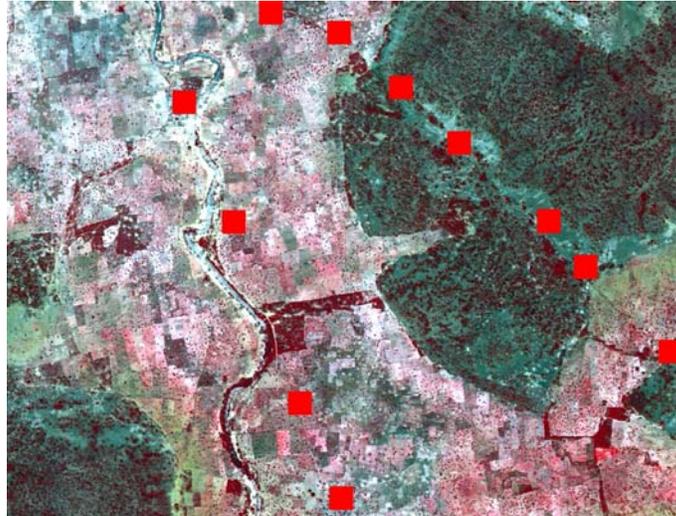


Figure 3. Extraction des indices images pour chaque point terrain sur un polygone carré centré sur ce point (zoom sur une image SPOT du site de Cinzana, les carrés rouges sont des polygones de $100\text{m} \times 100\text{m}$ centrés sur chacun des points terrain). © CNES 2009, Distribution SPOT Image

- date : estampille temporelle (constante dans cette expérience)
- site-étude : le nom du site d'étude
- type-de-culture : le type de culture
- nom-village : le nom du village le plus proche, pris parmi les relevés village
- distance-village : la distance du point au village le plus proche
- NDVI-100, variance-100, homogénéité-100, dissimilarité-100 et contraste-100 : les valeurs des descripteurs images calculés avec des polygones carrés de 100 m de coté
- NDVI-200, variance-200, homogénéité-200, dissimilarité-200 et contraste-200 : les valeurs des descripteurs images calculés avec des polygones carrés de 200 m de coté

Les domaines $Dom(D_i)$ de chaque dimension D_i sont ensuite discrétisés en découpant leurs intervalles de valeurs. Pour l'indice de végétation NDVI, le découpage de l'intervalle de valeurs $[-1, 1]$ est donné par l'expert. Les domaines de valeurs des indices de texture sont découpés en cinq classes de même effectif. Le tableau 5 résume l'ensemble des dimensions D_i utilisées pour décrire les points de l'ensemble d'apprentissage et le découpage de leurs domaines de valeurs.

dimension D_i	intervalles de valeurs				
id-pt	{1}, {2}, ..., {498}				
date	{1}				
site-étude	{cinzana, koutiala, sikasso}				
type-de-culture	{riz, sorgho, maïs, ...}				
nom-village	{dioforongo, tigui, sanando, ...}				
distance-village	proche	éloigne			
	[0,3000]	[3001,+∞[
NDVI-100	très faible	faible	modéré	élevé	
	[-1,0.2]	[0.2,0.3]	[0.3,0.5]	[0.5,1]	
NDVI-200	très faible	faible	modéré	élevé	
	[-1,0.2]	[0.2,0.3]	[0.3,0.5]	[0.5,1]	
variance-100	très faible	faible	modéré	élevé	très élevé
]-∞, 0.56]	[0.56, 1.22]	[1.22, 2.38]	[2.38, 4.00]	[4.00, +∞[
variance-200]-∞, 0.56]	[0.56, 1.22]	[1.22, 2.38]	[2.38, 4.00]	[4.00, +∞[
homogénéité-100]-∞, 0.67]	[0.67, 0.74]	[0.74, 0.80]	[0.80, 0.87]	[0.87, +∞[
homogénéité-200]-∞, 0.67]	[0.67, 0.74]	[0.74, 0.80]	[0.80, 0.87]	[0.87, +∞[
dissimilarité-100]-∞, 0.26]	[0.26, 0.40]	[0.40, 0.54]	[0.54, 0.72]	[0.72, +∞[
dissimilarité-200]-∞, 0.26]	[0.26, 0.40]	[0.40, 0.54]	[0.54, 0.72]	[0.72, +∞[
contraste-100]-∞, 0.27]	[0.27, 0.44]	[0.44, 0.72]	[0.72, 1.05]	[1.05, +∞[
contraste-200]-∞, 0.27]	[0.27, 0.44]	[0.44, 0.72]	[0.72, 1.05]	[1.05, +∞[

Tableau 5. Les dimensions D_i et le découpage de leurs domaines de valeurs $Dom(D_i)$ en intervalles.

Une hiérarchie de valeur H_i est construite pour chaque dimension D_i . Les hiérarchies considérées sont toutes de profondeur 1 sauf pour l'attribut `type-de-culture` (figure 4).

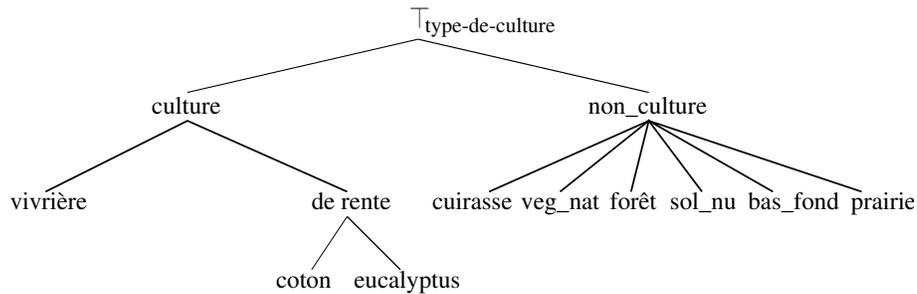


Figure 4. La hiérarchie de valeurs H_i pour la dimension `type-de-culture`.

Dans l'étape de fouille, l'algorithme de recherche d'itemsets séquentiels fréquents M^3SP est appliqué sur les données d'apprentissage formatées, en choisissant un seuil de support σ_{min} , un ensemble de dimensions de référence \mathcal{D}_R et un ensemble de dimensions d'analyse \mathcal{D}_A .

Les itemsets séquentiels extraits sont présentés à l'analyste pour interprétation et validation.

5. Résultats et discussion

Toutes les expériences ont été effectuées en prenant $\mathcal{D}_T = \{\text{date}\}$ et $\mathcal{D}_R = \{\text{id-pt}\}$. Comme la même estampille temporelle est associée à chaque point, les séquences obtenues ne sont constituées que d'un seul itemset. De plus, la dimension de référence étant l'identifiant des points, l'ensemble d'apprentissage est divisé en autant de blocs qu'il y a de points d'apprentissage. Par conséquent, dans ces expériences chaque itemset résultat est composé d'un seul item et son support correspond au nombre de points terrain qui le partagent.

Plusieurs expériences ont été menées en faisant varier les dimensions d'analyse \mathcal{D}_A et en ne fouillant que les points situés sur un site d'étude donné (par filtrage de l'ensemble d'apprentissage avant la fouille).

Site d'étude	Itemset	Support
Cinzana (223 points)	$s_1 = \langle\langle(\text{culture}, \top_{\text{distance-village}})\rangle\rangle$	138 (62%)
	$s_2 = \langle\langle(\text{culture}, \text{proche})\rangle\rangle$	121 (54%)
Koutiala (183 points)	$s_3 = \langle\langle(\text{culture}, \top_{\text{distance-village}})\rangle\rangle$	105 (57%)
	$s_4 = \langle\langle(\text{culture}, \text{proche})\rangle\rangle$	80 (44%)
Sikasso (92 points)	$s_5 = \langle\langle(\text{culture}, \top_{\text{distance-village}})\rangle\rangle$	46 (50%)
	$s_6 = \langle\langle(\text{culture}, \text{proche})\rangle\rangle$	27 (29%)

Tableau 6. Itemsets extraits pour $\mathcal{D}_A = \{\text{type-de-culture}, \text{distance-village}\}$.

Les itemsets présentés dans le tableau 6 ont été extraits en prenant comme dimensions d'analyse le type de culture et la distance au village le plus proche. Les résultats montrent que dans les 3 sites étudiés, les cultures sont généralement cultivées autour des villages, dans un rayon de 2 à 3 km pour la majorité. En effet, 88% des cultures de Cinzana (121/138, itemsets s_1 et s_2), 77% de celle de Koutiala (80/105, itemsets s_3 et s_4) et 59% de celles de Sikasso (27/46, itemsets s_5 et s_6) sont dans la couronne de 3 km autour des différents villages. Il n'apparaît pas possible de faire un lien entre la distance au centre du village et le type de culture au vu du nombre trop faible de points terrain disponibles par site. Il a cependant été déjà observé dans plusieurs villages d'Afrique de l'Ouest un aménagement en auréoles. Le village et les jardins occupent une position centrale. Une première auréole (soforo) est constituée par les champs « de case » cultivés en rotation annuelle. Une seconde auréole (kongo foro) est formée par les champs de brousse (mil, sorgho, arachides, coton...). Enfin, la brousse (kongo) fournit les produits de la chasse, de la cueillette, le bois d'œuvre et de feu. La distance entre ces 3 auréoles varie entre les villages.

Les itemsets présentés dans le tableau 7 ont été extraits en prenant comme dimensions d'analyse le type de culture et le descripteur NDVI calculé avec des polygones carrés de 100 m de côté. Nous pouvons signaler que le cycle de production végétale est particulier au Mali : les cultures étant pour la majorité des cultures pluviales, la croissance des plantes est étroitement liée à la pluviométrie (quantité et répartition).

Site d'étude	Itemset	Support
Cinzana (223 points)	$s_7 = \langle\langle\{\text{culture, très faible}\}\rangle\rangle$	74 (33%)
Koutiala (183 points)	$s_8 = \langle\langle\{\text{culture, modéré}\}\rangle\rangle$	56 (31%)
	$s_9 = \langle\langle\{\text{culture, faible}\}\rangle\rangle$	33 (18%)
Sikasso (92 points)	$s_{10} = \langle\langle\{\text{culture, faible}\}\rangle\rangle$	25 (28%)
	$s_{11} = \langle\langle\{\text{culture, modéré}\}\rangle\rangle$	20 (22%)

Tableau 7. Itemsets extraits pour $\mathcal{D}_A = \{\text{type-de-culture, NDVI-100}\}$.

Les supports des itemsets séquentiels sont plus fréquemment faibles concernant le NDVI à Cinzana, qu'à Koutiala, qu'à Sikasso. Ceci reflète bien le gradient bioclimatique au Mali. Il pleut moins au Nord qu'au Sud, et donc les plantes ont une activité photosynthétique inférieure à Cinzana qu'à Sikasso, en moyenne. D'autre part, pour le NDVI du mois de novembre, les cultures sont déjà entièrement récoltées à Cinzana, et partiellement récoltées à Koutiala et à Sikasso. Ceci explique les 54% de culture avec un NDVI « très faible » à Cinzana (74/138, itemsets s_1 et s_7), contre 98% et 94% de cultures avec un NDVI « faible » ou « medium » (itemsets s_3 , s_5 et s_8 à s_{11}) à Koutiala et Sikasso respectivement.

Site d'étude	Itemset	Support
Cinzana (223 points)	$s_{12} = \langle\langle\{\text{culture, } \top_{\text{variance-100}}, \top_{\text{homogénéité-100}}, \top_{\text{dissimilarité-100}}, \text{élevé}\}\rangle\rangle$	56 (25%)
Koutiala (183 points)	$s_{13} = \langle\langle\{\text{culture, } \top_{\text{variance-100}}, \top_{\text{homogénéité-100}}, \top_{\text{dissimilarité-100}}, \text{faible}\}\rangle\rangle$	35 (19%)
	$s_{14} = \langle\langle\{\text{culture, } \top_{\text{variance-100}}, \text{très élevé, très faible, très faible}\}\rangle\rangle$	23 (25%)
Sikasso (92 points)	$s_{15} = \langle\langle\{\text{culture, } \top_{\text{variance-100}}, \top_{\text{homogénéité-100}}, \top_{\text{dissimilarité-100}}, \text{faible}\}\rangle\rangle$	18 (20%)

Tableau 8. Itemsets extraits pour $\mathcal{D}_A = \{\text{type-de-culture, variance-100, homogénéité-100, dissimilarité-100, contraste-100}\}$.

Les itemsets présentés dans le tableau 8 ont été extraits en prenant comme dimensions d'analyse le type de culture et les quatre descripteurs de texture calculés avec des polygones carrés de 100 m de côté. La présence de quatre indices de texture induit une difficulté à interpréter la présence des itemsets fréquents. Si l'on analyse seulement le contraste, on observe dans 20 à 25% des cas, un contraste « élevé » pour Cinzana (itemset s_{12}), « faible » pour Koutiala (itemset s_{13}) et « très faible » pour Sikasso. Ce contraste qui diminue du Nord au Sud peut s'expliquer par une différence de densité d'arbres dans les champs cultivés. Il serait en effet plus commun de trouver des arbres comme le Balanzan, le Néré ou le Karitier dans des champs de la région de Cinzana, qu'à Koutiala ou Sikasso, ce qui expliquerait les brusques changements de radiométrie, et donc un indice de contraste élevé.

Enfin, les différentes tailles de polygones n'induisent pas de changements notables dans les résultats.

À la suite de cette étude, il apparaîtrait intéressant de faire un lien entre la distance au centre du village et le type de culture. Il doit exister un lien entre les espèces cultivées et la distance au village, que nous essaierons de déterminer à partir de données terrain supplémentaires. D'autre part, le NDVI et la texture pourraient être mis en relation avec des images du mois de septembre ou d'octobre. C'est la période pendant laquelle les cultures sont dans des phases de croissance différentes suivant les régions, et entre elles également, puisque le maïs par exemple, est récolté, alors que les autres cultures céréalières restent sur pied pour encore un mois voire deux. Nous essaierons également d'utiliser pour l'extraction d'indices des polygones plus grands, pour un calcul de texture optimum (la faible taille des polygones ne permet pas de détecter beaucoup de motifs de texture répétés) et enfin, d'utiliser des séries temporelles d'images MODIS pour prendre en compte le cycle de croissance de chaque occupation du sol et ainsi mieux les différencier.

6. Conclusion

Nous avons présenté la première étape d'extraction d'ensembles multidimensionnels d'une méthode de caractérisation des paysages ruraux et de leurs systèmes de culture. Nous avons mis en œuvre ce processus d'extraction sur des données du Mali et avons pu fouiller des données hétérogènes comme les relevés terrains avec des indicateurs obtenus des images satellitaires. Les premiers résultats sont prometteurs et laissent présager des perspectives d'analyse pertinentes. Il s'agit à présent de poursuivre ces travaux (1) afin de prendre en compte la partie séquentielle des données (séquences des images Modis) afin de trouver des motifs séquentiels qui seront le support du mécanisme de classification des types de culture puis (2) de proposer une évaluation (rappel, confiance) afin de mesurer l'efficacité d'une telle approche.

7. Bibliographie

- [AGR 95] AGRAWAL R., SRIKANT R., « Mining Sequential Patterns », YU P. S., CHEN A. L. P., Eds., *Proceedings of the Eleventh International Conference on Data Engineering, March 6-10, 1995, Taipei, Taiwan*, IEEE Computer Society, 1995, p. 3-14.
- [FAY 96] FAYYAD U. M., PIATETSKY-SHAPIO G., SMYTH P., « From Data Mining to Knowledge Discovery : an Overview », *Advances in knowledge discovery and data mining*, vol. 1, 1996, p. 1-34.
- [HAR 73] HARALICK R. M., SHANMUGAM K., DINSTEN I., « Textural Features for Image Classification », *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 3, n° 6, 1973, p. 610-621.
- [HAR 79] HARALICK R. M., « Statistical and Structural Approaches to Texture », *Proceedings of the IEEE*, vol. 67, n° 5, 1979, p. 786-804, IEEE-Inst Electrical Electronics Engineers Inc.

- [JUL 08] JULEA. A., MEGER N., BOLON P., « On mining pixel based evolution classes in satellite image time series », *Proc. of the 5th Conf. on Image Information Mining : pursuing automation of geospatial intelligence for environment and security (ESA-EUSC 2008)*, 2008, page 6.
- [MAR 06] MARASCU A., MASSEGLIA F., « Mining sequential patterns from data streams : a centroid approach », *Journal of Intelligent Information Systems*, vol. 27, n° 3, 2006, p. 291-307.
- [MAS 08] MASSEGLIA F., PONCELET P., TEISSEIRE M., MARASCU A., « Web usage mining : extracting unexpected periods from web logs », *Data Mining and Knowledge Discovery (DMKD)*, vol. 16, n° 1, 2008, p. 39-65.
- [PEI 00] PEI J., HAN J., MORTAZAVI-ASL B., ZHU H., « Mining Access Patterns Efficiently from Web Logs », TERANO T., LIU H., CHEN A. L. P., Eds., *Knowledge Discovery and Data Mining, Current Issues and New Applications, 4th Pacific-Asia Conference, PADKK 2000, Kyoto, Japan, April 18-20, 2000, Proceedings*, Lecture Notes in Computer Science, Springer, 2000, p. 396-407.
- [PER 09] PERERA D., KAY J., KOPRINSKA I., YACEF K., ZAÏANE O. R., « Clustering and Sequential Pattern Mining of Online Collaborative Learning Data », *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, n° 6, 2009, p. 759-772, IEEE Educational Activities Department.
- [PIN 01] PINTO H., HAN J., PEI J., WANG K., CHEN Q., DAYAL U., « Multi-dimensional sequential pattern mining », *CIKM '01 : Proceedings of the tenth international conference on Information and knowledge management*, New York, NY, USA, 2001, ACM, p. 81-88.
- [PLA 10] PLANTEVIT M., LAURENT A., LAURENT D., TEISSEIRE M., CHOONG Y. W., « Mining multidimensional and multilevel sequential patterns », *ACM Transactions on Knowledge Discovery from Data TKDD*, vol. 4, n° 1, 2010.
- [RAB 10] RABATEL J., BRINGAY S., PONCELET P., « Aide à la décision pour la maintenance ferroviaire préventive », YAHIA S. B., PETIT J.-M., Eds., *Extraction et gestion des connaissances (EGC'2010), Actes, 26 au 29 janvier 2010, Hammamet, Tunisie*, Revue des Nouvelles Technologies de l'Information, Cépaduès-Éditions, 2010, p. 363-368.
- [ROU 74] ROUSE I., « The explanation of culture change », *Science*, vol. 185, 1974, p. 343-344.
- [SAL 09] SALLE P., BRINGAY S., TEISSEIRE M., « Mining Discriminant Sequential Patterns for Aging Brain », COMBI C., SHAHAR Y., ABU-HANNA A., Eds., *Artificial Intelligence in Medicine, 12th Conference on Artificial Intelligence in Medicine, AIME 2009, Verona, Italy, July 18-22, 2009. Proceedings*, Lecture Notes in Computer Science, 2009, p. 365-369.
- [WAN 04] WANG K., XU Y., YU J. X., « Scalable sequential pattern mining for biological sequences », *CIKM '04 : Proceedings of the thirteenth ACM International Conference on Information and Knowledge Management*, New York, NY, USA, 2004, ACM, p. 178-187.

Extraction de motifs d'évolution dans les Séries Temporelles d'Images Satellites

François Petitjean * — Pierre Gançarski * — Florent Masseglia **

* *LSIIT (UMR 7005 CNRS/UdS)*
Pôle API – Bd Sébastien Brant
67412 Illkirch Cedex
France
{fpetitjean,gancarski}@unistra.fr

** *INRIA Sophia Antipolis*
2004 route des Lucioles
06902 Sophia Antipolis
France
florent.masseglia@sophia.inria.fr

RÉSUMÉ. La détection de changements dans les séries temporelles d'images satellites est un domaine important avec des applications variées en étude de l'occupation des sols. La plupart des méthodes existantes effectuent cette détection en étudiant les images deux à deux et en considérant par exemple, leurs différences. De fait, elles n'exploitent pas la totalité des données fournies par les images. De plus, elles ne proposent pas de caractérisation des changements observés. Cet article propose une approche d'extraction de motifs d'évolution de ces séries d'images. Notre méthode présente deux propriétés importantes. Premièrement, elle permet d'analyser la totalité des données fournies par les différentes images, sans sélection a priori. Deuxièmement, cette méthode a été spécifiquement conçue afin d'extraire des motifs d'évolutions non-majoritaires. Ce dernier point est particulièrement utile pour l'analyse de séries temporelles d'images satellites car le comportement de non-évolution est souvent majoritaire. Nos expérimentations démontrent la pertinence de notre approche pour l'extraction de motifs d'intérêt, et la signification des motifs d'évolutions extraits.

ABSTRACT. Change detection in satellite image time series is an important domain with various applications in land study. Most previous works proposed to perform this detection by studying two images and analysing their differences. However, those methods do not exploit the whole set of images that is available today and they do not propose a description of the detected changes. We propose a sequential pattern mining approach for these image time series with two important features. First, our proposal allows for the analysis of all the images in the series and each image can be considered from multiple points of view. Second, our technique is specifically designed towards image time series where the changes are not the most frequent patterns that can be discovered. Our experiments show the relevance of our approach and the significance of our patterns.

MOTS-CLÉS : Télédétection, Séries Temporelles d'Images Satellites, STIS, Extractions de Motifs Séquentiels

KEYWORDS: Remote Sensing, Satellite Image Time Series, SITS, Sequential Pattern Extraction

1. Introduction

Alors que le domaine de la télédétection a subi un saut technologique avec l'apparition d'images à haute résolution, il s'apprête à en connaître un autre avec l'apparition de satellites (ex : *Venus*, *Sentinel-2*) capables d'acquérir des séries temporelles d'images à haute fréquence d'acquisition (deux ou trois images par semaine voire plus). Ces Séries Temporelles d'Images Satellites (STIS) sont une source d'information importante pour l'analyse d'une scène. La principale utilisation actuelle de ces images consiste à sélectionner deux images de ces séries et à étudier leurs différences, afin d'étudier les changements dans la scène. Cependant, les changements dans une scène peuvent être disséminés sur une longue période (ex : l'urbanisation d'une zone peut prendre plusieurs années), ou peuvent être cycliques (ex : rotation des cultures).

Par conséquent, le nombre de combinaisons possibles d'évolutions est très important et l'analyse de la scène ne peut être réduite à l'analyse de deux images. Partant de ce constat, nous proposons une méthode permettant l'analyse d'une série temporelle d'images satellites dans sa globalité, et l'appliquons à une série de 35 images SPOT acquises sur une période de 20 ans. Notre approche s'appuie sur une méthode de fouille de données dédiée à l'extraction de motifs séquentiels [AGR 95, MAS 98]. Cette méthode est typique de l'extraction de connaissances et permet la découverte de motifs d'évolutions fréquents à partir d'un ensemble d'enregistrements. Dans cet article, un enregistrement sera les différentes valeurs prises par un pixel au long d'une série d'acquisitions (*i.e.*, son évolution dans le temps). Considérons par exemple un ensemble de 24 images satellites acquises sur une période de deux ans, avec une fréquence d'une image par mois. Un motif d'évolution fréquent qui pourrait être découvert dans ce type de données pourrait être "dans un premier temps, 15% des pixels ont des réponses correspondant à de la végétation, puis leurs réponses correspondent à des sols nus, et enfin à des toits". En d'autres termes, s'il existe un nombre suffisamment important de pixels ayant le même "comportement", *i.e.*, suivant la même évolution, alors ce comportement doit être découvert. Mentionnons ici que les relations de voisinage spatial des pixels ne sont pas utilisées dans cet article ; le but n'est pas d'extraire des pixels selon leurs formes, mais d'extraire des schémas d'évolution typiques de séries temporelles d'images satellites, et de fournir une représentation de cette évolution ayant un sens.

Une fois les motifs fréquents extraits, l'expert dispose de descriptions de zones ayant suivi la même évolution et d'un motif d'évolution la caractérisant. L'expert peut alors y associer un sens thématique, comme la création d'une route, ou encore une culture de blé. De plus, notre approche présente deux intérêts : elle supporte un échantillonnage irrégulier des images et permet automatiquement d'ignorer des valeurs aberrantes (nuage, étalonnage du capteur, etc.).

Dans cet article, la section 2 décrit un bref état de l'art des méthodes dédiées à l'analyse de STIS. La section 3 introduit les concepts principaux de l'extraction de motifs séquentiels. La section 4 décrit quant à elle les prétraitements effectués sur les données pour la découverte de ces motifs. La section 5 introduit ensuite notre méthode

d'extraction de motifs temporels dans des séries temporelles d'images satellites. La section 6 décrit les résultats obtenus sur une zone du bassin d'Arcachon. Finalement, nous concluons ce papier en section 7.

2. État de l'art : Analyse de STIS

La détection de changements dans une scène permet l'analyse de phénomènes terrestres, au travers d'observations. Ces changements peuvent être de différents types, de différentes origines et de durées variées. La détection de changements permet, entre autres, d'étudier l'occupation des sols ou d'établir une cartographie des dommages suite à une catastrophe naturelle. On peut distinguer trois grandes familles méthodologiques pour l'analyse de changements.

Les méthodes **bi-temporelles**, permettent de situer et d'étudier des changements abrupts ayant lieu entre deux observations. Le phénomène à caractériser est généralement court dans le temps, et les observations fournissent une information sur l'état de la scène avant le changement, et une après. Ce type de méthodes inclut la différenciation d'images [BRU 00, MEL 02], le ratio d'images [TOD 77] ou encore l'analyse par vecteurs de changement [JOH 98].

Une seconde famille de méthodes correspond à des techniques **mixtes**, principalement statistiques, s'appliquant généralement à deux images mais pouvant être combinées pour en analyser plusieurs. Cette famille inclut la comparaison postclassification [FOO 01], la transformation linéaire des données (ACP ou MAF¹) [NIE 98, BRU 99, HOW 06], la régression et l'interpolation [JHA 94, RID 98], ou encore l'analyse fréquentielle (ex : décomposition de Fourier ou en ondelette) [AND 94, CEL 09].

Enfin, la dernière famille correspond à des méthodes dédiées à l'étude de **séries temporelles** d'images. Ces méthodes sont généralement basées sur l'analyse de trajectoires radiométriques de pixels, afin de les comparer [KET 07] ou d'y détecter des ruptures [KEN 07, VER 10].

Quel que soit le type de méthodes utilisées pour l'analyse de séries d'images satellites, il existe un décalage entre la quantité d'information que représentent ces séries temporelles, et la capacité des algorithmes à les analyser. Ces algorithmes sont le plus souvent dédiés à l'analyse bi-date d'une scène et se concentrent sur la cartographie des zones de changements et non sur leur caractérisation. Les méthodes bi-date sont de plus liées à des thématiques d'études spécifiques et sont incapables d'appréhender des changements ayant lieu au travers d'une STIS. Quant aux méthodes multi-date, elles sont souvent difficilement interprétables et ne permettent pas de caractériser le changement. Elles se focalisent sur les données image par image et n'extraient pas l'information d'évolution disséminée au long d'une STIS.

Parallèlement, l'extraction de motifs séquentiels fréquents [AGR 95, MAS 98] permet d'extraire des motifs fréquents d'évolution d'une série de symboles. Ces mé-

1. Maximum Autocorrelation Factor

thodes permettent d'identifier les ensemble de séquences ayant suivi la même évolution. De plus, elles permettent de caractériser cette évolution, en fournissant le motif partagé par cet ensemble de séquences.

L'extraction de motifs d'évolution fréquents à partir de STIS a été introduit dans [JUL 06, JUL 08]. Les auteurs utilisent ces techniques pour la météorologie et l'agronomie. Cependant, leur proposition ne permet l'extraction de motifs que sur les séries d'images mono-bande, ce qui est facteur limitant bien connu pour la discrimination des objets observés.

L'étude que nous proposons, comme nous l'expliquons en section 4, porte sur des images multi-spectrale. Nous verrons que cette caractéristique, associée au grand nombre d'images que nous analysons, influe sur la complexité des motifs, leur pertinence et la difficulté de leur extraction.

3. Recherche de motifs temporels fréquents

Les motifs temporels sont extraits à partir d'un ensemble d'enregistrements (souvent très volumineux). Ces enregistrements contiennent des valeurs définies sur un vocabulaire précis et se présentent sous forme de séquences de données comme l'indique la définition 1, adaptée à partir des définitions de [AGR 95].

Définition 1 Soit $\mathcal{I} = \{i_1, i_2, \dots, i_m\}$, un ensemble de m valeurs (ou items). Soit $I = \{t_1, t_2, \dots, t_n\}$, un sous-ensemble de \mathcal{I} . I est appelé un itemset.

Une séquence s est une liste ordonnée, non vide, d'itemsets notée $\langle s_1, s_2, \dots, s_n \rangle$ où s_j est un itemset. Une séquence de données est une séquence enregistrée dans l'ensemble des données à analyser.

La définition 2 exprime les conditions de l'inclusion d'une séquence s_1 dans une séquence s_2 . En d'autres termes, s_1 est incluse dans s_2 si chaque itemset de s_1 est inclus dans un itemset de s_2 en respectant l'ordre des itemsets dans s_1 et dans s_2 . Cette définition est illustrée par l'exemple 1

Définition 2 Soit $s_1 = \langle a_1, a_2, \dots, a_n \rangle$ et $s_2 = \langle b_1, b_2, \dots, b_m \rangle$ deux séquences. s_1 est incluse dans s_2 ($s_1 \prec s_2$) si et seulement si il existe $i_1 < i_2 < \dots < i_n$ des entiers tels que $a_1 \subseteq b_{i_1}, a_2 \subseteq b_{i_2}, \dots, a_n \subseteq b_{i_n}$.

Exemple 1 La séquence $s_1 = \langle (3) (4\ 5) (8) \rangle$ est incluse dans la séquence $s_2 = \langle (7) (3\ 8) (9) (4\ 5\ 6)(8) \rangle$ (i.e $s_1 \prec s_2$) car $(3) \subseteq (3\ 8)$, $(4\ 5) \subseteq (4\ 5\ 6)$ et $(8) \subseteq (8)$. En revanche la séquence $s_3 = \langle (3\ 8\ 9) (4\ 5) \rangle$ n'est pas incluse dans s_2 car $(3\ 8\ 9)$ n'est inclus dans aucun itemset de s_2 .

Pour extraire des motifs temporels, la caractéristique retenue dans nos travaux sera la fréquence. La fréquence est basée sur le nombre d'apparitions par rapport

au nombre d'enregistrements, comme l'indique la définition 3. Enfin, pour simplifier l'interprétation des résultats, seuls les motifs temporels les plus longs sont gardés dans l'ensemble des motifs extraits (C.f. la définition 4).

Définition 3 Une séquence de données s_d supporte une séquence s (ou contribue au support de s) si $s \prec s_d$. Soit D un ensemble de séquences de données. Le support d'une séquence s dans D est calculé comme étant le pourcentage des séquences de données de D qui supportent s :

$$\text{support}(s) = \frac{|\{s_d \in D / s \prec s_d\}|}{|D|} \quad [1]$$

Soit min.Supp un support minimum fixé par l'utilisateur. Une séquence dont le support est supérieur à min.Supp est une séquence fréquente.

Définition 4 L'ensemble F^D des motifs temporels fréquents de D est constitué de toutes les séquences fréquentes sur D . L'ensemble L^D des motifs temporels fréquents maximaux de D est constitué de toutes les séquences s dans F^D telles que s n'est incluse dans aucune autre séquence de F^D .

4. Préparation des données (STIS)

Les images utilisées dans cet article sont extraites de la base de données Littoral du programme KALIDEOS² et se situent dans le sud-ouest de la France. Nous avons extrait de cette base de données une série de 35 images SPOT-1, SPOT-2 et SPOT-4 (illustrée en figure 1).

Cette STIS s'étend sur une période de 20 ans ; les distributions annuelles et mensuelles des images sont données par le tableau 1. Ce tableau montre que la fréquence d'acquisition des images n'est pas régulière. Disposer de données acquise irrégulièrement est un cas assez fréquent, pour lequel les méthodes d'analyses doivent être robustes. Nous montrerons que notre méthode est capable d'utiliser la totalité des données tout en étant robuste à la fréquence d'acquisition des images dans la série.

Années	1986–1989	1990–1994	1995–1999	2000–2004	2005–2006
Nb	7	13	4	7	4

Mois	J	F	M	A	M	J	J	A	S	O	N	D
Nb	2	3	3	3	1	6	5	4	1	4	3	0

Tableau 1. Répartitions annuelle et mensuelle de la série d'images étudiée.

Avant d'intégrer la bases de données KALIDEOS, les produits SPOT sont orthorectifiés (garantissant qu'un pixel (x, y) couvre la même zone géographique au long de

2. <http://kalideos.cnes.fr>

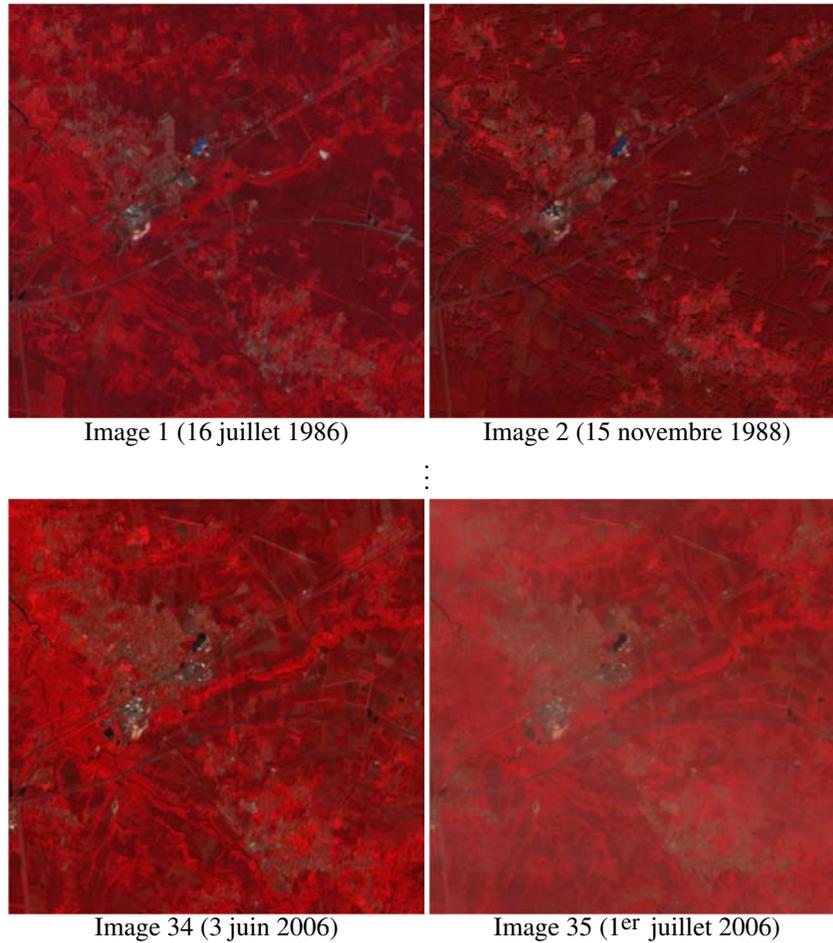


Figure 1. Extrait de la STIS utilisée provenant de la Base de données LITTORAL – KALIDEOS. © CNES 2010 – Distribution Spot Image

la séquence d'images) et subissent des traitements radiométriques afin que les valeurs des pixels soient comparables d'une image à une autre. Ces traitements consistent à convertir les comptes numériques fournis par le capteur en grandeurs physiques et à restituer la contribution propre à la surface par correction des effets de l'atmosphère. A partir du modèle radiométrique de l'instrument, les comptes numériques sont d'abord convertis en réflectances (grandeur physique normalisée de l'éclairement solaire). Les coefficients d'étalonnage absolu utilisés dans cette étape proviennent du suivi des capteurs SPOT effectué par le CNES. L'inversion de la réflectance de surface est faite ensuite par comparaison de la réflectance mesurée à des simulations au sommet de l'atmosphère, réalisées pour les conditions géométriques et atmosphériques de la me-

sure. L'élévation du terrain est prise en compte par la réalisation de simulations pour diverses altitudes, incluant une pondération de la pression atmosphérique et des quantités d'aérosols et de vapeur d'eau. L'état de l'atmosphère au moment de l'acquisition est quant à lui caractérisé à l'aide de sources météorologiques (NCEP pour la pression et l'humidité), de sources de données ozone (TOMS ou TOAST) et de données aérosols (SeaWIFS, Aeronet). A défaut, des valeurs climatologiques sont utilisées.

Une fois ces corrections effectuées, nous disposons de 35 images sur trois bandes : proche infra-rouge (PIR), rouge (R) et vert (V). À cela, nous ajoutons un quatrième attribut correspondant à l'indice de végétation NDVI³. Chaque séquence est ainsi construite comme la succession des quadruplet de valeurs (PIR, R, V, NDVI) pour chaque pixel (x, y) dans la série d'images.

Finalement, une étape de discrétisation est indispensable pour pouvoir analyser ces données avec une méthode d'extraction de motifs. Nous avons donc appliqué, sur chaque bande, une classification par l'algorithme K-MEANS [MAC 67]. Ainsi, pour chacune des bandes, nous avons rassemblé l'ensemble des valeurs de la série d'images afin de créer 20 tranches de valeurs par bande. Le fait de n'effectuer qu'un clustering par bande (et non un par bande et par image), permet de pouvoir comparer les tranches d'une image à une autre.

Pour plus de lisibilité, les numéros de clusters ont été réordonnés de façon croissante en fonction des valeurs de leurs centroïdes ; les séquences sont donc des suites de quadruplets (NIR,R,V,NDVI) de la forme :

$$(PIR_1, R_6, V_3, NDVI_{16}) \rightarrow \dots \rightarrow (PIR_{12}, R_3, V_{14}, NDVI_{19}) \quad [2]$$

où $(PIR_1, R_6, V_3, NDVI_{16})$ signifie que la valeur du pixel dans la première image se trouve dans la première tranche de proche infra-rouge, dans la 6^{ème} tranche de rouge, dans la troisième tranche de vert et dans la 16^{ème} tranche de $NDVI$.

5. Méthode d'extraction des motifs

Les étapes de pré-traitement décrites en section 4 permettent donc d'obtenir une série d'images dans lesquelles chaque pixel possède plusieurs valeurs. Considérons la série de 3 images réduites à 4 pixels (p_1 à p_4) illustrée par la figure 2 et où chaque pixel possède 3 valeurs (qui correspondent à 3 bandes, B_1 à B_3)⁴. Avec un support minimum de 100%, il n'existe aucun motif fréquent dans ces images (aucun comportement n'est suivi par l'ensemble des pixels). Avec un support minimum de 50%, en revanche, on trouve deux comportements fréquents :

1) $\langle (B_1, \text{blanc}; B_2, \text{blanc}) (B_1, \text{gris}; B_2, \text{rouge}) \rangle$. Ce comportement est suivi par les pixels p_2 sur les images 1 et 2 (ou 3) et p_3 sur les images 1 (ou 2) et 3.

3. Par abus de langage, nous parlerons de quatrième bande.

4. Nous utilisons des couleurs pour décrire les valeurs des bandes pour des questions didactiques et de visualisation.

2) $\langle (B1, \text{blanc}; B2, \text{blanc}) (B1, \text{blanc}; B2, \text{blanc}) \rangle$. Ce comportement est suivi par $p1$ et $p3$ sur les images 1 et 2.

Notons, dans l'exemple ci-dessus, que les motifs peuvent être fréquents malgré un décalage temporel dans les images où ils sont supportés.

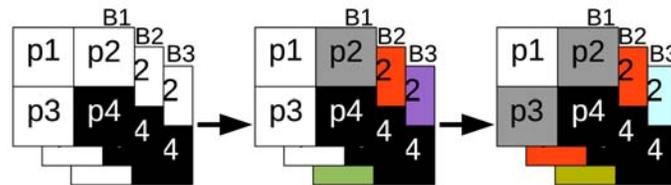


Figure 2. Une série de trois images composées de quatre pixels, chacun représenté sur trois bandes.

En général, la recherche de motifs consiste à trouver des motifs avec des supports élevés (très fréquents), la problématique est un peu différente pour l'application aux STIS. En effet, un motif ayant 80 % de support, s'interprète comme étant un motif suivi par 80 % des séquences et donc par 80 % des zones géographiques de la scène. Cependant, il est très rare qu'il existe un comportement couvrant une si grande partie de la scène qui soit intéressant. Il est certes possible de trouver que 80 % des séquences sont passées par la dixième tranche de la bande rouge, mais ce type de comportements ne correspond généralement pas à un comportement intéressant. Il est donc nécessaire d'appliquer des contraintes sur le processus d'extraction afin d'identifier des comportements locaux, *i.e.*, entre 1 % et 10 % des séquences. Sur notre scène couvrant 81 km², cela correspond à trouver des zones de moins d'un km² jusqu'à 8 km² ayant suivi une évolution similaire au long de la série d'images.

Cependant, rechercher des motifs moins fréquents complexifie le processus d'extraction. Considérons un motif m ayant un support supérieur à 80 %. Si ce motif a un support de 80 %, c'est que 80 % des séquences suivent ce motif. Mais si 80 % le suivent alors tout sous-ensemble de ces séquences suit le motif, et donc ce motif sera également identifié avec toutes les valeurs de support inférieures à celui-ci. Aussi, ces séquences fréquentes vont-elles inonder le processus d'extraction qui ne sera pas capable d'extraire les motifs ayant un support faible intéressant. C'est une des limitations principales dans la méthode mise en œuvre dans [JUL 06, JUL 08], celle-ci ne pouvant extraire que des motifs à supports élevés, empêchant l'identification de comportements minoritaires intéressants.

Ainsi, nous proposons dans cet article, un algorithme d'extraction de motifs fréquents basé sur [MAS 98] modifiés en deux points :

1) pendant la phase de découverte des valeurs fréquentes, nous ne considérons que les valeurs dont le support dans une image est compris dans un intervalle, ce qui permet d'éliminer un grand nombre de motifs non-pertinents.

2) pendant les étapes suivantes, nous éliminons les séquences contenant deux valeurs successives identiques sur une bande. Par exemple, la séquence candidate $\langle (PIR, 2) (R, 2) \rangle$ est conservée mais pas la séquence $\langle (PIR, 2) (PIR, 2) \rangle$.

6. Expérimentations / Résultats

Les images pré-traitées (voir section 4) ont une étendue de 202 500 pixels (450x450). Chaque pixel de chacune des 35 images étant décrit sur quatre attributs, cette série totalise 28 millions de valeurs à analyser.

La méthode détaillée en section 5 a été appliquée à la série d'images illustrée en figure 1. Plusieurs centaines de motifs ont été découverts par ce processus d'extraction ; une carte a été générée pour chaque motif afin de pouvoir les sélectionner manuellement. La figure 3 illustre trois motifs extraits de cette série.

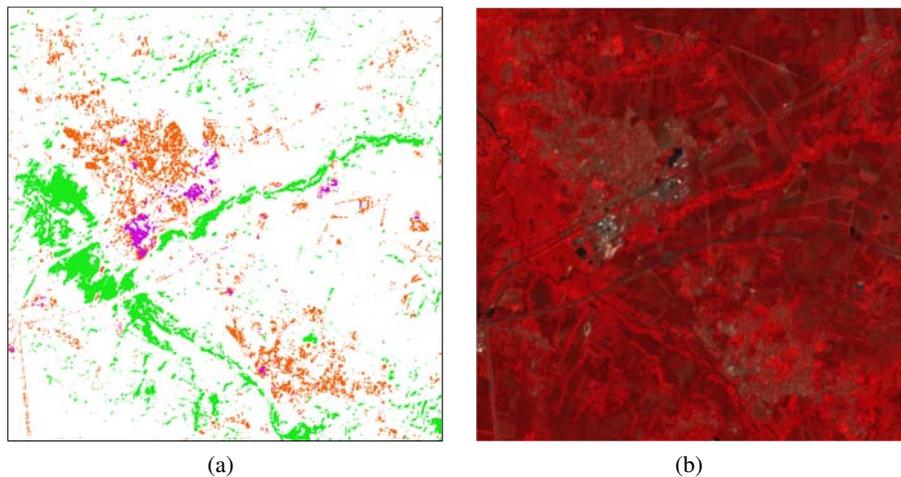


Figure 3. Présentation des résultats. La figure (a) illustre trois motifs extraits par la méthode proposée (une couleur par motif). L'image du 3 juin 2006 extraite de la STIS est illustrée en figure (b) pour faciliter l'interprétation de ces résultats.

Nous avons vu que chaque séquence du jeu de données est identifiée par ses coordonnées (x, y) dans les images. Un motif suivi par un ensemble de séquences peut donc être visualisé sur une image, comme un ensemble de pixels de coordonnées (x, y) . La figure 3(a) présente, sur ce principe, trois motifs extraits de la STIS (un par couleur). Nous détaillons ci-dessous ces trois motifs ainsi que l'interprétation thématique qui en découle :

1) le motif $\langle (PIR, 1) (NDVI, 20) \rangle$ est représenté en vert. Ce motif correspond à des zones marécageuses, dont le niveau oscille entre un niveau très faible de

proche infra-rouge quand les zones sont recouvertes d'eau en hiver et un fort niveau NDVI en été, car la végétation apparaît, et est de plus très irriguée et donc très chlorophyllienne. La représentation de ce motif est assez intéressante. En effet, ce motif est recentré sur l'évolution atomique de ces zones (deux états), alors que notre série couvre plusieurs années.

2) le motif $\langle (R, 17) (R, 18 ; NDVI, 3) \rangle$ est représenté en orange. Ce motif correspond à des zones urbaines s'étant densifiées. En effet, l'urbain répond fortement dans la bande rouge, d'où le niveau de départ élevé de 17. La densification urbaine de ces zones au long de la série d'images est indiquée par le niveau suivant dans la rouge, passant dans la 18^{ème} tranche, correspondant à une augmentation de 25 % de la réflectance. La présence d'urbain est de plus corroborée par le niveau final très faible de NDVI indiquant une quasi-absence de végétation.

3) le motif $\langle (NDVI, 2) (V, 20) (NDVI, 1) \rangle$ est représenté en violet. Ce motif correspond à des zones industrielles s'étant densifiées. La diminution de NDVI, passant de la deuxième à la première tranche, indique un recul du couvert végétal sur ces zones, d'autant que cette transition correspond à une diminution de 30 % de réflectance. Enfin, le niveau maximum atteint dans le vert corrobore la présence de toits plats (type tôle ondulée) caractéristiques des zones industrielles.

Plusieurs remarques peuvent être faites sur cette extraction. Tout d'abord, comme cette méthode permet d'éviter d'ignorer une valeur dans la série, les valeurs singulières (ex : nuage, bruit) sont automatiquement ignorés. De plus, seuls les pixels nuageux dans une image sont ignorés, ce qui permet d'utiliser toutes les données disponibles, sans supprimer de données non-bruitées. Par ailleurs, notre méthode a permis, dans le même temps, d'extraire des comportements saisonniers (marais) et donc cycliques, ainsi que des comportements transversaux à la série temporelle (densification). C'est une propriété intéressante de notre approche permettant d'identifier différents types de comportements. Enfin, en plus de fournir des groupes de séquences ayant évolué de la même façon, notre méthode fournit une description des motifs supportés par les différentes séquences. Cette description peut permettre à l'expert de qualifier les motifs extraits, et peut par exemple être réinjectée comme une connaissance dans un processus plus global d'étude de la scène.

7. Conclusion

Les Séries Temporelles d'Images de Télédétection sont des données très riches pour l'étude de l'occupation des sols, soit pour la discrimination d'un état d'occupation des sols, soit pour l'étude de leurs évolutions. Notre méthode d'extraction de motifs temporels nous a permis d'identifier des motifs d'évolutions pertinents que nous avons détaillés en section 6. Les motifs extraits correspondent aux états principaux d'évolution par lesquels passe une zone géographique (imagée par un ensemble de pixels). Ces motifs d'évolution ont permis d'extraire aussi bien des comportements cycliques (marais) que des comportements d'évolution transversaux à la série d'images (densification). Par ailleurs, comme les méthodes d'extraction de motifs fréquents per-

mettent de se focaliser sur les états principaux d'évolutions, elles sont robustes au bruit, qu'il soit lié au capteur ou à la météorologie. Enfin, les motifs extraits fournissent une représentation centrée sur les principales étapes d'évolution, ce qui les rend facile à lire et à interpréter.

L'adaptation des techniques d'extraction de motifs fréquents aux images de télé-détection ouvre de nouvelles perspectives. Ces motifs constituent une connaissance qui pourra par exemple être réinjectée afin de les utiliser comme des évolutions types dans des systèmes de prédiction. De plus, il serait intéressant d'étudier les résultats de cette approche sur une série d'images à l'échelle intra-annuelle (*e.g.*, agronomie). Il serait également intéressant de contraindre le processus afin d'influer sur la longueur des motifs extraits. Enfin, de la même façon que ces techniques permettent d'ignorer des valeurs aberrantes, elles pourraient permettre de traiter des séries de longueurs variées.

Remerciements

Ce travail est supporté par le Centre National d'Études Spatiales et par Thales Alenia Space sous le contrat de bourse de recherche n°1520011594.

8. Bibliographie

- [AGR 95] AGRAWAL R., SRIKANT R., « Mining Sequential Patterns », *Proceedings of the 11th International Conference on Data Engineering (ICDE'95)*, 1995, p. 3–14.
- [AND 94] ANDRES L., SALAS W., SKOLE D., « Fourier analysis of multi-temporal AVHRR data applied to a land cover classification », *International Journal of Remote Sensing*, vol. 15, n° 5, 1994, p. 1115–1121, Taylor & Francis.
- [BRU 99] BRUZZONE L., PRIETO D., SERPICO S., « A neural-statistical approach to multitemporal and multisource remote-sensing image classification », *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, n° 3, 1999, p. 1350-1359.
- [BRU 00] BRUZZONE L., PRIETO D., « Automatic analysis of the difference image for unsupervised change detection », *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, n° 3, 2000, p. 1171-1182.
- [CEL 09] CELIK T., « Unsupervised Multiscale Change Detection In Multitemporal Synthetic Aperture Radar Images », *17th European Signal Processing Conference (EUSIPCO)*, vol. 15, 2009, p. 47–51.
- [FOO 01] FOODY G., « Monitoring the magnitude of land-cover change around the southern limits of the Sahara », *Photogrammetric Engineering and Remote Sensing*, vol. 67, n° 7, 2001, p. 841–848, ASPRS American Society for Photogrammetry and Remote Sensing.
- [HOW 06] HOWARTH P., PIWOWAR J., MILLWARD A., « Time-Series Analysis of Medium-Resolution, Multisensor Satellite Data for Identifying Landscape Change », *Photogrammetric engineering and remote sensing*, vol. 72, n° 6, 2006, p. 653–663.
- [JHA 94] JHA C., UNNI N., « Digital change detection of forest conversion of a dry tropical Indian forest region », *International Journal of Remote Sensing*, vol. 15, n° 13, 1994,

- p. 2543–2552, Taylor and Francis.
- [JOH 98] JOHNSON R., KASISCHKE E., « Change vector analysis : a technique for the multispectral monitoring of land cover and condition », *International Journal of Remote Sensing*, vol. 19, n° 16, 1998, p. 411-426.
- [JUL 06] JULEA A., MÉGER N., TROUVÉ E., « Sequential patterns extraction in multitemporal satellite images », *17th European Conference on Machine Learning and the 10th European Conference on Principles and Practice of Knowledge Discovery in Databases (ECML / PPKD)*, 2006, page 96.
- [JUL 08] JULEA A., MÉGER N., TROUVÉ E., BOLON P., « On Extracting Evolutions from Satellite Image Time Series », *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 5, 2008, p. 228-231.
- [KEN 07] KENNEDY R. E., COHEN W. B., SCHROEDER T. A., « Trajectory-based change detection for automated characterization of forest disturbance dynamics », *Remote Sensing of Environment*, vol. 110, n° 3, 2007, p. 370–386.
- [KET 07] KETTERLIN A., GAŃCARSKI P., « Sequence similarity and multi-date image segmentation », *Fourth Intl Workshop on the Analysis of Multitemporal Remote Sensing Images*, Leuven, Belgique, july 2007.
- [MAC 67] MACQUEEN J., « Some Methods for Classification and Analysis of Multivariate Observations », *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*, 1967, p. 281-297.
- [MAS 98] MASSEGLIA F., CATHALA F., PONCELET P., « The PSP Approach for Mining Sequential Patterns », *Proceedings of the 2nd European Symposium on Principles of Data Mining and Knowledge Discovery*, 1998.
- [MEL 02] MELGANI F., MOSER G., SERPICO S. B., « Unsupervised change-detection methods for remote-sensing images », *Optical Engineering*, vol. 41, n° 12, 2002, p. 3288-3297, SPIE.
- [NIE 98] NIELSEN A., CONRADSEN K., SIMPSON J., « Multivariate Alteration Detection (MAD) and MAF Postprocessing in Multispectral, Bitemporal Image Data : New Approaches to Change Detection Studies », *Remote Sensing of Environment*, vol. 64, n° 1, 1998, p. 1–19.
- [RID 98] RIDD M. K., LIU J., « A comparison of four algorithms for change detection in an urban environment », *Remote sensing of environment*, vol. 63, n° 2, 1998, p. 95–100, Elsevier Science.
- [TOD 77] TODD W., « Urban and regional land use change detected by using Landsat data », *Journal of Research by the US Geological Survey*, vol. 5, 1977, p. 527–534.
- [VER 10] VERBESSELT J., HYNDMAN R., NEWNHAM G., CULVENOR D., « Detecting trend and seasonal changes in satellite image time series », *Remote Sensing of Environment*, vol. 114, n° 1, 2010, p. 106 - 115.

Classification des tissus urbains à partir de données vectorielles – application à Strasbourg

Anne Puissant*, Nicolas Lachiche**, Grzegorz Skupinski*, Agnès Braud**, Julien Perret***

* *Laboratoire Image, Ville, Environnement, ERL 7230 CNRS / Université de Strasbourg - 3, rue de l'argonne, F-67083 Strasbourg cedex, France.*
{anne.puissant; g.skupinski}@live-cnrs.unistra.fr

** *Laboratoire des Sciences de l'Images, de l'Informatique et de la Télédétection, UMR 7005 CNRS /Université de Strasbourg - Pôle API Bd Sébastien Brant 67412 Illkirch Cedex, France.*
{nicolas.lachiche; agne.braud}@unistra

*** *Laboratoire COGIT, Institut Géographique National 73 avenue de Paris, F-94165 Saint Mandé Cedex, France*
julien.perret@ign.fr

RÉSUMÉ. L'analyse morphologique des tissus urbains est un préalable important afin de mieux comprendre les dynamiques urbaines et afin de les simuler. Dans ce contexte, l'objectif de cet article est de tester un algorithme d'apprentissage supervisé symbolique (TILDE) permettant d'obtenir un modèle de classification représenté sous forme de règles de décision. L'objectif est d'automatiser le processus d'identification des tissus urbains à partir de bases de données vectorielles et d'extraire des connaissances morphologiques sur les tissus urbains. Différents tests, appliqués à l'agglomération de Strasbourg, sur des données historiques reconstruites sur une période de 50 ans, met en évidence l'intérêt de ce type d'algorithme pour classer les tissus urbains. Toutefois, l'identification automatique de règles caractérisant la morphologie des tissus urbains de manière univoque reste difficile.

ABSTRACT. Morphological analysis of urban fabric is a relevant step in order to perform the analysis of urban dynamics and to simulate their evolutions. In this context, the objective of this paper is to test a classification method based on a supervised and symbolic algorithm (tilde) to label urban blocks and to produce rules on their morphology. The objective is to automate the identification of urban fabrics from vector and historical database and to extract knowledge on urban morphology. Several tests applied on the Strasbourg area on historical database, are encouraging but rules are not relevant to extract urban blocks.

MOTS-CLÉS : tissu urbain, morphologie, bases de données vectorielles, classification, arbre de décision.

KEYWORDS: urban fabric, morphology, vector database, classification, decision tree.

1. Introduction

La ville est aujourd'hui au cœur des grandes préoccupations contemporaines. L'étalement urbain et la maîtrise de ses conséquences restent un défi important de l'aménagement du territoire. L'analyse morphologique des tissus urbains est un préalable important afin de mieux comprendre les processus d'urbanisation de nos agglomérations. A partir de cette connaissance, on peut imaginer de proposer des scénarios de développement urbain ou de développer des concepts différents pour trouver des réponses réalistes au défi de l'étalement urbain.

Le projet de recherche GeOpenSim (ANR Blanc 2007-2010) a pour objectif de proposer une plate-forme Open Source pour l'analyse et la simulation des dynamiques urbaines à l'aide de bases de données topographiques historiques. Ces dernières sont construites à l'aide de bases de données topographiques récentes ainsi que de photographies aériennes et/ou de cartes historiques (Puissant *et al.*, 2010). Ces bases de données vectorielles sont analysées afin d'améliorer les connaissances sur les objets spécifiques comme les bâtiments et afin d'identifier des règles d'évolution à différents niveaux méso-géographiques (e.g. la ville ou l'îlot) (Curie *et al.* 2010).

L'extraction de connaissances au niveau de l'îlot est une étape préalable à l'analyse de leur évolution. Celle-ci passe par une classification et une analyse morphologique des tissus urbains qui peut être assimilée à un problème de fouille de données. Dans ce contexte, l'objectif de cet article est de tester un algorithme d'apprentissage supervisé symbolique (Tilde) permettant d'obtenir un modèle de classification représenté sous forme de règles de décision. L'objectif est d'automatiser le processus d'identification des tissus urbains à l'échelle d'une agglomération de la taille de Strasbourg et d'extraire des connaissances morphologiques sur les tissus urbains permettant de suivre leur évolution dans le temps.

Après un état de l'art (section 2) sur les méthodes de classification des tissus urbains et sur les méthodes d'extraction de connaissances à partir de telles données, la zone d'étude ainsi que les mesures et les paramètres statistiques utilisés sont détaillés à la section 3. Les différentes étapes de la classification sont ensuite présentées à la section 4. Dans la section 5, les tests effectués ainsi que les résultats sont détaillés et discutés. Les conclusions et perspectives sont finalement exposées à la section 6.

2. Classification des tissus urbains et extraction de connaissances : état de l'art

A l'heure actuelle, la classification automatique ou semi-automatique des tissus urbains est rarement effectuée à partir de données vectorielles. Classiquement, ce type de classification est réalisé à partir d'images satellites et/ou de photographies aériennes et passe par une phase de photo-interprétation manuelle combinée à des vérifications terrain. Le résultat se présente sous la forme d'une base de données sur

l'occupation et l'utilisation des sols. La BD-OCS (base de données 'occupation du sol') proposée par la région Alsace peut être citée comme exemple (Geoméditerranée, 2004). Celle-ci est réalisée par photo-interprétation à partir d'images satellites, proposant une cartographie des tissus urbains à l'échelle du 1/10 000^e.

L'automatisation de ce type de cartographie en milieu urbain, fait l'objet de nombreux travaux de recherche en particulier à partir d'images satellites haute résolution (Donnay *et al.*, 2001) mais également depuis l'avènement des images à très haute résolution (Blaschke, 2009). Des méthodes de traitements d'image fondées sur un apprentissage supervisé (classification supervisée) sont largement utilisées dans ce domaine (Bauer *et al.*, 2001). Ces techniques basées sur les valeurs radiométriques des pixels des images, permettent d'identifier des zones qui combinées avec un découpage vectoriel (IRIS de l'INSEE par exemple) permettent d'obtenir une cartographie des îlots urbains (Holland *et al.*, 2006). Toutefois, les images satellites ne permettent pas d'analyser la composition et l'organisation spatiale des tissus urbains.

Avec l'arrivée des nouvelles images à très haute résolution (résolution spatiale métrique) et le développement de nouvelles méthodes de traitement d'images orientées-objets (Blaschke *et al.*, 2004), c'est-à-dire permettant de travailler directement sur des régions homogènes (agrégats de pixels homogènes), de nombreux travaux utilisent une segmentation multi-échelle pour extraire et classer des objets micros (bâtiment, végétation, route) et des objets méso (tissus urbains) (Jacquin *et al.*, 2010, Bouziani *et al.*, 2010). D'autres méthodes tentent d'utiliser simultanément les informations issues des images à différentes résolutions spatiales afin de reconstruire les objets méso (Kurtz *et al.* 2010). Malgré des résultats prometteurs, ces méthodes sont toutefois encore exploratoires et pas directement opérationnelles. Elles ne permettent pas non plus d'analyser et de caractériser la composition morphologique des tissus urbains.

L'analyse morphologique d'un tissu urbain peut être assimilée à un problème de fouille de données. En effet, la fouille de données spatiales est née du besoin d'exploitation dans un but décisionnel de données à caractère spatial produites ou importées, susceptibles de délivrer des informations ou des connaissances par le moyen d'outils exploratoires (Zeitouni *et al.*, 2000). Dans ce domaine, des méthodes classiques comme les arbres de décision ont été adaptées. Les arbres de décision ont pour but de trouver les attributs explicatifs et les critères précis sur les attributs donnant le meilleur classement vis-à-vis d'un attribut à expliquer. Dans Chelghoum *et al.* (2002), les auteurs pré-calculent les relations spatiales entre objets spatiaux et les stockent. Les critères utilisés dans l'arbre de décision construit sont alors des tests d'existence d'un objet vérifiant une relation spatiale donnée. Cet algorithme a été appliqué à l'analyse du risque d'accidents routiers. A la suite de ce travail, les auteurs se sont intéressés à un système de construction d'arbre de décision à partir de données relationnelles appelé TILDE (Blockeel and De Raedt, 1998), et l'ont étendu aux données spatiales avec S-TILDE qui intègre en plus au niveau du critère de

division les propriétés d'objets voisins (Chelghoum *et al.*, 2006). D'autres travaux ont porté sur la fouille de données à partir de données vecteurs. Il existe notamment un ensemble de travaux sur les règles d'association, comme (Salleb et Vrain, 2000) qui introduisent dans les règles d'association des relations spatiales entre les objets qui peuvent provenir de différentes tables. Ce travail étend celui de (Koperski et Han, 1995). Dans ce contexte, nous avons testé l'algorithme relationnel TILDE qui a une expressivité supérieure à l'algorithme attribut-valeur C4.5 utilisé couramment (Quinlan, 1995). Il s'applique également à nos données de départ (cf. section 3) contenues dans une base de données relationnelles, sous la forme de deux tables (une table pour les îlots et une table pour les bâtiments).

3. Zone d'étude, données et mesures morphologiques

La Communauté Urbaine de Strasbourg (CUS) est une agglomération qui s'est développée, comme la plupart des grandes villes françaises, selon le schéma de l'étalement urbain (Bessy-Piétry, 2000). Depuis sa création (en 1967) jusqu'aux années 1982, la surface bâtie a augmenté de 71.6% (ADEUS, 2003), contre 20.9% entre l'année 1988 et 2000. Depuis on assiste à une forte concentration de l'emploi tandis que la population se diffuse largement dans l'ensemble du département (Bas-Rhin), voire dans l'espace transfrontalier (Saib, 2009).

Dans ce contexte, les objets géographiques de la base de données topographique de l'IGN sur la CUS (IGN, 2002) sont utilisés comme données de référence. Les bases de données historiques sont ensuite construites sur quatre zones tests (Figure 1). Elles ont été sélectionnées en raison de leur localisation dans l'agglomération (zones périurbaines) et des processus d'évolution représentatifs des grandes agglomérations françaises au cours des 50 dernières années. Ces données historiques ont été produites, dans le cadre du projet, pour 5 dates distinctes : 1956, 1966, 1976, 1989, 2002 (référence) et 2008.

La typologie des tissus urbains proposée (Tableau 1) est adaptée afin de cartographier l'espace urbain d'une grande agglomération française à une échelle du 1/10 000^e. Ces classes, compatibles avec les nomenclatures existantes (e.g. Corine Land Cover, BDOCS (Alsace)), ne tiennent pas compte de la fonctionnalité des tissus urbains car cette information est impossible à reconstruire sur des données historiques. Elles permettent de se concentrer sur leur analyse morphologique. Parmi ces classes, trois d'entre elles ne sont pas prise en compte dans la suite de l'analyse en raison de leur faible potentiel d'évolution (classe n°1 – tableau 1) et d'interprétation (classe n°8 – tableau 1).

L'analyse morphologique des tissus urbains passe par une étape préalable de construction des îlots urbains. Ceux-ci sont construits automatiquement à partir des objets géographiques linéaires de la base de donnée topographique (objets de type 'réseaux de communication' regroupant le réseau routier, ferré et hydrographique).

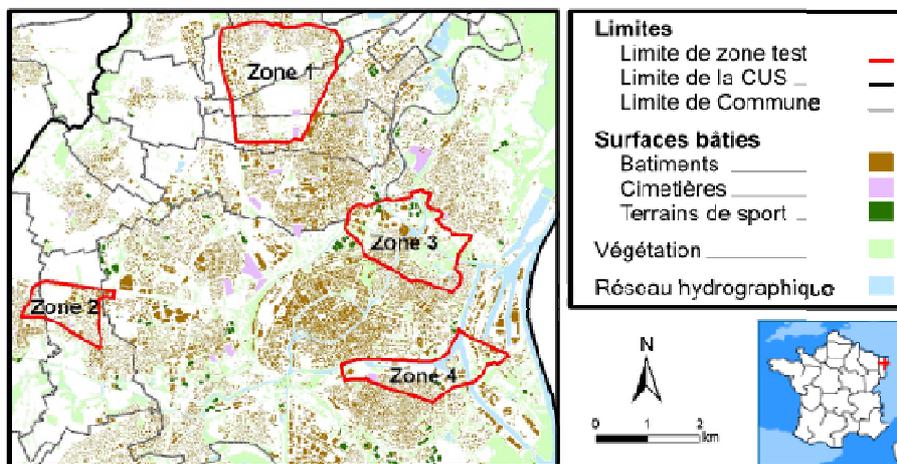


Figure 1. Localisation des quatre zones tests utilisées pour analyser l'évolution des tissus urbains (Source: BDTPOPO@2002).

Classes de tissu urbain
1. Habitat continu dense de type centre ville ou centre bourg
2. Habitat discontinu de type pavillonnaire ou individuel
3. Habitat discontinu de type collectif ou grand ensemble
4. Habitat discontinu mixte : <i>Mélange des classes 2 (pavillonnaire) et 3 (grand ensemble)</i>
5. Tissu urbain discontinu mixte <i>Mélanger de classes d'habitat (classe 1, 2, 3) et de la classe 6 (emprise spécialisées bâties)</i>
6. Emprises spécialisées bâties : <i>emprises militaires, zones industrielles, commerciales, emprises scolaires et universitaires, emprises hospitalières, emprise publique ou patrimoine</i>
7. Emprises spécialisées peu ou pas bâties : <i>emprises liées à l'activité d'extraction – carrières, surfaces associées aux réseaux : gare de triage, échangeur, emprises aériennes, ferroviaires et portuaires, équipements sportifs et de loisirs, espaces en mutation ou libres de construction (sols nus, artificiels ou végétalisés).</i>
8. Réseau de communication
9. Réseau hydrographique

Tableau 1. Liste des classes d'îlots urbains retenus

Des mesures (ou paramètres) sont ensuite calculées pour caractériser d'une part les objets géographiques surfaciques (e.g. l'aire, l'élongation et la convexité) et d'autre part l'organisation spatiale de ces objets dans les îlots (l'orientation des bâtiments et l'orientation relative des bâtiments par rapport à la route la plus proche

par exemple). D'autres mesures, telles que le nombre de bâtiments dans l'îlot, la surface construite (la somme des surfaces des bâtiments) et la densité (surface construite/surface de l'îlot) sont disponibles. Pour chacune d'entre elles, les valeurs du minimum, du maximum, de la moyenne, de la médiane et de l'écart type sont également calculées.

4. Etapes de classification et modules proposés

L'objectif de la classification est d'associer une des classes prédéfinies (Tableau 1) aux îlots urbains construits automatiquement à partir du réseau de communication. Cette étape de classification est une étape préliminaire nécessaire à l'analyse des évolutions proprement dite. De la qualité des résultats vont dépendre, la précision de l'analyse des évolutions. Cette classification effectuée par apprentissage supervisé se décompose en 3 étapes. Elle débute par la création d'une base d'exemples d'îlots urbains et par la construction d'un modèle de classification. Ce modèle est ensuite appliqué à l'ensemble de la base de données îlots. Le résultat est ensuite validé de manière quantitative et qualitative. La figure 2 schématise les différentes étapes ainsi que les 'modules' logiciels utilisés.

La base d'îlots urbains (destinés à l'apprentissage et à la validation) est créée à partir d'un module d'étiquetage open source fondé sur Géoxygène (Lesbegueries *et al.* 2009). Ce module permet à l'expert d'étiqueter manuellement des îlots urbains représentatifs d'une classe thématique (Tableau 1). L'analyse réalisée par l'expert est fondée sur les seuls critères morphologiques issus des bases de données topographiques et ne considère que les formes des objets topographiques.

Lors de la classification, comme les données sont contenues dans une base de données relationnelle sous la forme de deux tables (une table pour les îlots et une table pour les bâtiments), l'algorithme TILDE permet de construire un arbre de décision impliquant :

- des conditions sur l'îlot, par exemple sa superficie,
- des fonctions d'agrégation sur les propriétés des bâtiments contenus dans l'îlot, par exemple la superficie moyenne des bâtiments,
- des contraintes sur le nombre de bâtiments ayant telles propriétés, par exemple qu'il y ait au moins 5 bâtiments de superficie inférieure à 150 m².

Le modèle obtenu est ensuite appliqué à l'ensemble des îlots non-étiquetés afin de prédire leurs classes. Les résultats de la classification sont stockés dans la table attributaire et peuvent être visualisés dans un SIG.

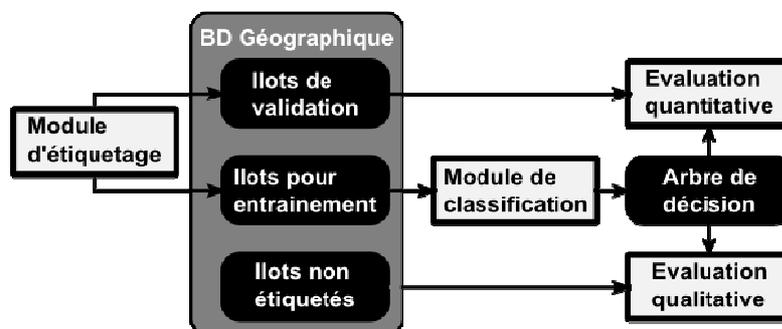


Figure 2. Etapes de la méthode de classification

Les résultats de la classification sont ensuite validés d'une part, qualitativement en analysant l'arbre de décision produit par TILDE et d'autre part quantitativement par validation croisée. Lorsque l'on dispose d'un seul jeu d'îlots étiquetés (apprentissage et entraînements confondus) (cf. test 1 – section 5), l'évaluation est faite par une validation croisée en 10 'paquets'. Cette procédure, standard en fouille de données, consiste à diviser le jeu de données en 10 sous-ensembles de données de taille semblable, en prenant soin que chaque groupe contienne un nombre d'îlot de chaque classe proportionnel au nombre présents dans le jeu complet (Witten et Frank, 2005). Le nombre de 'paquets', en général 5 ou 10, est un compromis entre le nombre d'exemples pour apprendre et le nombre de modèles à apprendre. Chaque paquet sert de jeu de validation, à son tour, et un modèle est construit à partir des 9 autres paquets. On construit donc 10 modèles, à partir d'un sous-ensemble des données, et on les évalue sur des données distinctes de celles ayant servi à construire chaque modèle. La moyenne des performances de ces modèles est un estimateur non-biaisé de la performance du modèle construit sur le jeu de données complet. Lorsque plusieurs jeux de données sont disponibles, par exemple les îlots des zones 1 à différentes dates (cf. tests 2 et 3 – section 5), un arbre de décision est construit à partir d'un jeu de données, par exemple les îlots de la zone 1 en 2002, puis ce modèle est évalué sur tous les autres jeux de données.

5. Tests de classification et analyse des résultats

La BDTopo@2002 est utilisée comme base de données de référence pour l'étape de création des îlots étiquetés (apprentissage et validation). Le premier objectif est d'identifier le type d'échantillonnage le plus approprié pour classer les îlots urbains sur l'ensemble de l'agglomération (section 5.1). Le second objectif est d'appliquer le modèle appris le plus pertinent, c'est-à-dire avec le taux de classement le plus élevé, à la base de données historique (section 5.2). Les résultats sont évalués quantitativement en calculant le taux de bon classement et qualitativement par l'interprétation de l'arbre de décision.

5.1. Résultats de la classification sur l'ensemble de l'agglomération

L'objectif est ici d'évaluer l'impact de la construction de la base d'exemples sur les résultats de classification. Deux tests de classifications sont appliqués à l'ensemble des îlots urbains de l'agglomération. La pertinence des règles produites sera évaluée uniquement sur le test fournissant le meilleur résultat de classement.

Le premier test de classification est effectué en créant une seule base d'exemple d'îlots urbains (Tableau 2). Ces îlots (apprentissage et validation confondus) sont sélectionnés par l'utilisateur comme morphologiquement représentatifs d'une classe d'îlots. Ils sont répartis sur l'ensemble de l'agglomération et représentent 20% du nombre d'îlots total. Le nombre d'exemples pour les six classes thématique comporte, pour ce test, au minimum 30 exemples par classes afin d'assurer la significativité statistique des mesures et des paramètres (Tableau 2). Pour les classes les plus représentées dans l'agglomération (habitat et tissu mixte), le nombre d'exemples est plus important.

Le second test de classification est effectué en utilisant une base d'exemples plus restreinte (10% du nombre total d'îlots) correspondant soit à la zone d'intérêt 1 (zone périurbaine - test 2a), soit à la zone d'intérêt 3 (zone en péricentre – test 2b). Le test 2c est effectué en tenant compte des exemples issus de ces deux zones d'intérêt (Tableau 3). L'objectif est également de vérifier si le nombre d'exemples influence le résultat de la classification.

L'analyse des résultats du taux de bon classement (Tableau 2) montre, comme on pouvait s'y attendre, que le nombre d'exemples influence le taux de bon classement puisque les tests 2a et 2b effectués à partir d'échantillons restreints sont plus faibles que les deux autres tests présentant un nombre d'échantillons plus important. Les tests effectués avec un plus grand échantillon obtiennent plus de 75% de bon classement. Ce résultat met en évidence que le choix et le nombre des échantillons doit être significativement représentatif des classes thématiques à identifier.

2002	Test 1	Test 2		
		Test 2a	Test 2b	Test 2c
Classes d'îlots urbains	Agglomération	Zone 1	Zone 3	Zone 1 et 3
Habitat pavillonnaire	40	56	22	78
Habitat collectif	39	21	21	42
Habitat discontinu mixte	71	15	16	31
Tissu discontinu mixte	60	14	7	21
Emprises spécialisées bâties	30	4	23	27
Emprises spécialisées peu ou pas bâties	34	11	25	36
<i>Nb total d'îlots*</i>	5940	133	201	354
<i>Taux global de bon classement</i>	78%	71%	63%	77%

Tableau 2. Récapitulatif du nombre d'îlots étiquetés (apprentissage et validation confondus) selon les tests effectués et les zones d'intérêts

* y compris les îlots correspondant à des réseaux de communication

L'analyse de la distribution des îlots urbains classés (Figure 3) combinée à l'interprétation visuelle (Figure 4), permettent toutefois de nuancer ce résultat. Sur l'ensemble de l'agglomération, le nombre d'îlots de type individuel et collectif est plus important à partir du test 2c. L'interprétation visuelle confirme ce résultat, plus proche. En effet, sur la figure 5, la majorité des îlots urbains de type individuel et collectif sont bien classés (par exemple au nord de la zone 1 – Figure 5). Le test 1 a plutôt tendance à surestimer les classes d'habitat et de tissu mixte. La surreprésentation de la classe d'emprise spécialisée peu ou pas bâtie était prévisible puisque aucun échantillon de cette classe n'a été pris en compte lors de la phase d'échantillonnage. Elle pourrait être corrigée en éliminant les îlots de cette classe selon un critère de forme combinée à un indice d'élongation, par exemple.

L'analyse des résultats de l'arbre de décision (obtenus à partir du modèle issu du test 2c) permet d'extraire 8 règles de caractérisation dont voici deux exemples :

- Règle 1 : Si densité îlot < 0,05 alors îlot urbain de type emprise spécialisée peu bâtie ;
- Règle 2 : Si densité îlot > 0,05 et nombre de bâtiments (dont la surface est inférieure à 105m²) est inférieur ou égal à 3 et que le nombre de bâtiments (dont la surface est inférieure à 185m²) est inférieur ou égal à 1 et la surface moyenne des bâtiments est inférieure à 1205 m², alors îlot urbain de type habitat collectif.

Deux règles pour les emprises spécialisées bâties ainsi que pour l'habitat mixte sont extraites. L'analyse détaillée de celles-ci si elle permet de repérer certains seuils (en terme de densité par exemple) mais n'a pas permis d'extraire une connaissance interprétable pour le géographe et susceptible d'être utilisée dans un modèle de simulation. D'autres tests sont en cours afin de vérifier si d'autres algorithmes tenant compte notamment du voisinage ou des relations spatiales permettraient d'extraire une connaissance réellement exploitable.

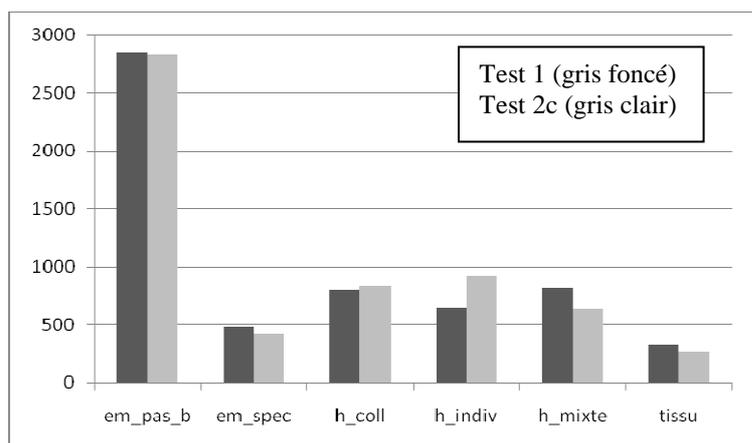


Figure 3. Répartition des classes thématiques sur l'ensemble de l'agglomération (5940 îlots urbains) selon le test 1 (gris foncé) et selon le test 2c (gris clair)

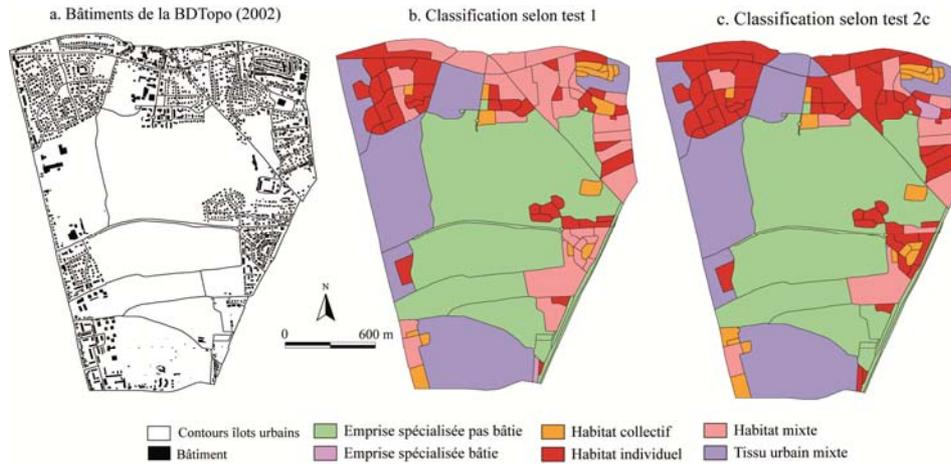


Figure 4. Comparaison des résultats de la classification issus des tests 1 et 2c (appliqué à l'ensemble de l'agglomération).

Le modèle construit à partir du test 2c le plus satisfaisant en terme quantitatif et qualitatif, a ensuite été appliqué à l'ensemble des données historiques sur les 4 zones d'intérêt (section 5.2). Les règles produites par l'arbre de décision sont également analysées.

5.2. Résultats de la classification sur la base de données historiques

Le modèle construit le plus satisfaisant (test 2c) est utilisé pour classer les îlots des bases de données historiques construites sur les zones 1 à 4 entre 1956 et 2008. Les résultats du taux de bon classement sont résumés au tableau 3. Les résultats de 2002 ne sont pas mentionnés pour les zones 1 et 3, puisque ces îlots ont contribué à construire le modèle appliqués aux autres dates.

Test 2c	Zone 1	Zone 2	Zone 3	Zone 4
2008	89	65	79	65
2002	/	63	/	58
1989	83	62	79	60
1976	84	71	76	56
1966	NA	NA	80	57
1956	79	93	83	53

NA – données non-acquises à cette date en raison de l'absence de photographies aériennes

Tableau 3. Taux de bon classement (%) par zone, selon les dates disponibles.

L'analyse ce tableau montre que le taux de bon classement aux dates antérieures à 2002 (et en 2008) est d'une manière générale proche de 80%. Ceci signifie que la composition d'un îlot urbain, par exemple, de type habitat individuel a peu varié dans le temps. Les différentes classes d'îlots définies initialement sont donc pertinentes pour l'analyse future de leur évolution. Deux exceptions cependant sont relevées : (1) un taux de bon classement plus faible (proche de 60%) sur les zones 2 et 4 et (2) un taux de bon particulièrement bon sur la zone 2 en 1956. La première anomalie s'explique par la spécificité de la morphologie urbaine de la zone 4, zone au contact du noyau urbain dense, qui a subi de nombreuses transformations durant les 50 dernières années en raison de la présence des instances européennes (Parlement européen, Conseil de l'Europe, etc.). L'interprétation visuelle de la morphologie de ces îlots dans le temps met en évidence de profondes restructurations (découpage, regroupement d'îlots). La seconde anomalie (93%) s'explique en partie par la taille de la zone (plus petite que les autres). En 1956, celle-ci est constituée d'un petit nombre d'îlots en raison de sa localisation en périphérie du noyau urbain (zone rurale en 1956).

La figure 5, qui présente les résultats de classification de la zone 1 dans le temps, permet de nuancer ces résultats d'un point de vue thématique. En effet, quelques anomalies de classement peuvent apparaître d'une date à l'autre, par exemple entre 2002 et 2008 (îlot central de grande dimension). Ce type de changement peut être dû à la manière dont les îlots sont découpés en fonction de l'interprétation du réseau de communication entre deux dates (chemin, par exemple).

6. Conclusion et perspectives

Dans cet article, une procédure de classification supervisée fondée sur un algorithme symbolique a été proposée afin de classer automatiquement les tissus urbains et d'extraire de la connaissance sur leur morphologie. Cette procédure utilise deux modules (étiquetage et classification) qui feront partie du module d'analyse spatiale proposé dans la plate-forme open-source GéopenSim. L'algorithme Tilde a permis de bien classer près de 80% des tissus urbains de l'ensemble de l'agglomération et des zones représentatives (en terme d'évolution) à différentes dates. Ceci signifie qu'il est pertinent d'analyser l'évolution des tissus urbains selon la typologie proposée. Toutefois, l'analyse des règles a mis en évidence la difficulté de caractériser de manière objective (mesures et paramètres statistiques) les tissus urbains. L'extraction d'une connaissance interprétable et exploitable dans un modèle de simulation n'est donc pas facilement automatisable. Ces règles permettent toutefois de relever des seuils pertinents, par exemple en termes de densité d'un îlot. Ces résultats pourraient être améliorés d'une part en créant des échantillons pour les classes plus spécifiques tels que le réseau de communication et d'autre part en utilisant d'autres algorithmes permettant de tenir compte des relations spatiales entre les bâtiments par exemple (Chelgoum *et al.*, 2002).

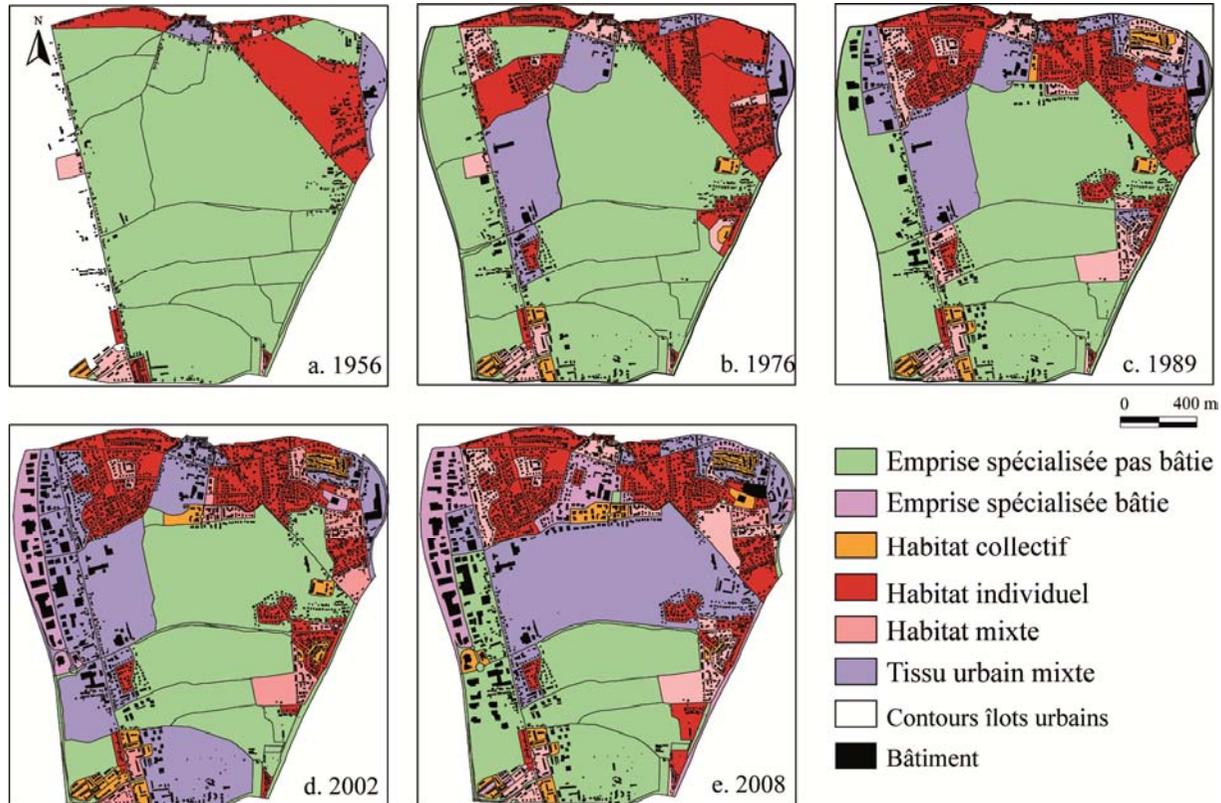


Figure 5. Classification des tissus urbains entre 1956 et 2008.

Ces premiers tests sur l'agglomération de Strasbourg ont également mis en évidence l'importance de l'échantillonnage. D'autres tests, appliquant le modèle le plus pertinent (Test 2c) appris sur Strasbourg, ont été appliqués à la ville d'Orléans (seconde villes de test dans le cadre du projet GéopenSim). Les premiers résultats ont montré des résultats de classification non concluant. On peut en déduire que ce type de procédure ne pourra être appliqué directement à d'autres agglomérations, et devra passer obligatoirement par une étape d'échantillonnage. Ceci reste cependant à confirmer par l'application du modèle de Strasbourg à une ville plus comparable (en termes de taille, de morphologie et de distribution des îlots urbains) comme par exemple Toulouse ou Montpellier.

6. Bibliographie

- ADEUS, 2003, « URBI et ORBI » L'étalement urbain : état des lieux, Dimension Villes & Territoires, n°39, 11 p.
- Baltasavias E.P., 2004, Object extraction and revision by image analysis using geodata and knowledge: current status and steps towards operational systems, *ISPRS Journal of Photogrammetry & Remote Sensing*, 58, 129-151.
- Bauer T., Steinnocher K. (2001): Per parcel land use classification in urban areas applying a rule-based technique. In: *GeoBIT/GIS* 6(2001), pp. 24-27.
- Bessy-Piétry P., 2000, Recensement de la population 1999 - Les formes de la croissance urbaine, *Insee Première*, n°701.
- Blaschke, T., Burnett, C., Pekkarinen, A., 2004. *New contextual approaches using image segmentation for object-based classification*. In: De Meer, F., de Jong, S. (Eds.), *Remote Sensing Image Analysis: Including the spatial domain*. Kluwer Academic Publishers, Dordrecht, pp. 211-236.
- Blockeel H., De Realt L., 1998, Top-down induction of first-order logical decision trees, *Artificial Intelligence*, vol. 101, n°1-2, 12p.
- Bondu A. and Lemaire V., 2007, Etat de l'art sur les méthodes statistiques d'apprentissage actif. RNTI, *Numéro spécial sur l'apprentissage et la fouille de données*, 2007.
- Bouziani M., Goïta K., Heb D., in press 2010, Automatic change detection of buildings in urban environment from very high spatial resolution images using existing geodatabase and prior knowledge, *ISPRS Journal of Photogrammetry and Remote Sensing*, in press, doi:10.1016/j.isprsjprs.2009.10.002
- Chelghoum N., Zeitouni K., Boulmakoul A., 2002, *Fouille de données spatiales par arbre de décision multi-thèmes*, EGC, 6 p.
- Curie F., Mas A., Perret J., Puissant A., Ruas A., s 2010, *Simuler la densification du tissu urbain au moyen d'un processus de peuplement*, SAGEO 2010, 15 p.
- Donnay, J.P., Barnsley, M.J., Longley, P.A., 2001. *Remote Sensing and Urban Analysis*. Taylor and Francis, London and New York., 315 p.
- Geomediterranée, 2004, BD OCS – Constitution d'une base de données 'occupation du sol' à partir d'images satellitaires. Rapport méthodologique, 34 p.

- Holland, D.A., Boyd, D.S., Marshall, P., 2006. Updating topographic mapping in Great Britain using imagery from high-resolution satellite sensors. *ISPRS Journal of Photogrammetry and Remote Sensing*, 60 (3), 212_223.
- IGN, 2002, Spécifications de contenu de la BDTOPO Pays, version 1.2., 65 p
- Jacquin, A., Misakova, L., Gay, M., 2008, A hybrid object-based classification approach for mapping urban sprawl in periurban environment. *Landscape and Urban Planning* 84 (2), 152-165.
- Kurtz C., Passat N., Gancarski P., Puissant A., in press 2010, Multiresolution region-based clustering for urban analysis, *International Journal of Remote Sensing*, 17 p.
- Lachiche N. and P. Flach. 2002: a true first-order bayesian classifier. In Springer-Verlag, editor, 12th International Conference on Inductive Logic Programming, pages 133–148, 2002.
- Lachiche N., 2010, Propositionalisation, In Encyclopedia of Machine Learning, Springer, 6p.
- Lesbueguerues J., Lachiche N., Braud A., Puissant A., Skupinski G., Perret J., 2009, A platform for spatial data labeling in an urban context, *International Opensource Geospatial Research Symposium*, 11 p.
- Nicot B., 1996, Une mesure de l'étalement urbain en France 1982-90, *Revue d'Economie Régionale et Urbaine*, n°1, pp. 71-98.
- Perret J., Ruas A., Boffet-Mas A., 2009, *Understanding urban dynamics: the use of vector topographic database and the creation of spatiotemporal databases*, International Cartographic Conference - ICC 2009, Santiago, Chile, November 2009, 11 p.
- Puissant A., Mas A., Skupinski G., Raimond A.-M., Perret J., Weber C., 2010, Méthodes de constitution de bases de données historiques pour le suivi spatiotemporel des tissus urbains, *Géomatique Expert*, Juillet-Aout 2010, 10 p.
- Quinlan J. R., 1993, C4.5 : Programs for machine Learning, Morgan Kaufmann.
- Saib A., 2009, Répondre à l'étalement urbain : construire un grand territoire métropolitain transfrontalier, *Dimension villes & territoires*, Adeus, n°63, 8 p.
- Salleb A. and Vrain C., 2000, *An Application of Association Rules Discovery to Geographic Information Systems*. In Springer-Verlag, editor, Proceedings of the 4th European Conference on Principles of Data Mining and Knowledge Discovery, pages 613 – 618.
- Witten I., Frank E., 2005, *Data mining: Practical machine learning tools and techniques*, second edition, Morgan Kaufmann, 50 p.
- Zeitouni K., Yeh L., 2000 Le data mining spatial et les bases de données spatiales, *Revue internationale de géomatique, Numéro spécial sur le Data mining spatial*, Vol 9, N°4.

Identification automatique de réseaux de voirie urbaine avec un algorithme d'optimisation par colonies de fourmis

Xavier Marsault*

* Laboratoire MAP-ARIA, FRE MAP 3315 du CNRS
ENSA de Lyon
3, rue Maurice Audin, F-69512 Vaulx en Velin cedex
xavier.marsault@aria.archi.fr

RÉSUMÉ. Le laboratoire MAP-ARIA développe – pour les phases de conception architecturale, urbaine et plus largement en aménagement du territoire – des dispositifs génératifs capables d’assister le concepteur ou le décideur dans l’élaboration de solutions formelles et fonctionnelles. Dans ce cadre, nous présentons ici une méthode générative innovante pour identifier des paramètres qualitatifs influant de manière générale sur une organisation de voirie, et calculer automatiquement des solutions optimisées. En partant de la seule donnée du plan du bâti urbain, nous parvenons à obtenir une hiérarchie émergente d’axes de voirie pour ces paramètres, qualitative (classification typologique: rocade, boulevards, avenues, rues et ruelles) et quantitative (caractéristiques du réseau). Nos recherches montrent qu’on peut se ramener à un problème de sélection de chemins, et un algorithme d’optimisation par colonies de fourmis permet alors de trouver de nombreuses solutions optimisées, cohérentes et pertinentes avec le choix des contraintes et des critères imposés.

ABSTRACT. At MAP-ARIA lab, we develop methods and tools for the generative design and assistance in the decision process in architecture and urban planning, that can assist the designer or decision maker in the development of formal and functional solutions. In this context, we present a new generative method for identifying qualitative parameters generally affecting a street network organization and automatically calculate optimized solutions. Only using urban building footprint data, we obtain emerging hierarchies of roads for these parameters, with qualitative (typology: ring roads / boulevards, avenues, streets and lanes) and quantitative (the network) aspects. Our research shows that we can reduce to a selecting paths problem, and an ant colony optimization algorithm can then find many optimized solutions, consistent and relevant to the selection criteria and constraints.

MOTS-CLÉS : réseaux de voirie, plus court chemin, optimisation par colonie de fourmis (ACO - SSP), classification et conception automatiques

KEYWORDS: street network, shortest path, ant colony optimization (ACO - SSP), automatic classification and design

1. Positionnement

L'analyse spatiale d'images aériennes ou satellitaires, à des fins de reconnaissance automatique d'éléments structurants du paysage urbain - comme par exemple le bâti et les axes de voirie - est un sujet de recherche vaste et passionnant. Ses applications sont multiples et variées en géographie, en urbanisme, en géomatique, pour le paysage, et ses enjeux plus actuels à l'heure de l'exploitation systématique à visée « tout automatique » de banques de données massives.

Or, il est souvent plus aisé de détecter dans ces images le bâti que la voirie (moins visible, partiellement occultée, aux contours hachés, et aux teintes moins tranchées). De plus, la thermographie urbaine est un procédé moderne qui donne accès directement à l'implantation du bâti, ce qui n'est pas le cas de la voirie. Par ailleurs, la plupart des problèmes d'ingénierie et de développement urbains utilisent des systèmes d'informations géographiques dans lesquels les plans numérisés (contours ou empreintes) du bâti sont un des éléments les plus significatifs. Enfin, même si pour des raisons historiques, la structure des rues a souvent imposé le découpage en îlots et contraint fortement le positionnement du bâti, des travaux de recherche récents en modélisation de villes virtuelles (Decoret *et al.*, 2002), (Marsault, 2005) se sont intéressés au problème inverse : comment retrouver des organisations plausibles de réseaux viaires en partant de la seule connaissance de l'implantation du bâti ?

Dans ce contexte caractérisé par deux logiques imbriquées, nous proposons dans cet article de montrer comment cette dernière voie de recherche a donné de bons résultats, et permis, pour un tissu urbain réel, d'approcher le réseau viaire existant et d'ouvrir des perspectives pour traiter d'autres problèmes de planification.

2. Précédents travaux

Dans la foulée de recherches menées en modélisation automatique de scènes urbaines avec les *Iterated Function Systems* (Marsault, 2005), nous nous sommes intéressés à un problème peu traité par la communauté scientifique : la génération automatique de voirie, de places et d'éléments de décor urbain, à partir d'images 2D½ représentant uniquement les empreintes au sol et la hauteur des bâtiments (en provenance du monde réel ou de calculs). Decoret est le premier chercheur à avoir proposé une méthode robuste permettant de calculer un réseau de rues à partir de contours du bâti vectorisés (Decoret *et al.*, 2002). Mais son approche, à visée purement géométrique, ne reposait sur aucune analyse typologique ou morphologique. Le sujet n'est même pas abordé dans (Larive *et al.*, 2004) et (Chen *et al.*, 2008), qui sont pourtant des articles de référence.

Nous avons proposé dès 2004 une approche par l'image, où les niveaux de gris peuvent coder différentes grandeurs comme l'empreinte, la hauteur ou la densité d'habitation. La correspondance pixel↔mètre, fixée par l'utilisateur pour chaque image, permet de tenir compte de l'échelle réelle des tissus analysés. Grâce à l'utilisation d'opérateurs provenant de la morphologie mathématique (Graffigne, 1995) et de la théorie des graphes (Lacomme, 2001), nous avons travaillé sur

l'extraction de caractéristiques structurantes du tissu urbain (Denis, 2004).

Les étapes fondamentales de ce travail étaient : la construction automatique d'enceintes plausibles d'un tissu urbain avec ou sans raccordement vers l'extérieur (systèmes de villes), et le calcul d'un graphe de voirie potentielle G (Figure 1) par squelettisation homotopique de l'espace inter-bâti (le sol) - qui est à l'image ce que le diagramme de Voronoï est à l'approche vectorielle (Decoret *et al.*, 2002). Nous avons aussi montré comment repérer automatiquement des cours d'eau d'une certaine importance qui traversent une ville de part en part. Mais notre plus forte contribution a été la détermination automatique des chemins et de leur type (résumée ci-après). Nous invitons le lecteur à consulter (Marsault, 2009) pour de plus amples détails.



Figure 1. a) tissu urbain test, b) enceinte, c) squelette morphologique homotopique.

Entre deux bâtiments proches, il existe toujours une jonction du graphe. Cette propriété du squelette morphologique (Fig. 1c) permet de cheminer dans ce graphe totalement connecté, entre deux points quelconques. Dorénavant, on appelle chemin $c(i,j)$ un groupe de jonctions consécutives entre deux nœuds i et j quelconques, de même largeur. L'ensemble des jonctions formant le chemin $c(i,j)$ adoptent ainsi pour largeur la plus faible largeur des composants du chemin. Ce nivellement au minimum a un sens physique fort issu du caractère homogène d'un tracé.

Nous avons mis au point une heuristique efficace pour déterminer des chemins plausibles, basée sur l'algorithme de plus court chemin de Dijkstra (Cormen, 2004), dont on a modifié la fonction de valuation « piéton » des jonctions. En prenant en compte la largeur maximale d'un chemin, sa longueur et sa rectitude, cette nouvelle valuation favorise le passage par les endroits où il y a le plus d'espace, avec le moins de méandres possibles (Fig. 2). Cela donne, pour une jonction j (longueur L_j , largeur R_j , rectitude $rect_j$) la valuation V_j [1], dépendant d'un paramètre α arbitraire :

$$V_j = L_j \cdot rect_j \cdot (\max \{ R_i \} + 1 - R_j)^\alpha \quad \text{avec } 0 \leq \alpha \leq 3 \quad [1]$$

Pour favoriser l'émergence d'une ceinture de boulevards et d'avenues dans le plan de voirie recherché, il suffit de choisir la distance qui sépare les jonctions d'enceinte du bâti suffisamment élevée pour privilégier leur utilisation dans l'heuristique [1].

On a alors recherché une méthode s'appuyant sur la robustesse de cette technique et l'exploration judicieuse du graphe G pour effectuer la reconnaissance automatique d'un réseau hiérarchisé de voirie. Une solution est apparue en découvrant l'intérêt de

l'algorithme d'optimisation par colonies de fourmis *Ant-SSP* (section 4). Avant d'entrer dans les détails (section 5), nous présentons la méthode mise au point.



Figure 2. Chemins dans le réseau viaire potentiel (rouge = Dijkstra ; vert = V_j).

3. Méthode

Le problème consiste à trouver un ensemble C de chemins de G constituant un plan de voirie optimal - c'est-à-dire maximisant un certain nombre de critères tout en satisfaisant un certain nombre de contraintes - et de les hiérarchiser en les classant par type de voie. L'analyse morphologique de la voirie urbaine (Mangin *et al.*, 1999) permet de dégager 4 critères caractéristiques, que l'on va combiner au sein d'un processus d'optimisation pour trouver « en moyenne » des chemins parmi les plus longs, les plus larges, les plus variés aux intersections et les moins tortueux possibles. On souhaite aussi que leur nombre soit plutôt faible, sans pour autant à avoir à le minimiser (ce qui n'aurait pas de sens physique).

Une typologie de voirie (rocade/boulevard, avenue, rue, ruelle) basée sur ces critères permet de classer chaque élément potentiel (Figure 3). En effet, à partir de la technique de cheminement exposée en section 2, il est simple de calculer pour un chemin donné ses caractéristiques morphologiques et typologiques en fonction de sa longueur L [3], sa largeur R [4] et sa rectitude $Rect$ [5]. La donnée d'une table de seuils d'admissibilité (Mangin *et al.*, 1999) permet d'obtenir à partir de ces valeurs la classification souhaitée. Pour certains éléments dont le rapport L/R est inférieur au seuil typologique fixé par la largeur R , on va ajuster R pour coller au type inférieur ou égal. Cette correction est essentielle pour ne pas perdre d'éléments indispensables. Enfin, bien que privilégiant les axes les moins tortueux possibles, définir un seuil de rectitude n'est pas souhaitable pour les boulevards, les rocades et les fleuves, naturellement courbes. Ainsi, on adapte les seuils par type de chemin.

Désormais, le problème n'est plus de calculer des chemins (section 2), mais d'en sélectionner un sous-ensemble parmi tous les possibles, pour former un réseau d'axes possédant une typologie, optimisant des critères, et obéissant à des contraintes (section 5). Mais quels chemins choisir, sachant qu'on ne dispose pas de bons critères locaux pour positionner leurs extrémités ? Nous voilà ramenés à un problème d'optimisation fortement combinatoire. D'emblée, soulignons que les

algorithmes génétiques (Renders, 2002) semblent peu adaptés ici, car la construction d'une solution, très contrainte, ne supporte pas de multiples divisions arbitraires. On aura donc recours à un procédé de construction incrémental, et l'algorithme *Ant-SSP* (section 4.3 et Figure 6) fournit une solution robuste et élégante à ce problème.

Type d'élément	min(L)	min(R)	max(R)	min(L/R)	Max(Rect)
ruelle	70 m	3,5 m	5,5 m	20	3
rue	110 m	5,5 m	20 m	20	2,5
avenue	600 m	20 m	80 m	30	1,4
boulevard/rocade	720 m	24 m	80 m	30	/
fleuve	diamètre (ville)	80 m	400 m	/	/

Figure 3. Paramètres typologiques de la voirie issus de mesures sur des villes.

Lors d'une étape de présélection, on calcule et l'on stocke (en mémoire et dans un fichier XML) pour chaque chemin admissible de G : sa longueur, sa largeur, sa rectitude, son type et la liste de ses noeuds N et jonctions J . Cela permet d'éliminer d'emblée un grand nombre par l'application des seuils (Figure 3), et de diminuer considérablement le nombre de chemins admissibles pour *SSP*, afin d'accélérer le processus de recherche de solutions.

4. Optimisation par colonies de fourmis (ACO)

4.1. Optimisation combinatoire

De nombreux problèmes de conception ou de décision - généralement multi-acteurs et multi-critères - sont traduits sous forme de programmes mêlant des phases de génération et d'optimisation combinatoire, notamment en synthèse de formes (Kicinger, 2006). Ce couplage, destiné à réduire drastiquement l'espace des solutions explorables, doit fournir un nombre de solutions optimales suffisamment faible pour être étudiées en pratique. Or, les techniques de l'optimisation exacte s'appliquent peu dans le champ de l'ingénierie conceptuelle ou décisionnelle, et on doit recourir à des heuristiques, voire des méta-heuristiques, qui calculent des solutions optimisées (approchant l'optimum) en temps raisonnable. Comme elles sont généralement assez efficaces, on se contente d'en parfaire le fonctionnement pour le problème traité. Inspirés par des systèmes naturels, les algorithmes génétiques (biologie de l'évolution), et les algorithmes de colonies de fourmis (éthologie) sont de plus en plus performants pour traiter des problèmes dans de nombreux domaines fort éloignés de la nature, comme le rappelle (Baquiast, 2006) : « une fois que la science commence à comprendre la façon dont la nature a sélectionné certaines formes et pas d'autres, aussi bien dans le monde physique que dans le monde biologique, elle s'applique à transposer les mécanismes correspondants en vue de résoudre des problèmes intéressant la fabrication d'artefacts, outils ou objets finaux ».

4.2. La méta-heuristique ACO (*Ant Colony Optimization*)

Dans la nature, les fourmis se déplacent et se dirigent en déposant des phéromones, substances olfactives volatiles qui jouent le rôle de mémoire collective. Ce concept forme la base de la méta-heuristique ACO (Dorigo *et al.*, 2004) utilisée en optimisation combinatoire (Figure 4), simulant sur plusieurs lancers (N_{run}) la coopération stigmergique entre insectes sociaux (fourmis, termites, guêpes). La stigmergie est une méthode de communication indirecte dans un environnement émergent auto-organisé, où le travail individuel est stimulé par l'activité antérieure de l'ensemble. Elle s'applique parfaitement à la recherche de cheminement ou d'ordonnancement dans un graphe, en utilisant des traces de phéromones virtuelles pour marquer les arêtes qui favorisent les meilleures solutions.

Choisir un nombre de runs N_{run} et n fourmis. Pour chaque run :

- initialiser** les traces de phéromones

- Jusqu'à atteindre un **critère d'arrêt (nombre de cycles N_c)** :

 - construction** de solutions par les n fourmis,
 - amélioration** des solutions par recherche locale (optionnelle),
 - sauvegarde** des meilleures solutions,
 - mise à jour** des traces de phéromones pour ces solutions.

Analyser les meilleures solutions trouvées

Figure 4. *Algorithme ACO générique.*

4.3. L'algorithme *Ant-SSP*

D'autres problèmes se ramènent plutôt à des processus de sélection, et non d'ordonnancement de composants. Le but est alors de trouver un sous ensemble d'objets dans un lot prédéfini, satisfaisant certaines contraintes et optimisant une ou plusieurs fonctions objectifs (ou critères). On appelle ces problèmes *SSP* (Subset Selection Problems), et le plus connu est le remplissage optimal du sac à dos (Cormen, 2004). Une adaptation d'*ACO* générique, *Ant-SSP*, décrit dans (Solnon, 2008), semble bien convenir à notre cas puisqu'il permet d'assembler incrémentalement des chemins dont l'ordre d'introduction dans la solution n'a a priori pas d'importance. Cette caractéristique permet aussi d'adapter le choix des chemins de manière à satisfaire les contraintes de consistance (section 5).

4.4. Résolution de problèmes multi-critères avec *ACO*

En présence de m objectifs partiels f_i qui ne sont pas complètement orthogonaux (indépendants), on peut définir un mono-objectif mixte par combinaison algébrique des f_i , souvent par pondération linéaire [2] avec des coefficients α_i positifs ou négatifs (car il y a des critères à maximiser et d'autres à minimiser).

$$\text{objectif_mixte} = \sum (\alpha_i \cdot f_i) \text{ (combinaison linéaire)} \quad [2]$$

Mais la plupart du temps, on se sait pas quantifier les poids α_i de chaque objectif, car ils sont de nature différente. On peut l'éviter en considérant une vraie

optimisation multi-objectifs. La difficulté est qu'il n'y a plus une « meilleure » solution correspondant à un seul objectif chiffré, mais un ensemble de solutions avec m objectifs chiffrés, qu'il s'agit de comparer (Agrawal *et al.*, 2000). Dans ce cas, on va calculer le « front de Pareto », qui est l'ensemble des solutions non dominées, c'est-à-dire au moins aussi bonnes que toutes les autres sur l'ensemble des objectifs, et meilleures sur un objectif au moins. Ces solutions, par définition, ne sont pas comparables entre elles : c'est à l'utilisateur, au final, de choisir, en faisant intervenir des critères non quantifiables. La mise en œuvre avec *ACO* ou *SSP* est décrite dans (Alaya *et al.*, 2007) et (Angus *et al.*, 2009). Elle associe en général sa propre trace de phéromones (ainsi qu'une fonction heuristique) à chaque objectif, pour tenter de l'optimiser indépendamment des autres. Pour cela, à chaque étape de la construction d'une solution, on ajoute un élément qui favorise aléatoirement l'un des m objectifs.

5. Détails de mise en oeuvre d'*Ant-SSP*

5.1. Critères à optimiser

L'objectif de maximiser la longueur globale L de tous les chemins c en même temps que leur largeur R et celui de minimiser leur rectitude sont déjà atteints au départ avec l'ensemble des jonctions-chemins de G . C'est donc en moyenne que l'on va calculer ces objectifs, pour faire émerger des solutions intéressantes, en pondérant chaque terme par la longueur du chemin, afin de tenir compte de l'étendue spatiale. C étant l'ensemble des chemins d'une solution et $l(k)$ la longueur d'une jonction k , on définit quatre critères simples :

- la longueur moyenne des chemins :

$$L(C) = \frac{\sum_{c \in C} L(c)}{|C|} \quad \text{avec} \quad l(c) = \sum_{k \in c} l(k) \quad [3]$$

- leur largeur moyenne :

$$R(C) = \frac{\sum_{c \in C} R(c) \cdot L(c)}{\sum_{c \in C} L(c)} \quad \text{avec} \quad r(c) = \min_{k \in c} l(k) \quad [4]$$

- leur rectitude moyenne :

$$Rect(C) = \frac{\sum_{c \in C} Rect(c) \cdot L(c)}{\sum_{c \in C} L(c)} \quad \text{avec} \quad rect(c(i, j)) = \frac{l(c)}{dist(i, j)} \quad [5]$$

- l'écart moyen $E(C)$ des largeurs aux intersections : ce paramètre à maximiser est calculé à partir de chaque noeud d'intersection de chemins et de chaque noeud terminal libre de la solution (cf. §5.4) :

$$E(C) = \sum_{(i, j) \in Sk \times Sk \mid ci \cap cj} |r(ci) - r(cj)| \quad [6]$$

La nature même de l'algorithme *Ant-SSP* évite d'introduire un critère de densité de voirie, car les solutions qui émergent optimisent naturellement le nombre de chemins pour un critère donné. En effet, tant que le réservoir de chemins admissibles n'est pas vide, on en choisit un élément via un critère probabiliste (Fig 6). Ce sont les

filtres initiaux (Fig 3) qui aident à contrôler la densité du réseau. On pourrait aussi introduire des objectifs optionnels visant à satisfaire par exemple un ratio de répartition des rues, une loi d'échelle. Nous avons laissé cela pour l'instant de côté.

On a conçu et testé 36 critères variés (Figure 5) à partir de ces 4 critères partiels, et en cohérence avec l'heuristique [1] pour certains : produits globaux, sommes de critères simples (avec normalisation, car on additionne des valeurs de nature différente) ou somme de produits de critères sur les éléments de la solution. A noter la nature purement géométrique et morphologique de ces choix, justifiée dans une première approche où seule l'empreinte au sol du bâti est vue comme signifiante.

5.2. Contraintes à satisfaire

La morphologie urbaine habituelle nous a déjà fait introduire trois contraintes d'admission, sous forme de filtres (Figure 3) sur la longueur, la largeur et le rapport L/R qui doit être supérieur à un seuil (plus une voie est large, plus elle est longue). La construction incrémentale d'une solution par *SSP* ajoute trois contraintes supplémentaires, dites « de consistance » :

- aucune jonction de la solution n'est commune à plusieurs chemins (principe de non superposition, même partielle).

- toute solution forme un sous-graphe connexe de G (possibilité de se déplacer n'importe où dans la ville). Pour ce faire, il suffit que tout chemin ajouté à chaque étape ait au moins un nœud d'extrémité (connexité faible) ou un nœud quelconque (connexité forte) en commun avec la solution en cours S_k . On obtient de meilleures solutions avec la seconde option, mais avec des temps de calculs réhaussés de 10%.

- minimisation du nombre d'impasses T_c (chemins libres à une seule extrémité, en partie générés par le processus de construction), tout en s'efforçant de garder les impasses naturelles du tissu urbain. Mais la satisfaction de cette contrainte à chaque étape est impossible car tout chemin ajouté à la solution ne peut pas toujours s'appuyer sur 2 nœuds déjà utilisés. Nous proposons de ne plus la considérer comme une contrainte, mais comme un critère d'optimisation propre à la construction

C	-C	L	-L	E	-E
R	-R	-Rect	Σl	$-\Sigma l$	$\Sigma r^a l$
$-\Sigma r^a l$	$\Sigma(r^a l/rect)$	$-\Sigma(r^a l/rect)$	$\Sigma(r^a l)/E$	$\Sigma(r^a l).E$	$\Sigma(r^a l)/Rect$
L+R-Rect	$LR^a/Rect$	R/Rect	$\Sigma(r^a l/rect)/E$	$\Sigma(r^a l/rect).E$	$L.R^a.Rect$
(R/Rect).E	$L.R^a.E/Rect$	R.E	-Rect.E	L.E	L+R-Rect+E
$L.R^a$	$L.R^a.E$	L+R	L+R+E	$\Sigma(r^a l).E/Rect$	$L.R^a/Rect+E$

Figure 5. 36 critères testés pour l'optimisation avec Ant-SSP.

5.3. Heuristiques locales

Le but des facteurs heuristiques h est d'évaluer l'intérêt de retenir des chemins,

connaissant la solution partielle S_k déjà construite. Ils concernent à la fois l'amélioration de la construction de la solution (h_c) et sa qualité (h_Q). Ainsi, l'heuristique finale $h = h_Q$. h_c est utilisée conjointement à une trace phéromonale pour définir la probabilité de sélectionner un chemin [8]. De plus, si les heuristiques statiques sont préstockées avec chaque chemin pour plus d'efficacité, l'évaluation des heuristiques dynamiques doit être faite le plus rapidement possible, puisqu'elle intervient des millions de fois durant la recherche. La définition de ces facteurs n'est pas toujours aisée, et demande d'effectuer de nombreux tests. Par exemple : $h_Q(L) = L$; $h_Q(R) = R$; $h_Q(E) = L / R$; $h_Q(LR, Rect) = LR / Rect$, sont des heuristiques de qualité statiques et performantes.

Une heuristique de construction h_c est utilisée lors de la mise en oeuvre de *Ant-SSP* pour résoudre progressivement l'objectif du minimum d'impasses T_c . Comme il faut placer 2 noeuds i et j et que l'un d'eux appartient à S_k (connecté), il y a 6 possibilités d'ajout d'un nouveau chemin à chaque étape, selon qu'on distingue les noeuds de terminaison libre (T , valence 1), interne à un chemin ou terminaison non libre (I , valence 2) ou externe à S_k (E , valence 0) : (TT), (TI), (TE), (II), (IE) et (EE). Le nombre total de chemins admissibles $C(k)$ se décompose, respectivement à ces 6 choix, en : $C_{TT}(k)$, $C_{TI}(k)$, $C_{TE}(k)$, $C_{II}(k)$, $C_{IE}(k)$ et $C_{EE}(k)$. On désigne par $\delta T_c(k)$ la variation de T_c à la fin de l'étape k . L'algorithme *Ant-SSP* s'exécute tant que $C(k)$ est non nul, et l'arbitrage dynamique de ces 6 possibilités a pour but de faire décroître T_c jusqu'à atteindre le minimum. Chaque choix influe de manière différente sur $T_c(k)$, mais seuls quatre d'entre eux réclament une régulation, car avec (EE) $\delta T_c = +1$, (IE) $\delta T_c = +1$, avec (TI) $\delta T_c = -1$ et avec (TT) $\delta T_c = -2$. La probabilité d'utiliser (TT) doit être plus faible que celle d'utiliser (TI), car ce choix n'explore pas de nouveaux noeuds. L'usage de (IE) ou (EE) doit bien sûr être limité, car il s'oppose à la décroissance de T_c . Nos expérimentations ont conduit à construire pour chaque cas les fonctions heuristiques h_c [7] avec d'excellents résultats.

$$\begin{aligned} h_c(TT) &= C(k) / C_{TT}(k) & \text{et} & \quad h_c(TI) &= C(k) / C_{TI}(k) \\ h_c(IE) &= C_{IE}(k) / (C(k).T_c(k)) & \text{et} & \quad h_c(EE) &= C_{EE}(k) / (C(k).T_c(k)) \quad [7] \\ h_c(\text{autre}) &= 1 \end{aligned}$$

5.4. Mise en oeuvre d'*Ant-SSP* – Construction des solutions

Deux modélisations phéromonales (vertex, clique) sont généralement proposées (Solnon, 2008) : soit associer à chaque composant i une trace t_i qui représente l'expérience passée de la colonie concernant l'intérêt de sélectionner l'objet i , soit associer à chaque paire de composants (i, j) une trace t_{ij} représentant l'intérêt de sélectionner i et j dans le même ensemble. Nous n'avons utilisé que la structure de type vertex, car l'ordre d'inclusion des chemins dans la solution n'a a priori aucune importance. De plus, la structure de type clique, plus consommatrice en temps de calcul, n'est pas plus efficace si les temps de calcul doivent être limités.

L'évaluation de chaque objectif lors de la construction est incrémental et très rapide : par exemple, quand on choisit un chemin candidat c , on vérifie rapidement qu'aucune jonction de c n'est déjà utilisée dans S_k , en maintenant une liste de marqueurs. Il en est de même pour la vérification des contraintes.

Le principe de sélection élitiste de *SSP* exige de faire concourir de nombreuses fourmis avant de mettre à jour les phéromones, seulement pour les meilleures fourmis du cycle, avec un dépôt inversement proportionnel à l'écart entre la combinaison construite et la meilleure combinaison obtenue depuis le départ [9]. Par rapport à une utilisation classique d'*ACO*, nous avons aussi constaté qu'il est préférable de récompenser à chaque cycle les meilleures solutions, avec une faible tolérance ($\epsilon=0,01$) par rapport à l'optimal du cycle. La vitesse de convergence est légèrement plus lente, mais les objectifs sont nettement supérieurs, et comme en plus ils sont atteints plus tôt, on retiendra cette astucieuse méthode.

```

Initialiser les facteurs phéromonaux à tmax
Sbest = ∅, Fbest = -∞

Tant que la condition d'arrêt n'est pas atteinte (nombre de cycles ou objectif seuil)
{
    // lancer un cycle d'optimisation (ici: maximisation)
    Sbest_cycle = ∅, Fbest_cycle = -∞
    pour chaque fourmi fk ∈ (f1...fn), construire une solution admissible Sk:
    {
        Sk = ∅
        candidats = { Oj ∈ S / Oj consistant dans S }

        tant que ( candidats ≠ ∅ )
        {
            mise à jour des heuristiques dynamiques h pour tous les candidats
            sélectionner Oi ∈ candidats avec la probabilité:
            
$$p(O_i) = \frac{h(O_i)^a \cdot t(O_i)^b}{\sum_i h(O_i)^a \cdot t(O_i)^b} \quad [8]$$

            ajouter Oi à Sk
            enlever Oi de candidats
            enlever de candidats tout Oj non consistant avec Sk
        }

        évaluer la fonction objectif F(Sk)
        si F(Sk) ≥ Fbest faire Fbest = F(Sk) et Sbest = Sk
    }

    // mettre à jour les phéromones
    appliquer la stratégie d'évaporation à tous les facteurs phéromonaux ti:
    ti = ti * (1 - ρ)
    récompenser les N meilleures fourmis (tolérance ε par rapport à Fbest_cycle):
    ti = ti + 1 / (1 + Fbest - Fbest_cycle)
    appliquer la stratégie Min-Max sur les facteurs phéromonaux ti :
    clamp ( ti, tmin, tmax)

    si Fbest_cycle ≥ Fbest faire : Fbest = Fbest_cycle et Sbest = Sbest_cycle
}
Retenir la solution Sbest et son objectif Fbest
    
```

Figure 6. Un « run » de l'algorithme Ant-SSP mono-objectif

Pour de nombreux problèmes de graphe, on cale souvent le nombre n de fourmis par cycle sur le nombre N_d de noeuds du graphe, lequel est souvent modeste. Dans notre cas, il peut y avoir des milliers de noeuds, donc des millions de chemins admissibles (N_{paths}). Nous avons constaté expérimentalement que le nombre de fourmis peut être fixé à $k\sqrt{N_{paths}}$ (largement plus faible que N_d , grâce au filtrage initial), sans pénaliser la qualité des solutions, tout en rendant le calcul plus rapide. Cependant, l'ajustement de k débouche en pratique sur un nombre de fourmis n peu dépendant de N_{paths} (entre 70 et 150 fourmis par cycle). On obtient ainsi un bon compromis entre le nombre de cycles N_c et le nombre de fourmis par cycle n , dont le produit est proportionnel au temps de calcul, ce qui permet de lancer plusieurs « runs » de *SSP* et d'obtenir plus de combinaisons optimales par laps de temps. Ainsi, les réglages optimaux des paramètres de l'algorithme *Ant-SSP* affectés à notre problème de reconnaissance de voirie ont été les suivants :

$$t_{min} = 0.01, t_{max} = 5 (\in [4 ; 10]), \rho = 0.01 \text{ (d'après [Solnon, 2008])}$$

$$a = 4, b = 3 \text{ (ajustement optimal), } n = 100 \pm 10\%, N_c \in [1000 ; 30000], N_{run} = 12$$

5.5. Algorithme *Ant-SSP* incrémental mono-objectif

On note S l'ensemble des chemins admissibles de G , ($f_1 \dots f_n$) n fourmis qui construisent respectivement les solutions ($S_1 \dots S_n$) par cycle d'optimisation, S_{best} la meilleure solution et F_{best} sa note d'objectif. La figure 6 décrit le déroulement de *SSP* mono-objectif. A noter qu'on ne peut améliorer les solutions par recherche locale (Figure 4), qui consiste, à partir d'une solution consistante, à appliquer des modificateurs de voisinage (ex: déplacements de noeuds) respectant les contraintes. Or, les chemins étant pré-définis, il est impossible d'en modifier un élément. La prise en compte de l'objectif (T) se traduit simplement à la fin en soustrayant $T_c(k)/20$ au critère mixte évalué (figure 5), et cela fonctionne très bien.

6. Tests et résultats

Pour illustrer les précédents exposés théoriques, et analyser quelles solutions émergent naturellement dans des temps de calculs raisonnables, pour les critères/contraintes retenus, on a choisi un exemple simple avec 208 noeuds et 7190 chemins admissibles (issus du filtrage de 89253 éléments potentiels), et un tissu plus dense (portion de la ville de Décines, avec 2580 noeuds et 194584 chemins). On a testé les 36 paramètres morphologiques de la Figure 5, et le calcul du « front de Pareto » des critères 30 et 36 éclatés. La Figure 7 montre les meilleurs plans de voirie obtenus après 10000 cycles de calcul (il n'est pas possible de montrer ici tous les résultats, y compris des comparaisons avec le tissu réel). Les 8 critères suivants, la plupart mixtes, donnent de bons réseaux, bien structurés, bien connectés aux endroits où dans la réalité il existe des connections jugées pertinentes, et évitant des groupements de chemins aléatoires. Il s'agit des critères : R (7), $\Sigma r^{\alpha} l$ (12), $\Sigma(r^{\alpha} l) / Rect$ (18), $L.R^{\alpha} / Rect$ (20), $L.R^{\alpha}.E / Rect$ (26), $L+R-Rect+E$ (30), $\Sigma(r^{\alpha} l).E / Rect$ (35) et $LR^{\alpha} / Rect+E$ (36), visiblement en concordance avec l'heuristique [1], mais avec des nuances. Par exemple, les solutions qui émergent optimisent naturellement le nombre C de chemins pour un critère donné, mais (20,26)

diminuent vraiment C , tandis que (12,18,35,36) l'augmentent mais préservent mieux les alignements ou des parallélismes naturels des axes des lotissements. (7,36) apparaissent ici comme un compromis entre ces deux tendances, mais dont la qualité est plus discutable. Nos expérimentations montrent aussi que l'objectif de minimisation de la rectitude $Rect$ est essentiel dans 6 des 8 meilleurs critères, même si l'heuristique [1] la favorise déjà. De plus, la minimisation de E privilégie systématiquement les solutions avec des chemins peu larges et peu longs, au détriment des autres critères. Par contre, sa maximisation favorise le raccordement de chemins de largeurs très différentes, et agit comme un régulateur de longueur (26,30,35,36). On a donc 8 critères typo-morphologiques mixtes permettant de retrouver et de classifier des axes de voirie existants, avec une petite préférence pour (12,18,35,36), même s'ils diminuent la longueur moyenne des chemins en augmentant leur nombre, car les réseaux qu'ils donnent sont plus réguliers et plus proches de ceux observés dans la réalité.

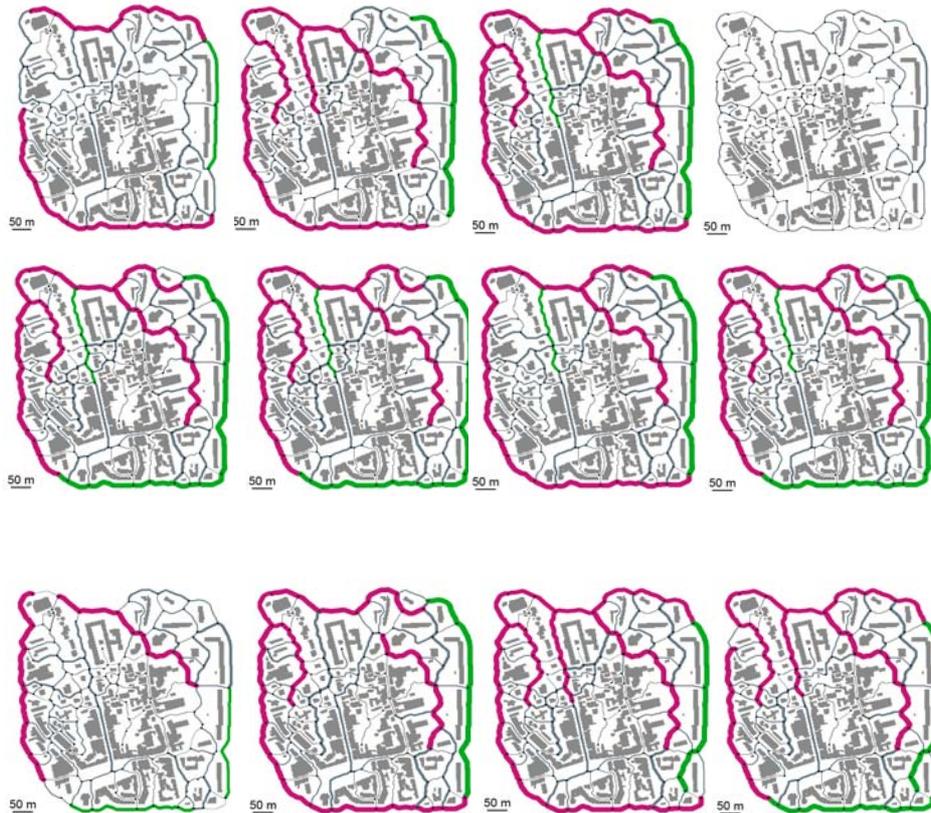


Figure 7. Optimisation des critères (2,5,7,9,12,18,20,28,26,30,35,36) sur un tissu test, portion de la ville de Saint-Genis. Meilleures solutions sur 10000 cycles de calcul – Légende: gris clair (bâti existant), rouge (boulevards ou rocade), vert (avenues), gris foncé (rues/ruelles) – rayons proportionnels aux largeurs des axes.



Figure 8. Optimisation des critères (18, 35, Pareto(30), Pareto(36)) sur une portion de la ville de Décines. Meilleures solutions sur 1000 cycles de calcul. Légende: gris clair (bâti existant), rouge (boulevards ou roades), vert (avenues), gris foncé (rues/ruelles) – rayons proportionnels aux largeurs des axes.

7. Conclusions et perspectives

Nous avons mis au point et testé une méthode générative originale pour tenter d'identifier expérimentalement des paramètres qualitatifs qui influent sur un problème difficile de reconnaissance de voirie urbaine. Elle fournit de bons résultats, encourageants, mais pas nécessairement une optimisation globale (la consolidation du choix des critères est loin d'être achevée). Qualitativement, on observe nettement sur un large tissu un découpage en îlots plausibles (Figure 8), ce qui traduit l'émergence d'un niveau morphologique supérieur à celui des seuls axes.

Le paramétrage du calcul des solutions optimisées s'est révélé assez long à mettre en oeuvre, et nous ne saurions conclure encore si l'optimisation de Pareto donne de meilleurs résultats que les bonnes pondérations de critères retenus (section 6). Curieusement, la contrainte de connexité peut émerger de la seule optimisation de certains critères mixtes (en cours d'analyse), au bout de plusieurs centaines de milliers de cycles. Son relâchement permet plus de liberté pour l'intersection des chemins, mais il est préférable de la maintenir pour accélérer la convergence.

L'optimisation poussée des codes C++ et un réglage fin des paramètres de l'algorithme *SSP* ont permis d'éviter des calculs superflus et contribué à diminuer massivement les temps de recherche. Notamment un bon compromis entre les nombres de cycles et de fourmis par cycle a été trouvé. Néanmoins, la complexité actuelle de notre algorithme en fonction du nombre de chemins c et du nombre de noeuds n demeure en $O(c.n^{1/2})$, impliquant des temps de calculs non négligeables pour des tissus urbains dont les noeuds du graphe se comptent en milliers. Sur un Intel QX6800, 3.16 Ghz, 10000 cycles de calcul prennent 18 minutes de calcul pour les résultats de la figure 7, alors que pour ceux de la figure 8, 15 heures sont requises pour 1000 cycles. Nous avons mis au point une version de l'algorithme dont la complexité est en $O(c)$, donc très rapide. Hélas, l'énorme quantité de mémoire nécessaire à son fonctionnement est de l'ordre de $c(c+1)/16$ octets, handicap qui la rend pour l'instant impossible à utiliser pour des tissus urbains de moyenne taille.

Il nous reste à analyser comment le paramètre α de l'heuristique [1] agit sur les solutions produites ($\alpha = 0.8$ semble meilleure pour l'échantillon présenté), et à chercher s'il existe une corrélation avec une mesure de la fractalité du réseau de rues. On améliorera aussi le calcul et le lissage des axes médians du squelette afin de prendre en compte des alignements plus réalistes de bâtiments et moduler la distance de l'axe à ces bâtiments avant de lancer les calculs d'optimisation. On effectuera enfin des tests sur des tissus urbains plus larges, ce qui nécessite d'accéder à plus de mémoire et certainement de paralléliser les recherches de solutions par les fourmis.

Notre démarche peut être transposée à d'autres problèmes basés sur l'exploitation d'un graphe spatial. En adaptant l'heuristique [1], le choix des critères et en utilisant les niveaux de gris de la carte des empreintes pour coder des données urbaines, sociales (ex: densité de population) ou environnementales (ex: nature du sol), transférables au graphe G , on peut envisager : d'améliorer un réseau viaire existant, de résoudre un problème économique (minimum de chemins déservant le maximum de lieux, asservis à la densité de population et à une fourniture énergétique...), ou encore de faire de la simulation de trafic sur un réseau.

Remerciements

A Christine Solnon, pour son accueil, sa gentillesse, ses conseils et sa patience. A Jean-Baptiste Denis pour les premiers codes C++ très structurés écrits en 2004, qui m'ont bien aidé à implémenter les algorithmes d'optimisation et de visualisation.

8. Bibliographie

- Agrawal S., Pratap A., Meyarivan T., A fast elitist non-dominated sorting genetic algorithm for multi-objective optimisation: NSGA-II, K. Deb, *Kangal report n° 200001*, 2000.
- Alaya I., Solnon C., Ghedira K., Ant Colony Optimization for Multi-objective Optimization Problems, *19th IEEE Int. Conf. on Tools with Art. Intel. (ICTAI)*, Patras, Greece, oct 2007.
- Angus D., Woodward C., Multiple objective ant colony optimisation, *Swarm Intelligence*, vol3, pp69-85, Springer, 2009.
- Baquiast J.P., La morphogénèse, <http://www.automatesintelligents.com/echanges/2004/jan/morphogenese.html>, 2006.
- Chen G, Esch G, Wonka P, Muller P, Zhang E, Interactive Procedural Street Modeling, *ACM Transactions on Graphics (TOG)*, Vol 27, Issue 3, Siggraph 2008.
- Cormen T, Leiserson E, Rivest R, Stein C, Introduction à l'algorithmique, Ed. Dunod, 2004.
- Decoret X, Sillion F, Street Generation for City Modelling, Architectural and Urban Ambient Environment, INRIA, 2002.
- Denis J.B., Calcul automatique de voiries et de places d'une ville représentée par l'image des empreintes de ses bâtiments, Magistère Informatique de Paris, MAP-ARIA, avril-juillet 2004.
- Dorigo M, Stützle T, Ant Colony Optimization, MIT Press, 2004.
- Graffigne C., Zerubia J., Masson, Analyse d'images : filtrage et segmentation, 1995.
- Kicinger R., Arciszewski T., Empirical analysis of memetic algorithms for conceptual design of steel structural systems in tall buildings, George Mason University, USA, 2006.
- Lacomme P, Prins C, Sevaux M, Algorithmes de graphes, Editions Eyrolles, 2001.
- Larive M, Gaildrat V, Génération Automatique de Zones Urbaines: Etat de l'Art, AFIG, 2004.
- Mangin D, Panerai P, Projet urbain, Editions parenthèses, 1999.
- Marsault X., Can Ants Build Urban Street Networks ?, *International Conference on Computers & Industrial Engineering (CIE'39)*, 6-8 juillet 2009, Troyes, France.
- Marsault X., Generation of textures and geometric pseudo-urban models with the aid of IFS, "Chaos in Art and Architecture", *Int. Journal of Dynamical System Research*, vol I, n°3, 2005.
- Renders J., Algorithmes génétiques et réseaux de neurones, Hermès, 2002.
- Solnon C., Optimisation par colonies de fourmis, Hermès, 2008.

**Session 4 : GeoWeb et cartographie
collaborative**

Cartographie et SIG à l'ère du Web 2.0

Vers une nouvelle génération de SIG participatifs

Boris Merickskay* — **Stéphane Roche****

* *Département de géographie*
Pavillon Abitibi Price, Université Laval
G1K7P4 Québec (Qc) Canada
boris.merickskay.1@ulaval.ca

** *Département des sciences géomatiques*
Pavillon Casault, Université Laval
G1K7P4 Québec (Qc) Canada
stephane.roche@scg.ulaval.ca

RÉSUMÉ. Avec le développement de l'Internet, la géomatique a profondément évolué. La convergence des technologies géospatiales et du Web se traduit aujourd'hui par l'émergence d'une nouvelle forme de cartographie reposant sur les techniques et les principes du web 2.0. Qualifiée de cartographie 2.0, elle se caractérise par une grande interactivité et des contenus géolocalisés générés par les utilisateurs. La mobilisation de ces technologies géospatiales orientées grand public tout comme l'information géographique volontaire dans le cadre de la participation publique constitue un nouvel enjeu pour les gestionnaires des territoires. Et plus largement, de nouvelles possibilités se dessinent pour les SIG participatifs, tant sur le plan des techniques que des méthodes. L'objectif de ce papier consiste précisément à mettre en perspective les potentialités participatives de la cartographie 2.0 avec nouvelles modalités de l'aménagement du territoire à l'heure débat public.

ABSTRACT. With the development of Internet, geomatics has greatly evolved. The convergence of geospatial technologies and Web is now resulting in the emergence of a new form of mapping based on techniques and principles of Web 2.0. Web-Mapping 2.0 is characterized by high interactivity and user-generated geographic content. The mobilization of geoweb technologies and volunteered geographic information in public participation represent a challenge for local authorities. And more broadly, new opportunities are emerging for PPGIS, both techniques and methods. The aim of this paper is precisely to put into perspective the potential of participatory of Web-Mapping 2.0 with the participatory planning.

MOTS-CLES: Géoweb, cartographie en ligne, Web 2.0, démocratie participative, information géographique volontaire, concertation territoriale, crowdsourcing

KEYWORDS: Geoweb, Web-Mapping, Web 2.0, Participatory Democracy, Volunteered Geographic Information, participatory planning, Crowdsourcing

Introduction

Traditionnellement réservée aux spécialistes, la géomatique s'est en quelques années largement démocratisée en évoluant au rythme de l'informatique et d'Internet. Caractérisée sous le terme de géoweb (Herring, 1994), la convergence du Web et des technologies géospatiales¹ est venue bouleverser la conception et l'usage de l'information géographique en offrant la possibilité au grand public d'interagir avec les cartes et les données. D'un côté, les technologies géospatiales se combinent dans une perspective de complémentarité (interopérabilité, systèmes de géolocalisation, etc.). De l'autre, les usages du Web 2.0² évoluent vers des formes plus matures de participation selon des logiques de partage de l'information et de travail collaboratif (Tapscott et Williams, 2007).

La mise en réseaux des bases de données, des applications et des usagers donne lieu à une grande diversité de pratiques tant au niveau de la consultation (calcul d'itinéraires, globes virtuels, services mobiles géolocalisés, etc.) que de la production de données géospatiales (cartographie personnelle, vectorisation des plans, géotagging, relevés GPS, etc.). Et dans la continuité du Web 2.0, un volume exponentiel de contenus géolocalisés émanant des utilisateurs a fait son apparition. Avec le développement de cette *cartographie 2.0* (Haklay *et al.*, 2008 ; Crampton, 2008), les logiques de production et de diffusion d'information géographique se transforment. Il serait illusoire de croire que cette évolution n'est pas dès à présent, porteuse d'enjeux sur la manière de lire et d'écrire (ensemble) la carte. Les concepts *d'information géographique volontaire (VGI)* et de *citoyens capteurs* (Goodchild, 2007) soulèvent toute une série de questions majeures sur les plans scientifique et sociétal (Jarnac, 2008 ; Sui, 2008 ; Goodchild, 2009, Mericskay et Roche, 2009).

Parallèlement à l'émergence de cette cartographie orientée grand public, le statut comme les formes de participation territoriale évoluent. En écho au développement de la démocratie participative, les modalités de la participation locale s'institutionnalisent. Et l'organisation de débats publics dans le cadre de la réalisation de projets ayant un impact sur le territoire s'impose comme désormais comme une étape obligatoire. Dans ce contexte, le recours à l'information géographique se généralise et son accès comme sa compréhension représente un nouvel enjeu démocratique. La cartographie dans sa forme 2.0 apparaît comme une des voies à explorer pour encourager la rencontre entre gouvernement et citoyens. D'une part, les techniques du géoweb sont de plus en plus utilisées par le grand public dans leur quotidien. D'autre part, l'utilisation d'informations géographiques volontaires dans des démarches formelles et encadrées se développe, notamment auprès des producteurs de données privés et institutionnels. Ceci pose de

¹ Les technologies géospatiales regroupent l'ensemble des outils permettant d'intégrer, d'analyser, de représenter et de diffuser des données géographiques (SIG, services Web cartographiques, CAO, DAO, télédétection, GPS, etc.).

² Le Web 2.0 renvoie à la seconde phase du Web, en termes d'architecture (plus flexibles), de contenus (générés par les usagers) et d'applications (dynamiques et interactives).

nombreuses questions quant à l'appropriation de ces contenus et techniques par les collectivités territoriales et les collectifs citoyens tant pour leurs missions d'intérêt public (diffusion d'informations et nouveaux supports de participation) que pour leurs besoins de fonctionnement (enrichissement et des bases de données).

C'est précisément sur cette problématique que se focalise le papier présenté ici, lequel fait un tour d'horizon des SIG participatifs à l'ère du géoweb et du Web 2.0. L'objectif étant d'expliquer et d'illustrer en quoi cette déclinaison grand public des SIG est en mesure d'encourager l'implication des citoyens dans la gestion territoriale. Pour ce faire, nous revenons préalablement sur les principes de la concertation territoriale en nous intéressant au rôle joué par l'information géographique et les technologies géospatiales (1). Nous présentons ensuite les principaux fondements du cadre socio-technique du géoweb du point de vue des technologies et des usages (2), afin de mettre en perspective les potentialités de la cartographie en ligne 2.0 pour les gestionnaires des territoires et les citoyens dans le cadre de la gouvernance territoriale (3).

1 L'aménagement du territoire à l'heure de la démocratie participative et des nouvelles technologies

La question de l'implication des citoyens est devenue en quelques années un élément constitutif de l'action publique en matière d'aménagement du territoire et d'urbanisme (Rey *et al.*, 2005 ; Gauthier *et al.*, 2008). Afin de répondre à ces nouvelles exigences, les collectivités locales mettent en place divers dispositifs participatifs mobilisant de plus en plus d'acteurs, d'informations et de technologies (consultation publique, conseil de quartier, concertation territoriale, enquête publique, atelier thématique, etc.). L'implication du public dans la prise de décisions territoriale quelque soit le niveau choisi (information, consultation, concertation) implique d'une part, la mise en place de politiques d'accès et de diffusion d'informations et d'autre part, le déploiement de dispositifs spécifiques d'accompagnement visant à impliquer les citoyens. L'enjeu informationnel est décisif, la revendication de la concertation porte d'ailleurs davantage sur le partage de l'information que sur le partage de la décision (Joliveau, 2001). Plus spécifiquement l'information géographique y est omniprésente sous diverses formes (cartes thématiques, croquis, photographies aériennes, vue 3D, etc.) (Roche, 2003). Elle permet de visualiser les impacts d'un projet d'aménagement sur l'espace, mais aussi de les expliquer par la localisation et description de la géométrie des objets et de leurs attributs. Dès lors, la constitution de bases de données localisées, accessibles et décrivant de façon approfondie et rigoureuse le territoire et le projet, apparaît comme un élément important pour favoriser la participation (CNIG, 2007).

L'information géographique est une ressource, mais elle est avant tout le produit de l'interprétation par les individus, de la signification de messages et de signes en lien avec l'espace qu'ils connaissent et pratiquent. La manière dont les acteurs appréhendent l'information géographique et son message influence leurs prises de

position et leurs décisions, ce qui peut avoir des conséquences non négligeables sur la compréhension d'un projet de planification en fonction des représentations spatiales mobilisées (Gumuchian, 1991 ; Lardon *et al.*, 2002 ; Debarbieux et Vanier, 2002). Pour mener à bien des procédures participatives, au-delà du partage de l'information, il paraît nécessaire de construire des représentations partagées du territoire en confrontant les points de vue de différents acteurs. Pour ce faire, la carte qu'elle que soit sa forme représente un objet performant en ayant la capacité d'associer prospective et citoyenneté (Debarbieux *et al.*, 2003 ; Roche et Caron, 2009). Sans porter par elle-même la neutralité, elle présente l'avantage de donner la possibilité à chacun de dire son mot à sa manière en offrant la possibilité de mettre sur le même plan, le présent, le prévisible et le souhaitable (Lévy *et al.*, 2004).

Plus spécifiquement, les technologies géospatiales représentent des instruments d'animation territoriale à part entière dans la mesure où elles permettent d'acquérir, de structurer, de gérer et de mettre à jour des données géographiques afin de produire et de véhiculer des représentations spatiales et des modèles (Caquard, 2001 ; Roche, 2003 ; Joliveau, 2004). Les processus participatifs basés sur le recours aux SIG se sont d'ailleurs multipliés depuis une vingtaine d'années (Craig *et al.*, 2002 ; Ramasubramanian, 2009). On peut définir les SIG participatifs³, comme des SIG développés en partie par (et pour) le public (des individus, des groupes locaux...), dans le but de favoriser la participation du public dans les processus de gestion territoriale locaux (collecte de données, opérations de cartographie, décision territoriale...). Théoriquement, les technologies géospatiales sont en mesure d'apporter une forte valeur ajoutée aux dispositifs de participation publique relatifs à la planification du territoire. Toutefois, il convient de ne pas tomber dans le déterminisme technologique et d'envisager ces technologies comme des dispositifs complémentaires aux autres canaux de participation, des leviers d'action au développement de la gouvernance territoriale (Blancher, 2004 ; Herbaut, 2007). Dans les faits, les technologies géospatiales sont davantage utilisées comme des outils techniques à usage interne qu'au niveau de la réflexion stratégique pour la conception de projets et la prospective. Le problème est avant tout d'ordre méthodologique, l'usage de l'information géographique et de la technologie en général n'est pas clairement défini ou tout simplement absolument pas envisagé (Péribois, 2008). Et dans la majorité des cas, les SIG participatifs mis en place se présentent comme des contre-projets politiques ou scientifiques (Joliveau, 2004). Se positionnant souvent à la marge des institutions et des processus décisionnels, ils manquent de soutien officiel et de légitimité, diminuant ainsi leur efficacité.

³ Le concept de *SIG participatifs* s'est développé au milieu des années 1990. Malgré le foisonnement d'idées et d'approches autour des *SIG participatifs*, le concept reste flou, ce qui engendre des difficultés à concevoir des méthodes de développement efficaces et à formaliser des critères d'évaluation efficaces. Il convient de différencier les *Participatory GIS (PGIS)* des *Public Participatory GIS (PPGIS)*, les premiers se destinant plus au développement local et à la gestion des ressources naturelles des pays du sud (contexte rural). À la différence des *PPGIS* qui concernent davantage les projets d'aménagement du territoire et d'urbanisme des pays du nord (Rambaldi *et al.*, 2005).

2 Le géoweb : une adaptation des SIG et de l'information géographique à l'environnement socio-technique du Web 2.0

2.1 Un environnement technique basé sur l'ouverture et le mélange

De manière corrélative avec les évolutions du réseau informatique Internet et du Web 2.0, la cartographie en ligne a évolué en termes de capacités de traitement, de diffusion et d'échange de données spatiales (Sample et al., 2007 ; Scharl et Tochtermann, 2007). Le géoweb est devenu une composante du Web et la carte par l'intermédiaire des hypercartes Web⁴ un navigateur du système hypertexte. La cartographie 2.0 repose sur plusieurs avancées technologiques relevant davantage du Web que de la géomatique (JAVA, SOAP, RSS, tag, etc.). Les technologies actuelles du Web offrent des architectures plus flexibles, une meilleure réactivité d'affichage, des interfaces plus conviviales, une plus grande interaction avec les contenus et une interopérabilité plus poussée. Le modèle émergent de « l'informatique dans les nuages » (*Cloud Computing*) permet le développement d'applications en ligne offrant des caractéristiques similaires aux logiciels installés sur un ordinateur (*software*). Basées sur le recours aux services Web, ces technologies permettent à des applications de dialoguer à distance indépendamment des plates-formes et des langages sur lesquelles elles reposent. Dans cette logique, les services Web géographiques actuels proposent un ensemble cohérent d'outils de manipulation de données spatiales (importation, catalogage, visualisation, création, traitement, documentation, diffusion, etc.).

C'est là toute la force du géoweb qui repose sur sa capacité à proposer au sein de l'environnement distribué du Web des applications géographiques composites mélangeant, selon la logique de *mashup*, différents services et contenus Web derrière une interface graphique unifiée. L'interopérabilité entre applications et données étant rendue possible par la standardisation des langages de dernière génération (XHTML, XML, PHP, JavaScript, etc.) et des formats d'échange spécifiques au géoweb (KML, CSV, GML, Land XML) qui s'inscrivent dans la logique des normes en matière de données géospatiales, en particulier celles de l'*OGC*⁵ (WMS, WFS, WCS, etc.). Les applications cartographiques composites reposent essentiellement sur l'utilisation des API (*Application Programming Interface*) comme celle de Google Maps, Bing Maps, ESRI ArcGIS ou celle du Géoportail de l'IGN. Ces interfaces de programmation permettent de recourir aux fonctions et contenus d'une application Web à partir de commandes externes en particulier au niveau de la visualisation de réseaux routiers, d'images satellitaires, de courbes de niveaux ou encore de nouveaux types de visualisations comme la trois

⁴ Les hypercartes Web désignent des applications cartographiques en ligne dynamiques et interactives associant des contenus hypermédias (hypertexte + multimédia) à des références géographiques (Google Maps, Google Earth, Bing Maps, Géoportail de l'IGN, etc.).

⁵ L'Open Geospatial Consortium a pour objectif de développer et promouvoir des standards ouverts dans le but de garantir l'interopérabilité des contenus, des services et des échanges.

dimensions ou les services photos réalistes. La production de contenus par les usagers s'appuie dans leur très grande majorité sur les « fonds de carte » fournis par les API. Ces dernières s'imposent comme les supports privilégiés du développement du géoweb et en constituent les nouveaux « référentiels cartographiques ». Au-delà de la consultation, les API fournissent également des utilitaires de manipulation et de création de données à l'image de la fonctionnalité *MyMaps* de Google. En complément des outils fournis par les API, les services d'éditions cartographiques comme [Zeemaps](#), [Umapper](#) ou [ScribbleMaps](#) (Figure 1) permettent la création de données, mais aussi l'importation et exportation de fichiers aux différents formats SIG traditionnels. D'autres services comme [Navx](#) ou [Listphile](#) permettent de gérer en ligne des bases de données géolocalisées. [Mapstraction](#), [Mapufecture](#) et [Map Channel](#) offrent quant à eux la possibilité de combiner au sein d'une même application plusieurs API cartographiques et différents modes de visualisation (fonds de cartes de différents services cartographiques, imagerie aérienne et satellitaire, trois dimensions, vues obliques, etc.).

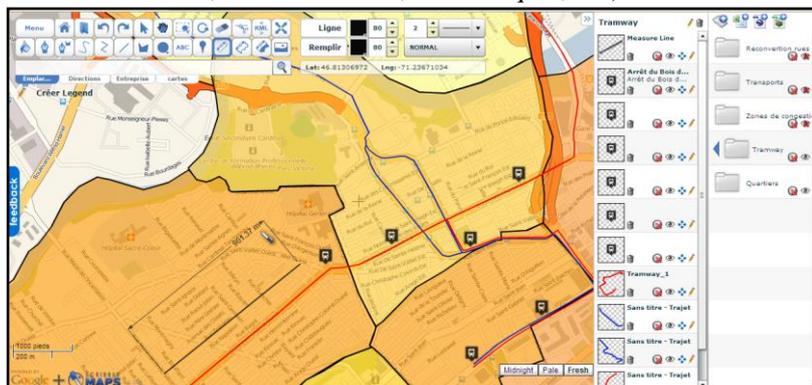


Figure 1. Interface de l'application de cartographie en ligne ScribbleMaps

2.2 De nouvelles modalités de production d'informations géographiques

En mobilisant les technologies 2.0, le géoweb devient un lieu participatif où l'utilisateur est appelé à créer des contenus, à les échanger et à les mélanger. Cette manière d'interagir avec les données géolocalisées vient modifier les modalités établies de production et de mises à jour de l'information géographique (Hudson Smith *et al.*, 2009). Le passage de la simple consultation d'informations à l'interaction avec les contenus permet la construction de réseaux ne se basant plus sur l'échange d'informations, mais sur le partage du savoir (Quoniam et Lucien, 2010). De fait, les formes, les thématiques comme le statut de l'information géographique évoluent tant au niveau de la composante graphique (personnalisation de l'iconographie), géométrique (possibilité de relocalisation) que descriptive (nouvelles modalités sémantiques de classification et diversification des contenus rattachés comme les commentaires, les photos ou les hyperliens). Plus globalement,

le renversement de paradigme dans la conception des données spatiales transforme la manière de produire les cartes. L'information géographique n'émane plus seulement de grands producteurs d'informations institutionnels et privés. Une partie est désormais produite, enrichie, mise à jour et diffusée par le grand public (communautés de pratiques, particuliers, citoyens) selon une logique ascendante (Budhathoki *et al.*, 2008 ; Coleman *et al.*, 2009). Du simple commentaire sur un restaurant à la production de référentiels en passant par la géolocalisation de photos, le grand public devient acteurs du géoweb en produisant et en enrichissant une grande quantité d'information. L'utilisation de l'information géographique volontaire par l'intermédiaire du *crowdsourcing* (externalisation par la foule) comme nouveau mode de production et de mise à jours des bases de données se généralise auprès des acteurs du géoweb (Google, Microsoft) et des producteurs institutionnels et privés (Tom-Tom, Télé Atlas, IGN, Ordnance Survey).

3 Quelles articulations entre participation territoriale et cartographie 2.0 ?

3.1 *Vers une nouvelle génération de SIG participatif basé sur les techniques du géoweb et du Web 2.0*

Comme nous venons de le développer, d'un côté l'information géographique est devenue une composante à part entière de la participation publique. De l'autre, la cartographie 2.0 offre aux citoyens une multitude d'outils pour lire et écrire la carte. Les applications cartographiques nouvelles générations à l'image de Google Maps ou Google Earth connaissent une forte popularité, favorisant ainsi l'interaction entre les citoyens et les données géographiques. En ce sens, il ressort que les solutions technologiques, mais aussi les approches et concepts, sur lesquels trouvent appui les SIG participatifs se sont multipliés au rythme du développement et de la démocratisation des outils du Web 2.0 et du géoweb. De fait, les SIG participatifs dont il faut bien avouer que le bilan demeure mitigé (Craig *et al.* 2002 ; Roche, 2003 ; Péribois, 2008) doivent être envisagés d'une manière renouvelée en prenant en compte les nouvelles fonctions de l'information géographique dans la société (Tulloch, 2008 ; Elwood, 2009). Car au-delà d'un artefact marketing, la cartographie 2.0 constitue un objet de réflexion pour les praticiens et les scientifiques de l'aménagement du territoire. Elle offre de réelles opportunités pour la mise en place de dispositifs participatifs visant à mieux impliquer les citoyens dans la prise de décision relative au territoire. D'une part, la multitude de services Web géographiques disponibles permet de supporter les fonctionnalités requises à la cartographie de planification (visualisation, création et système de gestion des données spatiales). Associé à la convergence croissante entre applications orientées grand public et logiciels SIG professionnels (applications, services et formats), les SIG sont en mesure de sortir des carcans professionnels et décisionnels pour toucher plus facilement le grand public. D'autre part, la mise en place de démarches visant à la mobilisation et la gestion de données (référentiels, thématiques, animation, etc.) est également renouvelée. La diffusion rendue possible par l'intermédiaire du réseau

Internet permet d'élargir le public concerné et d'élargir l'audience tant sur le plan quantitatif que qualitatif en intéressant des publics qui d'habitude ne s'impliquent que peu dans la vie locale (étudiant, jeunes actifs). Afin d'explicitier les potentialités participatives du géoweb, il convient de les différencier en fonction du degré d'implication des citoyens (information, consultation et concertation).

L'information consiste à « porter à connaissance », à donner des éléments de compréhension aux citoyens sur les projets liés à l'aménagement du territoire. Les cartes représentent des instruments d'information privilégiés, car elles permettent de synthétiser un ensemble d'indicateurs et de données. Mais dans la plupart des cas, les cartes mobilisées sont seulement accessibles par l'intermédiaire de rapports formalisés ou d'atlas cartographiques sous formes statiques. Néanmoins, les collectivités territoriales s'orientent de plus en plus vers la mise en place de portails cartographiques dynamiques et interactifs regroupant et structurant un ensemble cohérent de données relatif à un territoire. Mais ces applications restent souvent cantonnées à des fins de communication et les données disponibles y sont très généralistes (le zonage en vigueur n'étant généralement pas présent). En poussant la réflexion, il est pertinent d'envisager ce type d'application, mais dédiée à un projet propre, comme l'élaboration d'un SCOT ou d'un PLU par exemple. Elle pourrait par exemple centraliser et organiser divers documents relatifs au projet (rapport de présentation, PADD, document d'orientations générales, etc.), la réglementation, des données connexes (économiques, environnementales, démographiques, sociales, etc.) et une multitude de couches complémentaires (cadastre, plan de prévention des risques, zones inondables, anciens schémas directeurs, etc.) sous la forme d'une application cartographique en ligne. Cette perspective paraît particulièrement intéressante dans la mesure où les fonctions dynamiques, interactives et multimédias des hypercartes Web permettent une meilleure accessibilité aux données tout en améliorant leurs compréhensions (Caquard, 2001, Laurini, 2004).

À la suite de l'information, la consultation, processus à double sens a pour objectif de recueillir les avis des citoyens sur un projet donné. L'usage des cartes interactives à des fins consultatives se développe auprès des collectivités. L'objectif étant souvent de recueillir les retours et les propositions des citoyens sur des problématiques locales. La principale application de ces nouveaux services cartographiques consiste à alerter les services techniques pour solliciter leur intervention dans l'espace public (défauts de voiries et de signalisation, dégradations, mobiliers urbains inappropriés, éclairages défectueux, véhicules abandonnés, etc.). En d'autres termes, la carte interactive permet aux citoyens de signaler et de discuter des problèmes qu'ils rencontrent dans leur quartier. Ces observations localisées permettent l'enrichissement automatique des bases de données territoriales, offrant ainsi aux gestionnaires des territoires une prise plus directe avec la réalité. On peut citer comme exemples la [carte interactive](#) de la ville de Camden aux États-Unis, celle de la commune [d'Yverdon-les-Bains](#) en Suisse, celle de [Brisbane](#) en Australie ou encore l'application [FixMyStreet](#) utilisée par de

nombreuses villes britanniques et canadiennes (figure 2). Il ressort que la remontée, la formalisation et l'utilisation de ces informations provenant directement des citoyens sont facilitées grâce aux technologies du géoweb. En permettant : la mise en place d'applications cartographiques composites accessibles et simples d'utilisation ; la mise à jour automatique des bases de données par des architectures plus flexibles ; l'enrichissement des observations par des commentaires, des contenus multimédias (photos, vidéos) ; et le recours à des systèmes de classifications basées sur la sémantique du Web (catégories et tag).



Figure 2. Interface de l'application *FixMyStreet Canada*

Nous voyons dans la concertation un troisième niveau de participation où les citoyens en plus de consulter l'information et de fournir un retour pourraient participer à la production de contenus et de cartes. Le débat public est une occasion de discussions où se confrontent les points de vue et les connaissances des experts et des citoyens. Objets intermédiaires produits et échangés par des acteurs dans la durée d'un processus, les représentations spatiales mobilisées dans les procédures participatives sont dans la majorité des cas « fermées ». Elles ne laissent pas la possibilité de rajouter, de modifier ou de supprimer des éléments (Maurel, 2001). Il apparaît important de concevoir des moyens d'inclure aux représentations « officielles » de nouvelles informations s'appuyant sur l'expérience des citoyens. La connaissance qu'un individu ou un collectif possède de son environnement immédiat et quotidien, représente une source d'information exploitable et utile dans des processus de gestion et de planification territoriale. La réussite ou l'échec des SIG participatifs en matière d'apprentissage et d'appropriation dépend d'ailleurs largement de leur capacité à intégrer l'ensemble de ces connaissances (Ghose, 2004 ; Duncan et Lach, 2006). Décrites de manières subjectives et peu formalisées, elles diffèrent de celles des experts et des techniciens, mais sont en mesure de venir compléter et enrichir les données institutionnelles dans le but d'enrichir et de faire avancer le débat et le projet. Dès lors, il paraît important que les dispositifs participatifs mis en place dans le cadre de concertation territoriale encouragent, encadrent et formalisent la remontée d'informations provenant des citoyens et des collectifs (associations, conseils de quartiers, comités citoyens, etc.).

Les interfaces des SIG participatifs conventionnels étaient trop complexes pour les non-experts. Orientés vers des méthodes quantitatives alors que l'intégration, l'analyse et la représentation des connaissances locales reposent sur des méthodes qualitatives, ils n'étaient pas suffisamment interactifs pour supporter efficacement des processus participatifs (Mc Hugh *et al.*, 2009). Les évolutions techniques liées au géoweb et plus spécifiquement au Web 2.0 donnent naissance à une nouvelle génération de SIG participatifs plus accessibles et conviviaux favorisant l'interaction avec les données géographiques. La combinaison des fonctionnalités dynamiques et interactives associées aux nouvelles méthodes de gestion de l'information permet une formalisation et une organisation des connaissances locales plus efficace (prise en compte de l'incertitude des contributions et du profil des utilisateurs, discussion instantanée, argumentation spatiale, système de gestion des contenus, etc.). Toute une série d'expériences récentes ont ainsi démontré le potentiel du géoweb et de l'information géographique volontaire pour la mobilisation et la formalisation des connaissances locales (Seeger, 2008 ; Tulloch, 2008 ; Bugs *et al.*, 2009 ; De Longueville et Ostländer, 2009). Et les prototypes d'applications cartographiques à vocation décisionnelle se multiplient à l'image d'[Argoomaps](#) (Rinner *et al.*, 2008), de [GeoDeliberator](#) (Cai et Yu, 2009) ou de [MapChat](#) (Brent Hall *et al.*, 2010).

Les outils, les pratiques comme les formes d'informations géographiques ont évolué, mais la question de la qualité et de la confiance dans ces contenus demeure centrale. Les limites et difficultés liées à ce mode de production sont de tous ordres et ne sont bien entendu pas à ignorer (motivations variables des contributeurs, hétérogénéités des contenus, etc.). De plus, les applications cartographiques 2.0 existantes restent assez limitées tant au niveau de la précision que de la qualification des données : incapacité à assurer la traçabilité des objets géométriques et absence de métadonnées. D'où la nécessité de développer des applications cartographiques basées sur des méthodes de mobilisation et de formalisation des connaissances des citoyens renouvelées, à l'image de la géocollaboration ou de la wikification.

3.2 Géocollaboration et wikification, de nouveaux outils et de nouvelles méthodes pour la production de connaissance sur les territoires

La *géocollaboration* qui désigne les approches ou les situations collaboratives utilisant des données géographiques et des technologies géomatiques (MacEachren et Brewer, 2004 ; Pornon et Noucher, 2008) représente l'un des principaux axes de réflexion pour envisager de manière renouvelée la cartographie participative. Reposant sur l'interaction permanente des contenus et des usagers, ce type de situation permet de faire émerger de nouvelles connaissances localisées par le dialogue et la confrontation des savoirs, et ceux en s'affranchissant des contraintes de temps, de lieu, d'outils, et de procédures. Même s'il n'existe pas à proprement parler d'applications cartographiques collaboratives grand public mises en place par des collectivités à des fins de planification territoriale, on peut toutefois citer

l'exemple d'initiatives comme [OpenStreetMap](#)⁶ ou [Wikimapia](#)⁷ qui représentent les applications les plus abouties tant sur le plan des techniques mobilisées que des contenus produits (couverture, quantité et qualité). Dans la continuité de la géocollaboration, la *wikification* de l'information géographique renvoie au développement des applications cartographiques en ligne reposant sur un système de gestion des contenus (CMS) de type wiki (Sui, 2008). Les principes de base d'un wiki sont en rupture avec la production linéaire de l'information et le fonctionnement hiérarchique traditionnel. Ils adoptent une approche itérative et non cumulative de la production de connaissances et représentent de nouvelles pratiques de gestion de l'information. Deux fonctionnalités confèrent aux wikis géographiques une particularité collaborative avancée. D'une part, les contenus géolocalisés peuvent être modifiés, enrichis, mis à jour et supprimés par chaque utilisateur. D'autre part, toutes les contributions des utilisateurs concernant les objets et les attributs sont archivées et consultables par le biais d'un historique des actions (Ciobanu *et al.*, 2007 ; Guptill, 2007). Ce mode de production de contenus géographiques favorise une construction dynamique et collective des connaissances, tout en créant une transversalité entre les utilisateurs dans le but de partager et de produire de l'information spatiale de façon efficace et constructive.

Le prototype du [WikiSIG](#) développé au sein du Centre de recherche en géomatique de l'Université Laval illustre ce propos. Cette application cartographique collaborative basée sur architecture Web client/serveur classique s'appuie sur un système de gestion des contenus de type wiki. Côté client, l'interface utilisateur est construite autour d'une carte reposant sur l'API de Google Maps. Côté serveur, on distingue trois composantes : le serveur Google Maps, le serveur WikiSIG et la base de données WikiSIG. Les fonds de carte et les images satellites sont servis par le serveur Google Maps. Le stockage, et la traçabilité des composantes géométrique et attributaire de chacune des entités (points, polygones, polygones) constituant la carte est assurée par la base de données WikiSIG (de type « MySQL »). Ces entités sont servies au client, au format XML, par le serveur WikiSIG (de type « Apache/PHP »). Une des forces du WikiSIG par rapport aux autres applications existantes tient dans sa capacité à assurer efficacement la traçabilité documentée des composantes géométrique et descriptive des entités constituant la carte. Que ce soit un point, une polyligne, un polygone ou un commentaire, chaque action dans la base de données du WikiSIG (relocalisation des objets, modification de la géométrie, ajouts d'informations descriptives, etc.) est archivée, associée à un contributeur ainsi qu'à une date. L'archivage des actions permet de retracer l'évolution de la conception des objets de la carte en fonction des contributeurs à la manière d'un article de Wikipédia. Et la prise en compte du temps

⁶ OpenStreetMap est un projet de cartographie libre des réseaux routiers du monde. Toutes les contributions proviennent des utilisateurs (vectorisation, relevés GPS) et sont diffusées sous licence libre. Les données d'OSM sont de plus en plus utilisées par les professionnels.

⁷ Wikimapia est un service de cartographie collaborative ayant pour objectif de décrire la terre. Fusion entre Wikipédia et Google Maps, il regroupe actuellement plus d'un million d'utilisateur ayant créés treize millions de lieux (polygones associés à un wiki descriptif).

dans la conception collaborative des objets permet par exemple de visualiser l'évolution d'un tracé d'une ligne de bus ou d'un périmètre de manière temporelle. En combinant les principes d'itération et de traçabilité propres aux wikis et celui d'argumentation spatialisée des *Argumaps* (Rinner, 2001), la richesse de ce prototype repose principalement sur le fait qu'il donne à voir le processus de conception collaborative de connaissances spatialisées tout en assurant la traçabilité des actions. Ce qui est particulièrement intéressant dans un contexte de concertation où chacun apporte sa contribution au projet.

4 Conclusion

L'objectif de ce papier consistait, à travers l'exploration des techniques et des usages de la cartographie 2.0, à mettre en évidence les potentialités participatives du géoweb dans le contexte de généralisation du débat public. Ce faisant, nous avons pu expliciter en fonction du degré d'implication des citoyens, trois manières d'appréhender ces potentialités. Au niveau de l'information, les technologies du géoweb permettent un accès dynamique, interactif et multimédia à un grand nombre d'informations relatives au territoire et au projet. Au niveau de la consultation, elles permettent aux citoyens par l'intermédiaire de cartes interactives de s'exprimer et d'apporter leurs avis et contributions sur divers projets et thématiques. Enfin au niveau de la concertation, elles sont en mesure de supporter et d'appuyer des dispositifs visant à la fois à mobiliser et formaliser les connaissances locales des citoyens, tout comme la production coopérative, voire collaborative de connaissances relatives au territoire ou à un projet d'aménagement. Globalement le géoweb offre des technologies accessibles, simples d'utilisation et performantes, permettant ainsi d'éviter l'effet « usine à gaz » des SIG professionnels. Associée au Web comme support de diffusion de masse, l'accessibilité aux applications et aux données géographiques est démultipliée.

Néanmoins, il reste encore du chemin à parcourir pour faire des applications cartographiques 2.0 de véritables outils de participation territoriale. Les potentialités existent tant sur le plan des techniques que des pratiques des professionnels et du grand public. Mais il est important de réfléchir aux innovations socio-politiques qui accompagnent le développement de ces technologies selon la perspective des usages. La technologie n'a en effet pas de sens, de valeur et de conséquence en elle-même sur l'amélioration d'un processus social comme la démocratie participative (Roche, 2003). La variable déterminante réside dans la pratique individuelle et collective. Il convient donc de s'interroger avant tout sur les modalités d'appropriation (présentes et futures) de ces technologies par les collectivités et les citoyens (collectifs territorialisés, association, communauté de pratique, etc.). Cela passe par nécessairement par un travail d'observation des pratiques et de discussion avec les acteurs. Ceci afin de cerner leurs usages, leurs attentes et leurs besoins pour pouvoir proposer des solutions adaptées en termes de fonctionnalités, de données et de méthodes). C'est là tout l'enjeu « démocratique » du géoweb et des TIC en

général. Car au-delà des potentialités techniques, la question de la démarche à mettre en place est primordiale (sensibilisation, formation, accompagnement, transfert de compétence). Cette dernière devant d'une part se placer en complément des autres canaux de participations et d'autre part se construire en fonction du contexte socio-politique local dans le but de se placer au cœur du système décisionnel.

Bibliographie

- Blancher P., « Cyberspace, participation du public et mobilisations citoyennes », *Techniques, territoires et sociétés*, vol. n° 37, 2004, p. 281-308.
- Budhathoki N. R., Bruce B., Nedovic-Budic Z., « Reconceptualizing the role of the user of spatial data infrastructure », *GeoJournal*, vol. 72, n° 3, 2008, p. 149-160.
- Bugs G., Granell C., Fonts O., Huerta J., Painho M. « An assessment of Public Participation GIS and Web 2.0 technologies in urban planning practice in Canela, Brazil », *Cities*, vol. 27, 2010, p. 172-181.
- Cai G., Yu B., « Spatial annotation technology for public deliberation », *Transaction in GIS*, vol. 13, n°1, 2009, p. 123-146.
- Caquard S., Des cartes multimédias dans le débat public. Pour une nouvelle conception de la cartographie appliquée à la gestion de l'eau, Thèse de doctorat, Université J. Monnet de St Etienne, 2001.
- CNIG, « Citoyenneté et information géographique », Fiche maîtrise d'ouvrage n°101, 2007.
- Ciobanu L. D., Roche S., Badard T., Caron C., «Du Wiki au WikiSIG», *Geomatica*, vol. 61, n° 4, 2007, p. 455-469.
- Coleman D. J., Georgiadou Y., Labonte J., « Volunteered Geographic Information: the nature and motivation of producers », *International Journal of Spatial Data Infrastructures Research*, vol. 4, n° 2009, p. 332-358.
- Craig W. J., Harris T. M. Weiner D. (dir.), *Community participation and geographic information systems*, London, Taylor & Francis, 2002.
- Crampton J., « Cartography : Maps 2.0 », *Progress in Human Geography*, vol. 33, n° 1, 2008, p. 91-100.
- De Longueville B., Ostländer N., «Addressing vagueness in Volunteered Geographic Information (VGI) - A case study», *GSDI-11*, Rotterdam, 2009
- Debarbieux B., Lardon S. (dir.), *Les figures du projet territorial*, La Tour d'Aigues, Aube : Datar, 2003.
- Debarbieux B. Vanier M. (dir.), *Ces territorialités qui se dessinent*, La Tour d'Aigues / Paris, Éditions de l'Aube / DATAR, 2002.
- Duncan S. L., Lach D. H., « Privileged Knowledge and Social Change: Effects on different participants of using GIS technology in natural resource management », *Environmental Management*, vol. 38, n° 2, 2006, p. 267-285.
- Elwood S., « Geographic Information Science: new geovisualization technologies – emerging

- questions and linkages with GIScience research », *Progress in Human Geography*, vol. 33, n° 2, 2009, p. 256-263.
- Gauthier M., Gariépy M, Trépanier M.-O. (dir.), *Renouveler l'aménagement et l'urbanisme : planification territoriale, débat public et développement durable*, Montréal, Presses de l'Université de Montréal, 2008.
- Goodchild M. F., « Citizens as Voluntary Sensors: SDI in the World of Web 2.0 », *International Journal of Spatial Data Infrastructure Research*, vol. 2, 2007, p. 24-32.
- Goodchild M. F., « NeoGeography and the nature of geographic expertise », *Journal of Location Based Service*, vol. 3, n° 2, 2009, p. 82-96.
- Ghose R., « SIG et participation citoyenne : des projets de revitalisation urbaine aux États-Unis », in. Roche S. et Caron C. (dir.), *Aspects organisationnels des SIG*, Paris, Hermès Lavoisier, 2004.
- Gumuchian H., *Représentations et aménagement du territoire*, Paris, Economica, 1991
- Haklay M., Singleton A., Parker C., « Web Mapping 2.0: The Neogeography of the GeoWeb », *Geography Compass*, vol. 2, n° 6, 2008, p. 2011-2039.
- Herboux P., *Intelligence territoriale : repères théoriques*, Paris, L'Harmattan, 2007.
- Herring C., *An Architecture of Cyberspace: Spatialization of the Internet*, U.S: Army Construction Engineering Research Laboratory, 1994.
- Hudson-Smith A., Crooks A., Gibin M., « NeoGeography and Web 2.0: concepts, tools and applications », *Journal of Location Based Service*, vol. 3, n° 2, 2009, p. 118-145
- Jarnac G. (dir.), *Cartographie numérique et développement des territoires*, Etude IRIS à la demande de l'Observatoire des Territoires Numériques (OTEN), 2008.
- Joliveau T., « La participation à la décision territoriale : dimension socio-géographique et enjeux informationnels d'une question politique », *Géocarrefour*, vol. 76, n° 3, 2001, p. 273-279.
- Joliveau T., *Géomatique et gestion environnementale du territoire. Recherches sur un usage géographique des SIG, mémoire d'habilitation à diriger des recherches*, Département de Géographie, Université de Rouen, 2004.
- Lardon S., Maurel P., Piveteau V. (dir.), *Représentations spatiales et développement territorial*, Paris, Hermès Lavoisier, 2002.
- Laurini R., « SIG pour la participation des citoyens aux décisions relatives à l'aménagement du territoire », *Techniques, territoires et sociétés*, vol. n° 37, 2004, p. 209-219.
- Lévy J., Poncet P., Tricoire E., *La carte, enjeu contemporain*, Paris, La documentation française, 2004.
- MacEachren A. M., Brewer I., «Developing a conceptual framework for visually-enabled geocollaboration», *International Journal of Geographical Information Science*, vol. 18, n° 1, 2004, p. 1-34.
- Maurel P., «Les représentations spatiales : concepts de bases et éléments de typologie», in Lardon S., Maurel P., Piveteau V. (dir.), *Représentations spatiales et développement*

territorial, Paris, Hermès, 2001, 75-108.

McHugh R., S. Roche, Y. Bédard, 2009, « Towards SOLAP-Based Public Participation GIS », *Journal of Environmental Management*, vol. 90, n°6, p. 2041-2054.

Mericskay B., Roche S., «Cartographie numérique nouvelle génération: impacts de la néographie et de l'information géographique volontaire sur la gestion urbaine participative», *HyperUrbain.2*, Paris, Europia, 2009.

Péribois C., Usage de l'information géographique dans la gestion participative du territoire, Thèse de doctorat, Département de géographie, Université d'Angers, 2008.

Pornon H., Noucher M., « Bilan et perspective de 20 années de géomatique : Vers des SIG plus collaboratifs, la Géo-collaboration », *Géomatique Expert*, vol. 58, 2008, p. 56-60.

Quoniam L., Arnaud L., *Intelligence compétitive 2.0* Paris, Hermès Lavoisier, 2010.

Rey H., Bacqué M.-H., Sintomer Y. (dir.), *Gestion de proximité et démocratie participative : une perspective comparative*, Paris, La Découverte, 2005.

Ramasubramanian L., *Geographic Information Science and Public Participation*, New-York, Springer, 2010.

Rambaldi G., Kwaku Kyem P., McCall M., Weiner D., « Participatory Spatial Information Management and Communication in Developing Countries », *The Electronic Journal of Information Systems in Developing Countries*, vol. 25, 2006, p. 1-9.

Rinner C., «Argumentation maps - GIS-based discussion support for online planning.», *Environment and Planning B: Planning and Design*, vol. 28, n° 6, 2001, p. 847-863.

Rinner C., Kessler C., Andrulis S., « The use of Web 2.0 concepts to support deliberation in spatial decision-making », *Computers, Environment and Urban Systems*, vol. 32, n° 2008, p. 386-395.

Roche S., «Usages sociaux des technologies de l'information géographique et participation territoriale», in Debarbieux B., Lardon S. (dir.), *Les figures du projet territorial*, Paris, L'Aube, 2003, p. 61-82.

Roche S., Caron C. (dir.), *Organizational Facets of GIS*, London, John Wiley & Sons, 2009.

Sample, J., Shaw, K., Shengru, T., Mahdi, A. (dir.), *Geospatial Services and Applications for the Internet*, New York, Springer, 2008.

Scharl A., Tochtermann K. (dir.), *The Geospatial Web : How Geobrowsers, Social Software and the Web 2.0 are Shaping the Network Society*, London, Springer, 2007.

Seeger C., « The role of facilitated volunteered geographic information in the landscape planning and site design process », *GeoJournal*, vol. 72, n° 3, 2008, p. 199-213.

Sui D., « The Wikification of GIS and its consequences : Or Angelina Jolie's new tatoo and the future of GIS », *Computers, Environment and Urban Systems*, vol. 32, 2008, p. 1-5.

Tapscott D., Williams A. D., *Wikinomics : Wikipédia, Linux, You Tube : comment l'intelligence collaborative bouleverse l'économie*, Paris, Pearson/Village mondial, 2007

Tulloch D., « Is VGI participation ? From vernal pools to Video Games », *GeoJournal*, vol. 72, n° 3, 2008, p. 161-171.

Invasives Species Tracking in Ontario

Using Geomatics for Collaborative Innovation

Fred McGarry^a, Donald Cowan^b, Paulo Alencar^b

^aCentre for Community Mapping (COMAP)
Waterloo, Canada
mcgarry@comap.ca

^bDavid R. Cheriton School of Computer Science – University of Waterloo
Waterloo, Canada
{dcowan, palencar}@uwaterloo.ca

RÉSUMÉ. Nos capacités de bien gérer nos ressources naturelles dans le 21e siècle eut égard aux changements globaux et systémiques au niveau culturel, de la santé, l'eau, l'environnement et l'énergie, défient constamment nos communautés face à la gestion des ces ressources. Nos communautés doivent avoir la capacité d'adaptation tout en maintenant, voire accroître leur économie et qualité de vie, c'est en ce sens que le Forum Économique Mondial reconnaît le besoin essentiel d'approches favorisant collaboration inventive (CI) et tout activité connexe au sein des décideurs et de la population engagée. Ce document relate les concepts d'une approche à la CI basée sur la cartographie d'actifs et comment les bases de données spatiales, l'accès contrôlé des collaborateurs, la distribution et rapports, la gestion documentaire, les réseaux sociaux ainsi que d'autres outils de collaboration, le tout sécurisé dans un réseau WEB, d'où résulte, à titre d'exemple, "Invasive Species Tracking System" pour l'Ontario (Canada).

ABSTRACT. Twenty-first century systemic global change in cultural, health, water, environment, energy, business and socio-economic structures is challenging our communities' management of resources. For a community to adapt to this systemic change, while maintaining and even enhancing its economy and quality of life, the World Economic Forum has recognized the need for new approaches to enable collaborative innovation (CI) and related action among both the leadership and concerned members of the community. This paper outlines the concepts of an approach to CI based on asset mapping, how it has been supported, with participatory geomatics services related to temporal and spatial data, distributed role-based access controls, reporting, document management, community networks and other collaboration tools, and applied through a web-based technological framework to an Invasive Species Tracking System for Ontario (Canada).

MOTS-CLÉS : Le dépistage d'espèces indésirables, La collaboration inventive, La cartographie d'actifs, Le réseau WEB, Les méta-outils

KEYWORDS: invasive species tracking, collaborative innovation, asset-mapping, web-based framework, meta-tools

1. Introduction

Twenty-first century systemic global change in cultural, health, water, environment, energy, business and socio-economic structures is challenging our communities' management of resources. For a community to adapt to this systemic change, while maintaining and even enhancing its economy and quality of life, the World Economic Forum has recognized the need for new approaches to enable collaborative innovation (CI) and related action among both the leadership and concerned members of the community. Many see the web as an approach to CI and as a new form of creativity machine that can augment our intelligence.

The Centre for Community Mapping (COMAP.ca) in Waterloo, Ontario and the Computer Systems Group at the University of Waterloo are creating a suite of technological services that support collaborative innovation (CI). This paper outlines the concepts of an approach to CI based on asset mapping, how it has been supported and applied through a web-based technological framework and how it has been applied to an Invasives Tracking System for Ontario.

The focus of this paper is to examine experiences with collaborative innovation within the context of the new modes of communication that have arisen through the development of the Internet, Web and tools such as mediated community networks. The Web Informatics Development Environment (WIDE), a set of technologies designed and developed by the Computer Systems Group at the University of Waterloo (UWCSG), has been created to provide this type of service and can be easily tailored to support specific collaborative activity. WIDE is an example of the type of meta-tools that are needed. Lessons learned from WIDE are being applied in the creation of WIDE 2.0.

2. Understanding Collaborative Innovation

The World Economic Forum (WEF) [WEF] recognizes the need for an effective approach to collaborative innovation in the aftermath of the global economic crisis as local and international communities are becoming more and more interdependent. The WEF has recommended that communities both geographic and virtual re-examine their strategies based on new evolving networks and forms of collaboration. Thus public-private partnerships and smaller-scale region or city-driven initiatives must move away from one-size fits all arrangements and adapt to the re-design and rebuild their structures and processes related to both local and global community interactions, resource exploitation and governance.

Such related actions among participants in communities of practice and geographic communities must be supported by services providing an information infrastructure that is easily accessible through a medium such as the World Wide Web.

Collaborative innovation refers to a process that involves assembling a team of people to explore and act upon change in an idea or situation [CHE 06]. The span of collaboration is usually a virtual community of practice or interest, or a geographic community [WEN 98]. The individuals that compose the team can represent themselves, different departments in a single organization or different organizations. A team often works in a mediated environment where responsibility for actions can be devolved to specific team members. These responsibilities can be specified through one or more contracts among individuals and organizations.

Collaborative innovation teams have a collective vision and wish to work together by sharing ideas, information and work. Team members share directly rather than through a hierarchy, although each member may represent the views of the hierarchy to which they belong. Any collaboration must have four essential elements, which are :

- 1) sound ethical principles ;
- 2) trust and self-organization ;
- 3) universally accessible knowledge ; and
- 4) honest and transparent operation.

Although organized collaborative innovation has been recognized as early as Benjamin Franklin's "Junto" organization [Fra], today most work of this type, relies on modern information and communications technology (ICT) such as the Internet, e-mail, Web and more recently social networks.

What else is needed to make collaborative innovation effective ? Collaborative innovation requires shared knowledge of the resources or assets that are available within an application domain or context and then using that understanding to implement actions such as operational and structural decisions.

Collaborative innovation can be viewed as a two-step process where members of a geographic or virtual community

- 1) produce an inventory of assets and share this data across one or more communities ; and
- 2) collaborate and act on that asset knowledge to :
 - a) recognize additional "undiscovered" assets ;
 - b) produce value and change ; and
 - c) create new assets related to the change.

Basically the participants share knowledge and add value to that shared knowledge through a collaborative set of tools which are formed into community networks that operate under various constraints such as access, time and mobility.

In many fields, the concept of creating an asset inventory is called "asset mapping" and the new value and assets produced from the collaboration are often called emergent properties. Most work in this field to date has focused on "static" asset mapping and has only begun to investigate the methodology to keep the asset inventory current or "dynamic" and in taking action over the growing asset base. Further research is needed into an organized approach to on-going collaboration, dynamic asset mapping, and value production, although this lack of a theory of collaborative innovation should not stop its use. In this paper we outline a novel asset-mapping driven approach to web-based collaborative innovation based on our experience in designing and implementing over 50 such systems.

2.1. *Constituents of Asset Mapping CI*

2.1.1. Shared knowledge

There must be a collection of shared information or knowledge, often called a database, knowledge base, inventory or asset map, to provide a basis for collaboration that takes advantage of intelligent software agents to manage redundancy (Alencar et al., 2003).

2.1.2. The Collaborative Canvas

Information about these assets and input to and results of the collaboration can be presented through shared canvases or interfaces. The shared canvas can provide many different tools such as a data report, input to databases, e-mail, audio, video, pictures, text, maps (geomatics), short messages (Twitter tweets), blogs, and wikis to name some examples.

2.1.3. Access controls

Collaboration requires that a group forms around an idea or situation with the objective of working together. The group is self-limiting by expertise or interest although it may grow or change its composition as the collaboration forms and changes, and new expertise is needed or more individuals become interested. By its nature collaboration is not completely open. Therefore there must be moderators who manage the group composition and delegate authority to members of the group related to responsibilities. These moderators must be given a set of tools not only to enable the collaboration but to admit participants with responsibilities, so-called role-based access controls (Bertino et al., 2010).

2.1.4. Mediated Community Networks

Mediated Community networks can be viewed as a set of tools from the collaborative canvas that allow sharing and collaboration over the asset base with responsibilities usually delegated through role-based access controls. There can be many different types of tools that are used as components of the community network as already des-

cribed including maps (geomatics), input forms, reports, text, video, pictures, audio, wikis, blogs, and asset repositories.

2.2. Properties of Asset Mapping CI

2.2.1. Collaboration in Time

Collaboration can occur asynchronously or synchronously (real-time). Asynchronous collaboration is more common because contributors do not have to agree on a time to meet. The collaborators can contribute through email, maps, wikis, blogs, Web forms and other similar mechanisms. Synchronous or real-time collaboration requires the presence of two or more parties to the collaboration. Of course all the collaborators could be in the same room or share the map and associated communication channels over the Web.

2.2.2. Mobility

Mobility adds a new dimension to collaboration as now the collaboration and the creation of value can happen while participants are on the move.

2.2.3. Adding Value

Once the collaboration begins, the active collaborators through working together using the shared assets and collaborative canvas, add value since they create more information and knowledge or assets, which are usually added to the asset base. These additions are called emergent properties. Rich emergent properties often arise when groups form communities of practice.

3. Implementing Asset Mapping CI

There are many facets to collaborative innovation using ICT and based on an asset mapping approach. In this section we analyze this phenomenon and try to understand its constituents and related properties. Basically the participants share knowledge and add value to that shared knowledge through a collaborative set of tools under various constraints such as access, time and mobility.

The meta-tools in the WIDE include services related to temporal and spatial (mapping) data, role-based access controls, reporting, document management, mediated community networks and other collaboration tools, that is WIDE implements tools that support the concepts in the previous section. This joint research program has a number of goals namely to :

- 1) develop data models and corresponding databases for asset maps in different domains ;
- 2) create meta-tools to simplify the production of interfaces, access controls and mediated community networks ;

3) create interface and community network frameworks to asset maps to support powerful and purposeful collaboration ;

4) create interface frameworks to asset maps to assist with collaborative and dynamic asset map updating ;

5) create interface frameworks to interrelate information maintained by different communities of practice to enable synergistic benefits ; and

6) produce meta-tools that can be used and maintained by the collaborators.

Participating organizations are provided with mediated community network services that support self-organization among participants by promoting opportunities to share knowledge with other participants with common interests or goals. By combining dynamic asset mapping with community network services, communities of practice can collectively communicate to discern opportunities for asset management.

Since collaborative innovation requires advanced technologies to integrate known and novel web services, an efficient approach to building and testing such complex technologies with direction from community of practice participants is needed. The toolkit based on the WIDE technologies offers approaches that are simple to use and to change - programmers are rarely needed.

In the WIDE context, "programming" has effectively been replaced with a declarative methodology thus making it possible to provide a wizard or forms-based approach to building Web-based systems. This approach allows the technical team to develop web-based information systems about 10 times faster than more traditional methods.

COMAP and UWCSG and the users completely engage during the entire specification, design and implementation cycle and use an iterative approach within the context of the WIDE technologies to create complex web services. The approach is iterative in that once users operate a version of the system, they may actually change the specification on the fly, thus affecting further design and implementation.

4. Asset Mapping CI in Application

Different forms of collaborative innovation depend on factors such as community scope, degree of cohesion among community members, and type of knowledge exchanged and can lead to different types of communities [JAN 08]. The WIDE toolkit has been used to support many different forms of collaborative innovation with varying sizes of geographic and virtual communities and different degrees of community coherence based on common goals.

The Invasives Tracking System (ITS) was developed by COMAP and UWCSG for the Ontario Federation of Anglers and Hunters (OFAH) and will be launched in the summer of 2010. OFAH has been Ontario's 'hot-line' for the reporting of invasive species for the benefit of the Ontario ecosystem conservation community of practice since 1992. The ITS is intended to be a hub for citizen scientists and professional organizations involved in invasive species issues. The goal of the ITS project : to

monitor of invasive species presence and to track invasives control projects in support of appropriate decisions with regard to restoration priorities and techniques.

The Invasives Tracking System, will be further described to illustrate our current thinking and approaches. The example shows :

- 1) Collaborative geomatics (mapping) ;
- 2) Collaborative dynamic asset-mapping for communities of practice ;
- 3) An inclusive service framework for authoritative and volunteered content ; and
- 4) Collaboration through custodial distribution of authority to publish using Mediated Community Network services.

4.1. Collaborative Geomatics (Mapping)

Collaborative geomatics consists of a common mapping canvas over which where communities can have discussions and communications about evolving local issues and collaborate in using applications for local data capture, analysis and reporting. COMAP and UWCSG, have built a mapping interface and database (WIDE image server, with W3C and Open Geospatial Consortium services : WMS, WFS, VML, and SVG) which can serve thematic and highly resolute airphoto data for Southern Ontario. To ensure that all of Ontario is served and easily searchable COMAP uses GeoBase satellite data and a Lambert projection. Collaborative geomatics services have been deployed for the Invasives Tracking System (ITS).

The ITS is acknowledged to be the leading effort in shared information infrastructure for cross-scalar provincial reporting on invasives species management in Canada. WIDE mapping supports the capture of fine-scale information using standard protocols, which can then be used to augment existing GIS data to permit landscape GIS analyses. WIDE services can increase the value of information management resources by integrating numerous data sources in geographic space in order to provide enhanced direction for asset management decisions. These processes can be manual or automated.

4.2. Dynamic asset-mapping for communities of practice

Community participants will take the time to organize and share community perspectives, reporting and analyses using collaborative geomatics to populate maps with text, data and media, if the members of the community see value returning to their community through their active participation. This process known as "dynamic community asset mapping" is an extension of community 'asset mapping' as initially defined by John McKnight and John Kretzmann [KRE 93] and now widely practiced by communities in North America.

The ITS web mapping application was developed to enable citizen scientists as

well as conservation professionals to report sightings of both aquatic and terrestrial invasive species and accurately map their location directly on high resolution air photography. The resulting system has facilities for :

- tracking, mapping and reporting on invasive species in search, discovery, remediation and monitoring stages of invasives management,
- and for capturing, reporting and tracking of remedial action.

The ITS project is a database and a web-based set of applications that accesses spatial data and information in real-time from distributed sources over the Internet. The ITS was planned to permit the southern Ontario conservation community participants to work collaboratively by :

- entering spatial (polygon) and tabular data, photos and documents about their invasives species sightings and ecological restoration projects as well as exporting entered data to external geographic information systems ;
- querying the database to meet their needs for tracking invasives species sightings and specific restoration projects ;
- reporting and summarizing monitoring data about invasives species sightings and restoration projects by numerous parameters ; and
- implementing adaptive management of ecological restoration practices based on an ever-expanding base of knowledge about the factors that contribute to successful ecological restoration projects.

Expected benefits of the web mapping application include :

- Increased motivation by volunteer citizen scientists to report sightings of invasive species
- Increased accuracy of sightings, both for location and identity of species
- Accurate mapping of invasive species sightings to facilitate follow-up remediation or restoration activities by professional staff from resource management agencies
- Performance management system to monitor impacts of restoration activities (e.g., invasive species removal)
- Increased public awareness of the invasive species threat and enhanced motivated to take preventative action
- Increased public awareness of invasive species using information contained in invasive species thematic maps (verified sightings)

A field guide of invasives species is provided (which is editable by OFAH personnel) with a current listing for all recognized invasive Algae, Fishes, Invertebrates and Plants.

The ITS provides the user with the ability to select parameters (species common or Latin name, waterbody, 'geographic' county, Ministry of Natural Resources district, Conservation Authority, tertiary or quaternary watershed, drawn polygon, rectangle or circle search area, sighting date or sighting date range, confirmation status) and search

for, list and map all requested sightings. The mapping service, supporting the sightings search facility (See Figures 1 & 2), uses a novel application of Ajax technology whereby any parameter will list and map immediate results within a selected geography. The map bounding box and hence the selected zoom level is a search parameter. Additional parameters selected will further constrain the results.

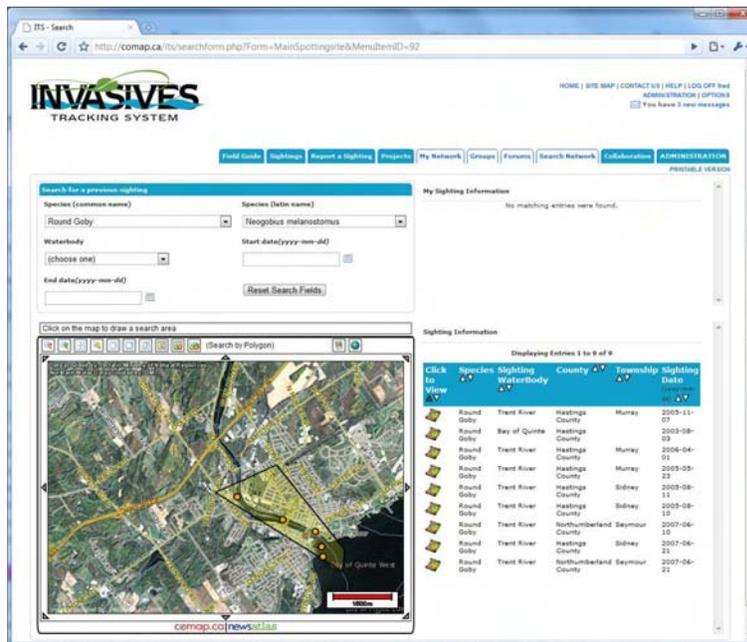


Figure 1. Hand drawn polygon-based spatial searching

The ITS is acknowledged to be the leading effort in shared information infrastructure for cross-scalar provincial reporting on invasives species management in Canada. WIDE mapping supports the capture of fine-scale information using standard protocols, which can then be used to augment existing GIS data to permit landscape GIS analyses. WIDE services can increase the value of information management resources by integrating numerous data sources in geographic space in order to provide enhanced direction for asset management decisions. These processes can be manual or automated.

4.3. Service frameworks for authoritative and volunteered content

Clay Shirky's [SHI 09] comments on 'cognitive surplus' suggest a huge unused capacity that could be channeled for social benefit by new social media frameworks, however, academics, professionals and other authorities strenuously resist the 'wisdom

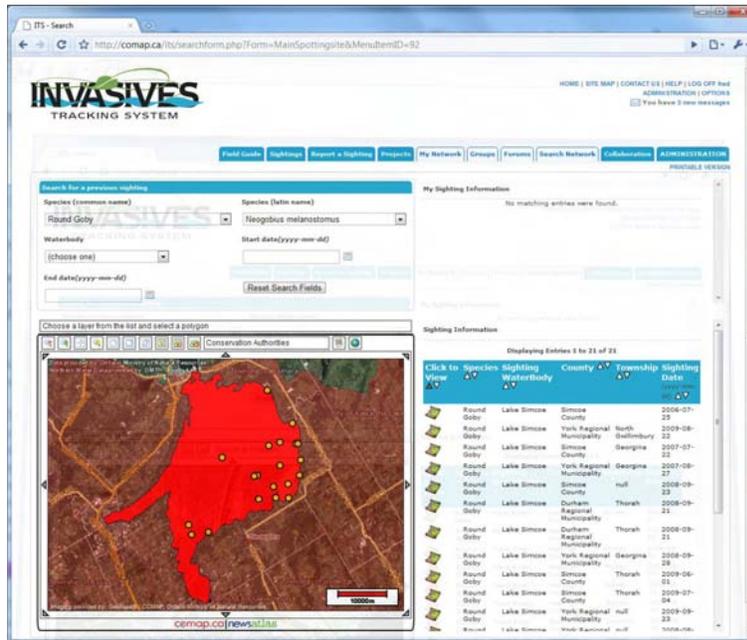


Figure 2. Thematic polygon-based spatial searching

of the crowd' for many and often good reasons. COMAP and UWCSG are implementing systems that provide for both authoritative and volunteered content in a manner that maintains clear distinctions. These novel approaches organize (generally chaotic) social network services, for community benefit. To encourage self-organized learning networks COMAP and UWCSG are developing user activated subscription services for content contributed by other participants. These services permit users to find out about activities that interest them and to learn from the members of their community of practice that lead these activities. Users are able to create subject based forums and invite selected members to discuss ideas and challenges.

The Invasives Tracking System (ITS) technical committee wanted a community network that is only open to the experts and related agencies. They reasoned that the general public may be confused and misconstrue volunteered content (such as species identification and remediation information) as being official content. Consequently, the ITS deploys a mediated community network system, consisting of groups, forums and authoring services, that the agencies can use to share their experiences, as a learning network and to collaboratively author and publish official and authoritative information used to help identify and control invasive species.

COMAP has developed a system with tools that permit the public and the ITS conservation community of practice to access an invasives species field guide, a sigh-

tings reporting facility and query service for approved sighting information. Only the ITS conservation community of practice, however, can report on a restoration project and export contributed tabular and spatial data.

All parties can access the 'Report a Sighting' service. Once a user has selected species type (Algae, Fishes, Invertebrates, Plants) a list for each species with thumbnail photo, a link to a Field Guide page and a 'Report a Sighting' link is presented for selection. The 'Report a Sighting' link presents a sighting report form tailored for context (Aquatic or Terrestrial) and pre-loaded with the selected species. Photos of the sighting can be uploaded. Map 'point' posting services allow the user to pinpoint the location of the sighting by locating a dot using high resolution (zoomable, pannable) aerial photo mapping. (See Figure 3) A sighting confirmation process and administration services are provided so that OFAH can manage confirmations.

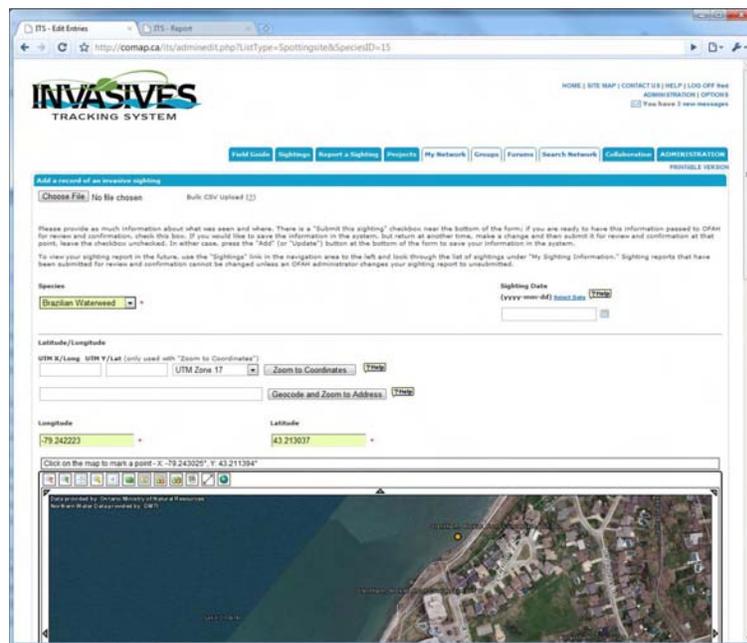


Figure 3. A dot indicating the species sighting placed accurately in a resolute air-photo using the COMAP map appliance

The ITS Community of Practice (CoP) can report on invasives management and control projects. Projects can be site (on-the-ground) related or public education or outreach in nature (off-the-ground). A project may have one or more sites where activities are performed. (See Figure 4) Project and site level photos and documents can be uploaded and site level activities and results can be recorded for the fiscal year of a project as attributes of a polygon entered over the web.

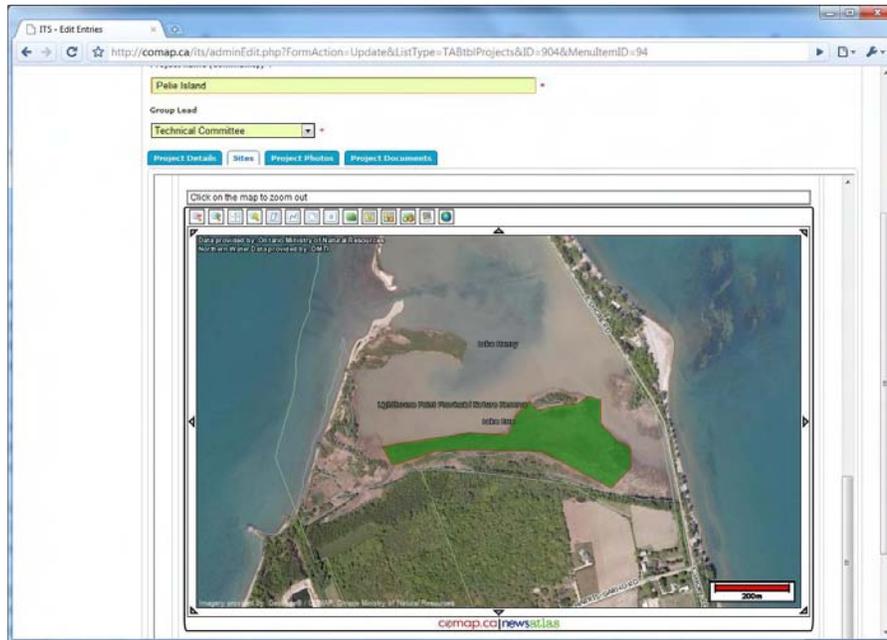


Figure 4. A polygon indicating the area of an invasive species control project placed accurately in a resolute airphoto using the COMAP map appliance

4.4. Custodial distribution of authority to publish using Mediated Community Network services

The Invasives Tracking System (ITS) service deploys a mediated community network system, (See Figure 5) consisting of groups, forums and authoring services, that the agencies can use to collaboratively author and publish official and authoritative information.

The right to participate in the ITS application community of practice mediated community network and contribute invasives control project information is distributed to participating community of practice organizations. System security is managed by the content custodian OFAH. OFAH distributes the right to publish and secures agreement from rights' recipients to conditions necessary for system security and integrity. Minimum conditions for the right to publish are : Agreement to appoint an organization's group moderator and maintain, in good standing, an organization's group membership and information for those members that publish on behalf of the organization.

Registered participants are consequently able to (or enable their membership or affiliates to) access ITS project content publishing. Groups and agencies, in turn, will



Figure 5. Group functionality within the ITS Mediated Community Network

have responsibility for : " granting access to their membership and affiliates to tools for publishing on their organization's behalf ; and " for maintaining the access rights of their membership.

A member of the general public will access the ITS via the OFAH web-site. She/he can access an invasives species field guide, sightings information and create a password to register and report a sighting.

To use the Report a Sighting service, the user will need to register and create a UserID and Password. Registered users can be invited by ITS mediated community network Group Moderators to participate. Only the custodian OFAH can create a new Group within the ITS mediated community network and appoint a Group moderator from a list of registered users. Once an ITS Group is set up, its moderator and members can access and participate in the community network services (group forums, private messaging, bookmarking and notifications) and publish ITS project information.

Once a Group is set up, its Moderator can invite members from the ITS registered users list. It is anticipated that each Group in the ITS will represent an organization that participates in the ITS community of practice and that a Group membership will publish with the authority of the participating organization.

5. Related Work

The asset-mapping CI approach proposed in this paper contrasts with closed community efforts, which involve limited interaction, and both restricted knowledge of community assets and their value chains [CHE 03]. In the context of service systems, it has been suggested that open service-oriented models could use novel paradigms based on innovation [MAG 06] and asset mapping.

Thinking frameworks have been proposed to help organizations focus their management attention, and enable users to participate in the innovation process [WAL 09]. In addition, modeling approaches have been defined to describe innovation networks from a services system perspective and to address inter-organizational interactions [JAN 08]. In comparison, our proposal focuses on a framework that supports the construction of web-based collaborative innovation systems based on asset mapping that can address both inter-organizational and general community interactions and resource exploitation and production.

In summary, the research described in this paper based on our experiences aims at defining asset-mapping driven approaches to the development of web-based intelligent collaborative innovation systems that promote explicit knowledge sharing needs among cohesive and global-scope communities. COMAP and UW CSG are engaged in the planning or development of CI services for

- equitable and timely consultation with Aboriginal First Nations for resource and infrastructure development
- community mediated restructuring of social service programs
- participatory heritage tourism
- shared stewardship of our ecosystems
- a community enterprise that leverage synergies between community service and local business.

6. Conclusion

This paper outlines the concepts of collaborative innovation based on asset mapping, its importance to the functioning of modern society and how modern web-based tools could be used to support this activity. Based on the experience gained in designing and building over 50 systems that incorporate collaborative innovation, some of which are described in some detail in this paper, it is clear that collaborative innovation takes many forms. Thus, it is not possible to build a single set of tools to support collaborative innovation. Rather a set of meta-tools is needed which can be used to build tailored systems to fit specific situations that arise when web-based collaboration is to occur. The WIDE technologies are an example of the type of meta-tools that are needed. Lessons learned from the construction of WIDE are being applied in the creation of WIDE 2.0.

Acknowledgment

The authors would like to thank the Ontario Ministry of Natural Resources (OMNR), Land Information Ontario, the Ontario Federation of Anglers and Hunters, GeoConnections, the Oak Ridges Moraine Foundation for their financial support and cooperation in implementing the projects described in this paper.

7. Bibliographie

- [CHE 03] CHESBROUGH H., *Open Innovation : The New Imperative for Creating and Profiting from Technology*, Harvard Business School Press, 2003.
- [CHE 06] CHESBROUGH H., VANHAVERBEKE W., WEST J., *Open Innovation : Researching a New Paradigm*, Oxford University Press, 2006.
- [Fra] « Junto », *Wikipedia*, <http://en.wikipedia.org/wiki/Junto>.
- [JAN 08] JANNER T., SCHROTH C., SCHMIDT B., « Modelling Service Systems for Collaborative Innovation in the Enterprise Software Industry - The St. Gallen Media Reference Model Applied », *IEEE International Conference on Services Computing*, IEEE Computer Society, 2008, p. 145–152.
- [KRE 93] KRETZMANN JOHN P. M. J. L., *Building Communities from the Inside Out*, ACTA Publications, Skokie Illinois, 1993.
- [MAG 06] MAGLIO P., SRINIVASAN S., KREULEN J., SPOHRER J., « Service systems, service scientists, SSME, and innovation », *Communications of the ACM*, vol. 49, n° 7, 2006, p. 81–85.
- [SHI 09] SHIRKY C., *Here Comes Everybody*, Penguin Books, 2009.
- [WAL 09] VAN DER WALT J. S., BUITENDAG A. A., ZAAIMAN J. J., VAN VUUREN J. J., « Community Living Lab as Collaborative Innovation Environment », *Issues in Informing Science and Information Technology*, vol. 6, 2009, p. 421–436.
- [WEF] « World Economic Forum Annual Meeting 2008 ; The Power of Collaborative Innovation », Available at <http://www.weforum.org/en/events/ArchivedEvents/AnnualMeeting2008/index.htm>.
- [WEN 98] WENGER E., *Communities of practice : learning, meaning, and identity*, Cambridge University Press, 1998.

Co-visualisation de données sur une carte

Aide à la cartographie de données multiples et hétérogènes sur un portail de données géographiques

RÉSUMÉ. Les récentes avancées technologiques offrent la possibilité aux utilisateurs non-experts de construire eux mêmes leurs cartes, adaptées à leurs besoins, sur le web. La liberté donnée à ces utilisateurs leur permet de représenter un grand nombre de données sur une seule carte. Mais la superposition de plusieurs thèmes sur une même carte diminue sa lisibilité et donc, son efficacité. Le travail présenté dans cet article correspond donc à une étude menée en thèse visant à proposer des solutions à cette problématique. Nous présentons en premier lieu quelques recherches menées dans ce domaine, puis nous expliquons nos premières propositions de solutions simples, comme la réorganisation de l'ordre des thèmes ou la création de plusieurs sous-cartes. Enfin, nous concluons en présentant les démarches futures de cette recherche.

ABSTRACT. Recent technological advances offer the possibility to non-experts users to build themselves their map, suited to their needs, on the web. The freedom given to these users allows them to represent a large amount of data on one map. But several overlapping themes on the same map decrease its readability and thus its efficiency. The work presented in this paper corresponds to a study conducted in a thesis, to propose solutions to this problem. We present, first, some research in this domain, and then we explain our proposals of simple solutions as the reorganization of the order of layers or the creation of several sub-maps. Finally, we conclude by presenting future steps of this research.

MOTS-CLÉS : geovisualisation, cartographie, légende, sémiologie, données hétérogènes, portail web, variables visuelles

KEYWORDS: geovisualization, cartography, legend, semiotics, heterogeneous data, web portal, visual variables

1. Introduction

La carte est depuis toujours un moyen de transmission d'informations très utile, ayant plusieurs champs d'actions, tels que se repérer dans l'espace, programmer des itinéraires, ou d'autres objectifs plus poussés. On retrouve des restes d'anciennes cartes remontant à 2300 ans avant JC. La carte s'est récemment démocratisée par l'explosion des sites web cartographiques. Elle est également passée d'une représentation papier à une représentation numérique visualisable sur écran, modifiable et manipulable par n'importe quel utilisateur.

Dans cet article, nous nous intéressons à la conception et à la visualisation de cartes personnalisées sur un portail cartographique sur le web. Dans la plupart des outils cartographiques récents, il est possible pour l'utilisateur de construire lui-même une carte sur mesure adaptée à ses besoins en superposant des couches de légende. Les données peuvent même provenir de sources hétérogènes. Ceci est actuellement possible par plusieurs outils, dont le GeoPortail, Google Maps, etc...

Cependant, les choix de sémiologie graphique pour chacune des données représentées sont souvent faits indépendamment des choix graphiques pour les autres données. Superposer des données aux choix sémiologiques incompatibles rend difficile la lecture de la carte (figure 1). En effet, la liberté permet aux utilisateurs de superposer un grand nombre de couches de légende, en revanche l'efficacité de la carte générée diminue avec l'augmentation des couches superposées. Or une carte ayant perdu son efficacité perd aussi son utilité. Cette mauvaise lisibilité est due au fait que les outils cartographiques n'offrent généralement pas d'analyses et de corrections des légendes. En ce sens ils ne reposent pas sur les travaux théoriques en cartographie (Bertin, 1967), (Morrison, 1974) et (Monmonier, 1991). S'ensuivent alors des problèmes de visualisation des cartes tels que le faible contraste entre deux signes voisins sur la carte, et l'opacité des signes surfaciques recouvrant d'autres signes qui deviennent alors, en partie ou en totalité, invisibles (selon leurs tailles et leurs positions vis-à-vis du signe recouvrant).

Ce constat motive notre recherche qui concerne l'amélioration de la représentation de cartes créées sur le web. Nous voulons concevoir un système de génération de cartes qui devra pouvoir analyser et reconstruire la carte dans le but de la rendre la plus compréhensible possible en respectant les règles de sémiologie, et ceci avant de la mettre à disposition de l'utilisateur. Le système devra également fournir des outils de manipulation de la carte pour faciliter la compréhension et la modification de celle-ci par l'utilisateur.

Cet article est composé de trois parties. Dans un premier temps nous expliquons les motivations de notre recherche, puis nous présentons une étude des principaux travaux précédents dans le domaine de la sémiologie graphique et de la superposition des couches de légendes. Dans la troisième partie de l'article, nous abordons l'approche choisie en proposant des solutions à la mauvaise lisibilité des

cartes surchargées. Enfin, nous concluons avec une présentation de quelques perspectives de recherche.



Figure 1. Carte représentant la région d'Ile-de-France, affichant les légendes des limites départementales, des surfaces bâties, du réseau routier et du réseau ferroviaire. (Juin 2010).

2. Motivations

Nous présentons dans cette section le domaine de la geovisualisation ainsi que les études et recherches le concernant. Nous expliquons ensuite quelques limites de certains outils de visualisation actuels, et nous présentons le but et l'orientation de notre recherche.

2.1. La geovisualisation

Depuis une décennie, nous assistons à une multiplication sensible du contenu cartographique sur internet – atlas de référence, atlas éducatifs, topographiques, nationaux et thématiques (Kraak, 2001). Ce contenu cartographique est propulsé par des sites web contenant des cartes. Chacun de ces sites web propose plus ou moins d'interactivité. Les limites constatées conduisirent à l'émergence d'un nouveau domaine de recherche : la geovisualisation.

Les cartes traditionnelles fournissent une capacité d'exploration limitée puisqu'elles sont statiques. Ces cartes peuvent être considérées comme des sources de contenu géographique puisque leur représentation graphique est étroitement liée à l'information géographique. Aujourd'hui, l'information géographique est stockée

dans des bases de données géographiques distinctes des cartes géographiques. A partir de ces bases de données, a lieu le processus de construction des cartes. Ceci offre à la carte les avantages des données numériques qui sont la facilité de manipulation, de diffusion, de stockage et de duplication. La carte n'a plus le rôle d'archivage exhaustif d'informations, l'ensemble des bases de données géographiques l'en libère (Mustière, 2001). De ce fait, les SIG actuels et la geovisualisation autorisent la construction de cartes plus interactives. Ils permettent par exemple de voir différentes couches de la carte, de zoomer ou de de-zoomer, de changer l'apparence visuelle de la carte en modifiant des couleurs ou des symboles (Jiang et al., 2003).

Ainsi, la geovisualisation est un nouveau domaine de recherche. Elle représente un ensemble de pratiques et de techniques cartographiques qui profite des possibilités des ordinateurs récents en matière de construction de cartes en temps réel, permettant ainsi à l'utilisateur de modifier les données de la carte en temps réel (MacEachren et al., 1997). Elle s'appuie sur des approches de plusieurs disciplines, dont, la cartographie, l'analyse d'image, la visualisation d'information, et l'analyse exploratoire des données, pour fournir des théories, des méthodes ainsi que des outils pour l'exploration visuelle, l'analyse, la synthèse et la présentation de données contenant des informations géographiques (Dykes et al., 2005).

Le portage des cartes sur écran permet donc d'avoir plus de libertés concernant les techniques de visualisation de ces cartes. Plusieurs études sont menées dans ce domaine. Fabrikant (2005) propose l'intégration de nouvelles variables visuelles, de variables sonores et de transitions visuelles dynamisant la visualisation de ces cartes. Robinson (2009) propose de nouvelles méthodes de mise en évidence visuelle (en utilisant la couleur, la transparence, les contours, etc...) qui peuvent être utilisées en geovisualisation.

2.2. Limites de la geovisualisation

Cependant, l'émergence de ce nouveau domaine nous amène à ressentir des besoins croissants dans plusieurs étapes, de la numérisation des informations géographiques et leur stockage, jusqu'à leur représentation sous forme graphique.

Ainsi, la séparation entre les données et leurs représentation, couplée à la liberté de l'utilisateur de construire sa carte, engendre souvent des erreurs d'ordre sémiologique et conventionnel telles que : le mauvais choix de styles pour un certain type d'objets, le faible contraste entre des signes voisins, le recouvrement de signes par d'autres signes opaques, l'utilisation de couches de données avec un niveau de détail différent pour chacune. Ces erreurs sont d'autant plus fréquentes en absence d'un cartographe traditionnel dont la tâche est d'optimiser la représentation des données cartographiques (Harrie et al., 2009). Afin d'accroître la lisibilité des cartes, des méthodes automatiques doivent être utilisées. A titre d'exemple, les outils

actuels comme le *geoportail*¹ permettent à l'utilisateur de construire une carte, mais ne proposent pas encore de méthodes d'analyses, d'application des règles sémiologiques et conventionnelles ni d'outils de corrections pour optimiser la visualisation de ces cartes.

Notre recherche s'oriente dans cette thèse vers l'automatisation de corrections de certaines erreurs dans le but d'améliorer le rendu visuel des cartes. Nous proposons donc de nouvelles techniques de visualisation de cartes, et nous les appliquons grâce à un outil qui nous permettra de tester et de valider nos différentes propositions.

3. Carte et légende

Nous présentons ici un rapide exposé de travaux déjà menés dans ce domaine. Nous définirons ce qu'est une carte et sa légende, puis nous parlerons de la sémiologie graphique, ensuite nous présenterons quelques travaux de recherche concernant l'amélioration de la lisibilité des cartes.

3.1. Principe de sémiologie graphique

Un document cartographique se compose d'une carte accompagnée généralement de sa légende qui met en correspondance un signe représenté sur la carte et sa sémantique.

Selon le (CFC, 1990), la carte est définie comme étant la « représentation géométrique conventionnelle, généralement plane, en positions relatives, de phénomènes concrets ou abstraits, localisable dans l'espace ; c'est aussi un document portant cette représentation ou une partie de cette représentation sous forme d'une figure manuscrite, imprimée ou réalisée par tout autre moyen ». Plusieurs types de cartes géographiques existent, on peut notamment citer : les cartes topographiques, les cartes géologiques, les cartes routières, ainsi que d'autres types de cartes.

La légende est la clef de la carte, son rôle est d'expliquer la signification de tous les signes que cette dernière utilise. Une bonne légende rend compréhensible la carte pour le lecteur car les correspondances signes-sens sont correctes, et les relations entre ses signes mettent en évidence des relations entre les sens de ces signes. Ainsi, un choix pertinent de légende aboutit à la construction d'une carte efficace.

Ainsi, l'étude des signes de la légende nous mène à nous intéresser et à présenter des notions de la sémiologie graphique.

La sémiologie est définie par (Saussure 1916) comme étant « la science qui étudie la vie des signes au sein de la vie sociale ». Est appelée sémiologie, toute

1. *Géoportail* : <http://www.geoportail.fr/>

science qui étudie les signes. Nous nous intéressons dans notre contexte à la sémiologie graphique.

La sémiologie graphique est une branche de la sémiologie. Elle offre un cadre théorique pour l'étude des signes, pour permettre de mieux comprendre les cartes.

Plusieurs travaux ont été faits dans ce domaine, notamment ceux de (Bertin, 1967) dans lesquels il a proposé des règles pour la sémiologie graphique en s'appuyant sur la formalisation de sept variables visuelles : la position géographique, la taille, la couleur, la valeur, le grain, l'orientation et la forme.

Ainsi, selon lui :

- La taille permet de percevoir les rapports de proportionnalité et donc de traduire des quantités et exprimer un ordre,
- La valeur est la variation de noir et de blanc dans une surface donnée, et est exprimée en pourcentage de noir. La valeur est adaptée pour traduire une relation d'ordre,
- Le grain conserve le rapport entre les éléments graphiques et le fond blanc. Sa variation différencie et ordonne les signes,
- La couleur (teinte) d'un signe est définie par son spectre. La teinte est une variable visuelle qui exprime la différenciation des figurés. La couleur seule n'est pas adaptée pour exprimer un ordre. L'ordre peut être perçu par dégradé de valeur dans une couleur,
- La forme est une variable exclusivement différentielle, elle consiste en la variation du tracé du contour des symboles,
- L'orientation est utilisée à but de différenciation, elle peut aussi indiquer une direction géographique. Elle consiste en la différence d'angles entre des champs créés par plusieurs signes parallèles.

Selon Bertin (1967), chaque variable visuelle est caractérisée par sa capacité à transmettre des relations existant entre les données et le nombre de possibilités de différenciation que cette variable visuelle offre. Les relations étant de trois types : relations de différence, d'association et d'ordre, qui lient entre elles les lignes de la légende, et qui peuvent caractériser n'importe quel ensemble de données représentées sur une carte (figure 2). La relation de différence entre des types de données géographiques est établie par la représentation de ces derniers à l'aide de signes qui se différencient par leur forme, leur type d'implantation (ponctuelle, linéaire, surfacique) et leur couleur (teinte). Nous appelons *thèmes*, ces différents types de données géographiques. Au sein d'un thème, les signes sont associés par leurs couleurs (teinte), leurs formes, et peuvent être ordonnés à l'aide de la couleur (valeur) et de la taille.

En combinant entre elles plusieurs variables, on risque fortement de diminuer l'efficacité visuelle. En effet, la mauvaise utilisation de ces variables peut entraîner des erreurs de représentation (Monmonier, 1991). Ainsi, une utilisation d'une variable visuelle peut être inadaptée à l'expression de la relation traitée, par exemple, l'utilisation de la variable forme en voulant exprimer un ordre. Un autre type d'erreurs fréquent est la mauvaise mise en œuvre d'une variable visuelle, un exemple parlant est l'utilisation de dégradés de valeur non réguliers. La couleur étant certainement la variable la plus délicate à utiliser.

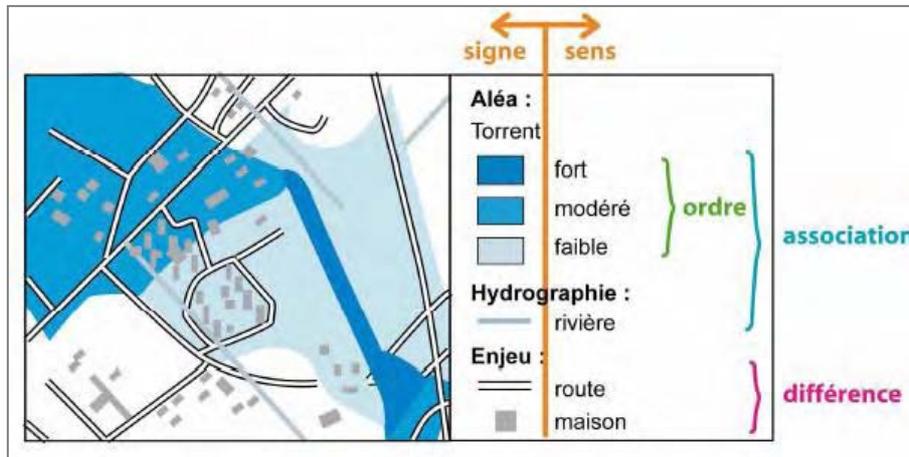


Figure 2. Mise en évidence des relations d'ordre, d'association et de différence dans une légende. Dissociation des signes et de leurs sens dans une carte. (Chesneau, 2007)

D'autres recherches ont été faites concernant ce domaine, certains auteurs considérant incomplètes les variables visuelles proposées par Bertin. Nous pouvons citer (Morrison, 1974) qui ajouta une nouvelle variable *saturation*, que Bertin incluait avec la teinte dans la variable *couleur*. En effet, grâce aux avancées techniques il devint important de séparer la variable *saturation* de la variable *couleur* puisque les outils de production de carte assistée par ordinateur permettent de contrôler indépendamment ces trois attributs de la couleur : la valeur, la teinte et la saturation.

D'autres recherches ont abouti à la proposition de nouvelles variables comme la variable *texture* de (Caivano, 1990), ou les variables *transparence*, *résolution*, et *netteté* qui sont proposées par (MacEachren, 1995).

Malheureusement, ces analyses ne sont pas souvent intégrées dans les outils standards (SIG) de géovisualisation, ce qui ouvre la porte à de nombreuses erreurs de représentation dans les cartes et donc à leur inefficacité.

3.2. Aide à la conception de légendes personnalisées

Plusieurs travaux ont été menés en vue d'améliorer l'efficacité des cartes. (Harrie et al., 2009) présentent des méthodes pour améliorer la qualité de visualisation des données cartographiques sur des geoportails. Le problème principal, selon leur étude, est l'absence de cartographe pour optimiser la présentation des données cartographiques. Leur recherche présente quatre types de méthodes pour améliorer la lisibilité des cartes : méthode sémantique, fusionnelle, de généralisation et de symbolisation. Le système décrit par (Harrie et al., 2009) doit proposer à l'utilisateur de spécifier les données qu'il souhaite afficher, l'ontologie sur laquelle ces données doivent être basées, et la symbologie que la carte finale doit utiliser. Le système pouvant faire une proposition automatique des deux derniers facteurs.

Brewer (2003) a proposé un outil qui aide les utilisateurs à trouver des couleurs efficaces pour leur carte choroplèthe. Elle produit une série de gammes colorées adaptées à leurs besoins pour l'appliquer sur leur carte. Les gammes colorées ont été construites en s'appuyant sur des relations de contrastes et sur le cercle chromatique (Itten, 1967). Dans son outil, l'utilisateur spécifie deux critères : Le nombre de classes thématiques qu'il souhaite cartographier ainsi que le type de relation entre ces classes (relation séquentielle, divergente ou qualitative). En retour, l'outil affiche les gammes correspondantes aux critères spécifiés par l'utilisateur. Cependant, cet outil est principalement destiné aux cartes choroplèthes et non aux cartes topographiques. Il ne prend pas en compte, non plus, les cas de superposition des données surfaciques, mais seulement une partition de celles-ci. En plus, la gamme colorée choisie par l'utilisateur est appliquée sur des données par défaut dans l'outil, et non sur ses propres données. Ceci ne permet donc pas d'estimer l'adéquation des couleurs avec les surfaces à représenter.

Plusieurs recherches ont été menées au laboratoire COGIT de l'IGN, elles proposent des modèles d'amélioration de la visualisation des cartes topographiques.

Chesneau (2006) a contribué à l'amélioration de la conception de légende en évaluant et en renforçant les contrastes de couleurs dans une carte donnée. Son outil évalue les contrastes de couleurs dans une carte et vérifie s'ils décrivent correctement les relations (association, différence, ordre) entre les thèmes. Le principe de son modèle consiste en l'analyse des éventuels conflits dans une carte puis leur correction. Le processus fonctionne ainsi par cycles (analyse et modification d'une couleur menant à l'amélioration des contrastes), à la fin d'un cycle, un nouveau cycle d'analyse est lancé jusqu'à ce que le résultat soit acceptable.

Buard et Ruas (2007) proposent une métrique des couleurs perçues à partir de tests d'utilisateurs. Cette métrique permet d'améliorer les calculs de contraste dans un but d'améliorer les performances des outils d'amélioration de légende. Elles proposent d'analyser les contrastes colorés entre chaque couple de ligne de légende indépendamment du positionnement du style dans la carte (Buard et al., 2009).

Les travaux de Buard et Ruas (2007) et de Buard et Ruas (2009) ont été menés dans une optique d'amélioration des résultats de Chesneau (2006). Les modèles précédents (Chesneau, 2006 et Buard et al., 2009) font leurs analyses et leurs corrections directement sur la carte, pour l'un, et sur la légende pour l'autre, sans prendre en compte les préférences des utilisateurs.

Christophe (2009) a contribué à la fabrication de légende en proposant un modèle de conception coopérative de légende : il s'agit d'aider l'utilisateur à construire des légendes cartographiquement correctes soit à partir d'échantillons de couleurs qu'il spécifie lui même, soit en lui proposant d'exprimer ses goûts concernant une palette de couleurs extraites de toiles de maîtres. Pour le second cas, la palette extraite du tableau sera appliquée sur la carte de l'utilisateur après d'éventuelles retouches. Son modèle formalise en outre les règles de cartographies conventionnelles et les règles de cartographie élémentaires (deux thèmes différents ont des teintes différentes), et prend en compte les préférences des utilisateurs.

Jolivet (2009) propose une application reposant sur des services web traduisant les étapes de réalisation d'une carte. L'utilisateur spécifie d'abord le contexte de la réalisation de la carte (zone géographique, échelle, type de carte), puis l'outil crée la légende en prenant en compte les préférences de l'utilisateur concernant les couleurs des symboles ainsi que les contraintes liées à une ontologie prédéfinie, Ontocarto, qui décrit les besoins spécifiques pour la modélisation et l'interrogation de l'information géographique. Puis il affiche la carte et la légende qui a été construite.

Notre recherche s'inscrit dans la continuité des recherches précédentes, et étudie le problème fréquent de superposition d'information.

4. Approche : Propositions d'améliorations

Dans cette partie nous présentons l'approche choisie pour proposer des premières solutions à l'éventuelle mauvaise lisibilité des cartes. Ces solutions sont une première étape dans une étude plus large concernant l'algèbre de légendes. Une des problématiques soulevées dans l'article concerne la superposition des données surfaciques provoquant alors des recouvrements de symboles. Nous proposons de tester quatre solutions simples de gestion des superpositions de couches.

4.1. Réorganisation des couches

Une première solution simple à ce problème : La réorganisation de l'ordre d'affichage des thèmes sur la carte de sorte à optimiser la surface visible de chaque couche simultanément aux autres. Nous faisons abstraction dans cette proposition de l'éventuelle sémantique des thèmes en les mettant tous au même niveau de priorité, et nous traitons exclusivement les données surfaciques. Pour cela, nous considérons que les données sont placées sur des calques invisibles superposés sur la carte. Celle-ci comportant autant de calques que de thèmes (figure 3).

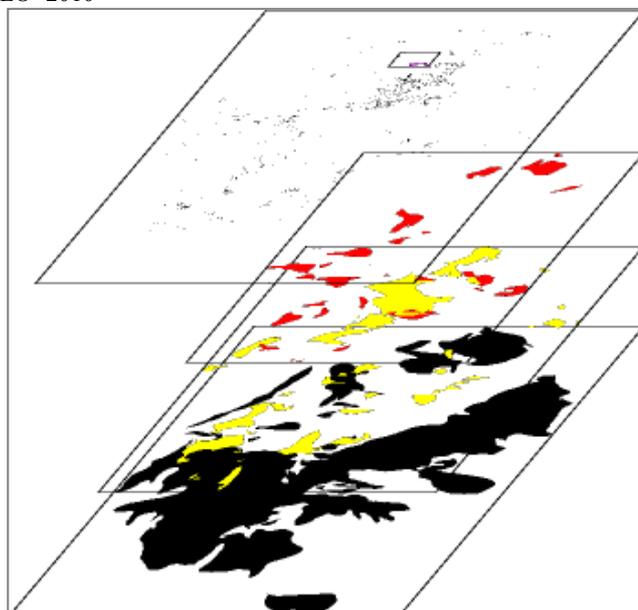


Figure 3. *Superposition de couches de légende. Chaque ensemble de données étant représenté sur un calque.*

Trois algorithmes de réorganisation ont été retenus :

- *Ordre selon l'objet ayant la plus grande surface dans chaque couche :* Dans ce cas, les couches sont organisées de telle sorte à mettre sur le calque le plus bas, la couche comportant l'objet ayant la plus grande surface en comparaison avec les objets des autres couches. Et ainsi de suite jusqu'à arriver à la dernière couche qui correspond à celle dont son plus grand objet est le plus petit comparé aux autres couches. Cette solution montre rapidement ses limites puisqu'une couche peut contenir plusieurs symboles à la fois, et l'existence d'un symbole de grande taille n'exclue pas la possibilité d'avoir d'autres symboles de petite taille susceptibles d'être recouverts par les couches disposées sur les calques plus hauts.

- *Ordre selon la surface totale de chaque couche :* Dans cet algorithme l'ordonnement des couches est déterminé par les surfaces totales respectives de celles-ci. La somme des surfaces de tous les symboles de la couche est calculée. La couche ayant la plus grande somme est placée sur le calque le plus bas, puis la couche suivante sur le calque directement au-dessus et ainsi de suite jusqu'à placer la couche ayant la plus petite somme sur le calque le plus haut.

De même que dans la proposition précédente, celle-ci présente des limites. En effet, dans le cas où on aurait deux couches, une composée de plusieurs éléments surfaciques de petites tailles et l'autre composée d'un seul élément mais de plus grande taille. Si la somme des aires des éléments de la première couche est plus grande que l'aire de la seconde, alors la seconde couche sera placée au-dessus de la première engendrant alors le risque de recouvrement des petits éléments de la première couche par le grand élément de la seconde couche.

- *Ordre selon la surface moyenne de chaque couche* : Pour cette proposition, le classement des couches se fait par la moyenne des surfaces des symboles composants chaque couche. Ainsi, la couche dont la surface moyenne de ses objets est la plus importante sera placée sur le calque le plus bas, et la couche dont la surface moyenne de ses éléments est la plus faible sera placée sur le calque le plus haut (figure 4).

La réorganisation des couches de légendes résout rarement, à elle toute seule, le problème de recouvrement d'une manière efficace. D'où la nécessité de mettre en œuvre d'autres modifications conjointement à cette première.

4.2. Automatisation du niveau de transparence

Ainsi, notre seconde proposition consiste en l'automatisation du réglage du niveau de transparence des couches de légendes. MacEachren (1995) propose la transparence comme une nouvelle variable visuelle. Cette nouvelle variable permet de visualiser un élément d'une couche au travers d'un élément d'une autre couche le recouvrant. Le problème qui en découle se rapporte à la couleur d'un élément ayant un certain niveau de transparence. En effet, la variation du taux de transparence d'un objet influence sa couleur, et des problèmes de contraste avec des objets voisins peuvent survenir. Le système devra donc pouvoir faire un compromis entre la visibilité de la couche recouverte et le taux de contraste entre les éléments.

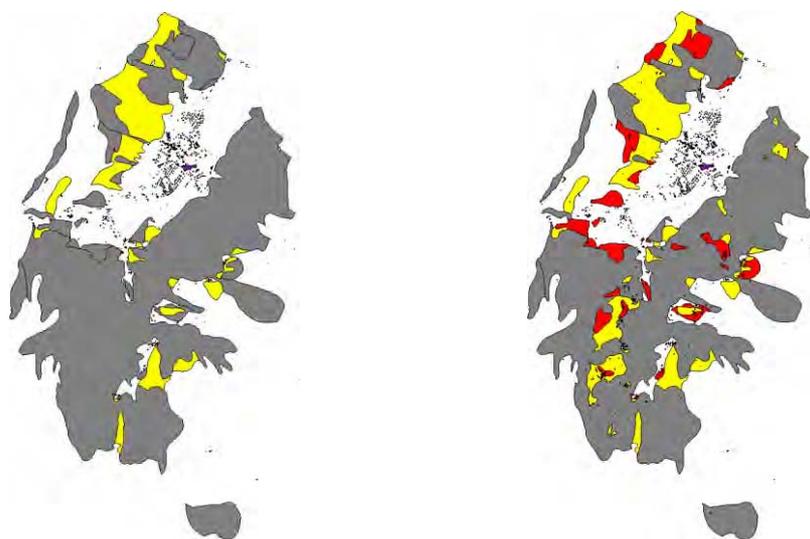


Figure 4. Carte initiale (à gauche), et la même carte après réorganisation des couches de données (à droite) en utilisant le troisième algorithme (réorganisation selon la surface moyenne de la couche).

4.3. Effet de lumière et d'ombrage

Une troisième suggestion consiste en l'utilisation de la capacité des systèmes informatiques à simuler un effet de lumière, et donc d'ombrage sur la carte. L'ombre des couches supérieures étant portée sur celles d'en-dessous, la carte s'affiche désormais avec un effet de relief. Grâce à cet effet d'ombrage, la superposition des couches devient plus facilement perceptible, et l'utilisateur aura donc plus d'aisance à la comprendre (figure 5).

Cette technique est différente de celle souvent utilisée en cartographie, l'estompage, puisque dans cette dernière l'effet d'ombrage est utilisée pour mettre en évidence un relief, donner une impression de volume, alors que la notre est utilisée pour mettre en évidence la superposition des couches de données indépendamment de la forme du terrain.

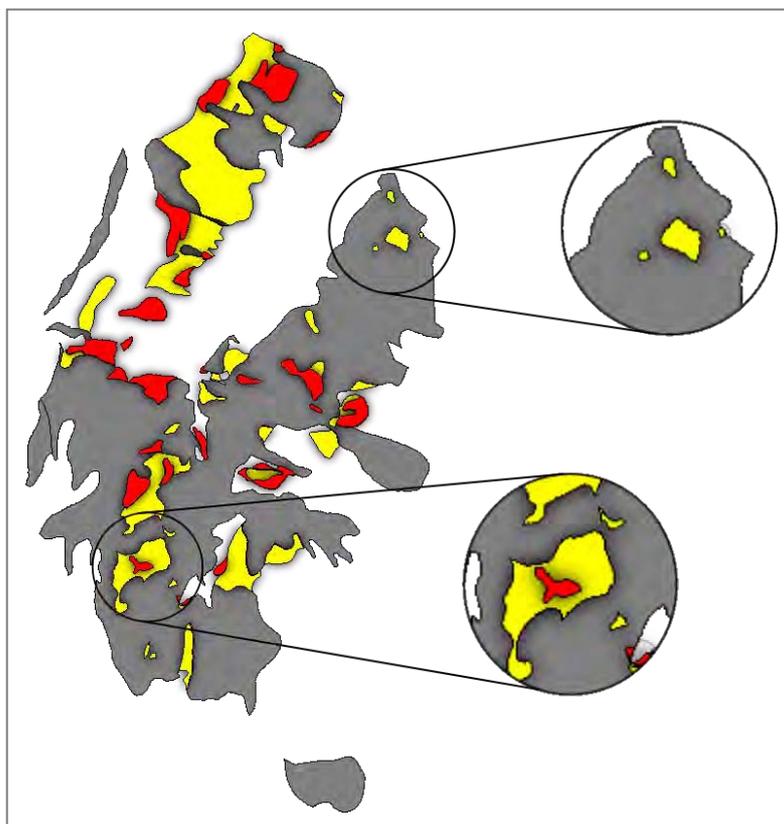


Figure 5. Utilisation d'effet d'ombrage sur une carte. La superposition des couches est mise en évidence

4.4. Construction de sous-cartes

Ces propositions peuvent apporter des solutions préliminaires au problème de co-visualisation de données sur la carte, mais dans les cas où l'utilisateur aurait choisi un grand nombre de couches de légende, il devient presque impossible de toutes les représenter sur la carte sans perdre son efficacité. Le système devrait pouvoir reconnaître ces cas extrêmes et décomposer la carte en deux ou plusieurs sous-cartes comportant chacune un sous-ensemble des couches de légende choisies par l'utilisateur et compatibles entre elles (figure 6).

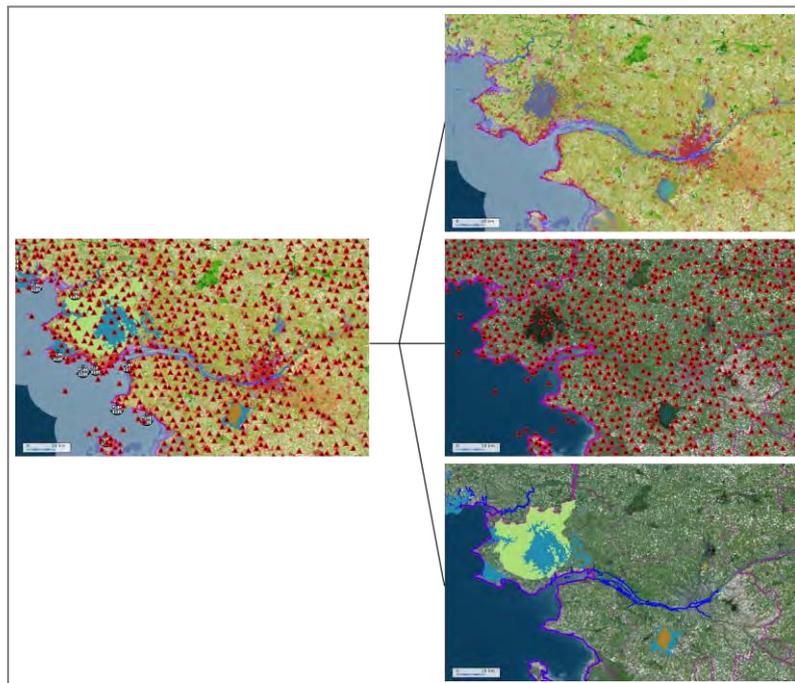


Figure 6. Décomposition de la carte initiale surchargée en n sous-cartes.

Ces propositions consistent en une première piste d'étude d'une réflexion sur la co-visualisation de légendes sur une carte. La carte porte un message cartographique, et ce message doit être compréhensible par les lecteurs de la carte. Le concepteur de la carte (l'utilisateur et le système d'aide à l'utilisateur) doit donc respecter un certain nombre de règles existantes sur l'usage de représentations graphiques et de signes cartographiques pour pouvoir être compris par le plus grand nombre de lecteurs. L'objet qui fait la correspondance entre les signes utilisés sur la carte et le sens, est la légende. En l'occurrence, lorsque l'utilisateur choisit plusieurs données à afficher sur la carte, il se peut que les légendes de ces données ne soient pas compatibles entre elles, d'où la nécessité d'une réflexion sur des opérations sur les légendes : Que peut-on avoir comme résultat en additionnant deux ou plusieurs légendes ? Et comment un nouveau thème peut s'insérer au thème existant ?

5. Architecture de test

Nous avons mis au point un outil logiciel dédié afin de tester ces différentes propositions. Notre optique étant de permettre aux utilisateurs de construire eux-mêmes leurs cartes, cet outil a été implémenté selon l'architecture client-serveur 3-tiers (figure 7). Cette architecture est facile à déployer et a l'avantage de centraliser les données et donc de pouvoir leur appliquer des mises-à-jour simplement et rapidement, ainsi que l'avantage du déchargement des postes clients vu la grande capacité de stockage disponible chez le serveur. Cet outil a été développé en java et est composé d'un serveur communicant via des sockets, et d'un client installé sur les ordinateurs des utilisateurs (figure 8). Une utilisation classique de cette architecture se fait comme suit : Le client envoie une requête au serveur, ce dernier interprète la requête et réagit en conséquence. Selon la requête, le serveur peut accéder aux bases de données mises à sa disposition, extraire les données demandées par le client, éventuellement les traiter, puis les envoyer au client. Dans notre cas, le serveur a accès aux différentes fonctions de la bibliothèque Geoxygene². Geoxygene étant une plateforme de développement 2D et 3D du COGIT. Il s'agit d'un projet open-source proposant les fonctionnalités des SIG classiques ainsi que différents travaux des chercheurs du COGIT. Le serveur récupère les données géographiques d'une base de données, les traite en faisant appel à des fonctions de Geoxygene, construit la carte, puis l'envoie par réseau au logiciel client qui l'affiche sur son interface. L'implémentation de nos propositions est faite au niveau du serveur et est donc intégrée au code de celui-ci.

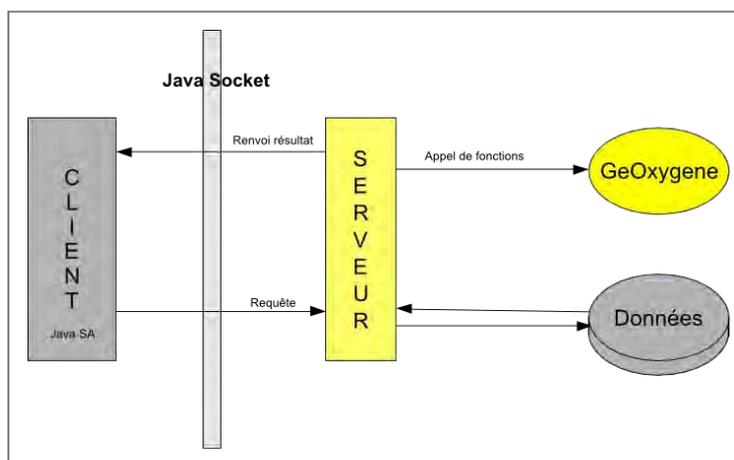


Figure 7. Schématisation de l'architecture de l'outil de test.

2. Geoxygene : <http://oxygene-project.sourceforge.net/>

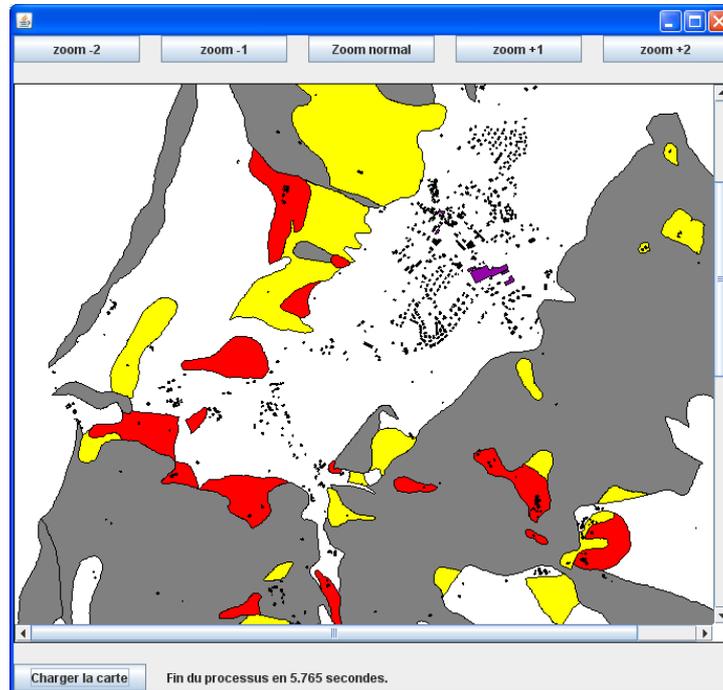


Figure 8. Capture d'écran du logiciel client affichant une carte envoyée par le serveur.

6. Conclusion et perspectives

Ce travail de thèse s'inscrit dans un processus d'amélioration de la carte pour une meilleure compréhension de celle-ci par le plus grand nombre de lecteurs. Le contexte principal étant la mise à disposition pour des utilisateurs non experts, d'un outil de conception d'une carte géographique en offrant la possibilité de combiner plusieurs données, éventuellement hétérogènes, à cartographier.

Ici, le projet principal était de mener une étude sur les travaux précédents dans ce domaine de la geovisualisation, puis d'étudier différentes méthodes d'amélioration de la lisibilité de la carte dans un but d'accroître son efficacité. Ces méthodes étant destinées à être intégrées dans un outil d'aide à la conception de cartes géographiques.

Désormais, deux axes sont étudiés en parallèle pour compléter et continuer cette recherche visant à automatiser l'amélioration de la lisibilité des cartes pour une plus grande efficacité :

- Etudier et formaliser la notion d'algèbre de légendes. La co-visualisation de données hétérogènes sur une même carte, ayant chacune leur propre symbologie, peut altérer la représentation graphique globale de la carte (et donc

son efficacité). Une piste de recherche est d'analyser et d'évaluer l'assemblage des légendes sur une même carte, ceci en essayant de répondre à la question : Qu'obtient-on en additionnant deux ou plusieurs légendes ? La notion d'opérateurs sur les légendes devra être étudiée ainsi que celle d'indices de visibilité destinée à mesurer le taux de lisibilité de la carte.

- Modéliser les propositions faites, en un outil d'aide pour l'utilisateur en ligne. Le système proposerait à l'utilisateur une interface web pour définir la région à cartographier et les données qu'il voudrait représenter ainsi que ses préférences graphiques. Le système analysera alors les éventuelles erreurs sémiologiques pour les corriger et améliorer la lisibilité globale de la carte.

5. Bibliographie

- Bertin J., *Sémiologie graphique: les diagrammes, les réseaux, les cartes*, Première édition en 1967 puis en 1973, 1998, Paris, Editions de l'EHESS, 1967.
- Brewer C. A., « A transition in improving maps: The colorbrewer example », *Cartography and Geographic Information Science*, vol.30, no.2, 2003, p. 159-162.
- Buard, E. et Ruas, A., « Evaluation of colour contrasts by means of expert knowledge for on-demand mapping », In: *23th ICA conference*, 4-10 August 2007, Moscow.
- Buard E., Ruas A., « Processes for improving the colours of topographic maps in the context of Map-on-Demand », *24th International Cartographic Conference (ICC'09)*, November 15th-21st, 2009, Santiago (Chile).
- Caivano, J. L., « Visual texture as a semiotic system », *Semiotica* 80(3/4), pp. 239-252, 1990.
- CFC, Comité français de Cartographie, 1990.
- Chesneau E., *Modèle d'amélioration automatique des contrastes de couleurs en cartographie*, Thèse de doctorat, Université de Marne La Vallée, 2006.
- Chesneau E., « Etude des contrastes de couleur pour améliorer la lisibilité des cartes », *8èmes rencontres ThéoQuant*, 10-12 janvier 2007, Besançon.
- Christophe S., *Aide à la conception de légendes personnalisées et originales : Proposition d'une méthode coopérative pour le choix des couleurs*, Thèse de doctorat, Université de Marne La Vallée, 2009.
- Dykes J., MacEachren, A.M. and Kraak, M.J., « Exploring Geovisualization », In: *Exploring geovisualization*, 2005, p. 3-19.

- Fabrikant S.I., « Towards an understanding of geovisualization with dynamic displays ». *Proceedings, American Association for Artificial Intelligence (AAAI) 2005 Spring Symposium Series: Reasoning with Mental and External Diagrams: Computational Modeling and Spatial Assistance*. Stanford University, Stanford, CA, March 21-23, 2005, p. 6-11.
- Ferdinand de Saussure, *Cours de linguistique générale*, 1916.
- Harrie L., Mustière S., Stuckenschmidt H., Stigmar H., « Cartographic aspects of geoportals». *Presenting Spatial Information: Granularity, Relevance, and Integration*, Workshop at COSIT'09, september 21st, 2009, Aber Wrac'h (France)
- Itten J., *Art de la couleur*. Edition originale en 1961 : *Kunst der Farbe*. Traduction en français en 1967, puis en 2001. Ed. Dessain et Tolra, 1967
- Jiang, B., Huang, B., and Vasek, V., « Geovisualisation for Planning Support Systems », In: *Planning Support Systems in Practice*, Geertman, S., and Stillwell, J. (Eds.). Berlin: Springer, 2003.
- Jolivet L., « Characterizing maps to improve on-demand cartography - the example of European topographic maps ». *Actes de la conférence GISRUK*, 2009, Royaume-Uni, Durham.
- Kraak M. J., « Web maps and atlases », in Kraak M. J., Broon A. (eds.), *Web cartography. Developments and prospects*, 2001, New York, Taylor & Francis.
- MacEachren A.M. and Kraak M.J., « Exploratory cartographic visualization: advancing the agenda ». *Computers & Geosciences*, Volume 23, Issue 4, May 1997, Pages 335-343
- MacEachren A., *How maps work : Representation, Visualization, and Design*. Guilford Publications, New York, 1995.
- Monmonier M., *How to lie with maps*, University of Chicago Press, 1991.
- Morrison, J. L., A theoretical framework for cartographic generalization with emphasis on the process of symbolization. *International Yearbook of Cartography*, 14, pp. 115-127, 1974.
- Mustière S., Apprentissage supervisé pour la généralisation cartographique, Thèse de doctorat, Université Pierre et Maris Curie, Paris VI, 2001
- Robinson A. C., « Visual highlighting methods for geovisualization », *24th International Cartographic Conference (ICC'09)*, November 15th-21st, 2009, Santiago (Chile).

Les randonneurs définissent leurs cartes

Exploitation d'une enquête semi-directive à questions ouvertes avec des outils statistiques et linguistiques

Catherine Domingues*, Priscille Baldit-Schneller**

**Laboratoire COGIT – Institut Géographique National
73 avenue de Paris
F-94160 Saint-Mandé
catherine.domingues@ign.fr*

***TIM/INaLCO
49 bis avenue de la Belle Gabrielle
F-75012 Paris
priscille.schneller@yahoo.fr*

RÉSUMÉ. Dans le contexte de la conception de carte sur mesure, cet article présente l'exploitation d'une enquête réalisée auprès de différents groupes de randonneurs afin qu'ils définissent les spécifications de cartes adaptées à leurs pratiques de randonnée. Les différents entretiens portent sur la sélection des données et leurs modes de représentation. Ils ont été contrastés afin de dégager les points, notions et concepts importants pour les randonneurs ainsi que les attentes et critiques de ceux-ci par rapport à la conception de supports de randonnée adaptés. Ils ont finalement conduit à la définition de profils-types de randonneurs. L'exploitation de cette enquête combine des outils statistiques simples : recherche des termes significativement fréquents, et des outils linguistiques qui permettent de préciser les résultats statistiques.

ABSTRACT. In the context of on-demand map making, this paper presents how to use interviews carried out with trekkers. The aim of the survey was to specify maps suitable to trekkers. The different interviews are about data selection and data portrayal. A comparative study has been made on the basis of the interviews in order to extract the items and the concepts considered as important for trekkers and their expectations and criticisms about trekking maps. Finally, four different trekkers' profiles have been defined. In this survey's analyse, simple statistical tools (i.e. search for abnormally frequent words) combine with natural language processing tools which make it possible to specify the statistical results.

MOTS-CLÉS : cartographie, TALN, modélisation des connaissances, carte de randonnée.

KEYWORDS: cartography, NLP, knowledge modelling, trekking map.

Introduction

Le besoin de carte sur mesure se manifeste dans de nombreuses occasions. Dans ce contexte, l'IGN souhaite proposer des cartes adaptées à différentes pratiques de la randonnée : promenades occasionnelles ; randonnées à pied, en vélos, en VTT. Il s'agit de spécifier, par la sélection des données représentées sur la carte et par les choix de représentation de ces données, une carte adaptée aux besoins de ces quatre types de randonneurs. Dans ce but, une enquête a été commanditée par l'IGN.

Nous avons souhaité, dans le cadre d'un stage de master 2, tester des méthodes et des outils de traitement automatique du langage naturel (TALN) pour exploiter cette enquête et en particulier explorer l'expression du ressenti exprimé par les randonneurs. Il s'agissait d'extraire les objets, notions et concepts importants pour les randonneurs selon leurs pratiques de la randonnée ainsi que leurs commentaires, attentes et critiques quant à la représentation de ces objets sur les supports de randonnée qu'ils utilisaient.

La démarche mise en œuvre ici s'appuie sur un corpus constitué de la transcription des entretiens menés lors de l'enquête et sur l'exploitation de ce corpus à l'aide d'outils de TALN. L'objectif de cet article est de présenter cette démarche et les résultats obtenus. Les différentes étapes sont schématisées sur la figure 1. L'enquête ainsi que la méthode de transcription pour la constitution d'un corpus textuel sont développées dans la partie 1. Dans la partie 2, sont détaillés la méthode et les outils employés pour identifier les objets, notions et concepts importants pour les randonneurs selon leurs pratiques de la randonnée ainsi que les résultats obtenus. Différentes méthodes ont été utilisées pour mettre en évidence les critiques, attentes et commentaires des randonneurs ; elles sont exposées dans la partie 3. Leurs résultats sont présentés dans la partie 4. La synthèse de ces commentaires permet de dégager des profils-types de randonneurs selon leur pratique de la randonnée qui sont proposés dans la partie 5, avant les conclusions (partie 6).

1. Présentation de l'enquête et constitution du corpus de travail

Les randonneurs interrogés sont regroupés en fonction de leur mode de randonnée : randonneurs/promeneurs occasionnels (moins de huit randonnées par an quel que soit le mode de randonnée) ; randonneurs pédestres fréquents (sans aucune indication du nombre de randonnées) ; randonneurs à vélo, randonneurs en VTT (au moins cinq randonnées par an pour les participants de chacun de ces groupes). Quatre groupes de neuf à dix participants sont ainsi constitués.

Chaque groupe est réuni pour un entretien collectif d'environ trois heures (étape de la figure 1), conduit par un animateur, le même pour les quatre groupes. Les entretiens sont guidés par l'animateur ; celui-ci dispose d'un guide d'animation qui

récapitule l'ensemble des questions auxquelles l'entretien est censé apporter des réponses. Cependant le format des réponses est libre : les participants peuvent adopter le vocabulaire et la syntaxe qui leur conviennent ; ils peuvent aussi aborder des thèmes qui ne figurent pas explicitement dans les questions. Questions et réponses composent une enquête qualitative à partir de questions ouvertes (Brugidou et al., 2004 ; Gélinas-Chebat et al., 2004).

Les entretiens sont enregistrés puis retranscrits par écrit par l'animateur du groupe ; cette retranscription constitue le corpus de travail.

Pour que ce corpus puisse être manipulé avec des outils de TALN, il doit être prétraité. Ce prétraitement (Heitz, 2006) recouvre différentes tâches, en particulier :

- la correction des interprétations et des fautes de frappe et d'orthographe engendrées lors de la transcription des entretiens oraux ; par exemple :

<i>coordonnées lombaires</i>	doit être corrigé en	<i>coordonnées Lambert</i>
<i>ballade</i>		<i>balade</i>
<i>point côté</i>		<i>point coté</i>
<i>déclinité</i>		<i>déclivité</i>
<i>bonen</i>		<i>bonne</i>

- la normalisation de la graphie. Pour que les outils de TALN puissent opérer, il est nécessaire d'unifier les différentes désignations d'un même objet. Par exemple, les randonneurs font souvent référence à l'échelle de la carte mais utilisent différentes formulations comme : *la 25*, *la carte au 1/25* qui sont alors remplacées dans le corpus par la forme unique *la carte au 1:25 000*.

- la segmentation du texte. Pour permettre la comparaison entre les réponses des différents locuteurs à la même question ou les réponses d'un même locuteur à l'ensemble des questions, des jalons textuels sont insérés qui repèrent et segmentent chacune des questions et les réponses correspondantes.

Ces tâches sont réalisées manuellement mais guidées par des outils lexicaux. Par exemple, les mots contenant des fautes de frappe ressortent, le plus souvent, comme inconnus pour les dictionnaires électroniques utilisés par les outils de TALN. C'est de cette manière que peuvent être repérées les erreurs de frappe comme *déclinité* ou *bonen*. Les erreurs qui reviennent à mettre un mot à la place d'un autre : *lombaire* pour *Lambert* ou *point côté* pour *point coté* sont plus difficiles à détecter automatiquement parce qu'elles sont liées au sens des mots dans un contexte, notion plus difficilement accessible à des outils automatiques. Cette phase correspond à l'étape ② de la figure 1.

Finalement, le corpus prétraité se compose de quatre sous-corpus (un par entretien et donc par type de randonnée) et compte plus de 46 000 mots, soit environ une centaine de pages.

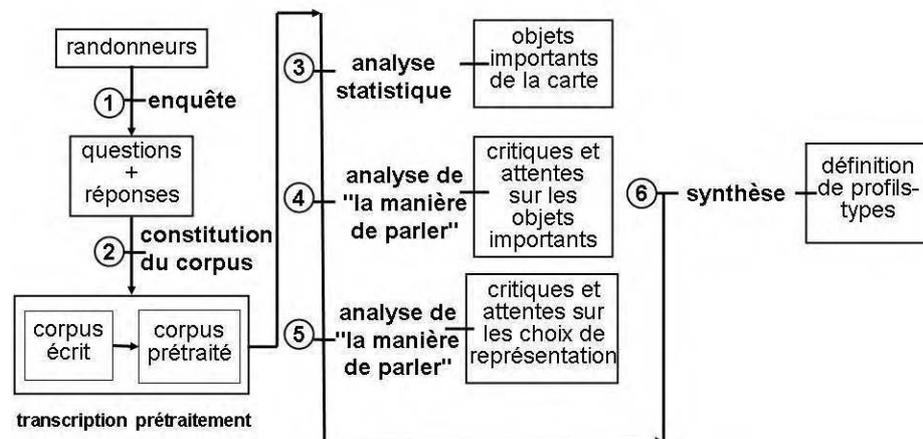


Figure 1. Les différentes étapes de la démarche.

2. Recherche des objets importants

L'enquête fournit des sous-corpus différenciés ; le repérage des objets, notions et concepts importants pour les randonneurs selon leurs pratiques de la randonnée est donc possible.

Pour cela, une analyse statistique et contrastive du vocabulaire employé (Tomasetto et al., 2008) est mise en place comme le montre l'étape ③ de la figure 1. Cette méthode relève de l'analyse lexicale descriptive (Labbé, 2000 ; Beaudoin, 2000) et s'appuie ici sur l'outil Lexico¹. Les mots « significativement fréquents » sont recherchés dans chaque sous-corpus : un mot est dit caractéristique d'une partie si sa fréquence est « anormalement élevée » dans cette partie. Cette fréquence s'observe par rapport à un seuil de spécificité et à un corpus de référence.

Ici, les parties étudiées sont successivement chacun des sous-corpus et le corpus de référence est formé de l'ensemble des quatre sous-corpus. La structure grammaticale des sous-corpus et celle du corpus de référence étant comparables, les mots-outils tels que les prépositions, déterminants, auxiliaires grammaticaux, conjonctions de coordination ou de subordination qui sont fréquents aussi dans le corpus de référence ne peuvent apparaître dans les sous-corpus comme des mots significativement fréquents. Enfin, le corpus de référence ayant pour thème la randonnée, les mots appartenant au domaine de la randonnée mais non spécifiques d'une pratique sont aussi éliminés.

¹ <http://www.cavi.univ-paris3.fr/Ilpga/ilpga/tal/lexicoWWW/>

La comparaison statistique repère, dans chaque sous-corpus, les termes significativement fréquents qui sont présentés en figure 2 puis commentés dans la partie 4.

Pédestres	Vététistes	Cyclistes	Occasionnels
[carte au] 1:50 000	[carte au] 1:25 000	chevron/chevrons	balade/balades
chargée	cartes	équipements	compliqué <e s es>
chemin	CD-rom	forêt/forêts	facile
dénivelé/dénivelés	clarté	Internet	fouillis
kilométrage	confiance	kilomètres	itinéraire/itinéraire
[une vue] macro	courbe/courbes [de niveau]	logiciel	jeune
pédestre	GPS	Michelin	légende
randonne	GR	pente	massif
repérage	information	route/routes	nous
sentier/sentiers	Libris	vélo	points [de repère]
terrain	végétation	ville/villes	sympa
vous	VTT	visible	trop

Figure 2. Termes spécifiques de chaque sous-corpus. Ils ont été regroupés thématiquement (en utilisant le guide d'animation ou des connaissances en cartographie, externes à l'enquête). Les cases relevant d'un même thème sont représentées avec la même texture.

	définition de l'activité
	support de randonnée
	repérage, balisage
	déclivité
	distance
	aspect visuel de la carte

3. Méthodes et outils pour identifier critiques et attentes des randonneurs

Pour repérer les critiques et les attentes des différents types de randonneurs, différents outils peuvent être mis en œuvre isolément ou combinés, qui sont présentés ici. Les résultats concernant les notions importantes ainsi que leur représentation cartographique sont présentés en partie 4.

3.1. Recherche du contexte des termes pertinents

Dans le contexte de spécifications de cartes adaptées à la randonnée, identifier les commentaires des randonneurs concernant les objets, notions et concepts importants pour leur pratique de la randonnée est pertinent. Ces éléments ont été identifiés dans la partie 2 ; cependant une étude en contexte est nécessaire pour

préciser leurs voisinages, cette étude s'appuie sur la construction de patrons morpho-syntaxiques².

Ces patrons sont construits à partir des lemmes³ désignant les notions importantes et permettent d'extraire leurs contextes droit et gauche sous forme de concordances⁴. Dans l'exemple de la figure 3, le patron repère les termes correspondant à la notion de déclivité. Les termes recherchés sur ce thème sont issus de l'étude de la partie 2. Les chevrons qui encadrent les termes indiquent que ce sont les lemmes qui sont recherchés. Par exemple <courbe> repère les formes : *courbe* et *courbes*. La première ligne de l'automate de la figure 3 repère les suites de termes où <courbe> est suivi à une distance maximale de deux mots de <niveau> (i.e. *niveau* ou *niveaux*).

Les concordances suivantes sont extraites du corpus des vététistes :

*{S}si on va en montagne les courbes de niveau sont indispensables. {S} La v
nnotations, on peut sortir le dénivelé et la vue en 3D. {S}Pour randonner {
que la distance des courbes de niveaux à 10 mètres ne s'impose pas pour un c
ur la GR, l'avoir sur la B {S}Pentes {S}6 ils ont copié les Michelin {S} s*

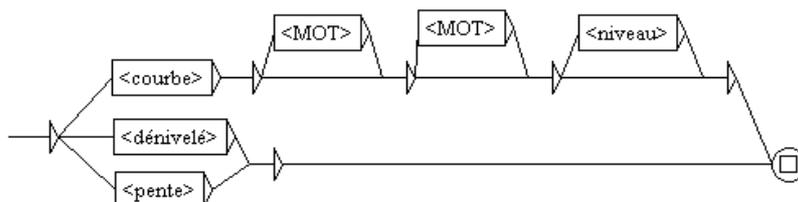


Figure 3. Patron morpho-syntaxique construit sur le thème de la déclivité (représentation graphique proposée par le logiciel Unitex⁵).

3.2. Recherche des marqueurs modaux d'opinion

Les modalités élocutives « précisent la manière dont le locuteur révèle sa position vis-à-vis du propos qu'il énonce » (Charaudeau, 1992). Pour identifier les critiques et les attentes des randonneurs, une méthode consiste à repérer les marqueurs modaux

² Un patron morpho-syntaxique est un automate qui permet de repérer une suite de termes dans un texte.

³ Un lemme est une forme graphique choisie conventionnellement comme adresse d'un mot dans un lexique, en particulier les dictionnaires.

⁴ Une concordance est un index de mots présentés avec leurs contextes droit et gauche tronqués à un nombre arbitraire et fixe de caractères.

⁵ www-igm.univ-mlv.fr/~unitex/

qui indiquent une prise de position du locuteur (étape ④ de la figure 1). Les marqueurs pertinents sont :

- des verbes psychologiques (Mathieu, 1997) : *agacer, contrarier, décevoir, déranger, ennuyer, étonner, excéder, inquiéter, intéresser, passionner, rassurer, satisfaire, surprendre* ;
- ainsi que des verbes d'opinion (Généreux *et al.*, 2007) : *apprécier, demander, désirer, manquer, nécessiter, préférer, satisfaire, souhaiter*.

Ces marqueurs ont été repérés dans les corpus à l'aide de patrons. Les formes les plus fréquemment relevées sont les suivantes :

- pour introduire une opinion : *je pense que, je trouve que, pour moi* ;
- pour exprimer une préférence : *je préfère, <avantage>* ;
- pour exprimer un souhait : *j'aimerais bien que, si ... ce serait ...* ;
- pour exprimer une critique : *<être> trop, pas assez* ;
- pour exprimer une évaluation : *j'aime bien, c'est bien* pour une satisfaction complète ; pour une satisfaction moyenne : *c'est pas mal, il n'y a pas* ; pour un manque de satisfaction : *(il) manque, j'aime pas/je n'aime pas*.

Pour limiter le bruit (i.e. l'extraction d'informations non pertinentes) et le silence (i.e. l'échec dans le repérage d'informations pertinentes), la mise en évidence des termes pertinents et l'identification des marqueurs modaux ont été combinées : rechercher des séquences qui contiennent un marqueur modal d'opinion et dont l'objet désigne un objet ou une notion important pour les randonneurs (et repéré dans la partie 2).

3.3. Recherche des modificateurs de <carte>

Repérer les modificateurs associés au lemme <carte> permet d'identifier les commentaires des locuteurs sur la carte en général. Les modificateurs ont été recherchés à l'aide de patrons permettant de reconnaître des modificateurs de structures complexes comme par exemple *carte <adjectif> de <nom>* ou contenant des coordinations : par exemple, *carte <adjectif> et <adjectif>* (Dominguès, 2001). Cependant, les concordances obtenues montrent que les modificateurs associés à *carte* dans le corpus sont de structure simple, ce qui est cohérent avec le fait que le corpus d'enquête est la transcription écrite d'une conversation orale. Les concordances suivantes ont été extraites du corpus des randonneurs occasionnels à l'aide du patron de la figure 4 :

*s {S}1 un ami avait sorti une carte en temps et non pas en kms, ça n'a pas
{S}J'aime bien. {S}8 ça fait carte pour débutant {S}Du coup peut-être qu'i
de chez IGN, qui sont des cartes anciennes, sur lesquelles il n'y a pas d'infor
me, ça fonctionne ça gêne {S} carte simplifiée {S} ça monopolise mon attenti
, on s'en sert beaucoup de la carte papier, si on l'enlève il va rester quoi*

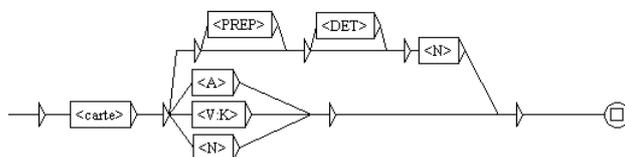


Figure 4. Patron permettant de repérer les modificateurs de <carte> de la forme : préposition nom (<PREP><N>) pour la première ligne, adjectif (<A>) pour la deuxième ligne, participe passé (<V:K>) pour la troisième, nom (<N>) pour la dernière.

4. Résultats de l'extraction d'information

Les méthodes et outils décrits dans la partie précédente ont été appliqués au corpus pour identifier les critiques, les attentes et les commentaires des randonneurs concernant les objets, notions et concepts importants, ainsi que leur représentation.

4.1. Critiques, attentes et commentaires des randonneurs sur les objets importants

Les commentaires des randonneurs sur les objets, notions et concepts importants identifiés (à l'aide des outils présentés dans la partie 3) sont présentés en utilisant la thématique mise en place dans la figure 2.

4.1.1. Définition de l'activité

Les randonneurs occasionnels *font de la balade* ou *des balades*, alors que les randonneurs pédestres *randonnent* ou *font de la randonnée*. Les vététistes et les cyclistes parlent d'une manière significativement fréquente de leur vélo.

4.1.2. Support de randonnée

Les randonneurs pédestres utilisent la carte au 1:25 000 et éventuellement celle au 1:50 000. Ils font, beaucoup plus que les autres randonneurs, allusion à ces deux échelles (la référence fréquente à l'échelle 1:50 000 est probablement due au fait que l'enquête était orientée principalement sur les usages de cette carte). Ce type de randonneur a l'habitude d'utiliser les cartes : il y a trois fois plus d'occurrences concernant l'échelle dans leur sous-corpus que dans celui des randonneurs occasionnels.

4.1.3. Repérage, balisage

Les termes identifiés sur ce thème ne renvoient pas tous au même objet mais sont sémantiquement assez proches. La notion de balisage s'ajoute à celle d'infrastructure utilisée pour la randonnée ; elle est inhérente à des termes comme *circuit*, *GR*, *circuit cyclotouristique*, *piste* ou ajoutée sous la forme du modifieur *balisé* à des termes comme *sentier* ou *chemin* pour former les composés : *sentier balisé* ou *chemin balisé* qui figurent dans le corpus.

Le syntagme⁶ *point de repère* est particulièrement utilisé chez les randonneurs occasionnels, ce qui s'explique par le peu d'expérience qu'ils ont de la lecture des cartes. Ils cherchent à avoir des repères cartographiques qui soient clairement reconnaissables sur le terrain. Un randonneur occasionnel parle des « *grands randonneurs* » en disant : « *[ils ont] besoin de moins de points de repère que nous* ». Le repérage sur la carte est utilisé par les randonneurs aussi bien pour la préparation de la randonnée que pendant son déroulement.

Les quatre catégories de randonneurs pensent que le quadrillage est inutile. Seul un randonneur cycliste y trouve une utilité, mais conseillerait d'y ajouter un lexique des noms de villes et de lieux-dits : *ce quadrillage sert juste à repérer, et de rajouter un lexique de lieux-dits, ça serait bien.*

4.1.4. Déclivité

La déclivité d'un parcours constitue un des éléments les plus importants pour les cyclistes. Elle est identifiée par la pente qui, sur les cartes IGN, n'est pas indiquée mais figurée par l'ombrage des reliefs et des courbes de niveau. Le terme *pente* apparaît dans le corpus des cyclistes, justement parce que ceux-ci déplorent cette absence ; en parallèle, ils emploient avec une fréquence significativement élevée le terme *Michelin*. En effet, sur ses cartes, Michelin utilise les chevrons pour figurer le sens et la raideur de la pente.

4.1.5. Distance

Les randonneurs qui pratiquent sur de grandes distances, cyclistes, vététistes et quelques randonneurs pédestres, soulignent le besoin d'indiquer des distances sur les cartes. Cette information ne figure pas pour le moment sur les cartes IGN. Une indication de la difficulté permettrait également de guider les randonneurs dans le choix de leur parcours : [corpus des cyclistes] *il faudrait la difficulté et les distances, donner un max d'indications sur les distances.*

4.1.6 Les cartes d'autres éditeurs

Les randonneurs fréquents citent le plus souvent les cartes IGN, contrairement aux cyclistes qui évoquent davantage les cartes Michelin. Les vététistes mentionnent

⁶ Ici, un syntagme est une combinaison de mots qui se suivent.

également les cartes Libris⁷ et Didier Richard⁸. Les randonneurs occasionnels les utilisent beaucoup moins mais nomment les cartes IGN, Didier Richard et Michelin :
léatoire. {S}R. Support. {S}9 Michelin a un support CD rom, je l'avais pour s. {S}5 elle me parle plus la Didier Richard, c'est plus facile d'accès. {S

4.2. Résultats de la recherche des critiques, attentes et commentaires des randonneurs sur la représentation des objets importants

L'aspect visuel de la carte et sa lisibilité sont évoqués par tous les groupes (cf. symbole  dans la figure 2). D'après les travaux de (Bertin, 1967), la lisibilité d'une carte est, notamment, améliorée par le choix des couleurs et de leurs contrastes, et d'un lettrage adapté (taille et police) ; elle est amoindrie si la quantité d'informations est trop importante. Ces différents aspects sont évoqués par les randonneurs ainsi que la composition de la légende.

4.2.1. Sélection des informations

Une des principales critiques formulées à l'égard de la carte au 1:50 000 est le fait qu'elle soit trop chargée : *elle me fait mal aux yeux la carte au 1:50 000. Il y a trop d'informations, c'est trop condensé*

Une légende trop chargée ne plaît pas aux randonneurs qui souhaitent pouvoir accéder à l'information pertinente rapidement (ils se trouvent la plupart du temps sur le terrain, et même sur un vélo pour les cyclistes et les vététistes). Cependant l'expérience des lecteurs influe aussi sur leur perception de la légende. Pour des randonneurs occasionnels (qui ont donc peu l'habitude de lire des cartes), améliorer la lisibilité signifie le plus souvent diminuer la quantité d'information cartographiée. Au contraire, des lecteurs plus expérimentés apprécient une plus grande quantité d'information à condition que les choix de représentation mettent en évidence les objets cartographiques importants pour la randonnée :

[cyclistes] là l'autoroute est moins mise en valeur [...] , elle est comme une bonne route, c'est une bonne chose.

4.2.2. Les couleurs

Les randonneurs émettent des critiques sur le choix des couleurs et son efficacité quant à la compréhension de la carte. Ces commentaires paraissent davantage liés à la facilité de lecture qu'à des préférences individuelles des randonneurs.

Pour beaucoup de randonneurs, le défaut de la carte au 1:50 000 est justement son manque de couleurs :

[cyclistes] c'est gris ce n'est pas lisible au premier abord

⁷ Les cartes Libris sont très détaillées et contiennent les circuits pédestres en couleur, avec des icônes indiquant les hébergements et des informations touristiques.

⁸ Les cartes Didier Richard ne délivrent pas d'informations touristiques, mais les abris sont indiqués et les passages difficiles mis en valeur.

[pédestres] *C'est hyper chargé, la présentation d'aspect est vachement fade*

L'amélioration des contrastes de couleurs (Buard et al., 2009) permettrait d'améliorer la lisibilité de la carte :

[pédestres] *1 Là, c'est pareil, tout ce qui est églises, châteaux, mettre de couleurs différentes, pour que ça accroche*

[occasionnels] *pour les sentiers, là c'est un petit trait noir tout fin, moi il me faudrait un trait rouge, qu'on les voie tout de suite*

[cyclistes] *c'est fouillis parce qu'il y a trop de vert. Il faut vraiment visualiser pour voir.*

[occasionnels] *sur Didier Richard : Il y a moins de rouge que sur l'autre. C'est plus lisible*

[pédestres] *on confond les sentiers et les nationales, pour moi rouge c'est les nationales, il faut que ce soit nouveau, un violet, on risque de confondre après avec le relief*

Les randonneurs apprécient aussi l'utilisation des couleurs conventionnelles (le bleu associé à l'hydrologie, le vert à la végétation, ...) pour la représentation des objets cartographiques :

[cyclistes] *Pour la symbolisation s'il faut se reporter à chaque fois à la lecture de la légende, on ne le fait plus ..., il faudrait que le jeu des couleurs fasse penser à ce qu'on va trouver.*

4.2.3. Composition de la légende

La lecture de la légende pose des problèmes aux randonneurs occasionnels par la complexité des libellés et la taille des caractères :

C'est un peu du chinois la légende.

On se dit quoi quand on voit des termes pointus ? ça nous concerne pas.

La légende est écrite trop petite.

Quand on regarde la légende, il y a 5 points noirs qui ont une signification différente.

Pour les cyclistes aussi, les caractères de la légende sont trop petits :

Si la légende était plus lisible, ce serait un avantage, en caractères plus gros.

Les randonneurs pédestres préconisent de modifier l'ordre de présentation des objets cartographiques dans la légende afin de mieux tenir compte de l'importance des objets dans le contexte de la randonnée :

Il faut changer l'ordre de la légende, ne pas parler de la route en premier.

Ca fait trop carte routière

5. Synthèse : définition de profils types

La mise en évidence, dans chaque sous-corpus, des objets importants pour les randonneurs ainsi que leurs attentes et critiques permet d'élaborer un profil type selon la pratique de la randonnée.

5.1. Les randonneurs pédestres

Ces randonneurs connaissent le vocabulaire technique de la cartographie et utilisent souvent des cartes, en particulier celles de l'IGN. Ils empruntent ou non les circuits de randonnée existants.

Ils utilisent principalement les cartes au 1:25 000 sur lesquelles les chemins de randonnée sont indiqués et les cartes 1:50 000 pour le repérage préalable de leur randonnée dont ils ont ainsi une vue d'ensemble.

La version actuelle de la carte au 1:50 000 est perçue comme chargée, ce que certains de ces randonneurs apprécient toutefois puisque cela signifie que la carte contient beaucoup d'informations. Ils souhaiteraient même en avoir davantage : sur la randonnée (et non pas uniquement l'indication des GR), des informations touristiques (auberges, sites à visiter, églises...) et des indications pratiques (emplacement des parkings, points de repères, ...). Les choix de symbolisation devraient alors garantir la mise en valeur des objets cartographiques concernés.

5.2. Les randonneurs à vélo

Les randonneurs à vélo pratiquent principalement sur route ou sur les chemins en forêt.

Ils utilisent des cartes routières comme la carte 1:100 000 de l'IGN ou la carte Michelin au 1:250 000 puisqu'ils parcourent de grandes distances. Certains utilisent la carte cyclotouriste de l'Ile de France, à l'échelle 1:100 000, dont la mise en page recto-verso est très appréciée. Les cyclistes utilisent également des cartes plus précises et détaillées, comme la 1:25 000 (IGN ou Michelin), notamment pour avoir une vue de détail lorsqu'ils ont déjà une vue d'ensemble du trajet complet ou pour les parcours moins longs. L'échelle la plus adaptée serait le 1:50 000 mais elle est assez difficile à trouver dans le commerce. En outre, la carte IGN actuelle à cette échelle n'est pas assez lisible. Il faudrait alléger la légende et accentuer le contraste entre les couleurs. Les coordonnées Lambert ne sont pas utiles.

Les cyclistes apprécient les indications de relief, comme les courbes de niveau, mais ne les trouvent pas toujours assez lisibles sur les cartes IGN. Ils préconisent d'indiquer le pourcentage de la pente à l'aide de chevrons, comme sur les cartes Michelin. Ils apprécient les indications de distance, comme sur certaines cartes au

1:100 000. La carte idéale devrait aussi indiquer la fréquentation des routes, afin de pouvoir éviter celles très fréquentées ; la légende permettrait de distinguer les voies carrossables des voies semi-carrossables, et les autoroutes des routes secondaires. L'épaisseur du trait pour les autoroutes et les nationales pourrait être diminuée afin de mettre en valeur les chemins et circuits de randonnée. Toutefois les *points de raccordement* (nœuds routiers) devraient être représentés avec précision, *parce qu'il faut les franchir*. Les réparateurs de vélo pourraient aussi être indiqués.

La carte doit donc délivrer des informations variées tout en restant lisible, et sa légende mettre en valeur les infrastructures routières, le relief et les informations touristiques.

5.3. Les randonneurs en VTT

Les vététistes suivent généralement les circuits balisés, tels que les GR ou les circuits VTT jalonnés par la Fédération Française de Cyclisme, même si le VTT permet de rouler sur tous les terrains.

Ils utilisent plutôt la carte au 1:25 000, celle au 1:50 000 de l'IGN étant brouillon et beaucoup trop chargée.

Les vététistes qui pratiquent la randonnée en montagne souhaitent que les courbes de niveau soient plus lisibles. Ceux qui pratiquent la randonnée à la campagne s'intéressent plutôt au type de végétation. Par ailleurs, les vététistes sont plus attentifs au paysage que les cyclistes ; ils pratiquent aussi bien la *randonnée contemplative* que la *randonnée sportive*. Ils apprécient donc que les informations touristiques soient indiquées ainsi que les niveaux de difficulté et les distances.

Pour améliorer la lisibilité de la carte; les vététistes préconisent également de choisir des couleurs vives et contrastées ; il serait aussi nécessaire de limiter la quantité d'information. Le quadrillage et les coordonnées Lambert sont inutiles, comme les traits de limite des communes (qui peuvent être confondus avec la représentation des chemins). Un randonneur souligne cependant l'intérêt de préciser le type de randonneurs autorisé à emprunter les circuits pour ne pas se retrouver en VTT sur un parcours réservé aux randonneurs pédestres.

5.4. Les randonneurs occasionnels

Les randonneurs occasionnels pratiquent la randonnée d'une manière moins sportive que les randonneurs fréquents et préfèrent parler de *balade*.

Ils utilisent assez peu les cartes IGN qu'ils jugent trop *compliquées*, voire trop *fouillis*, et s'appuient davantage sur des *cartes plus simples* distribuées par les offices de tourisme, ou simplement sur le fléchage des itinéraires sur le terrain. Pour eux, la carte idéale serait *aérée*, *allégée* et où les sentiers et circuits pédestres seraient mis

en valeur grâce à une légende *précise* mais *simple* à comprendre. La couverture actuelle n'est ni attractive ni explicite ; il serait préférable d'y indiquer la cible : randonneurs pédestres, cyclistes, automobilistes, ...

Ils préfèrent des couleurs plus *appuyées* et trouvent que la carte au 1:50 000 est *trop pastel*. Ils plébiscitent l'utilisation dans la légende des couleurs conventionnelles, voire *proches de la nature*, afin de pouvoir repérer les objets géographiques correspondants plus vite. Les symboles doivent être discriminants (ceux représentant par exemple ceux désignant les églises et les cimetières sont trop proches sur la carte au 1:50 000) et *ludiques*.

6. Conclusions

Cette étude avait pour objectif d'explorer la faisabilité et l'intérêt d'utiliser des outils de TALN pour exploiter une enquête semi-directive à questions ouvertes sur le thème de la randonnée. L'enquête a été réalisée auprès de groupes constitués selon des pratiques de randonnées considérées a priori comme pertinentes, afin qu'ils définissent une carte adaptée à leur pratique. L'originalité de la démarche réside dans l'utilisation conjointe d'outils statistiques et linguistiques. L'analyse statistique automatisée a permis d'identifier, sans a priori, des termes pertinents selon les différentes pratiques de la randonnée. Une étude manuelle des concordances réalisées à l'aide de patrons morpho-syntaxiques et guidée par des connaissances linguistiques a été nécessaire afin de préciser le contexte sémantique de ces termes.

L'étude des commentaires des randonneurs permet de mieux prendre en compte leur ressenti qui s'ajoute à des connaissances académiques en cartographie pour concevoir des cartes sur mesure. Les attentes exprimées par les randonneurs renvoient à des propriétés antagonistes de la carte : être riche, lisible et adaptée à son contexte d'utilisation, en l'occurrence différentes pratiques de la randonnée puisque l'enquête montre que les besoins et les attentes varient sensiblement selon la pratique. Il faut noter qu'une adaptation fine au contexte d'utilisation, telle qu'elle est imaginée par les randonneurs, ne peut être proposée qu'au travers d'un service de conception de carte à la demande, par exemple par l'intermédiaire d'un service internet dont plusieurs réalisations ont été déjà présentées (Jolivet, 2009 ; Christophe *et al.*, 2009).

Les résultats de cette étude devront être formalisés et intégrés à une ontologie de la cartographie (Dominguès *et al.*, 2009) afin de contribuer à des applications de conception de cartes sur mesure pour différents publics. Le savoir-faire et les ressources terminologiques électroniques construites à l'occasion de ce travail (dictionnaires, patrons, grammaires locales) pourront être réutilisés pour l'exploitation d'autres enquêtes déjà réalisées dans le but d'explorer le ressenti des lecteurs de cartes.

7. Bibliographie

- Beaudouin V., « Statistique textuelle : une approche empirique du sens à base d'analyse distributionnelle », *Texte !*, 2000, [En ligne], URL : <http://www.revue-texto.net/index.php?id=642>
- Bertin J., *Sémiologie graphique: les diagrammes, les réseaux, les cartes*. Rééditions en 1973, puis en 1998, Paris, Editions de l'EHESS, 1967.
- Brugidou M., Mandran N., Moine M. et Salomon A-C., « Méthode et stratégie d'analyse des questions ouvertes du panel électoral français », *Bulletin de méthodologie sociologique*, 82|2004, Mis en ligne le 09 juillet 2008, URL : <http://bms.revues.org/index1068.html>
- Buard E. et Ruas A., « Processes for improving the colours of topographic maps in the context of Map-on-Demand », *Actes de 24th International Cartographic Conference (ICC'09)*, 2009, Santiago (Chile).
- Charaudeau P. *Grammaire du sens et de l'expression*. Hachette Education, 1992.
- Christophe S., Ruas A., « Conception de légendes créatives sur mesure », numéro spécial de la revue *Le monde des cartes* du Comité Français de Cartographie sur les publications de la conférence ICC'09, n°202, décembre 2009.
- Dominguès C., Etude d'outils informatiques et linguistiques pour l'aide à la recherche automatique d'information dans un corpus documentaire, Thèse de doctorat, Université de Marne-la-Vallée, 2001.
- Dominguès C., Christophe S. et Jolivet L., « Connaissances opérationnelles pour la conception automatique de légendes de cartes », *Actes des 20^{èmes} Journées Francophones d'Ingénierie des Connaissances (IC2009)*, 2009, Hammamet (Tunisie), p. 253-264.
- Gélinas-Chebat C., Daoust F., Dufresne M., Gallopel K. et Lebel M-E., « Analyse exploratoire d'entrevues de groupe : les jeunes Français et le tabac », *Actes de JADT'2004 (7èmes journées internationales d'analyse statistique des données textuelles)*, 2004, Louvain-la-neuve, p. 479-487.
- Généreux M., Santini M., « Défi : classification de textes français subjectifs », *Actes de l'atelier de clôture du 3^{ème} DEfi Fouille de Textes*, 2007, Grenoble, p. 85-94.
- Heitz T., « Modélisation du prétraitement des textes », *Actes de JADT'2006 (8èmes Journées internationales d'Analyse statistique des Données Textuelles)*, 2006, Besançon, p. 499-506.
- Jolivet L., « Characterizing maps to improve on-demand cartography - the example of European topographic maps », *Actes de GISRUK*, Royaume-Uni, Durham, 2009.
- Labbé D., « Analyse des données textuelles et statistique lexicale », *Actes de JADT'00 (5èmes Journées internationales d'Analyse statistique des Données Textuelles)*, 2000, Lausanne.
- Mathieu Y.Y., Un classement sémantique des verbes psychologiques. *Cahiers du CIEL* 1996-1997. Publications Paris 7.
- Tomasetto C., Matteucci MC., Selleri P., Cavrini G., Scalone L., « Analysis of open answers in the validation of a health-related quality-of-life questionnaire for children », *Actes de JADT'08 (9èmes Journées internationales d'Analyse statistique des Données Textuelles)*, 2008, Lyon, p 1093-1102

Vers la conception d'un service web de contexte spatial dédié aux téléphones intelligents dans le cadre de jeux éducatifs

Elodie Edoh-Alove *,** - Frédéric Hubert * - Thierry Badard *

**Centre de Recherche en Géomatique,
Pavillon Louis-Jacques Casault local 0744, Université Laval
Québec (Québec) G1K 7P4 Canada
{frederic.hubert, thierry.badard} @scg.ulaval.ca
**djogbenuyekpe.edoh-alove.1@ulaval.ca*

RÉSUMÉ. GéoEduc3D est un projet qui vise à élaborer des jeux éducatifs interactifs, pour téléphones intelligents, basés sur les sciences géomatiques et exploitant la réalité augmentée pour rendre ces jeux plus immersifs. Pour améliorer l'aspect immersif et interactif de ces jeux, nous nous sommes concentrés sur l'exploitation du contexte dans ce cadre applicatif. Des travaux de recherche se sont déjà orientés vers la définition et l'élaboration de modèles de contexte et de plateformes de gestion dudit contexte. Cependant, le cadre applicatif diffère et l'interopérabilité n'a pas été vraiment explorée. Nos travaux ont pour objectif de concevoir et d'implémenter une solution informatique dédiée à l'acquisition et à la diffusion de contexte spatial dans un environnement multi-joueurs sur et pour des téléphones intelligents. Dans cet article, nous présentons notre définition de contexte spatial, une modélisation du contexte, l'architecture de notre système orienté service pour la gestion du contexte spatial et enfin une discussion sur de futurs travaux.

ABSTRACT. GeoEduc3D is a project that aims to develop interactive serious games for smart phones based on Geomatics and that uses augmented reality techniques in order to make these games more immersive. To improve the immersive and interactive aspect of those games, we focused on the exploitation of context in this application framework. Research has already been orientated towards the definition and development of context models and platforms for managing that context. However, the application framework is different and interoperability has not really been explored. Our work aims to design and implement an IT solution that focuses on the acquisition and dissemination of spatial context in a multi-players environment on and for smart phones. In this paper, we present our definition of spatial context, a context modeling, the architecture of our service-oriented system for managing spatial context and finally a discussion on future work.

MOTS-CLÉS : contexte spatial, service web, mobilité, jeux éducatifs, réalité augmentée, multiutilisateurs.

KEYWORDS: spatial context, web service, mobility, serious games, augmented reality, multi-users.

1. Introduction

Actuellement, nous constatons l'émergence d'un nombre grandissant d'applications pour appareils mobiles qui prennent en compte de façon dynamique les informations et les changements de l'environnement des utilisateurs. Le but ultime est de satisfaire encore plus les utilisateurs en leur offrant des applications riches qui s'adaptent à leur environnement (position, besoins, préférences, etc.), tant au niveau des interfaces proposées que des informations et services offerts. L'exemple du Métro de Paris¹ sur iPhone est une application très complète, interactive et explicite qui fournit, selon la localisation de l'utilisateur, diverses informations sur les stations de métro ou des enseignes connues (ex. Galerie Lafayette) à proximité et sur les orientations pour s'y rendre. De telles applications exploitent donc le contexte de l'utilisateur pour servir les informations appropriées et ciblées. Plusieurs travaux (Dey, 2000 ; Schilit *et al.*, 1994 ; Schwinger *et al.*, 2008) ont été menés sur ce sujet et ont conduit à des définitions, des approches, des modèles et/ou des outils pour des applications, mobiles ou pas, conscientes du contexte dans différents domaines comme la santé, le tourisme ou les loisirs.

Nos travaux s'inscrivent dans le cadre du projet GéoÉduc3D², financé par le réseau des centres d'excellence GEOIDE. Ce projet vise à proposer des jeux interactifs éducatifs, s'appuyant sur les technologies géospatiales et explorant des questions thématiques auxquelles le jeune public des adolescents est sensible : les changements climatiques, le développement durable. Ainsi, il vise à concevoir et mettre en œuvre un ensemble d'outils d'apprentissage innovants qui vont permettre d'enrichir - « augmenter » - l'expérience du joueur en rendant le jeu plus immersif, réactif, interactif. De plus, ce projet a choisi les téléphones intelligents (smartphones) comme support de travail et les « serious game » comme type de jeu éducatif. Les téléphones intelligents de type iPhone ou OS Android, très répandus de nos jours offrent une puissante et impressionnante combinaison de caractéristiques techniques (taille de l'écran, puissance de calcul, écran tactile, GPS intégré, accéléromètre, accès aux réseaux sans fil, ...) intéressantes pour le déploiement de tels jeux. Concernant les jeux de type « serious game », leur sélection pour le projet résulte de leur vocation pédagogique, et du fait qu'ils permettront de rendre très ludique et attractif l'acquisition de notions ayant trait à l'environnement et au changement climatique par le jeune public.

Au final, le projet GéoÉduc3D vise à fournir des jeux éducatifs mobiles, situés dans des environnements réels, et donc exploitant la localisation du joueur et tout ce qui l'entoure, et qui offre une haute immersion à travers l'utilisation de techniques de réalité augmentée. Le contenu de tels jeux et les opportunités d'interactions du joueur dépendent de sa position courante dans le monde réel. En outre les jeux mobiles utilisant la réalité augmentée procurent une expérience de jeu plus riche permettant aux joueurs de bouger et d'interagir dans leur environnement physique

¹ Métro de Paris sur iPhone ou iPod Touch - <http://www.metroparisiphone.com/>. Accédé le 23 juin 2010.

² Projet GéoÉduc3D - <http://geoeduc3d.scg.ulaval.ca>, Accédé le 25 juin 2010.

avec du contenu 3D. Cette interaction, faisant partie intégrante du jeu, peut être faite entre le joueur et des objets, virtuels ou non, se trouvant dans son environnement, ou même entre différents joueurs qui doivent coopérer ou être adversaires de jeu partageant ainsi une certaine zone réelle (Broll et al., 2008).

Une attention particulière est ainsi accordée au contexte spatial du joueur et de l'exploitation que l'on peut en faire, pour le type de jeu visé, dans un environnement de jeu multi-joueurs. À titre d'exemple, la position et l'orientation d'un joueur, avec un profil de maçon, pourraient être utilisées pour offrir à ce dernier une visualisation et une interaction avec des objets de réalité augmentée géo-référencés comme des sacs de ciment, ou une truelle, objets qu'un autre joueur avec un profil d'éclaireur ne serait pas en mesure de voir ou d'utiliser. Les possibilités d'exploitation de ce contexte sont multiples dans le cadre d'applications de ce type. Cependant, il n'existe pas actuellement de modèle de contexte spatial applicable aux jeux éducatifs interactifs avec l'usage de téléphones intelligents dans un environnement de jeu multi-joueurs. Dans un tel environnement, il n'existe pas, non plus, de système distant et interopérable pour échanger des données de contexte entre les plateformes mobiles (capteurs et applications clientes). En outre, la dissociation entre gestion du contexte et application est rarement faite et nous assistons à un manque indéniable d'accès à des données de contexte spatial (RA, Jeux, Mobile) centralisées.

Il se pose alors une question fondamentale : comment rendre disponible les données et informations de contexte spatial à toute application, n'importe où et en tout temps, en tenant compte des aspects de réalité augmentée, de jeu éducatif et de mobilité ? Notre objectif général consiste donc à concevoir et implémenter une solution informatique dédiée à l'acquisition et à la diffusion de contexte spatial dans un environnement multi-joueurs sur et pour des téléphones intelligents. Pour atteindre cet objectif, nous nous sommes orientés vers les services web (qui permettent de rendre disponible, au travers d'un simple langage d'interrogation des données et/ou informations de manière indépendante et distribuée). Dans notre situation, un service web offre d'une part la possibilité de centraliser toutes les informations de contexte de joueurs dans un environnement multi-joueurs sur un système distant en employant des opérations d'acquisition de données contextuelles. D'autre part, les données de contexte rendues disponibles dans un tel service web sont accessibles, échangeables et interrogeables par des applications clientes du jeu sur téléphones intelligents de manière totalement transparentes.

La suite de cet article présente, en section 2, un bref état de l'art sur les notions de contexte spatial et de contexte en mobilité. Ensuite, nous proposons, en section 3, une définition du contexte spatial en tenant compte des aspects de réalité augmentée, jeux, mobilité et environnement multi-joueurs. Une modélisation descriptive du contexte spatial est également offerte dans cette même section. La section 4 présente l'architecture de notre système orienté service de gestion des données du contexte spatial. Une discussion est enfin proposée, en section 5, sur différents choix pour la suite des travaux.

2. Définitions de contexte, contexte spatial et contexte en mobilité

Plusieurs définitions de la notion de contexte existent depuis les années 90. L'une des plus pertinentes a été celle donnée par (Dey, 2000). Selon lui, le contexte est « toute information qui peut être utilisée pour caractériser la situation d'une entité, une entité pouvant être une personne, un lieu ou un objet qui est considéré comme approprié dans l'interaction entre un utilisateur et une application (les deux inclus) ». Ici, le « lieu » correspond à des espaces géographiques comme les chambres, les bureaux, les immeubles ou les rues ; une « personne » peut être un individu ou un groupe d'individus géographiquement dispersés ou réunis ; un « objet » fait référence soit à un objet physique, soit à un composant logiciel. Pour (Schilit *et al.*, 1994), le contexte d'un utilisateur se définit par le lieu, les identités des objets et personnes voisines et tous les changements sur ces objets. (Brown *et al.*, 1997) caractérise le contexte par le lieu, l'identité des personnes autour de l'utilisateur, le moment de la journée, la saison, la température, etc.

La notion de conscience du contexte (ou sensibilité au contexte) a été introduite par (Schilit *et al.*, 1994) comme étant l'habilité qu'ont les applications mobiles à découvrir et réagir aux changements de l'environnement où elles sont localisées. Pour (Hristova, 2008), un système (ou une application) est dit sensible au contexte s'il est conçu pour réagir aux changements constants dans l'environnement. Pour (Dey, 2000), un système est sensible au contexte s'il utilise le contexte pour fournir des informations pertinentes et/ou des services à l'utilisateur, la pertinence dépendant de la tâche de l'utilisateur. Nous pouvons donc retenir qu'un système (ou une application) est conscient(e) du contexte s'il (ou elle) est conçu(e) pour acquérir, stocker, interpréter les données contextuelles afin de fournir les services et les informations répondant aux besoins de n'importe quel utilisateur à tout moment et n'importe où, utilisateur pouvant désigner une personne, une application ou un appareil. Dans la suite de cette section, nous nous concentrons sur différents types de contexte plus proches de nos intérêts : le contexte spatial et le contexte en mobilité.

2.1. Le contexte spatial

Dans la plupart des travaux menés sur les applications sensibles au contexte, le contexte spatial est réduit à la variable localisation (localisation de l'utilisateur); dans le meilleur des cas, nous pouvons identifier les places se trouvant dans l'environnement relatif de l'utilisateur (George *et al.*, 2009). Nous ne retrouvons d'ailleurs pas le terme « contexte spatial » en tant que tel dans la littérature existante, mais on peut retrouver plutôt la notion de « contexte d'usage » qui fait référence à toute information caractérisant l'utilisateur, l'environnement et la plateforme (Petit, 2010). (Li, 2007), quant à lui, avance une définition du contexte spatial (ou contexte géographique) en trois dimensions : (1) le contexte spatial statique, (2) le contexte spatial dynamique et (3) le contexte spatial interne.

Le *contexte spatial statique* (1) représente des informations géospatiales numériques et stockées qui peuvent avoir un impact sur l'environnement de l'utilisateur. Par exemple, ces informations peuvent être : le réseau routier, les bâtiments, les magasins, les stations d'essence, les structures intérieures (salles, bureaux, appartements, ...), les lots de stationnements.

Le *contexte spatial dynamique* (2) correspond aux informations géo-référencées collectées par différents types de capteurs. Par exemple, nous pouvons obtenir des données issues de capteurs de vitesse sur les routes, ou des capteurs sur le taux d'occupation des lots de stationnements.

Le *contexte spatial interne* (3) réfère aux informations ou données directement accessibles à partir des appareils locaux (ex. GPS). Par exemple, un automobiliste pourrait accéder directement à sa position, sa vitesse, son orientation, sa destination avec son GPS de voiture, à des données dans son agenda électronique, ainsi qu'au niveau d'essence avec un capteur intégré à son véhicule.

2.2. Le contexte en mobilité

Avec la prise en compte de la mobilité, différents travaux ont principalement conduit à la modélisation du contexte et à la réalisation de solutions informatiques sensibles au contexte dans plusieurs domaines applicatifs comme les maisons intelligentes, la santé, le tourisme ou les loisirs (George *et al.*, 2009 ; Tan *et al.*, 2009 ; Meng *et al.*, 2005 ; Filho *et al.*, 2008 ; Abowd *et al.*, 1997 ; Li *et al.*, 2006 ; Schwinger *et al.*, 2008 ; Repo *et al.*, 2004 ; Cai *et al.*, 2006 ; Cheverst *et al.*, 2000 ; Lopez-Velasco *et al.*, 2006). Parmi ces travaux, nous retenons (George *et al.*, 2009) qui a élaboré un modèle de contexte pour des utilisateurs d'applications de communication en mobilité. Ce modèle est basé sur la définition de six dimensions clés représentées sous la forme d'un vecteur < profil utilisateur, activité, appareil, environnement, localisation, temps >. À un moment donné, on obtiendrait par exemple le vecteur < 'Étudiant', 'Visiter le Louvre', 'PDA', 'oui', 'Mona Lisa', '19-09-2008' >. Le modèle de contexte en mobilité, nommé TILES (Temporal, Identity, Location, Environmental and Social) proposé par (Tan *et al.*, 2009) est aussi d'intérêt. TILES est un modèle de classification et d'utilisation d'informations contextuelles dans des applications mobiles conscientes du contexte. 42 propriétés de contexte ont été classifiées à la suite de lectures d'études sur l'exploitation du contexte dans le domaine du tourisme en mobilité. Une typologie, dressée après analyse de différentes applications rencontrées dans la littérature (contexte identifié dans plusieurs domaines : médical, apprentissage ludique, bureautique et tourisme), est composée de dix catégories de propriétés d'informations contextuelles : position, temporel, identité, environnemental, social, réseau, activité, appareil, physiologie et cognitif. Après enquête et validation des propriétés utiles dans le domaine du tourisme auprès d'un groupe d'utilisateurs, 5 catégories ont été retenues pour constituer le modèle TILES (Temporal, Identity, Location, Environmental, et Social).

Au regard de ces travaux de recherche, nous remarquons que la notion de localisation n'implique pas uniquement les coordonnées de l'utilisateur et que l'environnement peut faire référence à des objets à référence spatiale (places ou personnes avoisinantes) ou non (température, niveau de bruit). D'un autre côté, la mobilité a rendu pertinente et enrichissante les éléments de contexte comme la localisation physique, l'appareil, l'environnement relatif à l'utilisateur, c'est-à-dire l'ensemble des places proches de la position courante de l'utilisateur. Selon nous, ceci implique qu'il ne serait pas intéressant de considérer le contexte spatial comme une catégorie isolée dans le cadre de notre projet. En effet, le modèle de contexte que nous proposons de définir doit prendre en compte des éléments purement spatiaux, mais aussi d'autres critères comme le profil de l'utilisateur – soit le joueur dans le cadre du projet GéoÉduc3D –, l'appareil – un téléphone intelligent –, les aspects temporels, et les personnes à proximité. Dans la section suivante, nous allons nous attarder sur la notion de contexte spatial en proposant une définition et un modèle du contexte spatial propre au cadre de notre projet.

3. Définition et modélisation du contexte spatial (réalité augmentée, jeux, mobilité)

Selon nous, la définition du contexte à adopter doit offrir une description complète du contexte de l'utilisateur en tenant compte des propriétés que peuvent apporter la réalité augmentée, les jeux éducatifs, l'environnement multi-joueurs et la mobilité. De plus, à la section 2.2, nous avons présenté la définition du contexte spatial par (Li, 2007) avec trois types de contexte spatial : contexte statique, contexte dynamique et contexte interne. Nous nous sommes inspirés de cette définition pour proposer une nouvelle définition du contexte spatial selon le cadre spécifique à notre travail (cf. figure 1).

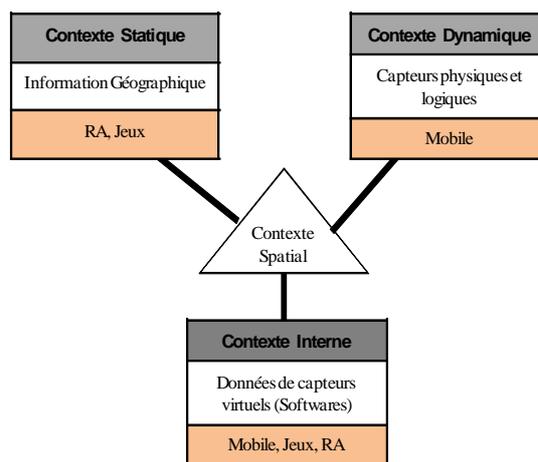


Figure 1. Définition du contexte spatial (RA, Mobile, Jeux), adaptée de (Li, 2007)

Le contexte spatial est identifié en soi par l'ensemble des informations, d'ordre spatial ou connexe, caractérisant les contextes statique, dynamique ou interne d'une entité qui peut être une personne, un endroit ou un objet considéré comme pertinent pour l'interaction entre un utilisateur et une application, les deux inclus.

Afin de mieux décrire les contextes dynamiques et internes, différentes sources d'informations contextuelles, soit des capteurs, doivent être prises en considération. Ces capteurs peuvent être identifiés selon trois types : capteurs physiques (ex. capteurs de température, capteurs de position), capteurs virtuels (ex. performance appareil, profil utilisateur) et capteurs logiques (création d'information à partir de plusieurs sources) (Chaari, 2007). Les données peuvent être recueillies directement en étant fournies par l'utilisateur, ou indirectement, en étant fournies par d'autres services ou par des capteurs physiques. Plus précisément, nous proposons les définitions de contexte suivantes :

- **Contexte statique:** Toute information spatiale qui peut être pertinente pour la description de l'environnement de l'utilisateur et de son interaction avec l'application de jeu. Par exemple, il pourrait s'agir d'information spatiale permettant l'augmentation de la vue de la réalité de l'utilisateur avec des éléments ponctuels d'intérêts de réalité augmentée (Points Of Interest-POI) et/ou une scène de jeu avec ses limites géographiques.
- **Contexte dynamique:** ce contexte fait référence à toutes les données issues des capteurs physiques et logiques. Il s'agit notamment d'informations fournies par des capteurs comme le capteur de température, de lumière, et l'accéléromètre pour ne citer que ceux là. Par exemple, on pourrait retrouver ici l'information concernant la position courante (GPS) des joueurs, qui ont différents profils, ce qui permettrait de les faire interagir avec les objets de réalité augmentée , ou d'offrir une visualisation des positions des autres joueurs à chacun d'entre eux en temps réel.
- **Contexte interne:** ce type de contexte prend en compte les informations issues de capteurs virtuels et qui décrivent en l'occurrence le profil de l'utilisateur, l'appareil mobile, l'environnement de jeu, l'environnement social de l'utilisateur, et le réseau. On retrouvera ici par exemple les informations concernant le profil général du joueur (ex : nom, date de naissance, préférences) ou son profil dans le jeu (ex: rôle dans le jeu), ou encore celles concernant les autres joueurs appartenant ou non à son équipe. Ces informations pourraient permettre d'adapter les interactions entre les joueurs, et entre les joueurs et l'appareil.

Avec un contexte spatial ainsi défini, nous constatons que ce dernier a un aspect multidimensionnel et multi-échelle que nous devons prendre en compte lors de sa modélisation. Cette définition ne se concentre pas uniquement sur l'utilisateur et sa position, mais aussi sur l'appareil mobile, l'environnement physique, le temps, l'environnement du jeu, etc. De plus, il ne s'adresse pas seulement à une échelle locale, mais aussi à une échelle globale, relativement à un environnement multi-joueurs. Cela nous amène donc à proposer une représentation formelle du contexte

spatial selon 8 catégories d'informations contextuelles. Le tableau 1 présente ces 8 catégories selon des sous-catégories et des propriétés, qui correspondent ainsi à une modélisation descriptive du contexte spatial dans le cadre spécifique du projet GéoÉduc3D. Les descriptions des différentes catégories sont :

- *Environnement d'exécution* : cette catégorie prend en compte des informations sur la date et l'heure associées à un contexte donné pour la session utilisateur courante. Elle est importante car elle permettra de lier d'une part les différents contextes (donc informations de contexte) captés par notre système au joueur concerné, et d'autre part d'associer un temps, donc une durée de vie aux contextes, ce qui rendrait la gestion d'historique effective.
- *Joueur* : cette catégorie prend en compte des propriétés contenant des informations personnelles et professionnelles du joueur comme son nom, son âge, sa langue préférée ou son programme d'études. Cette catégorie est l'une des plus importantes car le fait de connaître des informations contextuelles liées au profil général du joueur avant et pendant le jeu permettront de produire un contenu et une interface de jeu personnalisés, et enrichiront les interactions entre les joueurs.
- *Localisation* : cette catégorie indique aussi sous forme de propriétés, la situation géographique de l'utilisateur, les éléments spatiaux qui se trouvent dans son voisinage etc. Elle est aussi primordiale puisqu'elle peut avoir un impact fort sur le déroulement des jeux et les contenus visuels et interactifs proposés dans ces derniers, dans la mesure où suivant les positions du joueur au cours du jeu, des objets peuvent lui apparaître en réalité augmentée, des actions peuvent être faites seulement en certains endroits pour avancer dans le jeu, des interactions peuvent être proposées entre les joueurs etc.
- *Environnement*: cette catégorie doit être prise en compte car durant le jeu, les joueurs interagissent constamment avec leur environnement. Nous faisons référence ici à des informations contextuelles propres à l'environnement ambiant (température, luminosité, climat etc.) ainsi qu'à des données sur les conditions de trafic ou le type de localité (rural, urbain, industriel).
- *Social*: connaître les personnes qui entourent (physiquement ou logiquement) un joueur ou quels joueurs constituent une équipe ou sont adversaires dans un jeu est crucial dans un environnement multi-joueurs. Cette catégorie va permettre de sélectionner à tout moment les joueurs dont les informations, comme la position par exemple, intéressent un ou plusieurs autres joueurs, ainsi par exemple un joueur ne pourrait voir, sur son écran de jeu, que les joueurs de son équipe ou ses adversaires qui sont à une certaine distance de lui.
- *Réseau*: connaître l'état du réseau, les ressources disponibles et leur état est une source d'information importante pour tout ce qui concerne le jeu et son déroulement dans la mesure où certains jeux peuvent nécessiter l'exploitation d'un serveur de jeu et/ou d'une connexion sans fil.
- *Appareil*: cette catégorie prend en compte les informations comme le type de l'appareil, les capteurs disponibles sur l'appareil (ex. accéléromètres, GPS, microphone), les possibilités de connexion, la vitesse de la connexion, l'espace mémoire, la taille de l'écran, le type d'entrée/sortie, etc.

- *Jeu*: Cette dernière catégorie est celle qui nous relie au côté jeu des applications qui devront exploiter toutes ces informations de contexte. Elle indique en effet toutes sortes d'informations contextuelles liées au déroulement du jeu comme l'heure de début, le nombre de joueurs actuels ou les limites géographiques de l'espace de jeu.

Catégories	Sous-catégories	Propriétés
Environnement d'exécution	Aspect temporel	Date ; Heure
	Session courante	État (connecté, déconnecté, en attente)
Joueur	Profil	Nom ; Prénom ; Date de naissance ; Sexe ; Langues préférées; Humeur
	Aspect éducation	Domaine ou Programme d'étude ; Niveau d'études
	Aspect jeu	Profession du joueur ; Connaissance préféré ; Score ; Ressources et outils disponibles
Localisation (proximité)	Localisation générale	Pays ; Région ; Ville ; Quartier ou Zone
	Localisation ponctuelle	Position actuelle ; Vitesse de déplacement ; Orientation (accéléromètre) ; Direction de déplacement
	Localisation dans le jeu	Limites géographiques de la scène de jeu ; Points Of Interests (POI) pour la RA ; Objets géographiques créés par l'utilisateur
	Par proximité	Objets géographiques avoisinants ; Distance entre la position et les objets avoisinants
Environnement	Aspect général	Température ; Climat ; Saison ; Luminosité
	Aspect spatial	Type de localité (zone urbaine ou rurale) ; Conditions de trafic
Jeu		Heure début du jeu ; Nombre de joueurs autorisés ; Nombre de joueurs actuels ; Meilleurs scores ; Score de l'équipe ; Statistiques de jeu ;
Appareil mobile	Capacités	Mémoire vive ; Mémoire de stockage ; Niveau Batterie
	Fonctionnalités	Écran tactile ; Reconnaissance vocale ; Caméra ; Microphone ; Téléphone activé ; GPS ; Accéléromètre
Social	Équipe	Nom de l'équipe ; Nom des équipiers
	Général	Joueurs actuels; Personnes avoisinantes
Réseau		État du réseau ; Connexion (WiFi / GSM)

Tableau 1. *Modèle descriptif du contexte spatial (RA, Jeux, Mobile)*

Les catégories Joueur et Jeux relèvent plus de notre cadre applicatif tandis que toutes les autres sont plus génériques. Nous considérons donc que notre proposition de modèle descriptif du contexte spatial peut être utilisée dans notre propre cadre applicatif de tout en permettant d'y intégrer de nouvelles propriétés d'informations contextuelles. La liste des propriétés actuellement incluses dans le tableau 2 n'est évidemment pas exhaustive, mais elle en reflète les plus attendues. Aussi, nous y avons mis quelques unes de celles qui nous intéressent dans notre cadre applicatif. Ce modèle est aussi utile à la définition, la conception et la mise en œuvre d'un service web de contexte spatial, intégré dans un système plus global de gestion des données contextuelles, dont l'architecture fera l'objet de la section suivante.

4. Architecture du système de gestion des informations contextuelles

Le système, orienté service, de gestion des données de contexte que nous proposons a pour vocation d'acquérir, rendre disponible et gérer des données ou informations de contexte selon notre cadre applicatif. Ces données proviennent de différents capteurs intégrés aux appareils mobiles ou des capteurs externes comme des stations météorologiques (Benazzouz *et al.*, 2010). Ainsi, elles peuvent être partagées entre les différentes applications mobiles de jeu. La figure 2 présente l'architecture fonctionnelle de ce système de gestion des données de contexte avec comme élément central un service web de contexte spatial (WSCS - Web Spatial Context Service) s'appuyant sur les recommandations de la spécification OWS³ (de l'OGC (Open Geospatial Consortium) qui sert de référence à l'implantation de services web dans le domaine géospatial. Cette architecture se décompose comme suit :

- *Tiers ressources* : Les données de contexte proviennent des différents capteurs intégrés ou non aux appareils mobiles se trouvant dans l'environnement de jeu multi-joueurs. Ces capteurs sont de différents types comme mentionné dans la section 3 : capteurs physiques ; capteurs virtuels ; capteurs logiques. D'autres ressources peuvent être exploitées comme d'autres services web de contexte qui pourraient être utiles. Un autre composant essentiel à ce tiers est la base de données qui va être directement connectée au service web de contexte spatial pour stocker très régulièrement durant une phase de jeu les données contextuelles, dont la quantité peut devenir conséquente. Cette base de données doit être en mesure de gérer les données en fonction de notre modèle descriptif défini dans la section précédente pour pouvoir répondre à nos attentes.
- *Tiers serveur* : Ce tiers est l'élément central de notre architecture car il correspond à notre service web de contexte spatial dont le but est de rendre disponible une multitude de données ou informations contextuelles à des applications clientes mobiles. Ce service web traite les différentes requêtes en provenance des applications clientes pour stocker des données et pour en retourner à n'importe quel moment et de n'importe où. Pour effectuer le

³ OGC Web Service Common Implementation Specification-
<http://www.opengeospatial.org/standards/common>, Accédé le 14 Octobre 2010

stockage et la gestion des données de contexte, nous avons besoin d'utiliser la base de données du tiers ressources. Ce service doit également être capable d'exploiter les autres ressources disponibles pour compléter les informations contextuelles à stocker.

– *Tiers client* : Les clients sont des applications mobiles de jeu déployées sur différents téléphone intelligent (iPhone ou sous OS Android). Ils utilisent les réseaux sans fil (WiFi, 3G) pour accéder à notre service web. Via cet accès, des requêtes d'acquisition peuvent être envoyées au service web avec de nouvelles données de contexte. Ces clients peuvent aussi produire des requêtes pour interroger le service web afin de récupérer des données de contexte selon différents niveaux de détails dans le but final de les traiter côté client et de servir l'utilisateur en conséquence.

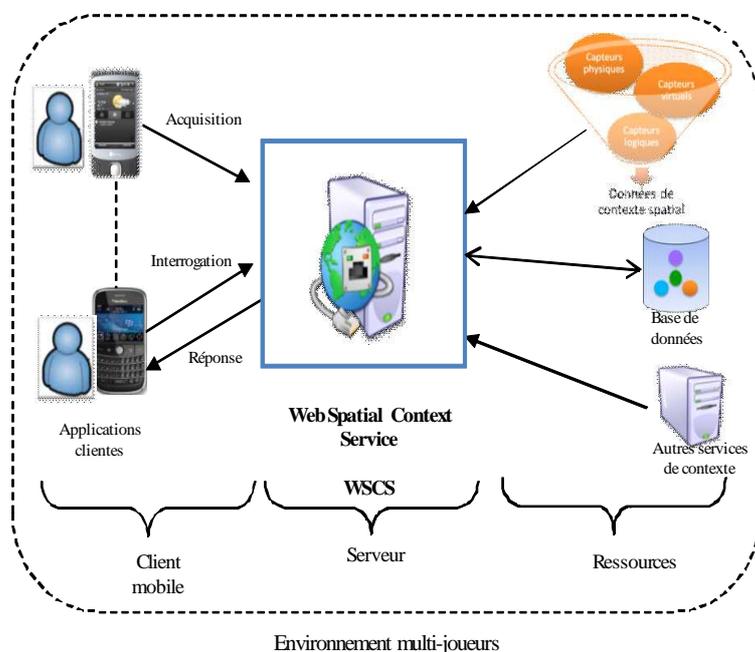


Figure 2. Architecture du système de gestion de données contextuelles spatiales.

Pour mieux comprendre le fonctionnement de ce système, prenons l'exemple d'un appareil comme l'iPhone qui dispose de multiples capteurs physiques comme le GPS pour la position, l'accéléromètre pour l'orientation, la caméra pour le flux vidéo, ainsi que des capteurs virtuels comme un formulaire web pour la saisie d'un profil joueur. Les fréquences d'envoi des requêtes d'acquisition des informations contextuelles fournies par ces capteurs ne sont pas les mêmes : le profil du joueur est envoyé une seule fois au début de la session de jeu, alors que sa position et son orientation sont fournies à intervalles de temps réguliers. Le service web de contexte

spatial, en recevant ces requêtes, récupère ces informations d'intérêt et les stocke dans la base de données, tout en les complétant avec les informations issues des capteurs ou autres services de contexte. De part le fait que cette information soit rendue accessible, toute application de jeu peut ainsi la récupérer au travers de requêtes d'interrogation, envoyée au WSCS.

5. Discussions

Au regard de l'architecture présentée dans la section précédente, différents éléments dans la conception de notre système sont encore en discussion. Le *stockage des données de contexte spatial*, comme premier élément de discussion, est un élément central de notre service web en fonction de la structuration de nos données pour en faciliter les requêtes et aussi les gains de temps. Nous avons défini dans la section 3 un modèle descriptif du contexte spatial dédié aux spécificités et attentes du projet GéoÉduc3D. Les catégories, proposées dans ce modèle sont reliées entre elles par différentes relations sémantiques. Par exemple, plusieurs utilisateurs (catégorie joueur) jouent à un jeu (catégorie jeu), ou encore mon téléphone intelligent sur Android (catégorie appareil mobile) est localisé dans le pavillon Casault de l'Université Laval (catégorie localisation). Nous envisageons donc de concevoir un modèle de données générique et flexible supportant cette sémantique, ainsi que le modèle descriptif, tout en tenant compte d'autres contraintes comme l'aspect temps réel inhérent à la notion de contexte dans un jeu, ou la constitution de requêtes efficaces étant donné le grand nombre de données. Différentes stratégies devront être explorées selon la définition d'un modèle de classe ou bien la définition d'un modèle basé sur des graphes RDF (Resource Description Framework) pour les relations sémantiques (Benazzouz *et al.*, 2010).

Les *spécifications du service web de contexte spatial*, comme second élément de discussion, doivent être définies pour pouvoir concevoir et mettre en œuvre ce service. Outre la définition d'un contrat de service, un format d'échange et un langage d'interrogation sont à définir et à adopter pour une pleine efficacité de notre service. Le contrat de service devra offrir différentes opérations d'acquisition et de diffusion de données de contexte comme `setContextData`, pour enregistrer de nouvelles données contextuelles ou `getContextData` pour récupérer des données de contexte spatial. Le format d'échange et le langage d'interrogation pourront aussi être définis selon le modèle final que nous aurons adopté pour la modélisation des données de contexte évoquée précédemment et s'appuieront sur le langage XML pour permettre une interopérabilité entre systèmes communicants avec ce service.

6. Conclusion et perspectives

L'objectif de cet article était de présenter les premières réflexions et contributions quant à la conception d'un service web de contexte spatial dédié à des applications de jeu éducatif interactifs et immersifs pour et sur des smartphones dans le cadre du projet GéoÉduc3D. Après avoir effectué une brève revue de littérature sur différentes notions du contexte spatial, nous avons proposé une définition du

contexte spatial qui permet de déterminer les données de contexte à prendre en compte dans notre cadre applicatif. Une autre contribution a été d'élaborer un modèle descriptif facilement extensible et réutilisable du contexte spatial, en se basant sur cette définition. Enfin, nous avons proposé une architecture pour un système de gestion du contexte spatial avec comme élément central un service web de contexte spatial, ce qui nous a conduit à engager des discussions sur le choix du type de modélisation pour le stockage des données de contexte et sur les spécifications de notre service web.

Dans la suite de nos travaux, nous envisageons donc de concrétiser la modélisation des données de contexte spatial en proposant un modèle de stockage générique et flexible qui offre une gestion complète des dites données et de leur sémantique. Les spécifications de notre Service Web de Contexte Spatial (WSCS) devront également être définies, conçues et implémentées selon le choix judicieux d'opérations à proposer par le contrat de service, un format d'échange interopérable et un langage de requête facile d'utilisation. Enfin, une phase de tests et validation devra être effectuée dans un environnement de jeu multi-joueurs sur téléphones intelligents.

7. Bibliographie

- Abowd, G.D., Atkeson, G.C., Hong, J., Long, S. Kooper, R., Pinkerton, M. , *Cyberguide: A mobile context-aware tour guide*, Wireless Networks, Volume 3, Number 5, October 1997, pp 421-433.
- Benazzouz, Y., Beaume, P., Ramaparany, F. et Boissier, O. (2010) *Modeling and Storage of Context Data for Service Adaptation, Enabling Context-Aware Web Services, Methods, Architectures and Technologies*, CRC Press Taylor & Francis Group, p. 469-490
- Broll, W., Lindt, I., Herbst, I., Ohlenburg, J., Braun, A-K., Wetzel, R. (2008) *Toward Next-Gen Mobile AR Games*, IEEE Computer Graphics and Applications, IEEE CS Press.
- Brown P. J., Bovey, J. D. & Chen, X. (1997). *Context-aware applications: From the laboratory to the marketplace*, IEEE Personal communications, 4(5), p.58-64.
- Cai, G. et Xue, Y. (2006), *Activity-oriented Context-aware Adaptation Assisting Mobile Geospatial Activities, Proceedings of the 11th international conference on Intelligent user interfaces*, Sydney, Australia, 354-356.
- Chaari, T (2007) *Adaptation d'applications pervasives dans des environnements multi-contextes*. Thèse de doctorat, INSA de Lyon, 2007.
- Cheverst, K., Davies, N., Mitchell, K., Friday, A., Efstratiou, C. (2000), *Developing a Context-aware Electronic Tourist Guide: Some Issues and Experiences. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (The Hague, The Netherlands)*. CHI'00. ACM Press.
- Dey A. K. (2000). *Providing Architectural Support for Building Context-Aware Applications*. Unpublished PhD Thesis, Georgia Institute of Technology.

- Filho, J.B. et Martin, H. , *QACBAC : an owner-centric QoC-aware context-based access control model for pervasive environments*, *Proceedings of the SIGSPATIAL ACM GIS 2008 International Workshop on Security and Privacy and LBS*, Irvine, California, pp 30-38.
- George, S. et Lekira, A.S. (2009) *MeCoCO : A Context-Aware System for Mediated Communications*, *International Journal of Interactive Mobile Technologies (iJIM)*, Volume 3, Issue 3, DOI: 10.3991/ijim.v3i3.748, p.26-33 .
- Hristova, A. (2003) *Conceptualization and Design of a Context-aware Platform for user-centric Applications*.
- Li, Ki-Joune. (2007) *Ubiquitous GIS, Part I: Basic Concepts of Ubiquitous GIS*, Lecture Slides, Pusan National University. <http://stem.cs.pusan.ac.kr/UBGIS/UBGIS.html>
- Li, C. et Willis, K., *Modeling Context Aware Interaction for Wayfinding using Mobile Devices*, *Mobile HCI'06*, September 12-15, 2006, Helsinki, Finland.
- Lopez-Velasco C., Carrillo Ramos A., Villanova-Oliver M., Gensel J., Martin H., *Sélection de services Web adaptés au contexte d'utilisation*, 24^o congrès INFORSID 2006, Hammamet, Tunisie, 31 mai – 3 juin, 2006.
- Meng, Liqiu, Zipf Alexander and Reichenbacher Tumasch, (2005) *Map-based Mobile Services. Theories, Methods and Implementations*. Heidelberg: Springer Verlag, 2005
- Petit, M. (2010), *Approche spatiale pour la caractérisation du contexte d'exécution d'un système d'information ubiquitaire*, Thèse de Doctorat, Ecole Nationale Supérieure d'Arts et Métiers
- Repo, P. et Riekk, J., *Middleware Support for Implementing Context-Aware Multimodal User Interfaces*, *MUM 2004*, October 27-29, 2004 College Park, Maryland, USA.
- Schilit, B., et Theimer, M. (1994) *Disseminating active map information to mobile hosts*, *IEEE Network*, 8(5), p. 22-32.
- Schwinger, W., Grün, C., Pröl, B., Rasinger, J. Et Retschitzegger, W., 2008, *Context-Awareness in Mobile Tourist Guides*, Handbook of research in mobile multimedia, 2nd edition, Khalil-Ibrahim Ismail (ed.), IGI Global, USA.
- Tan, E.M., Foo, S., Goh, D.H. et Theng, Y., (2009) *A Model for Classifying and Using Contextual Information for Context-Aware Applications*, *Electronic journal Aslib Proceedings* Volume 61, Issue 6 p. 565-586, Emerald Group Publishing Limited.

**Session 5 : Qualité, métadonnées et
ontologies**

DL based automated consistency checking of spatial relationships

Catherine Roussey— Francois Pinet

Cemagref, 24 Av. des Landais, BP 50085, 63172 Aubière, France

first-name.last-name@cemagref.fr

Université de Lyon CNRS, Université Lyon 1, LIRIS UMR5205, Villeurbanne, France

RÉSUMÉ. les contraintes d'intégrité spatiales permettent de spécifier les relations autorisées entre des objets spatiaux des bases de données. Dans ce contexte, il est nécessaire de vérifier la cohérence d'un jeu de relations spatiales. Ce contrôle permet de détecter si des relations spatiales spécifiés dans les contraintes d'intégrité par des utilisateurs sont incompatibles. Cet article propose une nouvelle méthode de détection d'incohérences spatiales basée sur des traductions en Logiques de Description (LD) des relations RCC8 entre objets spatiaux. Nous proposons d'utiliser un moteur d'inférence LD traditionnel pour vérifier la cohérence d'un ensemble de relations RCC8. Les traductions LD proposées dans cet article pourront être étendues pour vérifier la cohérence entre des types évolués d'objets spatiaux (les régions composites, les régions complexes, etc).

ABSTRACT. In using integrity constraints, one can specify the relations allowed between spatial objects of databases. In this context, the control of the consistency of a set of spatial relations is an interesting issue. This control aims at checking if all the spatial relations specified in integrity constraints by users are compatible. In this paper, we propose a new method based on Description Logics (DL) to identify inconsistencies in RCC8 relations. We propose to use traditional DIG reasoners to reach this goal. The DL specifications proposed in this paper can be extended in order to allow checking consistency between different types of spatial objects (composite regions, complex regions, etc.).

MOTS-CLÉS : contrainte d'intégrité, relations spatiales, RCC8, Logiques de Description, moteur d'inférence spatial

KEYWORDS: integrity constraints, spatial relations, RCC8, Description Logics, spatial reasoner

1. Introduction

The formalization of spatial relationships is an active research field. Numerous researchers are interested in the identification of a pertinent set of qualitative spatial relationships. These relations model the possible configurations between spatial objects. For example, the method called RCC8 (Region Connection Calculus) models 8 qualitative spatial relations between two regions [RAN 92]. Let X, Y be two simple regions i.e., two non-empty regular subsets of some topological spaces. These two regions are represented in Figure 1 by closed connected point set without hole in a 2-dimensional space. The 8 possible relations proposed by RCC8 are defined as follow :

- 1) DC Disconnected : X is disconnected from $Y = DC(X, Y)$
- 2) EC Externally Connected : X is externally connected from $Y = EC(X, Y)$
- 3) EQ Equal : X is equal to $Y = EQ(X, Y)$
- 4) PO Partially Overlapping : X is partially overlapping $Y = PO(X, Y)$
- 5) TPP Tangential Proper Part : X is a tangential proper part of $Y = TPP(X, Y)$
- 6) TPPi Tangential Proper Part Inverse : X is a tangential proper part inverse of $Y = TPPi(X, Y)$
- 7) NTPP Non-Tangential Proper Part : X a non-tangential proper part of $Y = NTPP(X, Y)$
- 8) NTPPi Non-Tangential Proper Part inverse : X is non-tangential proper part inverse of $Y = NTPPi(X, Y)$

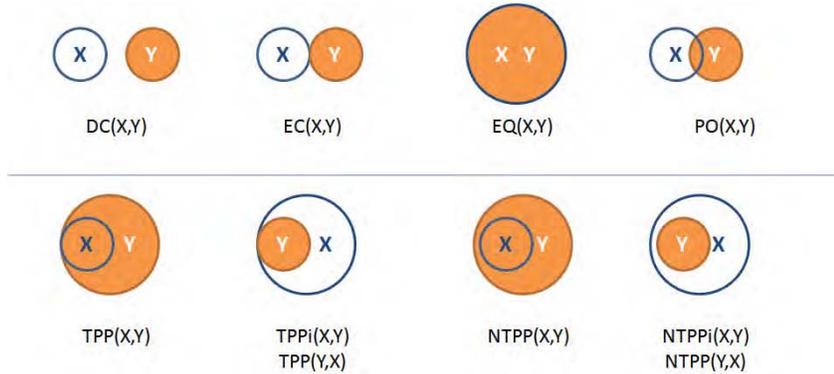


Figure 1. A graphical representation of the RCC8 relations

Notice that all these relations are mutually exclusive, that is to say between two regions only one RCC8 relation is applicable.

In parallel of the spatial relationships formalization, an important research field concerns the modelling of spatial integrity constraints in databases ([COC 97] [COC 98] [COC 01] [COC 04] [SER 00]). These constraints are expressed by a set of relations between spatial objects defined by users. These relations are supposed to be satisfied

by the data. If it is not the case, the data are considered as inconsistent and should be detected thanks to *data consistency* checking mechanisms. At present, some tools allow producing data consistency checking mechanisms from specifications of integrity constraints based on RCC8 relations. For instance, SQL code (queries or triggers) can be produced from a specification of spatial constraints [PIN 07]. This code can be used to check if a database verifies constraints or to forbid inserting data that do not verify them. Figure 2 sums up this approach. Figure 3 shows an example of spatial data that complies with the integrity constraints defined in Figure 2.

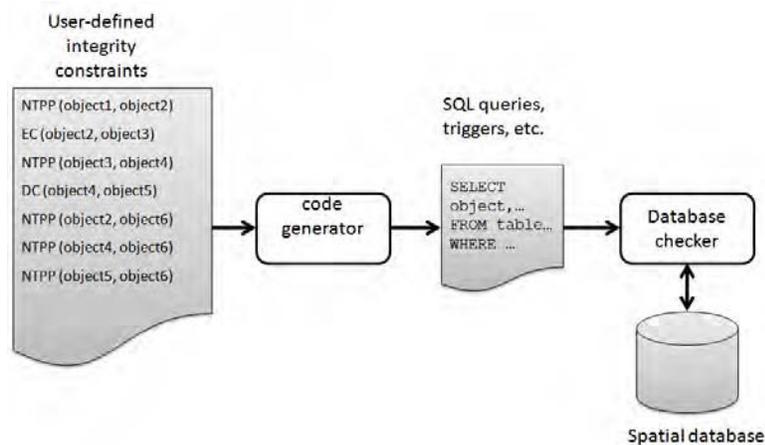


Figure 2. Data consistency checking

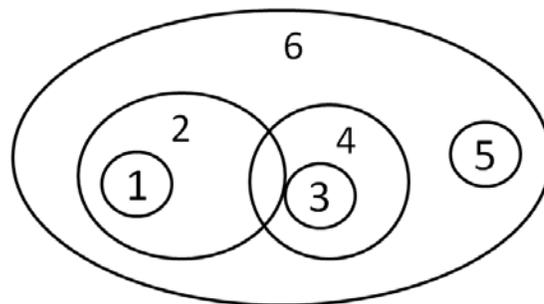


Figure 3. A spatial configuration that complies with the integrity constraints defined in Figure 2

In addition to the data consistency checking, the control of the *specification consistency* of a set of spatial integrity constraints is a complementary issue. This control aims at checking if all the spatial integrity constraints specified by users are compatible and consistent. For instance, suppose that user adds the relation EC(object3, object5) to the list of spatial integrity constraints presented in Figure 2. This new relation

leads to an inconsistency because $\text{NTPP}(\text{object3}, \text{object4})$ and $\text{DC}(\text{object4}, \text{object5})$. Thus, if the spatial integrity constraints are inconsistent, none data can fulfill these constraints. Other examples of relations that are incompatible with the set of relations presented in Figure 2 are :

- $\text{DC}(\text{object2}, \text{object4})$ because $\text{EC}(\text{object2}, \text{object3})$ and $\text{NTPP}(\text{object3}, \text{object4})$
- $\text{EC}(\text{object1}, \text{object3})$ because $\text{NTPP}(\text{object1}, \text{object2})$ and $\text{EC}(\text{object2}, \text{object3})$

In [STO 98], an operational method is proposed to help users to specify a consistent set of RCC8 relations. This method provides an algorithm that allows deducing all the possible relations from a finite set of relations specified by a user. This can be easily used to control the consistency of the relations. Only the RCC8 relations between 2 simple regions are considered in this method.

In the present paper, we propose an alternative method based on Description Logics (DL) to detect inconsistency in a set of RCC8 relations between simple regions. We introduce DL specifications of RCC8 relations (Figure 1). This method enables using traditional DIG reasoners (e.g., Pellet or Fact++) to check the consistency of relations. The proposed DL specifications could be extended in order to allow checking consistency between different types of spatial objects (composite regions, complex regions, etc.). This constitutes the main advantage of our approach. These extensions could be based on the DL translation techniques presented in this paper. In our opinion, the DL modelling of RCC8 relations opens a new and promising research field.

Representations of RCC8 relations between two spatial regions have been already proposed in Modal Logics [NUT 99] and DL [KAT 05]. Unfortunately, several errors occur in these representations. We highlight these limits in the present paper and we present correct translations of the RCC8 relations in DL.

The paper is organized as follow. Section 2 presents related works. Sections 3 introduces our proposal i.e., our translations of RCC8 relations in DL. Section 4 describes our prototype and a set of experiments.

2. Related works

There exist several logical languages for reasoning over qualitative/symbolic spatial relations. For example, the first workshop on temporal and spatial reasoning at the IJCAI conference was held in France in 1993 [GUE 98]. We do not intend to describe all of them, we just highlight the languages and tools that we thought to be most promising. [HAA 98] present a new Description Logic (DL) called $\text{ALCRP}(D)$ to combine thematic reasoning (the usual DL reasoning) and spatial reasoning. This approach is based on concrete domain that will define predicate for representing spatial relations over polygone. At that time, we do not know any reasoner available for free to implement directly this logic. [MIR 07a] [MIR 07b] propose a knowledge representation language based on object called AROM and its extension AROM-ST,

AROM-ONTO. AROM languages are implemented in a tool called ONTOAST. Indeed ONTOAST proposes a quantitative reasoning. The authors argue that AROM languages and associated tools are a complement to a qualitative spatial reasoning. [GRÜ 08] use OWL axioms to simulate RCC5 topological relations. The goal of this work is to approximate possible part-of relations and not to determine exactly which spatial objects are linked with a part-of relation. Thus the result obtains by [GRÜ 08] can not be used to check consistency in integrity constraints.

The authors of [STO 09] propose a new reasoner tool called Spatial Pellet. Spatial Pellet contains a separate RCC8 reasoner in order to provide spatial inference engine. This RCC8 reasoner is based on RCC8 composition table and uses a specific path consistency algorithm. It can be used to check the consistency of a set of spatial relations. Unfortunately, this tool does not provide information about the spatial objects responsible for inconsistencies. The outputs of Spatial Pellet indicate if a spatial configuration is inconsistent or not, but the list of inconsistent spatial relations (or objects) is not provided. Moreover, Spatial Pellet is based on a specific extension of traditional DL. This extension is not supported natively by traditional DIG reasoners and by tools such as *Protégé*.

The works of [NUT 99] propose a translation of RCC8 relations in Modal Logic. Based on this work, the authors of [KAT 05] propose a translation of the RCC8 relations in OWL-DL axioms (with SHOIN(D)). The RCC8 relations are translated into a list of axioms where a spatial region is represented by a class. This translation contains errors. We notice that the translation of the RCC8 relations do not keep the property to be mutually exclusive. We made an experiment using the translations proposed by [KAT 05] to check that if two RCC8 relations between two regions are defined, then a DL reasoner should find inconsistencies. Over 15 combinations of two RCC8 relations, we found that 6 combinations do not raise inconsistency using the Fact++ reasoner. Our work is based on the correction of the [KAT 05] translations.

Because DL reasoners are no more prototypes but have achieved the state of final products, our goal is to reuse a DL reasoner for reasoning over RCC8 relations. Moreover, this approach enables identifying precisely which subsets of relations constitute a source of inconsistency.

3. A new OWL-DL translation of RCC8 topological relations

[KAT 05] propose a first attempt to translate RCC8 relations in OWL-DL based on the Modal Logic (ML) translation proposed by [NUT 99]. We notice several drawbacks in these translations. First of all, the authors of [KAT 05] do not provide any explanation about their translation. They use a role R to express the constraints of RCC8 relations but without giving the semantics of this role. Secondly, the ML and DL translation do not keep the RCC8 relations property to be mutually exclusive.

The main problem comes from the fact that in [NUT 99] each RCC8 relation is expressed by a conjunction of two sets of axioms :

- 1) $X \cap Y = \perp$
- 2) $X \cap Y \neq \perp$

But in DL it is not easy to translate the first set of axioms $X \cap Y = \perp$. To overcome this problem we propose to define a new formal translation based on a third set of axioms $X \sqsubseteq Y$. We define our new formal translation using an extension of the well known 9-IM model of RCC8 relations. Thus we organize the description of our proposition as follow : First of all, we present the 9-IM model (section 3.1) and our extension (section 3.2). Thus, we propose in section 3.3 our new formal translation for RCC8 relations. Then we explain our DL translation in section 3.4 .

3.1. 9-IM translation of RCC8

A well-known method that can be used to model RCC8 relations is the 9-Intersection Model (9-IM) [EGE 92]. In 9-IM, each RCC8 relation is represented by a matrix. This matrix represents the intersections of boundary, interior and exterior of two spatial regions. The result of these 9 intersections might be empty (0) or not (1). Thus each RCC8 relation between two regions X and Y is represented by a 3x3 matrix whose coefficients correspond to the results of the intersection between X° (i.e., the interior of X), ∂X (i.e., the boundary of X), \bar{X} (i.e., the exterior of X) and $Y^\circ, \partial Y, \bar{Y}$.

$$M = \begin{pmatrix} X^\circ \cap Y^\circ & X^\circ \cap \partial Y & X^\circ \cap \bar{Y} \\ \partial X \cap Y^\circ & \partial X \cap \partial Y & \partial X \cap \bar{Y} \\ \bar{X} \cap Y^\circ & \bar{X} \cap \partial Y & \bar{X} \cap \bar{Y} \end{pmatrix}$$

In theory, there are $2^9 = 512$ matrixes. However, some of them are incoherent ; they cannot be drawn in a 2-dimensional space. For two simple regions, 8 meaningful configurations have been identified which lead to the following matrixes.

DC(X, Y) $\begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}$	EC(X, Y) $\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$	EQ(X, Y) $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$
PO(X, Y) $\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$	TPP(X, Y) $\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$	NTPP(X, Y) $\begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}$

Tableau 1. *Matrixes of 9-IM*

3.2. New extension of 9-IM

We will adapt the 9-IM model in order to represent the inclusion relation between regions. Thus the matrix represents the relations between boundaries, interiors and exteriors of two spatial regions.

$$\mathbf{M} = \begin{pmatrix} Relation(X^\circ, Y^\circ) & Relation(X^\circ, \partial Y) & Relation(X^\circ, \bar{Y}) \\ Relation(\partial X, Y^\circ) & Relation(\partial X, \partial Y) & Relation(\partial X, \bar{Y}) \\ Relation(\bar{X}, Y^\circ) & Relation(\bar{X}, \partial Y) & Relation(\bar{X}, \bar{Y}) \end{pmatrix}$$

The relation means that :

- (2) one of the region is included in or equal to the others $X \subseteq Y$ or $Y \subseteq X$,
- (1) there is no inclusion relation between the two regions but the intersection between two regions is not empty $X \cap Y \neq \emptyset$,
- (0) the intersection between two regions is empty $X \cap Y = \emptyset$.

The previous table becomes :

DC(X, Y) $\begin{pmatrix} 0 & 0 & 2 \\ 0 & 0 & 2 \\ 2 & 2 & 1 \end{pmatrix}$	EC(X, Y) $\begin{pmatrix} 0 & 0 & 2 \\ 0 & 1 & 1 \\ 2 & 1 & 1 \end{pmatrix}$	EQ(X, Y) $\begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$
PO(X, Y) $\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$	TPP(X, Y) $\begin{pmatrix} 2 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 2 \end{pmatrix}$	NTPP(X, Y) $\begin{pmatrix} 2 & 0 & 0 \\ 2 & 0 & 0 \\ 1 & 2 & 2 \end{pmatrix}$

Tableau 2. *Matrixes of 9-IM extension*

3.3. New formal translation of RCC8

Using our new 9-IM model, we propose a RCC8 translation in a formal language using the following operators : Let s and t be two variables , the formal axioms are build up according to the syntax rules $\top \mid \perp \mid s \sqcap t \mid s \sqcup t \mid s \sqsubseteq t \mid \bar{s} \mid \partial s \mid s^\circ$. Our goal is to replace all the formula $X \cap Y = \perp$ of the formal translation proposed by [NUT 99] with an inclusion relation $X \sqsubseteq Y$ according to our 9-IM extension model. We will have to check that set of axioms are mutually exclusive. That is to say that one component of axiom (e.g., $\partial X \cap \partial Y \neq \perp$ or $X^\circ \sqsubseteq Y^\circ$) is false when combining two formal translation of RCC8 relations.

$$DC(X, Y) \quad :- \quad X^\circ \sqsubseteq \bar{Y} \wedge \tag{1}$$

$$\partial X \sqsubseteq \bar{Y} \wedge \tag{2}$$

$$Y^\circ \sqsubseteq \bar{X} \wedge \tag{3}$$

$$\partial Y \sqsubseteq \bar{X} \tag{4}$$

$$EC(X, Y) :- X^\circ \sqsubseteq \bar{Y} \wedge \quad (5)$$

$$\partial X \cap \partial Y \neq \perp \wedge \quad (6)$$

$$Y^\circ \sqsubseteq \bar{X} \quad (7)$$

$$EQ(X, Y) :- X^\circ \sqsubseteq Y^\circ \wedge \quad (8)$$

$$Y^\circ \sqsubseteq X^\circ \wedge \quad (9)$$

$$\partial X \sqsubseteq \partial Y \wedge \quad (10)$$

$$\partial Y \sqsubseteq \partial X \quad (11)$$

$$PO(X, Y) :- X^\circ \cap Y^\circ \neq \perp \wedge \quad (12)$$

$$X^\circ \cap \bar{Y} \neq \perp \wedge \quad (13)$$

$$\partial X \cap \bar{Y} \neq \perp \wedge \quad (14)$$

$$\bar{X} \cap Y^\circ \neq \perp \wedge \quad (15)$$

$$\bar{X} \cap \partial Y \neq \perp \quad (16)$$

$$TPP(X, Y) :- X^\circ \sqsubseteq Y^\circ \wedge \quad (17)$$

$$\partial X \cap \partial Y \neq \perp \wedge \quad (18)$$

$$\bar{X} \cap Y^\circ \neq \perp \quad (19)$$

$$NTPP(X, Y) :- X^\circ \sqsubseteq Y^\circ \wedge \quad (20)$$

$$\partial X \sqsubseteq Y^\circ \wedge \quad (21)$$

$$\bar{X} \cap Y^\circ \neq \perp \quad (22)$$

3.4. New DL translation of RCC8 and associated explanation

Then we use the following assumptions to translate the formal axioms in DL ones (see Figure 4) :

– X and Y are two non empty regions. This assumption is translated in DL as : X and Y are two satisfiable classes.

– p is a point. This assumption is translated in DL as : p is an individual.

– A point x_i belongs to the region X . This assumption is translated in DL as : an individual x_i is an instance of the class X .

– \bar{X} is the exterior of X and is equivalent in DL as the complement of X \bar{X} : $-\neg X$.

– Two individuals x_i and y_j are linked with the role R , if the point x_i is defined to be close to y_j . In point set topology, we could say that two points x_i and y_j are close if for all open set O_i such as $x_i \in O_i$ then there exist an O_j such as $y_j \in O_j$ and $O_i \cap O_j \neq \emptyset$. The close relation is linked to touching point definition [NUT 99].

– The interior of X , named X° in our formal language, is defined in DL as $iX : \neg X \sqcap \forall R.X$. iX is the set of points of X that are close to only points of X .

– The boundary of X , named ∂X in our formal language, is defined in DL as $bX : \neg X \sqcap \exists R.\neg X$. bX is the set of points of X that are close to at least one point of the complement of X .

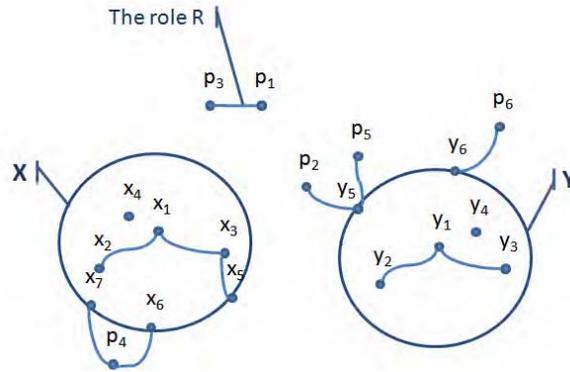


Figure 4. A graphical representation of the role R

The translation proceeds in several steps : First, we generate a class for each RCC8 relations, called RCC8 classes. For each RCC8 classes, we generate two child classes that represents spatial regions, called region classes. For each region class, we generate their child classes that represent its components : interior and border following the equations 23 to 26. Depend of RCC8 class, we generate their specific subclasses following their DL translation (presented in equations 27 to 42) : like $bXbY$ for EC relation. Then if a region is implied in several RCC8 relations, we defined equivalence relations between its region classes, subclasses of different RCC8 classes.

$$iX \equiv X \sqcap \forall R.X; \quad (23)$$

$$iY \equiv Y \sqcap \forall R.Y; \quad (24)$$

$$bX \equiv X \sqcap \exists R.(\neg X); \quad (25)$$

$$bY \equiv Y \sqcap \exists R.(\neg Y); \quad (26)$$

$$DC(X, Y) : - \quad X \sqsubseteq \neg Y; \quad (27)$$

$$Y \sqsubseteq \neg X; \quad (28)$$

$$EC(X, Y) : - \quad iX \sqsubseteq \neg Y; \quad (29)$$

$$bXbY \sqsubseteq bX; bXbY \sqsubseteq bY; \quad (30)$$

$$iY \sqsubseteq \neg X; \quad (31)$$

$$EQ(X, Y) : - \quad X \equiv Y \quad (32)$$

$$PO(X, Y) : - \quad iXiY \sqsubseteq iX; iXiY \sqsubseteq iY; \quad (33)$$

$$iXeY \sqsubseteq iX; iXeY \sqsubseteq \neg Y; \quad (34)$$

$$bXeY \sqsubseteq bX; bXeY \sqsubseteq \neg Y; \quad (35)$$

$$eXiY \sqsubseteq iY; eXiY \sqsubseteq \neg X; \quad (36)$$

$$eXbY \sqsubseteq bY; eXbY \sqsubseteq \neg X; \quad (37)$$

$$TPP(X, Y) : - \quad iX \sqsubseteq iY; \quad (38)$$

$$bXbY \sqsubseteq bX; bXbY \sqsubseteq bY; \quad (39)$$

$$eXiY \sqsubseteq \neg X; eXiY \sqsubseteq iY; \quad (40)$$

$$NTPP(X, Y) : - \quad X \sqsubseteq iY; \quad (41)$$

$$eXiY \sqsubseteq \neg X; eXiY \sqsubseteq iY; \quad (42)$$

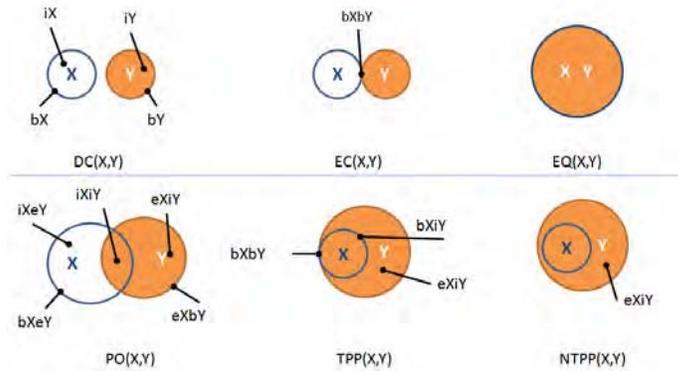


Figure 5. A graphical representation of our DL translation

4. Evaluation

In order to validate our proposition of DL translation of RCC8 relations we have implemented a code generator and made several experiments.

4.1. Prototype

We have implemented a tool supporting our approach. This tool allows producing automatically an OWL-DL specification (an ontology) from a set of RCC8 relations defined by a user. The OWL-DL specification corresponds to the DL translations introduced in this paper. This OWL-DL representation can be loaded in *Protégé* in order to check the coherency and consistency of the ontology with a DIG reasoner.

This code generator is fully adaptable. All transformation rules are defined in configuration files. One can add new topological relations and new transformation rules by modifying these configuration files. The tool has been implemented in Java and is fully compatible with MS Windows and Linux. We will make it possible to download this tool on Internet soon.

Figure 6 presents the main inputs and outputs of the tool.

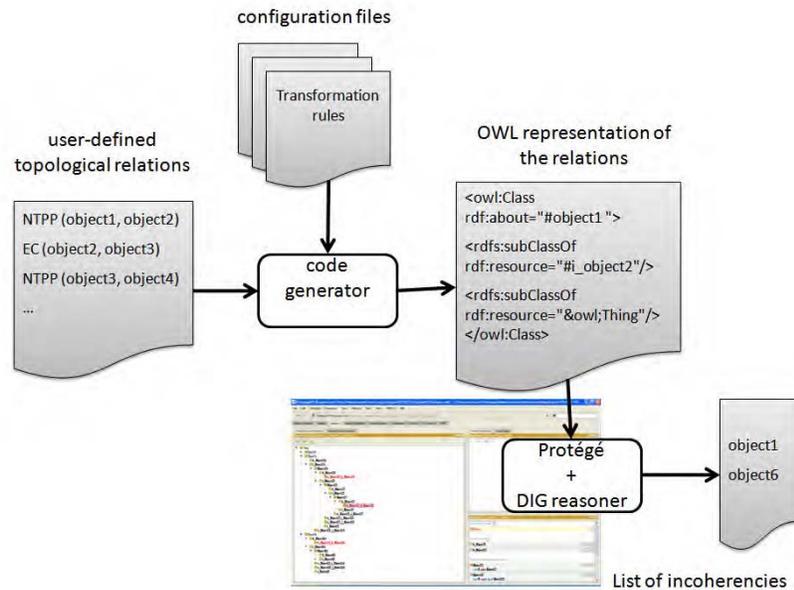


Figure 6. Architecture of our approach

In the future, we will continue to enrich the configuration files of the tool when we will propose new qualitative spatial relations and their DL representations.

4.2. Tests on two RCC8 relations

The DL axioms representing RCC8 relations should be mutually exclusive. That is to say if an ontology is composed of the set of DL axioms representing two different

RCC8 relations between two regions X and Y , this ontology should be consistent and incoherent according to [STU 08] definitions. In [STU 08], an ontology O is said to be consistent if there is an interpretation that is a model for O . An ontology O is called incoherent if there is at least one class C of O and C is unsatisfiable. A class C is said to be unsatisfiable if for all models of the ontology O , it has no instance. This type of classes is deduced by a DL reasoner.

The following table presents the results we obtain when testing that DL axioms representing two RCC8 relations are mutually exclusive. We used *Protégé* and Fact++ reasoner to check the coherency of our OWL representation. The OWL representation was built by our code generator. Each cell represents a combination of two RCC8 relations between two regions X and Y . The cell contains the unsatisfiable classes found by the Fact++ reasoner. As you can see, each combination of two different RCC8 relations raise unsatisfiability of classes.

For example when combining TPP(X,Y) and NTPP(X,Y) set of DL axioms, Fact++ infer that the class $bXbY$ is unsatisfiable because the $bXbY$ can not be contained in the interior of Y (even if the boundary of X is contained in the interior of Y) and in the boundary of Y .

relations	DC	EC	EQ	PO	TPP	NTPP
DC		bXbY	X,Y	iXiY	bXbY iX	X
EC	bXbY		iX,iY	iXiY	iX	iX,bXbY X,bX
EQ	X,Y	iX,iY		iXeY,bXeY eXiY,eXbY	eXiY	eXiY bX,bY
PO	iXiY	iXiY	iXeY,bXeY eXiY,eXbY		iXeY	iXeY,bXeY
TPP	bXbY iX	iX	eXiY	iXeY		bXbY
NTPP	X	iX,bXbY X,bX	eXiY	iXeY,bXeY	bXbY	

Tableau 3. Results of our DL translations on 2 regions

4.3. Tests on 3 or 4 RCC8 relations

In order to test our DL translations we execute some tests using 3 and 4 regions. These tests are presented in the tables 4 and 5. We used *Protégé* and Fact++ reasoner to check the coherency of our OWL representation. The OWL representation was built by our code generator. The visual editor of *Protégé* enables us to locate which regions are involved in spatial inconsistencies. In the tables 4 and 5, the first column contains the set of RCC8 relations between regions. The second column presents the problematic RCC8 relations and third column presents the unsatisfiable classes detected by

Fact++. As you can see in these tables when the set of spatial relations is consistent, the reasoner does not detect unsatisfiable classes.

Topological Relations	Incoherent Topological Relations	Detected Incoherences
NTPP(A,B) EC(B,C) EC(A,C)	EC(A,C)	bAbC
NTPP(A,B) EC(B,C) PO(A,C)	PO(A,C)	iAiC
NTPP(A,B) EC(B,C) NTPP(A,C)	NTPP(A,C)	A bA iA
NTPP(A,B) EC(B,C) TPP(A,C)	TPP(A,C)	bAbC iA
NTPP(A,B) EC(B,C) NTPP(C,A)	NTPP(C,A)	C bC, bBbC iC
NTPP(A,B) EC(B,C) TPP(C,A)	TPP(C,A)	bCbA iC
TPP(A,B) EC(B,C) EC(A,C)		
TPP(A,B) EC(B,C) DC(A,C)		
TPP(A,B) EC(B,C) NTPP(A,C)	NTPP(A,C)	bAbB iA
TPP(A,B) EC(B,C) TPP(A,C)	TPP(A,C)	iA
TPP(A,B) EC(B,C) NTPP(C,A)	NTPP(C,A)	C bC, bBbC iC
TPP(A,B) EC(B,C) TPP(C,A)	TPP(C,A)	iC

Tableau 4. Results of our DL translations on 3 regions

Topological Relations	Incoherent Topological Relations	Detected Incoherences
NTPP(A,B) NTPP(B,C) NTPP(C,D) NTPP(A,C) NTPP(A,D) NTPP(B,D)		
NTPP(A,B) NTPP(B,C) NTPP(C,D) TPP(A,C)	TPP(A,C)	bAbC
NTPP(A,B) NTPP(B,C) NTPP(C,D) TPP(A,D)	TPP(A,D)	bAbD

Tableau 5. Results of our DL translations on 4 regions

5. Conclusion

In this paper, we use a method based on Description Logics (DL) to detect inconsistency in a set of qualitative spatial relations between simple regions. We introduce DL specifications of the well-known RCC8 relations. The advantage is that traditional DIG reasoners can be used to check the consistency of relations. Representations of RCC8 relations between two spatial regions have been already proposed in Modal Logics [NUT 99] and DL [KAT 05] but several errors occur in these representations.

We have implemented a tool that produces the DL representation of RCC8 relations and made several experiments that validate our proposition.

In the future, we will extend our DL specifications proposed in order to allow checking consistency between other types of spatial objects (composite regions, complex regions, etc.). In our opinion, the DL modelling of RCC8 relations opens a new and promising research field.

6. Bibliographie

- [COC 97] COCKCROFT S., « A Taxonomy of Spatial Data Integrity Constraints », *GeoInformatica*, vol. 1, n° 4, 1997, p. 327-343.
- [COC 98] COCKCROFT S., « User Defined Spatial Business Rules : Storage, Management and Implementation - A Pipe Network Case Study », *10th Colloquium of the Spatial Information Research Centre, University of Otago, Dunedin, New-Zealand, 16-19 novembre 1998*, 1998, p. 73-81.

- [COC 01] COCKCROFT S., « Modelling Spatial Data Integrity Rules at the Metadata Level », *6th International Conference on GeoComputation, Brisbane, Australia, 2001 september*, 2001, p. 24-26.
- [COC 04] COCKCROFT S., « The Design and Implementation of a Repository for the Management of Spatial Data Integrity Constraints », *GeoInformatica*, vol. 8, n° 1, 2004, p. 49-69.
- [EGE 92] EGENHOFER M., HERRING J., « Categorizing Binary Topological Relationships between Regions », *Lines and Points in Geographic Databases, Department of Survey Engineering, University of Maine*, , 1992.
- [GRÜ 08] GRÜTTER R., SCHARRENBACH T., BAUER-MESSMER B., « Improving an RCC-Derived Geospatial Approximation by OWL Axioms », *International Semantic Web Conference*, 2008, p. 293-306.
- [GUE 98] GUESGEN H. W., ANGER F. D., LIGOZAT G., RODRÍGUEZ R. V., « Introduction to the Special Issue of CONSTRAINTS on Spatial and Temporal Reasoning », *Constraints*, vol. 3, 1998, p. 127-128, Springer Netherlands, 10.1023/A :1009761308492.
- [HAA 98] HAARSLEV V., LUTZ C., MÖLLER R., « Foundations of Spatioterminological Reasoning with Description Logics », *KR*, 1998, p. 112-123.
- [KAT 05] KATZ Y., GRAU B. C., « Representing Qualitative Spatial Information in OWL-DL », GRAU B. C., HORROCKS I., PARSIA B., PATEL-SCHNEIDER P. F., Eds., *OWLED*, vol. 188 de *CEUR Workshop Proceedings*, CEUR-WS.org, 2005.
- [MIR 07a] MIRON A. D., GENSEL J., VILLANOVA-OLIVER M., MARTIN H., « Towards the Geo-spatial Querying of the Semantic Web with ONTOAST », *W2GIS*, 2007, p. 121-136.
- [MIR 07b] MIRON A. D., GENSEL J., VILLANOVA-OLIVER M., MARTIN H., « Relations spatiales qualitatives en ONTOAST pour le Web sémantique géospatial », *Colloque International de Géomatique et d'Analyse Spatiale (SAGEO2007)*, Clermont-Ferrand, France, 2007.
- [NUT 99] NUTT W., « On the Translation of Qualitative Spatial Reasoning Problems into Modal Logics », BURGARD W., CHRISTALLER T., CREMERS A. B., Eds., *KI*, vol. 1701 de *Lecture Notes in Computer Science*, Springer, 1999, p. 113-124.
- [PIN 07] PINET P., DUBOISSET M., SOULIGNAC V., « Using UML and OCL to maintain the consistency of spatial data in environmental information systems », *Environmental Modelling and Software*, vol. 22, n° 8, 2007, p. 1217-1220.
- [RAN 92] RANDELL D. A., CUI Z., COHN A. G., « A Spatial Logic based on Regions and Connection », *KR*, 1992, p. 165-176.
- [SER 00] SERVIGNE S., UBEDA T., PURICELLI A., LAURINI R., « A Methodology for Spatial Consistency Improvement of Geographic Databases », *GeoInformatica*, vol. 4, n° 1, 2000, p. 7-34.
- [STO 98] STOCK O., *Spatial and temporal reasoning*, Springer-Verlag New York Inc, 1998.
- [STO 09] STOCKER M., SIRIN E., « PelletSpatial : A Hybrid RCC-8 and RDF/OWL Reasoning and Query Engine », *OWLED*, 2009.
- [STU 08] STUCKENSCHMIDT H., « Debugging OWL Ontologies - A Reality Check », GARCIA-CASTRO R., GÓMEZ-PÉREZ A., PETRIE C. J., VALLE E. D., KÜSTER U., ZAREMBA M., SHAFIQ M. O., Eds., *EON*, vol. 359 de *CEUR Workshop Proceedings*, CEUR-WS.org, 2008.

Vers un système d'aide à l'évaluation de la qualité de données spatio-temporelles

Dounia Azzi*, **Christine Plumejeaud***, **Marlène Villanova-Oliver***, **Jérôme Gensel***

* UMR 5217 Laboratoire d'Informatique de Grenoble,
681 rue de la passerelle,
38402 Saint Martin d'Hères
<mailto:{prenom.nom}@imag.fr>

RÉSUMÉ. *La mise en place de processus d'évaluation et de documentation de la qualité des données statistiques – des données à référence spatiale et temporelle –, est fondamentale pour rendre ces données réutilisables. Un nombre important de méthodes statistiques existent pour établir des rapports de qualité sur les données, mais les résultats de ces méthodes ne sont pas toujours concordants. Ce travail porte sur la définition d'un système qui permette de visualiser les métadonnées associées à des jeux de données statistiques, mais aussi de les enrichir avec des rapports produits par le pilotage interactif des méthodes ayant servi à les produire. Ce système propose en outre un affichage cartographique des métadonnées produites par différentes méthodes visant à en accroître la lisibilité et à faciliter leur comparaison. L'article montre notamment comment le système est utilisé pour intégrer et paramétrer des méthodes de détection de valeurs exceptionnelles.*

ABSTRACT. *The ability to measure and document the internal quality of data inside a statistical information system is essential for the usability of data. Metadata collection is a first step, but is not sufficient. Expert analysis reports may be enclosed within metadata, yet they are often in unstructured text files and difficult to automatically exploit. A great number of statistical methods can help to establish those reports, but their results are not always converging. This work is about the definition of a system that could display collected metadata on a selected subset of data, inside an interactive cartographic tool; furthermore, it can also produce interactively new reports and display their results, in order to compare the various analyses. The study case is the integration of outlier detection methods above a spatio-temporal database.*

MOTS-CLÉS : *Estimation, Géostatistique, Valeur exceptionnelle, Métadonnées, Qualité.*

KEYWORDS: *Estimation, Geostatistic, Outlier, Metadata, Quality.*

1. Introduction

L'organisation de l'activité sociale et économique d'une société requiert la connaissance d'un grand nombre d'indicateurs de développement sur des périodes temporelles étendues. Ces indicateurs (démographiques, économiques, financiers, etc.) sont en effet utilisés pour établir des scénarios d'évolution à plus ou moins long terme sur des espaces géographiques d'intérêt : l'Europe ou le Maghreb, par exemple. Des données peu fiables, incomplètes, avec de nombreuses erreurs de saisie ou simplement issues d'une méthodologie biaisée peuvent produire des scénarios d'anticipation approximatifs voire faux. Dans ce contexte, la qualité des données statistiques joue un rôle important (Chrisman, 1984). En vue de documenter cette qualité, les jeux de données sont souvent accompagnés de métadonnées qui donnent des informations sur la provenance des valeurs, sur la méthode de calcul utilisée pour les obtenir, ou encore sur une estimation de la qualité par le fournisseur lui-même. Mais ces rapports d'expertise dans un format très descriptif ne sont pas forcément exploitables de façon automatique (Dean, 1996). Nous nous intéressons ici aux autres moyens permettant de vérifier la qualité interne¹ des données.

Cette évaluation de la qualité peut reposer sur une étape de détection de valeurs exceptionnelles, qui vise à identifier les valeurs qui sont très différentes de leur voisinage (temporel, spatial et thématique). Plusieurs méthodes (géo)statistiques permettent l'évaluation de la qualité d'un jeu de données en qualifiant chacune des valeurs comme étant exceptionnelle ou non. Cependant, suivant le type de méthode employée, et/ou son paramétrage, les résultats d'évaluation d'une même valeur ne sont pas forcément concordants. Il faut alors envisager d'exécuter plusieurs méthodes afin de vérifier si ces méthodes convergent vers l'attribution du qualificatif « exceptionnelle » à la valeur. Néanmoins, cela n'est pas toujours suffisant pour qualifier la valeur d'erreur de mesure : par exemple, la valeur du Produit Intérieur Brut (PIB) du Liechtenstein est très haute par rapport à son voisinage spatial, sans être une erreur. Donc, à ce point de l'analyse, seul un expert peut trancher et décider du niveau de fiabilité de chaque valeur remarquée. Pour l'aider dans son analyse, il nous paraît pertinent de présenter la valeur suspecte avec ses métadonnées, expliquant la provenance de la valeur par exemple, mais aussi le niveau de confiance que le fournisseur lui-même accordait à cette valeur.

Dans ce cadre, un système permettant de paramétrer et d'exécuter des méthodes sur des sous-ensembles de données choisis par un utilisateur, et lui permettant de visualiser aisément leurs résultats dans une interface cartographique, serait un apport certain pour l'analyse de la qualité des données. En effet, via ce type d'interface, nous pourrions répondre aux questions suivantes: quels sont les lieux géographiques pour lesquels toutes les méthodes révèlent la présence d'une valeur exceptionnelle pour l'indicateur choisi ? Ces valeurs suspectes proviennent-elles du même fournisseur ? Via cette interface, l'expert serait amené à se questionner sur l'exactitude et la véracité d'un sous-ensemble réduit de valeurs.

¹ la qualité interne des données définit le niveau d'adéquation entre une base de données et ce qu'elle aurait dû être si elle avait été parfaite.

La première section de l'article compare notre proposition aux travaux existants, et décrit quelques méthodes de détection de valeurs exceptionnelles, destinées à illustrer l'approche proposée. Une seconde partie présente en détail le processus d'évaluation de la qualité et l'outil associé que nous développons. Appelé *QUALESTIM* (pour « Qualité, Espace, Source, Temps, Indicateur et Métadonnées »), notre système offre un support au processus d'évaluation de la qualité par l'affichage et l'exploitation conjoints des résultats produits par les méthodes statistiques et des métadonnées connues sur les données. Une troisième partie vient illustrer l'approche avec l'évaluation par l'outil d'un jeu de données. La dernière section présente les perspectives de ce travail.

2. Travaux relatifs à l'évaluation de la qualité des données

2.1. Les outils similaires

À notre connaissance, il n'existe pas d'outils visant des objectifs similaires aux nôtres. En revanche, certains outils proposent des versions approchantes de quelques-unes des fonctionnalités de *QUALESTIM*. Par exemple, des outils d'analyse spatiale comme *SADA* (SADA, 1997), *Geoda* (Geoda, 1998), *CrimeStat* (CrimeStat, 2009) proposent des fonctions d'analyse statistique spatiale, couplées à des fonctions de visualisation et d'exploration de données, qui sont assez intéressantes lorsque l'on souhaite faire de l'estimation de valeur manquante : interpolation par krigeage complexe, simulation spatiale. En effet, estimer pour compléter une valeur manquante est une opération souvent similaire à la détection de valeurs exceptionnelles : si la valeur mesurée est très loin de la valeur estimée, alors elle peut-être qualifiée d'anormale. Cependant, en dehors de *Geoda*, qui est open-source, le code réalisant les fonctionnalités de ces outils n'est pas réutilisable par d'autres développeurs. De plus, ces logiciels importent des fichiers décrits dans des formats propriétaires (*Shapefile*, *DBF*, *Excel*, etc.) sans fournir de connexion vers des bases de données spatiales, ni temporelles.

D'autres outils, comme *QuantumGis* (QuantumGis, 2002), ou bien *GRASS GIS* (Grass, 2010) proposent une lecture de données depuis des bases de données libres (*PostgreSQL*, ou *MySQL*), couplées à des fonctions de visualisation des données, par interface cartographique. Cependant, les fonctions d'analyse spatiale sont beaucoup moins évoluées que celles des outils précédemment cités.

Toutefois, il faut noter que, si pour *GRASS GIS* la liste des fonctionnalités d'analyse spatiale disponibles immédiatement pour l'utilisateur est plus restreinte, elle peut être étendue. En effet, le logiciel donne la possibilité d'intégrer des scripts pour l'analyse statistique programmés avec *R*. La librairie open-source *R* (*Projet R*, 1997) se révèle être un langage d'expression privilégié pour de nombreux statisticiens (Templ *et al.*, 2009). Par exemple, *Geoda* est devenu libre, et il s'appuie sur *R* : ses contributeurs enrichissent la bibliothèque *R* avec de nouvelles méthodes régulièrement. De même, certains services Web pour l'interpolation de données spatiales comme, par exemple, *INTAMAP* (Pebesma *et al.*, 2010), sont aujourd'hui

basés sur R. Si l'idée de services de calculs distribués nous semble très pertinente, l'interface d'utilisation nous apparaît comme trop sommaire.

Toutefois, aucun de ces outils ne propose une interface simple et intuitive du point de vue de l'utilisation des fonctions d'estimation. En effet, pour comprendre leur utilisation, il est nécessaire de parcourir très attentivement la documentation qui les accompagne, lorsque cette dernière existe. De plus, le choix des paramètres nécessaires au bon fonctionnement des méthodes de calculs n'est pas évident, un utilisateur novice serait confronté à de nombreuses difficultés.

Nous avons également constaté que ces outils ne fournissent pas d'informations sur les métadonnées sous un format non textuel, par exemple, au moyen de cartes et de représentations interactives qui permettraient à l'utilisateur de mettre facilement en relation les informations collectées sur le jeu de données qu'il analyse avec les résultats calculés. Ils ignorent tout à fait la présentation des métadonnées associées aux données, puisque le schéma d'importation des données n'intègre pas l'import des métadonnées : les données spatiales, en particulier, sont simplement réduites à l'association d'un fond de carte (l'ensemble des géométries) et d'attributs thématiques, sans aucune information sur la provenance des valeurs, ou description des méthodes d'évaluation des données. L'affichage des fiches de métadonnées se répand dans les SIG actuels, mais encore sous une forme qui reste très primitive (un document textuel séparé), et leur consultation est séparée des données : les valeurs continuent d'apparaître dans les interfaces de façon assez dépouillée.

Ainsi, le développement d'un outil d'évaluation de la qualité des données, basé entre autres sur des méthodes statistiques, s'avère nécessaire. Cet outil doit fournir une interface simple, dont le paramétrage serait explicite, et qui permettrait de confronter et comparer les résultats des méthodes statistiques aux métadonnées existantes. De l'étude des travaux existants, nous retenons que l'usage de R comme outil d'analyse, statistique et géostatistique, s'avère un choix pertinent.

2.2. Méthodes statistiques de détection de valeurs exceptionnelles

Une valeur exceptionnelle (*outlier* en anglais) est définie comme une observation qui dévie de la valeur moyenne de l'échantillon dans lequel elle est observée (Grubbs, 1969). Elle est détectée via des méthodes (géo)statistiques cherchant à mesurer le niveau de ressemblance de la valeur avec l'échantillon selon des critères qui peuvent tenir compte de la dimension thématique, spatiale ou temporelle. Ainsi, dans le cas des méthodes spatiales, les indices de ressemblance sont calculés en fonction d'un voisinage spatial, alors que dans le cas de méthodes thématiques, le calcul porte sur la corrélation entre les variables. Une implémentation open-source en R d'un sous-ensemble de méthodes connues pour la détection d'*outliers*, (Rousseeuw *et al.*, 1996), ainsi que de méthodes géostatistiques plus originales nous a été fournie par le *National Centre for Geocomputation*, (Charlton *et al.*, 2010). Afin d'illustrer notre proposition, en particulier du point de vue *i)* de l'intégration de méthodes de détection de valeurs exceptionnelles à un système, et *ii)* de la problématique du paramétrage, nous fournissons ici une description plus complète

de ces méthodes qui seront à terme intégrées à *QUALESTIM*. Les méthodes sont présentées selon le type de dimension explorée.

Parmi les méthodes aspatiales, une des méthodes les plus simples est celle de la boîte à moustaches (*Boxplot*), inventé par John Tukey en 1977. Elle permet de situer rapidement le profil d'une donnée, appartenant à un ensemble univarié de valeurs, en la comparant à des constantes statistiques calculées pour ce même ensemble (médiane, quartiles, minimum, maximum). Par exemple, une valeur est dite exceptionnelle si elle est supérieure (respectivement inférieure) à la valeur du quartile supérieur (resp. inférieur) augmentée (resp. diminuée) de trois fois l'écart interquartile. La méthode peut être généralisée pour un jeu de données multivarié (*Bagplot*). Elle ne nécessite pas de paramétrage particulier, contrairement à une autre méthode proposée, basée sur la distance de Mahalanobis, qui s'utilise aussi pour la recherche de valeurs exceptionnelles dans un ensemble multivarié. Cette métrique permet de mesurer la distance d'un point à la moyenne de la distribution des valeurs. Une forte valeur (supérieure à un seuil prédéfini) montre que l'observation considérée est en périphérie de nuage, et doit donc être considérée avec attention, voire avec suspicion. La dernière méthode aspatiale proposée est l'Analyse en Composantes Principales (ACP) qui, tout en déterminant les principaux axes d'une distribution multivariée, mesure l'écart de chaque valeur à ces axes, sans autre paramétrage que le choix des indicateurs supplémentaires qui entrent dans l'analyse.

En ce qui concerne les méthodes spatiales, nous disposons d'une implémentation de la technique de Hawkins, qui permet la détection d'*outliers* pour un jeu de données univarié. Son principe est de considérer, dans un premier temps, toute donnée $z(x_i)$ comme étant anormale, puis, de la considérer réellement comme une valeur exceptionnelle si le test du Chi2 [1] est vérifié :

$$N \left(\frac{(z(x_i) - m_l)^2}{(N + 1)s_l^2} \right) \geq X_{crit-1}^2 \quad [1]$$

Pour i variant de 1 à n , x_i étant la position spatiale de la valeur $z(x_i)$, N représente la distribution gaussienne des valeurs $z(x_i)$ centrées et réduites sur les voisinages spatiaux considérés, m_l étant la moyenne locale, et s_l^2 la variance moyenne. X_{crit-1}^2 est une valeur critique du Chi2 de degré de liberté 1. Il faut ici définir un voisinage spatial approprié pour le jeu de données considéré, via la portée, un paramètre de type flottant, toujours positif, et dont les valeurs minimale, maximale et optimale dépendent des propriétés de l'espace d'étude sélectionné (nombre d'unités spatiales, forme et surface). La portée permet de déterminer quelles unités territoriales sont considérées pour le calcul de la moyenne locale et de la variance. Des algorithmes simples de calcul basés sur les enveloppes des unités sont envisagés. Le second paramètre de cette méthode définit la valeur critique du Chi2 par un type énuméré prenant ses valeurs dans l'ensemble suivant : {95%, 99%, 99,9% }.

Par le biais des méthodes de régression, nous pouvons aussi déterminer si une valeur appartient à la moyenne locale m_i . Le modèle de prédiction utilisé pour calculer m_i est appelé le prédicteur localisé moyen. Dans cette catégorie, nous disposons de l'implémentation de trois types de méthode: une méthode paramétrique par régression linéaire multiple, (MLR), une méthode non paramétrique par régression locale sur les attributs (LR) et la régression géographiquement pondérée, ou *Geographically Weighted Regression*, (GWR) par (Fotheringham *et al.*, 2002). Pour le bon fonctionnement de ces méthodes, il est encore nécessaire de spécifier la portée qui permet à l'utilisateur de contrôler le caractère local des poids.

3. QUALESTIM, un support pour le processus d'évaluation de la qualité

Suite à l'étude des outils approchant nos besoins, nous avons retenu R comme langage d'expression des méthodes statistiques de détection de valeurs exceptionnelles. Nous articulons cette fonction de calcul avec *i)* des fonctions d'extraction de données d'une base de données spatio-temporelles, accompagnées de leurs métadonnées, *ii)* la production de rapports d'analyse, et *iii)* l'affichage des résultats basé sur une visualisation cartographique appropriée. La figure 1 présente une vue globale de l'architecture de l'application.

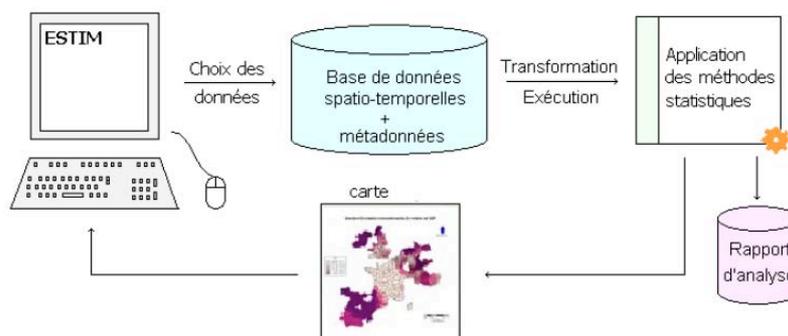


Figure 1. Architecture de l'application.

Les étapes du processus d'analyse sont les suivantes :

- L'utilisateur choisit le jeu de données qu'il souhaite analyser, via une interface qui interroge le contenu de la base de données spatio-temporelle.
- Les données sont extraites de la base de données et formatées pour les méthodes de calcul statistique.
- L'exécution de chaque méthode sélectionnée et paramétrée par l'utilisateur produit un rapport d'analyse, qui est conservé et peut être combiné avec d'autres. Ce rapport est affiché sous la forme de cartes et de diagrammes.
- Dans cette interface, l'utilisateur consulte simultanément les métadonnées et les données.

L'interface graphique de *QUALESTIM* (voir Figure 2) est structurée selon la logique de ce processus. Une première zone (zone 1 de la Figure 2) de cette interface est dédiée à la sélection des données à analyser : l'utilisateur doit choisir un espace d'étude, un niveau de zonage, un jeu de données, l'indicateur dont il souhaite évaluer la qualité, et la période temporelle. Dans une seconde zone, il choisit la méthode d'évaluation qu'il souhaite appliquer à l'échantillon sélectionné et, spécifie les valeurs des paramètres des scripts (si nécessaire). Dans la troisième zone, les résultats de ses requêtes et analyses sont affichés. La quatrième zone est dédiée aux métadonnées associées au contexte de cette analyse.

Espace d'étude et maillage

Jeu de données
 Aire
 Maillage

Indicateur principal

Indicateur
 Jeu de données
 Date

Visualisation

Indicateur | Outliers | Histogramme | Boxplot | Bagplot

Suspected outliers - weak to strong (yellow to dark red) evidence

3

Indicator sum (max.: 7)

□	under 1
□	1 to under 3
□	3 to under 5
□	5 to under 7
■	exactly 7

0 0.5 1 1.5 2 x1000 km

1900 ← → 2000

Métadonnées

Indicateur | Valeur | Expertise

4

Fiabilité

Valeur estimée
 Qualité

Source

URL
 Extraite le

Fournisseur

Officiel
 Nom
 Code

Evaluation de la qualité des données

Variables auxiliaires

Univarié Bivarié Multivarié

Caractérisation par rapport aux dimensions

Thématique Spatiale Temporelle

Méthodes et paramètres

Méthodes
 Paramètre
 Autres indicateurs

Figure 2. Maquette de l'interface de QUALESTIM.

3.1. Choix, extraction et transformation des données

Les données statistiques sont extraites d'une base de données spatio-temporelles dont le schéma est établi à partir d'un modèle identitaire évolutif (Plumejeaud *et al.*, 2009). Ce modèle est centré sur les *unités géographiques*, qui possèdent une période de validité, une identité (nom, code, centre, etc.), une extension spatiale (la

géométrie des unités, qui est versionnée) et une partie thématique qui décrit les indicateurs statistiques disponibles sur ces unités, les valeurs associées, avec leurs différentes périodes de validité. Le modèle est structuré de façon à rendre compte de la dimension hiérarchique de l'organisation spatiale, qui sera réutilisée lors de l'interrogation du modèle. En effet, il décrit les zonages (et leurs versions) comme des sous-ensembles d'unités formant un découpage d'un certain niveau du territoire, découpage valide durant une certaine période. Par exemple, la NUTS (Nomenclature des Unités Territoriales Statistiques) comprend six versions de nomenclature (1980, 1988, 1995, 1999, 2003, 2006), qui définissent cinq niveaux administratifs chacune (des communes aux états). Or, les méthodes d'analyse spatiale tiennent compte des géométries des unités sur lesquelles sont collectés les indicateurs. Ces géométries changent suivant les versions de nomenclature. En termes d'interface et de requêtes, il est donc essentiel de sélectionner de façon transparente pour les méthodes la version de zonage correspondant à chaque date de la période de traitement de l'indicateur. De même, l'utilisateur peut vouloir restreindre ses requêtes à des espaces d'étude (l'Europe des 15, l'Arc Atlantique, par exemple), qui sont des sous-ensembles d'unités sur lesquels tous les niveaux de découpage ne sont pas valides, ni tous les indicateurs disponibles. Ainsi, le choix de l'aire d'étude est à synchroniser avec le choix du zonage, comme avec le choix d'indicateur et de la période d'analyse. Les unités du jeu de données qui ne sont pas évaluées sont écartées automatiquement de l'analyse.

Par ailleurs, cette base de données conserve également un ensemble de métadonnées qui ont été collectées suivant un profil de la norme ISO19115 adapté pour les données statistiques (Plumejeaud *et al.*, 2010), et qui apportent des informations sur la provenance (ou lignage) des valeurs. En effet, les mêmes indicateurs (dont nous conservons le nom, la description, l'unité de mesure, et la classification) peuvent provenir de jeux de données différents, avec des valeurs différentes pour une même unité à un moment donné. Par ailleurs, toutes les valeurs d'un indicateur issues d'un certain jeu de données ne sont pas forcément collectées ni calculées de façon homogène sur l'espace géographique ou sur la période temporelle considérés. En conséquence, nous devons présenter des métadonnées renseignant à la fois sur le lignage de chaque valeur, et sur le lignage de l'indicateur.

Nous avons donc développé un module de connexion à la base de données permettant d'interroger et d'extraire les données sur la base des critères définis plus haut concernant le contexte d'analyse. Le modèle de la base de données doit être conforme au schéma spatio-temporel qui a été décrit dans (Plumejeaud *et al.*, 2009).

Les données extraites de la base de données doivent être converties dans le format d'entrée du langage R, le langage d'implémentation des méthodes dans *QUALESTIM*. En effet, pour ce qui concerne les données spatio-temporelles, c'est-à-dire les géométries des unités géographiques, R utilise comme attribut spatial associé aux valeurs statistiques une structure spéciale de données, détaillée dans (Bivand *et al.*, 2008).

3.2. Choix, paramétrage et exécution de méthodes statistiques

Dans la zone 2 de l'interface graphique, l'utilisateur est amené à choisir le type de méthode d'évaluation de la qualité qu'il souhaite exécuter sur le jeu de données sélectionné. Une première partie de cette zone propose à l'utilisateur de définir le nombre de variables auxiliaires qu'il souhaite mettre en relation avec l'indicateur principal. Ces variables auxiliaires sont d'autres indicateurs de développement. Le champ « univarié » signifie que seul l'indicateur principal est étudié, alors que les champs « bivarié », respectivement « multivarié », annoncent la prise en compte d'un seul, respectivement plusieurs, indicateurs. Par exemple, dans le cadre d'une étude bivariée, l'utilisateur pourrait choisir comme indicateur principal, le Produit Intérieur Brut (PIB), et comme variable auxiliaire, le taux de chômage (sur une même période temporelle). Une fois ce choix effectué, la dimension de l'analyse est à spécifier. L'utilisateur peut choisir entre une ou plusieurs dimensions : thématique, spatiale ou temporelle. Pour la dimension spatiale, les méthodes vérifient la variabilité des valeurs en fonction de voisinages spatiaux, alors que pour les méthodes temporelles, c'est la variabilité dans le temps qui est mesurée.

Ainsi, en fonction du nombre de variables auxiliaires sélectionnées et des dimensions choisies, la liste des méthodes disponibles est mise à jour. En effet, une correspondance a été définie dans le modèle des méthodes entre le type d'analyse (dimensions et nombre de variables) et les méthodes implantées dans QUALESTIM. Par exemple, la méthode du Boxplot prend en paramètre un jeu de données univarié et analyse suivant la dimension thématique, alors qu'une technique par Régression Linéaire Multiple considère la variabilité d'un jeu de données multivarié selon la dimension spatiale. De plus, chaque méthode requiert des paramètres qui lui sont spécifiques et l'interface est mise à jour en fonction de la méthode sélectionnée.

Si l'utilisateur choisit d'effectuer une analyse spatiale de l'indicateur « Produit Intérieur Brut » sur un jeu de données multivarié, alors, parmi les méthodes proposées, il pourra choisir entre la régression linéaire multiple, la régression locale ou encore la régression géographiquement localisée. Si son choix se porte sur la méthode de régression linéaire multiple, alors il devra spécifier plusieurs paramètres, dont la portée de son analyse, par exemple 80 kilomètres, et les indicateurs qu'il souhaite lier à l'indicateur principal : par exemple, le taux de chômage et les subventions européennes pour l'agriculture.

3.3. Exploitation et Interprétation des résultats

Les résultats de l'analyse statistique apparaissent dans la partie « Visualisation » de l'interface (zone 3). L'onglet "Indicateur" représente la carte avec les données brutes choisies par l'utilisateur. L'onglet "Outliers" permet la visualisation de la

dernière évaluation effectuée par la méthode sélectionnée. Les autres onglets comportent toutes les autres formes de graphiques calculés par cette méthode (histogramme, *boxplot*, etc.).

L'interface est prévue pour autoriser la comparaison de plusieurs résultats de méthodes. Pour cela, nous associons un drapeau à chaque rapport d'analyse produit par l'exécution d'une méthode. La couleur du drapeau est choisie de manière arbitraire (aléatoire) et ne correspond pas à une sémantique particulière. L'utilisateur peut revoir un ancien rapport en sélectionnant le drapeau correspondant. En effet, au survol du drapeau par la souris (voir figure 3), une bulle apparaît et contient des informations identifiant le rapport, comme sa date d'exécution, et des informations plus descriptives : la méthode et les paramètres utilisés.

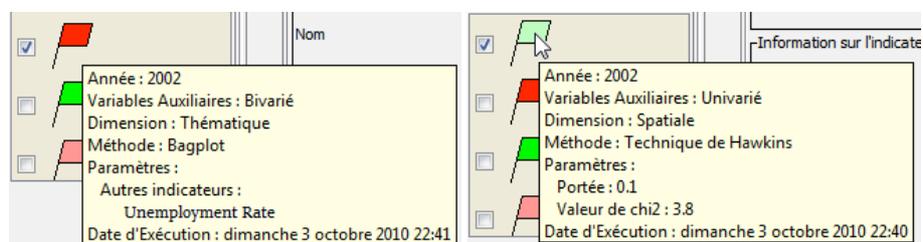


Figure 3. Choix de rapports d'analyse associés au contexte courant de l'analyse.

Les rapports d'analyse de chaque méthode n'étant pas forcément concordants pour une même valeur, nous proposons plusieurs méthodes de combinaison des résultats, afin de restituer une vue d'ensemble de ces analyses. La combinaison des résultats consiste à superposer les cartes dans l'onglet « Outliers » suivant trois méthodes, qui sont proposées à l'utilisateur. L'« agrégation » consiste à assombrir l'aplatissement de couleur d'une unité territoriale lorsqu'augmente le nombre d'exécutions de méthodes considérant sa valeur comme exceptionnelle. Le « minimum » représente les unités territoriales dont la valeur d'indicateur n'est jamais évaluée comme exceptionnelle par l'ensemble des analyses exécutées, et le « maximum », celles dont la valeur d'indicateur est toujours considérée comme exceptionnelle.

Le *slider* temporel (ou barre de progression temporelle) permet à l'utilisateur de définir la date à laquelle il souhaite visualiser le résultat de l'analyse pour l'indicateur choisi. En effet, pour un indicateur annuel couvrant une période temporelle de dix ans par exemple, la méthode fournit une analyse des valeurs pour chaque année et les résultats de l'analyse peuvent varier en fonction des années. L'utilisateur peut déplacer le curseur du *slider* et visualiser les résultats de l'analyse de l'indicateur pour chaque année de l'intervalle de temps choisi.

3.4. Exploitation des métadonnées

Pour rappel, l'objectif est d'enrichir les métadonnées qui accompagnent les valeurs statistiques à partir des résultats de l'évaluation de la qualité des données. Dans un premier temps, l'utilisateur accède aux métadonnées fournies avec le jeu de données dans la zone 4 de l'interface. Lorsque l'utilisateur parcourt la carte avec la souris, les onglets de métadonnées sont mis à jour en fonction de l'unité géographique survolée. Ils présentent des informations sur les indicateurs ou sur les valeurs (identité du fournisseur, description, etc.), cf. Figure 2. Concernant les indicateurs, nous affichons le nom de l'indicateur, le code de l'indicateur, un résumé, une unité de mesure, le nom du jeu de données et la date d'acquisition. Ces informations sont importantes car elles indiquent à l'utilisateur la provenance des valeurs analysées, selon plusieurs niveaux d'information (jeu de données, indicateur et valeur elle-même).

Cette exploitation des métadonnées issues de la base de données permet de mettre en relation ces connaissances avec l'estimation faite par les méthodes de la qualité des valeurs observées. Par exemple, l'utilisateur peut observer que, très fréquemment, une source de données particulière apporte de nombreuses valeurs exceptionnelles, et ceci peut aussi l'amener à réviser le niveau de qualité estimé d'un groupe de valeurs, à la baisse ou à la hausse. Ainsi, suite à l'exécution d'une ou plusieurs méthodes sur un jeu de données, l'utilisateur a la possibilité de réévaluer le niveau de qualité des données dans l'onglet Expertise de la zone 4, en y associant le rapport d'analyse généré.

Les rapports d'analyse des différentes exécutions de méthodes de détection de valeurs exceptionnelles sont conservés, durant la session de l'utilisateur, dans des tables temporaires (voir figure 4), et qui sont associées à un contexte d'analyse défini par les choix de l'utilisateur dans la première zone (cf. Figure 2, zone 1). À chaque exécution (*Run*) est associée un utilisateur (*userId*), et une date d'exécution (*date*), mais aussi un contexte d'analyse qui est défini par le niveau de zonage (*zoningLevel*), le nom de l'aire d'étude, (*studyAreaName*), la période de l'analyse représentée par l'intervalle [*startTimePeriod*, *endTimePeriod*], les indicateurs (*Indicator*) impliqués et les jeux de données (*Dataset*) dont ils sont issus chacun. On qualifie le rôle de chaque indicateur dans l'exécution, afin de signaler si les valeurs de l'indicateur étaient l'objet de l'étude (*inspected*), ou bien des variables auxiliaires (*ancillary*). La méthode de calcul employée (*Method*) et les valeurs de paramètres utilisés (*ParameterInstance*) sont associées à cette exécution. Les résultats de l'exécution (*Analysis*) donnent lieu à une notation de la valeur inspectée (*Value*) de type numérique entier: plus la note est élevée, plus la méthode considère comme suspecte la valeur. À zéro, la valeur est considérée comme normale durant cette exécution.

Un modèle des méthodes est introduit ici, pour décrire les instances des méthodes utilisées. Chaque méthode est modélisée par son nom (*name*) et sa description (*description*), le nombre d'indicateurs auxiliaires qu'elle peut prendre

en compte (*nbAncillaryVar*), et les dimensions explorées : *timeDim*, *spaceDim* ou *thematicDim* qui valent vrai si la dimension temporelle, spatiale ou thématique respectivement sont explorées. De même, les paramètres (*Parameter*) que la méthode doit utiliser sont décrits par leur nom (*name*), et leur type (*EnumType*). Le type peut être un type de base parmi l'énumération suivante : booléen, entier, réel ou chaîne de caractères. Lorsque le paramètre est de type numérique, sa valeur minimum (*min*), maximum (*max*) sont précisées, et une formule stockée dans une chaîne de caractères, *optimumFormula*, décrit comment spécifier une valeur adaptée au jeu de données sélectionné.

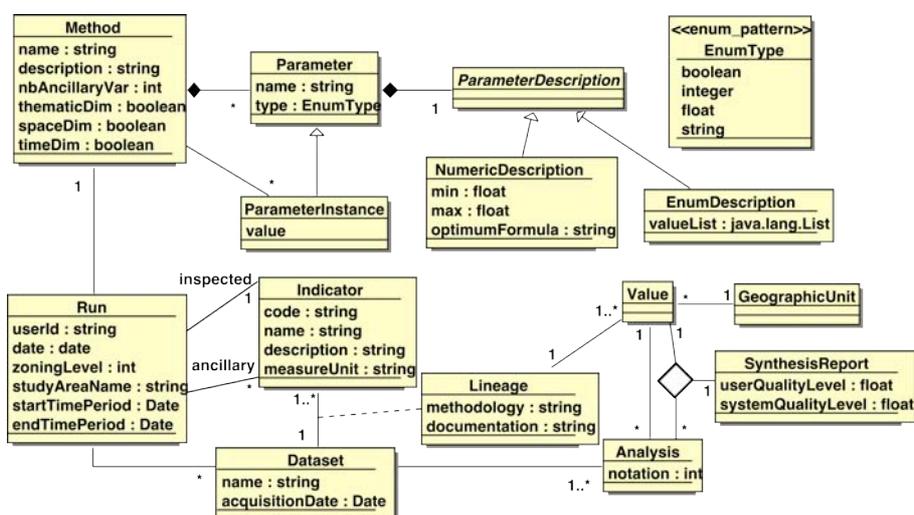


Figure 4. Diagramme UML des tables temporaires pour le stockage des résultats des scripts.

Après l'exécution d'un ensemble de méthodes de détection, une synthèse des méthodes d'estimation peut être établie. Elle est représentée sous la forme d'une classe *SynthesisReport*, associée à la valeur inspectée (*Value*), et à l'ensemble des rapports d'analyse (*Analysis*) qui servent de base à cette synthèse. La synthèse produit automatiquement un indicateur, *systemQualityLevel*, qui représente le rapport entre la somme des notations appliquées à la valeur, et le nombre d'exécutions considérées. Ce pourcentage augmente avec le nombre de méthodes qui estiment la valeur exceptionnelle. Néanmoins, l'utilisateur peut ensuite rectifier ce pourcentage avec sa propre estimation de la qualité, établie d'après ses connaissances (*userQualityLevel*). L'objectif des valeurs *systemQualityLevel*, qui sont affichées dans l'interface sur une carte choroplèthe, est d'attirer l'attention de l'expert, qui peut trancher après analyse avec sa propre notation sur la valeur. Nous enrichissons ainsi progressivement les métadonnées avec des rapports d'expertise documentés. Il est ainsi possible de savoir quelles méthodes ont produit les rapports d'analyse, et dans quelles conditions (liste des paramètres utilisés).

4. Illustration par un cas d'étude

Nous présentons un cas d'étude permettant d'illustrer notre proposition, et qui montre comment QUALESTIM permet de réaliser l'évaluation de la qualité d'un jeu de données. L'évaluation porte sur l'évolution du Produit Intérieur Brut (PIB) sur les pays de l'Union Européenne de 2000 à 2005, dans la version de nomenclature des NUTS 2006 au niveau 3. L'indicateur observé est un pourcentage et a été calculé en effectuant un ratio du PIB par habitant de 2005 à 2006 par le PIB par habitant de 2000 à 2005.

Dans la zone 1 de l'interface, nous choisissons donc les paramètres correspondant à ce jeu de données. L'interface est mise à jour automatiquement : une carte choroplèthe apparaît désormais dans la zone de visualisation, et les unités sont colorées en fonction de la valeur du PIB. Au survol de l'unité dans la carte, les métadonnées s'affichent à droite : par exemple, alors qu'une majorité de valeurs sont issues d'Eurostat, en Bulgarie, pour Kyustendil, la valeur (initialement manquante) est issue des estimations menées par le projet ESPON 2013 Database.

Dans ce scénario, nous avons choisi d'analyser ce jeu de données suivant plusieurs méthodes univariées, spatiales ou thématiques, qui sont : le *Boxplot* standard et ajusté, la technique de Hawkins, la moyenne locale, la régression linéaire multiple, la régression locale et la régression géographiquement pondérée. La carte des valeurs exceptionnelles issues de la technique de Hawkins (méthode spatiale) indique notamment que la valeur d'indicateur de l'unité territoriale « Kyustendil » en Bulgarie est exceptionnelle. Par l'usage du *Boxplot* standard, la valeur d'indicateur des unités territoriales « Latgale » en Lettonie ou, de nouveau, « Kyustendil » en Bulgarie sont considérées comme étant également exceptionnelles.

La Figure 5 présente une vue de l'interface positionnée sur la carte des *outliers*, qui synthétise le résultat des analyses pour ce jeu de données.

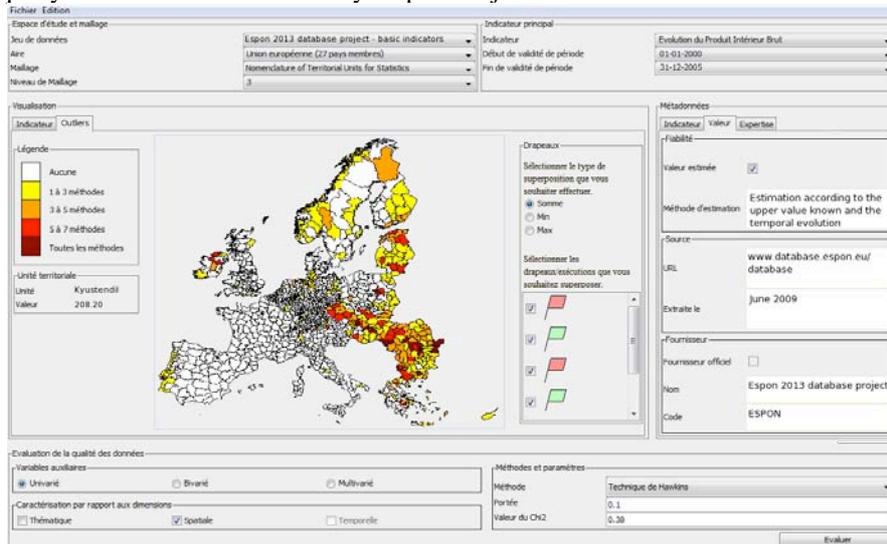


Figure 5. Carte des outliers dans le prototype de Qualestim.

En choisissant le type « agrégation », la carte de synthèse des analyses présente, l'unité territoriale de « Latgale », en jaune car sa valeur d'indicateur est considérée comme exceptionnelle par au moins une évaluation, alors que l'unité territoriale « Kyustendil » est en rouge foncé, sa valeur d'indicateur ayant été considérée comme exceptionnelle par toutes les évaluations. Ainsi, l'utilisateur, amené à s'interroger sur la valeur de Kyustendil, remarque que la valeur résulte d'un processus d'estimation (information issue des métadonnées). En sa qualité d'expert, l'utilisateur peut alors revoir à la baisse la fiabilité accordée à cette valeur, et par la suite réévaluer son niveau de qualité dans l'onglet « expertise ». Cette correction est automatiquement associée au rapport produit par les méthodes.

5. Conclusion et Perspectives

Nous avons décrit dans cet article une application permettant d'évaluer la qualité interne des données d'une base de données spatio-temporelles via une interface de type cartographique spatio-temporelle. De façon concrète, cette interface permet de paramétrer et d'exécuter des méthodes statistiques pour la détection de valeurs exceptionnelles, sur des données choisies par l'utilisateur. Elle offre ensuite le moyen de comparer plusieurs analyses dans une carte, et de les interpréter à la lumière des métadonnées collectées sur les données. Un premier prototype fonctionnel de *QUALESTIM* a été développé.

Nous travaillons actuellement à la mise en place d'un protocole de tests, qui impliquera des utilisateurs de niveaux d'expérience variés (débutant à expert), afin de mesurer leur satisfaction en termes d'apport et d'efficacité de l'outil, de pertinence des résultats produits ou encore, de simplicité d'utilisation.

Parallèlement, une extension du système *QUALESTIM* est à l'étude visant à le doter de fonctionnalités de complétion de valeurs manquantes, en intégrant des méthodes d'estimation de données manquantes, comme le proposent d'autres travaux basés sur R (Templ *et al.*, 2009). Une perspective complémentaire consisterait à proposer une description des méthodes, via le langage XML (Pebesma *et al.*, 2010) par exemple, au même titre qu'un service web géographique de traitement de données (WPS, 1994). Un utilisateur aurait donc la possibilité d'implémenter et décrire de nouvelles méthodes, étendant ainsi progressivement les capacités d'estimation de l'outil.

Remerciements

Nous remercions Martin Charlton et Paul Harris du National Centre for Geocomputation (NCG, <http://ncg.nuim.ie/>), de l'Université Nationale d'Irlande, pour avoir mis gracieusement à notre disposition les scripts R mettant en oeuvre les méthodes de détection d'*outliers*.

6. Bibliographie

- Bivand, R. S., Pebesma, E. J., Gómez-Rubio, V. Applied Spatial Data Analysis with R, XIV, 2008, 378 p., Springer
- Chrisman, N. R., The role of quality information in the long-term functioning of a geographic information system. *Cartographica*, 21, 1984, pp. 79-87.
- Dean P., Sundgren B., Quality Aspects of a Modern Database Service. In: Proc. of the 8th Int. Conf. on Scientific and Statistical Database Management, SSDBM'96, 1996, pp. 156-161
- Fotheringham A., Brunson C., and Charlton M., *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*, Chichester: Wiley, 2002.
- Grasland C., Martin H., Vincent J.M., Gensel J., Mathian H., Oulahal S., Cuenot O., Edi E., Lizzi L., *Le projet Hypercarte : analyse spatiale et cartographie interactive*, Colloque International de Géomatique et d'Analyse Spatiale , SAGEO, 2005
- Grubbs, F. E., Procedures for detecting outlying observations in samples. *Technometrics* 11, 1–21, 1969.
- Harris P. and Charlton M., “*Spatial analysis for quality control, phase 1: The identification of logical input errors and statistical outliers*” ESPON, Tech. Rep., 2010.
- Pebesma E. J., Cornford D., Cornford G., Heuvelink G. B., Hristopoulos D., Pilz J., Stohler U., Morin G., and Skoien Jon O., “*INTAMAP: the design and implementation of an interoperable automated interpolation web service*, *Computers & Geosciences*, vol. 2, pp. 12–30, 2010.
- Plumejeaud, C., Gensel, J., Villanova-Oliver, M., Opérationnalisation d'un profil ISO 19115 pour des métadonnées socio-économiques, INFORSID Marseille, May 25-28, 2010.
- Plumejeaud, C., Gensel, J., Villanova-Oliver, M., Ben Rebah, M., Vergnaud G., Modélisation de hiérarchies territoriales multiples évolutives, Colloque International de Géomatique et d'Analyse Spatiale (SAGEO), Paris, November 25-27, France, 2009
- Templ M., Filzmoser P., and Hron K., *Robust imputation of missing values in compositional data using the r-package robcompositions*, Proceedings of the NTTS Conference, 2009, talk: NTTS Conference, Brüssel ; 2009-02-19.
- Rousseeuw, P. and Leroy, A.: 1996, *Robust Regression and Outlier Detection*. John Wiley & Sons, 3rd edition.
- CrimeStat, <http://www.icpsr.umich.edu/icpsrweb/CRIMESTAT>, 2004.
- GeoDa, <http://geodacenter.asu.edu>, 1998
- GRASS GIS, <http://grass.itc.it/index.php>, 2010
- QuantumGis, <http://www.qgis.org>, 2002.
- SADA, <http://www.tiem.utk.edu/sada/index.shtml>, 1997.
- Projet R, <http://www.r-project.org>, 1997.
- Web Processing Service, <http://www.opengeospatial.org/standards/wps>, 1994

Integrating many legends through ontology for cartographic symbols

Karam Roula*, *****- Favetta Franck*** - **Laurini Robert*** -
Chamoun Rima**

**INSA de Lyon, LIRIS CNRS UMR5205, Université de Lyon
F-69621 Villeurbanne, France*

{roula.karam, franck.favetta, robert.laurini}@insa-lyon.fr

***Université Saint Joseph, Ecole Supérieure d'Ingénierie de Beyrouth
B.P :11-514, Beyrouth, Lebanon*

{rima.kilany@usj.edu.lb}

**** Ecole Navale, Institut de Recherche de l'Ecole Navale IRENav
29240 Brest cedex 9, France*

{roula.karam@ecole-navale.fr}

Résumé. *Plusieurs domaines ont recours aux ontologies comme solution pour présenter, partager et enrichir leur connaissance à citer le web sémantique, le commerce électronique, l'intelligence artificielle et les systèmes d'information géographiques. Cependant, les concepts et les relations sont identifiés textuellement alors que certains concepts ont un aspect visuel surtout dans le domaine cartographique. Par exemple, le symbole d'un point d'intérêt dans une légende, est identifié comme un concept avec une icône et /ou abréviation, une couleur, une texture, une police de caractère, une orientation ou un certain numéro. Notre approche concerne le développement et l'intégration de ces ontologies visuelles, représentant des légendes, vers une ontologie référence du domaine (fond de carte et symboles provenant de plusieurs fournisseurs). L'objectif de cet article est de 1) proposer un nouveau type d'ontologie dans laquelle les concepts sont visuels et de l'appliquer aux symboles cartographiques et de 2) développer une application prototype pour tester l'idée tout avec une extension de Web Ontology Language (CartOWL).*

Abstract. *The use of ontology is common for many domains such as semantic web, e-commerce, artificial intelligence and geographical information systems. However, the vocabulary used in ontologies is always textual; concepts and relations are identified and labelled by words. However, some concepts include a visual aspect especially in the cartographic domain. For example, in a cartographic legend, a point of interest symbol is identified as a concept with its icon and/or abbreviation, color, texture, font style,*

orientation or number. The overall concerns of our research are to integrate different legends as visual ontologies towards a unique reference one (base map and symbols from many providers). The objective of our article is to 1) propose a new type of ontology where the concepts are visual and to apply it to cartographic symbols then to 2) develop an application prototype in order to handle these kinds of ontologies with an extension of Web Ontology Language (CartOWL).

MOTS-CLES: SIG, LBS, Cartographie, Ontologie, OWL.

KEYWORDS: GIS, LBS, Cartography, Ontology, OWL.

1 Introduction

Ontology is a specification of a domain conceptualization. It deals with issues concerning what entities exist and how such entities can be grouped, related within a hierarchy and subdivided according to similarities and differences. From a computing point of view, ontologies are tools or solutions used to represent a corpus of information specific to a certain domain in order to be able to share, queries, extract and update information among many providers through Internet. Nowadays, many examples are dealing with ontology concepts such as semantic web, artificial intelligence, knowledge engineering, genetics and geographical information systems (GIS). [EUZ 07] [UIT 99]

In [KAR 10], we proposed and implemented a GIS framework for mobile in order to solve the issues related to LBS integration, at the data level (position addresses, place names and semantic details for homologous objects). We studied the case where a user wants to use its mobile in order to find the nearest restaurant in his area, and to know how to get there. The figure below shows that the same Italian restaurant is listed by two different providers and visualized by two different cartographic symbols, which are not exactly located at the same place due to GPS precision. In this article, we will study the same integration problem, but at the map conflation level. This approach could be applied, not only with legends for LBS domain, but for historical [VAR 06] or maritime domains as well.

Map conflation is the process of producing a new map by integrating two existing digital maps and their cartographic overlaid symbols for homogenous objects in case of conflicts. [BEE 04]

The current types of ontologies (i.e. taxonomic or descriptive ones) [CUL 03] are dealing with textual vocabulary to list concepts and relations. In order to be able to achieve the LBS integration of the visual concepts which are related to the cartographic domain, we propose in this paper an extension to the geographical

ontology concepts. Take as an example, the restaurant symbol in [Figure. 1] which is presented as a concept with its icon and/or abbreviation, colour, texture, font style and orientation in two different ways.

In this paper, section 1 presents a simple case study and the related integration issues. Section 2 details the solution we propose, by extending the web ontology language OWL [BEC 03]. Section 2 also gives an explicit description of both the building/matching application we developed, as well as the visual domain ontology to design in order to validate the proposition. Finally, conclusions and future work are elaborated in Section 3.

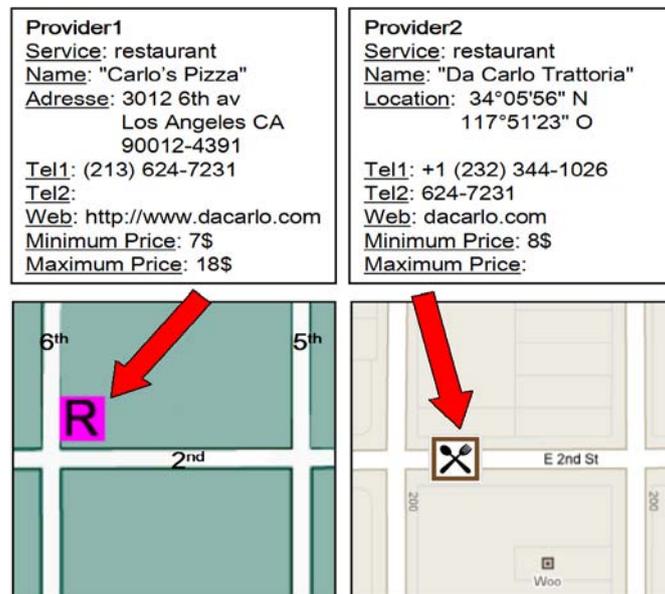


Figure. 1. Example of same LBS object from two different providers (Candidates for integration)

2 Integration issues for a simple case study

In this section, we will first present a simple example case, which we will use in order to explain the issues related to the integration of the cartographic symbols. For this purpose, we choose to collect three touristic legends from three different cartographic organizations: Ordnance Survey (UK), RandMcNally (USA) and IGN (Institut Geographique National, France). We should note that legends from different LBS providers will be used in the implementation of a full prototype.

Once the legends we need to integrate are collected, we begin by identifying the visual aspect of each concept via Microsoft Visio software (i.e. tourist information center is identified only by its icon without any texture or number, etc.). The next step would be to use an XML editor like Altova XML Spy [ALT], in order to create an XML Schema describing this information for each organization's legends.



Figure 2. Excerpts from three touristic legends: Ordnance Survey [ORD], RandMcNally [RAN] and IGN [IGN]

Having in hand all the elements needed in order to begin the integration process, many issues have to be dealt with:

1. The choice of the colors of the background map and the icons could have an important impact on the visibility of the map. Based on the graphical semiology of J. Bertin [BER 99], many schemas for intelligent color distribution were elaborated. Other algorithms based on the chromatic circle [CHE 05] [ITT 91], the colour intensities [BRE 03] [MER 90] and their meanings [BER 99] in different countries/cultures were also taken into consideration (i.e. Red implies danger in western countries or good luck in Asia, Blue implies nobility in France and immortality in China).

2. Location based fusion may seem to be an easy task, since locations could be constructed as global identifiers. This is not true, however, for several reasons. First, measurements may introduce errors. These errors in different databases are independent of each other. Second, each organization has its own requirements and approach, and hence uses different measurements techniques and may record spatial properties of entities using a different scale or a different structure. For example, one organization might represent buildings as points, while another could use polygonal shapes for the same purpose. While an estimated point location can be derived from a polygonal shape, it may not agree with a point-based location in another database (i.e. 0D (point), 1D (line), 2D (surface), 3D (Volume)).

3. Finally, the most critical argument to cover in this article is related directly to the presence of different spatial attributes such as icons and/or abbreviation, color, label, number, orientation and font style. Building a graph of visual concepts instead of textual ones could be a good suggestion for GIS or any other visual application where the extraction and fusion of objects is based on their visual aspects and not the semantic ones. For example, Metro is symbolized by letter M in a country or S (Subway) or T (Tube) in others. This means that the same service could have different shapes or spatial attributes from one provider or another. Object fusion and visual ontology matching should be applied from the table below [Figure.3].

Cartographic Symbols	Ordnance Survey	Rand McNally	IGN
Museum	 icon	 icon	 icon
Park	 icon	 icon  color	 icon
Tourist Info. Center	 icon	 icon	 icon
Picnic Site	 icon	 icon	 icon
Parking, Park ride	 icon	 icon	 icon
Selected places of Tourist Interest	 texture	 icon/color	 icon
Main Road	 Abb./Number texture/color	 icon	 Abb./Number color

Figure. 3. Visual aspects collection from the three legends

Several software tools for building ontologies like Protégé Plugin [NOY 01] and for merging them, two by two, like Mafra [MAE 02] or for producing alignments like S-Match [GIU 07], mainly need human interaction to insert the textual concepts. Janus [BED 08] is another software tool that can automatically build textual ontologies and align them towards one reference knowledge base (domain ontology) and thus by extracting the concepts, properties and relationships from XSD files (XML Schema de facto standard), Janus generates output in RDFS or OWL standard format.

Nowadays, all these tools can extract concepts from textual corpus only. What if we have visual aspects (i.e. icons) to describe geographic locations that need to be extracted and imported as XML files? The current XML Schema file or ontology standards (RDF/RDFS, OWL), do not support visual tags like icon.jpg. To overcome this limitation and to improve the extraction time, the complexity of

alignment and the difficulty in the validation process, we propose another building/matching application and an extension of the de facto standard OWL (CartOWL) for the treatment of visual aspects via ontology as detailed below in section 3.

3 Proposed Solution

Sharing geospatial data is difficult due to diverse conceptual schemata and semantics [KAV 00]. From the above case study and its limitations, we understand clearly the need of visual spatial ontologies for the integration of heterogeneous information sources (map conflation [UIT 99] [BEE 04]).

This section describes the prototype we developed in order to meet two requirements in the LBS integration process : 1) it provides a LBS provider with a way to describe its map legend by building an ontology with symbols, and 2) it offers a way to realize the matching of several ontologies from several providers. As a result, the prototype merges ontologies from several providers in order to obtain the domain ontology.

To achieve these goals, we aimed at developing a user-friendly interface with a simple handling [Figure. 4]. The interface is divided into three “columns”: 1) on the right, in the building column, an ontology may be built, 2) on the left, in the I/O column, an ontology may be loaded and saved with an ad hoc format (CartML) and exported to CartOWL format, and 3) in the central column, matching relations may be added to link classes from both ontologies. The central column displays also visual representations of the classes of both ontologies. Built ontologies may be shifted to the left column in order to be saved, and loaded ontologies on the left side may be shifted to the right in order to be modified.

3.1. Building an ontology with symbols

In the building column, the user creates a new ontology by clicking the “New Right Ontology” button. The user has then to enter an ontology name and to choose a language for it. He also has to enter an abbreviation for this ontology that will be later its identifier in the domain ontology. The user adds new subclasses by selecting a class and clicking the “New Subclass” button and may rename them. An existing class may also become the subclass of a second parent class if the user drags an existing class and drop it on another. The user may modify the visual representations of the currently selected class. The user may change its colour or its font by choosing one in a panel. He may enter an abbreviation. Check boxes indicate if the class is represented by a number.

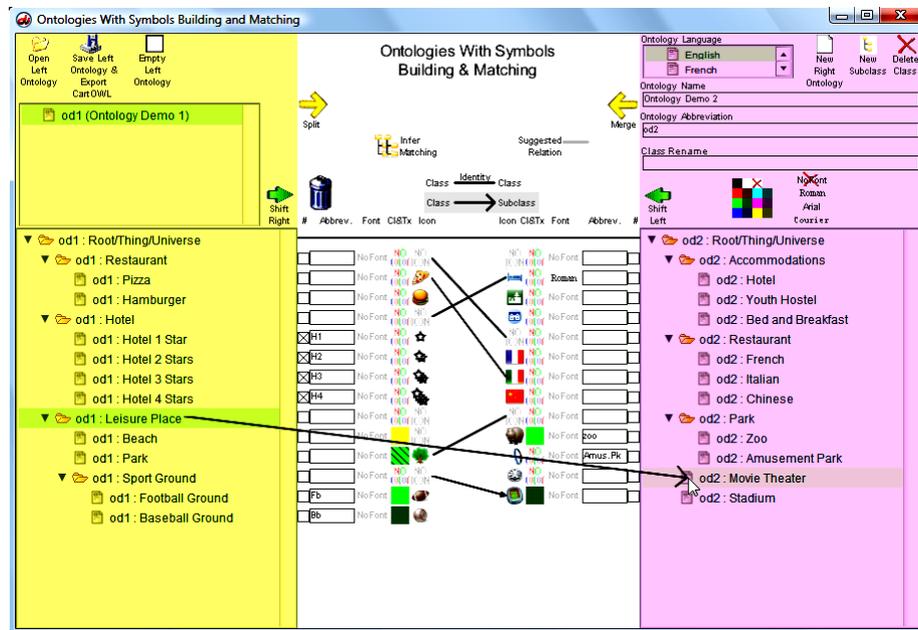


Figure 4. Building part of the Application

3.2. Matching ontologies with symbols

To link two classes from both ontologies, the user has to drag a class from one ontology and to drop it on another class in the opposite ontology. Two kinds of links are available: identity links represented by lines and class-subclass links represented by arrows. To switch from one kind to the other, the user clicks on the corresponding button at the top of the interface. Links may be deleted by drag'n dropping them to the bin.

When linking process is done, the user clicks on the “Merge” button in order to merge both ontologies into one resulting ontology, on the left side [Figure. 5]. The merged ontologies are listed above. The user may incrementally redo this process with another ontology and the resulting one. The user may extract one of the merged ontologies by clicking on the “Split” button and then modify it on the right side.

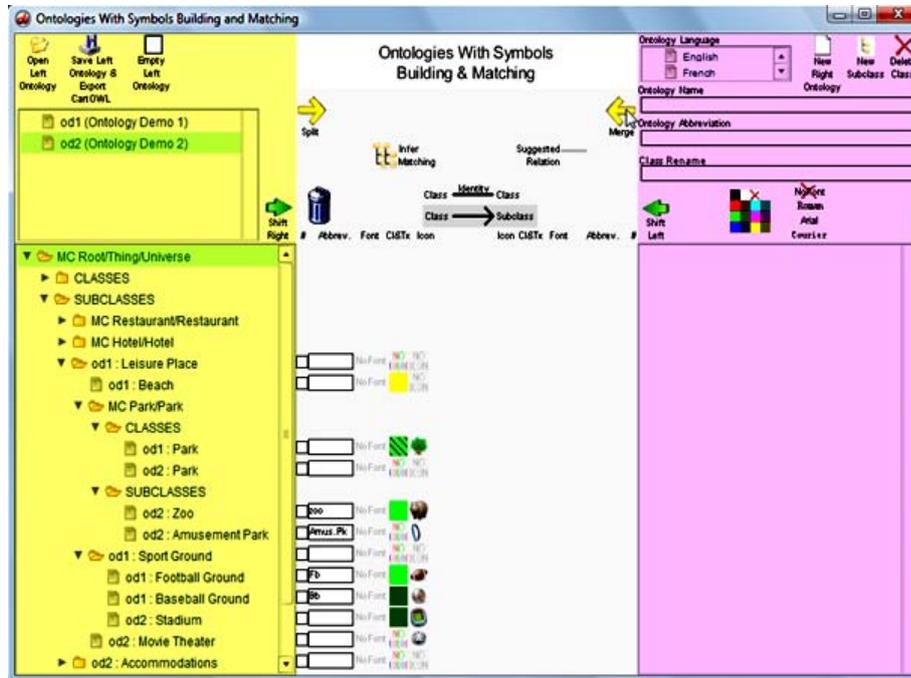


Figure 5. Matching part of the Application (focusing on the “Leisure Place” class)

3.3. Web ontology language dialect (CartOWL)

In order to facilitate the automatic build of local ontologies and their integration towards domain reference ontology, we propose CartOWL (Cartographic OWL) as an extension to the Web Ontology Language OWL [BEC 03]. Building visual ontologies would become much simpler by generating the corresponding CartOWL file through our application. The full prototype would be able to parse the CartOWL output files and align them towards one reference knowledge base (domain ontology) so that we can ensure map conflation result on a mobile device for example.

The cartographic symbols of the spatial ontologies are detailed below:

- An icon, described by its URL, e.g., <http://www.example.net/TouristInfoCenter.png>
- A color, described by its RGB model value, e.g., #FF8000
- A texture, described by its URL, e.g., <http://www.example.org/ParkTexture.png>
- An abbreviation, e.g., “info”

- A number, e.g., “A 30”
- A font, described by its name, e.g., “Times”

We propose an approach to deal with symbols in OWL as below:

OWL can handle only textual concepts, it needs to be revised in order to describe visual aspects as well. In the <Class> tag, the human readable textual representation of the concerned class is done with the <Label> tag. One appropriate way seems to extend this <Label> tag with new attributes or to add a new <Symbol> tag that includes two parameters:

- The symbol type: the parameter value may be “iconURL” or “color” or “texture” or “abbreviation” or “number” or “font”.
- The symbol value: the parameter value may be one of the following string values: the icon URL, the RGB color, the texture URL, the abbreviation’s text, the presence of a number indicated by the boolean values “yes” or “no”, or the font name.

Below is an example. We define the class “Tourist Information Center” and its three attached symbols: an icon, a color and an abbreviation. The first part of the example shows the DTD definition of the new <symbol> tag.

```
<!DOCTYPE rdf:RDF [
<!ELEMENT cartowl:Symbol EMPTY >
<!ATTLIST
      cartowl:Symbol
      cartowl:symbolType
      (iconURL|color|texture|abbreviation|number|font)#REQUIRED
      cartowl:symbolValue CDATA #REQUIRED
      xmlns:cartowl CDATA #FIXED
      "http://www.example.net/CartOWL.owl# ... >
]>
<rdf:RDF ... ..
xmlns:cartowl="http://www.example.net/CartOWL.owl#" >
...
<owl:Class rdf:ID="Tourist Information Center">
<cartowl:Symbol
      cartowl:symbolType="iconURL"
      cartowl:symbolValue="http://www.example.net/TouristInfoCenter.png" />
<cartowl:Symbol cartowl:symbolType="color" cartowl:symbolValue="FF8000" />
<cartowl:Symbol
      cartowl:symbolType="abbreviation"
      cartowl:symbolValue="info" />
</owl:Class>
```

To ensure ontology matching, relations such as equivalence and inclusion have to be previously set at the level of the domain ontology. The CartOWL ontology

reasoner would then take the statements encoded (asserted) in this reference ontology as input and derive (infer) new statements from them.

We should note that some subjectivity could interfere in case of conflicts due to linguistic and cultural differences while drawing the relations via the matching application. This step would allow us to keep and/or to get only the visual concepts aligned with the reference ontology, which we will use when showing to the mobile user the results of his search.

Different other elements should be taken into consideration in order to achieve a coherent and complete visual integration:

- The profile of the user (culture, age, map preferences, etc.)
- The geographical zone and the graphical semiology rules

4 Conclusions and future work

Due to the fact that we aim to integrate cartographic symbols' ontologies from many LBS providers as a use case study, our approach in this paper was to 1) develop an application that is able to handle the visual aspects of different cartographic ontologies and to 2) propose an extension of OWL standard (CartOWL). In the next OWL revision, these visual aspects of concepts could be taken into consideration by the W3C community. However, a variety of issues remain to be solved such as 1) to test the effect of visual ontology components, besides semantics and hierarchies, on the potential for information integration when combining two ontologies, 2) the widespread analysis of other legends from different providers such as: Via Michelin, DeAgostini, Google Maps, Bing, Mappy, etc., in order to visualize any kind of spatial information, 3) Psycho-cognitive tests should be conducted to determine the user's ability to understand without legends, any cartographic symbol so we can adjust a degree of preference among icons that are representing the same service and finally 4) For scalability issue, some improvements should be done to support semi or full automatic building/ matching application.

Until now, the current ontologies are matched semantically even for GIS, but we think that visual aspects could be more adequate than texts in many domains such as roads' coding/ urban applications[LAU 07], historical map catalogues, or LBS integration. Besides, we could propose "audio ontology" for disabled people.

5 References

[ALT] Altova XML Spy Editor: <http://www.altova.com/xml-editor/>

[BEE 04] Beerli C., Kanza Y., Safra E., Sagiv Y.: "Object Fusion in Geographic Information Systems", *Proceedings of the 30th International Conference on Very Large Databases (VLDB), Toronto (Canada), (2004).*

[BER 99] Bertin J. : "*Sémiologie graphique: Les diagrammes -Les réseaux -Les cartes*", Book from : edition de l'école des hautes études des sciences sociales, Paris, (1999).

[BEC 03] Bechhefer S., Horrocks I., Patel-Schneider P.: "*Tutorial on OWL*", <http://www.cs.man.ac.uk/~horrocks/ISWC2003/Tutorial/>, (2003).

[BED 08] Bedini I., Nguyen B., Gardarin G.: "Janus: Automatic Ontology Builder from XSD files", *17th International World Wide Web Conference (WWW2008), Beijing, China*, (2008).

[BRE 03] Brewer C., Hatchard G., Harrower M.: "ColorBrewer in Print: A Catalog of Color Schemes for Maps", *Journal Cartography and Geographic Information Science CaGIS, Vol.30, n.1, p. 5-32*, (2003).

[CUL 03] Cullot N., Parent C., Spaccapietra S., Vangenot C.: "Des SIG aux ontologies géographiques", *International journal of 'géomatique', Hermès Science Edition, Vol. 13, n. 3, p. 285-306*, (2003).

[CHE 05] Chesneau E., Ruas A., Bonin O.: "Colour Contrasts Analysis For A Better Legibility Of Graphic Signs For Risk Maps", *International Cartographic Conference: Mapping Approaches Into A Changing World*, (2005).

[EUZ 07] Euzenat J., Shvaiko P.: "*Ontology Matching*", Book from Springer edition, (2007).

[FON 02] Fonseca F., Egenhofer M., Agouris P.: "Using Ontologies for Integrated Geographic Information Systems", *Journal Transactions in GIS Vol.6, n.3, p. 231-257*, (2002).

[GIU 07] Giunchiglia F., Yatskevich M., Shvaiko P.: "Semantic matching: Algorithms and implementation", *Journal on Data Semantics, Vol. 1*, (2007).

[HOR 04] Horridge M., Knublauch H., Rector A., Stevens R., Wroe C.: "*A Practical Guide to Building OWL Ontologies Using the Protégé-OWL Plugin and CO-ODE Tools*", Tutorial from The University of Manchester Edition 1.0. Matthew Horridge1, (2004).

[IGN] IGN Legend:
http://professionnels.ign.fr/DISPLAY/000/526/620/5266206/SCAN25_specification.pdf;
<http://www.ign.fr>

[ITT 91] Itten J.: "*Art de la couleur*", Book from dessain et tolra edition, (1991).

[JOH 90] Johann G.: "*Traité des couleurs*", Book Edition Triades, (1990).

[KOK 08] Kokla M., Kavouras M.: "Advances in Geospatial Interoperability", *First Mediterranean CEN/TC 287 Workshop, Athens, Greece*, (2008),

"Semantic Integration of Geographic Ontologies: the Onto Geo Approach", *Second CEN/TC 287 Mediterranean Workshop, UN FAO, Rome, Italy*, (2008) and

"Geo Ontology Integration: Identifying Issues, Dimensions and Developing Guidelines", <http://www.ikg.uni-hannover.de/>, (2006).

[KAV 00] Kavouras M., Kokla M.: "Ontology-Based Fusion of Geographic Databases", *Spatial Information Management, Experiences and Visions for the 21st Century, International Federation of Surveyors, Commission 3-WG 3.1, Athens, Greece*, (2000).

[KAR 10] Karam R., Favetta F., Laurini R., Kilany Chamoun R.: "Integration of Similar Location Based Services Proposed by Several Providers", *NDT2010, 2nd International Conference on Network and Digital Technologies, Prague*, (2010).

[LAU 09] Laurini R.: "GeoWeb: Internet Géographique, Chapitre 5: Ontologies pour les applications géographiques", <http://liris.cnrs.fr/robert.laurini/coursMRI.html>, (2009).

[MER 90] Mersey J.: "Colour and Thematic map Design: The Role of Colour Scheme and Map Complexity in Choropleth Map Communication", *Journal Cartographica, vol.27, n.3, University of Toronto Press, Cartographica Monograph, p.41-182*, (1990).

[MAE 02] Maedche A., Motik B., Silva N., Volz R.: MAFRA : "Mapping Distributed Ontologies in the Semantic Web", *Proceedings 13th European Conference on Knowledge Engineering and Management (EKAW 2002), Springer-Verlag, p. 235-250*, (2002).

[NOY 01] Noy N., Ferguson R., Musen M.: "The knowledge model of Protege-2000: Combining interoperability and flexibility", *Proceedings of the 12th European Workshop on Knowledge Acquisition, Modeling and Management, Springer publisher, p. 17- 32*, (2001).

[ORD] Ordnance Survey Legend:

<http://magazine.ordnancesurvey.co.uk/magazine/tscontent/editorial/mapfacts/2009/mapsymbolguides.html>

[RAN] Rand McNally Legend: <http://www.randmcnally.com/pdf/smpCityExpress.pdf>

[UIT 99] Uitermark H., Oosterom P., Mars N., Molenaar M.: "Ontology-Based Geographic Data Set Integration", *Proceedings of the International Workshop on Spatio-Temporal Database Management, LNCS Springer edition, Vol. 6176, p.60-78*, (1999).

[WAC 01] Wache H., Vogeles T., Visser U., Neumann H. et al.: "Ontology-Based Integration of Information-A Survey of Existing Approaches", *IJCAI-01. Workshop: Ontologies and Information Sharing, Seattle, USA, p. 108-117*, (2001).

[WEI 04] WeiBenberg N., Voisard A., Gartmann R.: "Using Ontologies in Personalized Mobile Applications", *Proceedings of the 12th annual ACM international workshop on Geographic Information Systems, GIS04, New York*, (2004).

[VAR 06] Varanka D.: "The 20th-Century Topographic Survey as Source Data for Long-Term Landscape Studies at Local and Regional Scales", *Open file report for US Geological Survey*, (2006).

[LAU 07] Laurini R.: "Pre-consensus Ontologies and Urban Databases", *In "Ontologies for Urban Development"*, Edited by Jacques Teller, John R. Lee and Catherine Roussey, Springer Verlag, *Studies in Computational Intelligence*, 61, pp. 27-36 ,(2007).

Evaluer, améliorer la visualisation des données spatiales, le cas des données urbaines

Construction d'un test psycho-cognitif sur la représentation visuelle des indicateurs de flux

Rosa Marina Donolo^{1,2}, Franck Favetta¹, Robert Laurini¹

¹Laboratoire d'InfoRmatique en Images et Systèmes d'information - LIRIS, Institut National des Sciences Appliquées -INSA- Lyon, 20, Avenue Albert Einstein, 69621, Villeurbanne, Cedex, France

²Dipartimento di Ingegneria Civile, Università degli Studi di Roma Tor Vergata, 1, via del Politecnico, 00133, Roma
{Rosa-Marina.Donolo, Franck.Favetta, Robert.Laurini}@insa-lyon.fr

RÉSUMÉ. La visualisation de l'information s'occupe des méthodes de présentation graphique des données et de la représentation visuelle des données spatiales. La conception de la visualisation de l'information est particulièrement intéressante dans les applications des Systèmes d'Information Urbains, étant donné que la grande quantité de données en milieu urbain peut être synthétisée et visualisée grâce à l'utilisation d'indicateurs spatiaux. La visualisation de l'information des zones urbaines est encore trop adressée à un public d'experts. Par la proposition d'un test sur la représentation visuelle des indicateurs spatiaux, cet article explore de nouvelles formes de représentation de données et les méthodes grâce auxquelles la visualisation de l'information urbaine peut devenir plus simple et permettre à tous, experts ou non, d'interpréter des cartes thématiques.

ABSTRACT. The domain of information visualization concerns the methods of graphic representation of spatial data. Information visualization design is very interesting in applications of Urban Information Systems, as the great amount of urban data can be synthesized and visualized thanks to the use of spatial indicators. Information visualization in urban areas is still too much designed for expert users. In this article, we propose to realize a test in order to experiment visual representation of spatial indicators. By this way we explore new ways of data representation and methods that can make urban data visualization and interpretation easier for everybody (experts or not).

MOTS-CLES : Visualisation d'information, Systèmes d'Information Urbains, tests psycho-cognitifs, indicateurs spatiaux, représentation visuelle

KEYWORDS: Information Visualization, Urban Information Systems, Psycho-Cognitive Tests, Spatial Indicators, Visual Representation

1. Introduction

Cet article présente une démarche pour évaluer l'efficacité de la représentation d'informations spatiales. Il propose un test qui interroge des utilisateurs potentiels sur la lecture de cartes thématiques avec un niveau grandissant de difficulté d'interprétation. Le test proposé se réfère aux travaux de Jacques Bertin concernant le modèle des *variables visuelles* qu'il définit. L'article est divisé en deux parties principales : la première traite de la visualisation de l'information urbaine et sa conception ; la seconde présente la construction du test que nous proposons. L'idée de faire ce test est née dans un contexte de recherche mixte, avec la collaboration de chercheurs de différentes compétences, mais avec la nécessité commune de comprendre les représentations visuelles. Ce test est né avec la perspective d'être composé en différentes parties et exemples pour évaluer les représentations visuelles des données spatiales en 2D et 3D. La première partie du test proposé concerne la représentation en 2D du flux de passagers dans un métro. Les limites du test sont : la nécessité de sélectionner un grand nombre d'exemples de représentations visuelles pour valider les résultats et trouver les combinaisons qui génèrent les meilleurs pourcentages de bonnes réponses ou fausses réponses.

2. La visualisation de l'information

Cette partie décrit les méthodes de visualisation des données spatiales en général, puis sont ensuite spécifiées les caractéristiques de la visualisation des données en milieu urbain.

2.1. Définition et description

McCormick [1987] définit ainsi la visualisation de l'information : *“La visualisation est une méthode de calcul. Elle transforme le symbolique en géométrique, permettant aux chercheurs d'observer l'invisible ou le caché. La visualisation enrichit le processus de la découverte scientifique et dévoile des vues profondes et inattendues”*.

Un nombre grandissant d'études concernent l'utilisation des représentations de données dans le but d'améliorer l'interaction utilisateur-carte-territoire. Pour étudier les divers aspects de la visualisation de l'information, une analyse multidisciplinaire et multidimensionnelle est nécessaire. Actuellement, les progrès de la visualisation de l'information sont principalement dus aux études sur la sémiologie graphique [Bertin77], aux études sur le système perceptivo-cognitif humain [Card83, Marr87], et surtout à l'évolution de l'informatique graphique qui permet d'obtenir des représentations de données très sophistiquées et expressives. Le développement des Systèmes d'Information Géographiques sur Internet a également favorisé la recherche dans ce domaine pour permettre un accès facile aux données mémorisées. Dans cet article, nous souhaitons montrer en particulier comment les études sur la perception visible humaine et ses méthodes de représentation de l'information peuvent amener à des visualisations plus efficaces des données territoriales.

Les avantages principaux de la visualisation de données sont les suivants :

1. Réduire le temps de compréhension de l'information chez les personnes expertes ou non ; Réduire le temps de gestion et d'élaboration de l'information,
2. Faciliter le repérage de caractéristiques cachées et de relations entre données (repérage de points critiques et/ou favorables et/ou anomalies),
3. Synthétiser l'information extraite d'une grande quantité de données,
4. Favoriser l'interaction utilisateur-système d'information.

L'avantage principal de la visualisation des données est de communiquer rapidement à l'observateur d'éventuelles situations d'anomalies ou de risque pour faciliter les processus décisionnels. Grâce à la visualisation des données, il est possible d'obtenir une première évaluation d'un phénomène, et exploiter le fait que les caractéristiques du phénomène sont immédiatement observables en relation avec l'espace. [Friedman08] répète cependant l'importance de ne pas compliquer ni « d'embellir » trop la représentation visuelle des données spatiales, pour ne pas perdre l'intuitivité et l'immédiateté dans la communication de l'information. En outre, les instruments de visualisation fournissent à l'utilisateur un cadre général de données ainsi que la possibilité de les interroger à travers leurs attributs principaux et de voir représentés les résultats de la requête spatiale. Une représentation des variables spatiales bien réalisée, et ainsi d'une bonne visualisation des données, peut réduire le temps de gestion et de compréhension de l'information, car elle réduit le temps de toutes ces opérations qui se font communément par l'analyse spatiale : sélection, intégration, homogénéisation, archivage, interrogation, élaboration des données.

Les instruments de visualisation sont nécessaires pour identifier les relations et les points critiques qui ne sont pas évidents parmi les données numériques. La présentation des informations de façon visuelle est parfois indispensable pour mettre en évidence certaines caractéristiques des données territoriales. Cet avantage dérive du fait que l'information, étant visualisé dans l'espace, permet à l'utilisateur de déplacer l'élaboration de l'information du système cognitif au système perceptif. Ce passage obligatoire dans l'analyse des données visuelles, contraint l'utilisateur à conceptualiser l'information dans son esprit et à la connecter visiblement avec d'autres phénomènes territoriaux qui interviennent au même endroit et/ou dans le même temps. La visualisation de l'information est définie comme un processus de transformation de l'information vers une forme visuelle qui permet à l'utilisateur d'observer l'information [Card97]. La difficulté de visualiser et mettre en relation les informations sans support graphique spatial, est surmontée grâce aux représentations spatiales. L'utilisation de la couleur permet de mettre en évidence des agrégations entre données, l'utilisation de différents niveaux de classification et de différentes échelles de représentation permet de rendre différents niveaux de précision de l'information, alors que l'utilisation de l'animation permet d'alterner les représentations de paramètres variables dans l'espace et le temps.

L'utilisateur peut visualiser et gérer de grandes quantités d'information de façon

rapide, efficace et stimulante en utilisant certains paramètres comme le voisinage spatial, la couleur, la forme, la dimension, la classification et le choix de l'échelle de restitution. Parfois les techniques de visualisation peuvent être utiles lors de la sélection de grandes quantités de données. Dans ce sens, la visualisation de l'information peut être considérée comme une technique de *fouille de données visuelle* puisqu'ainsi la visualisation devient un instrument de synthèse des données.

La visualisation est une des techniques qui permettent une meilleure interaction entre utilisateurs et système. Les données dans un système visuel peuvent être manipulées à travers des opérations comme l'animation, l'interrogation ou le zoom. Elles ne produisent pas seulement un résultat statique, mais dynamique et interactif.

2.2 Conception de la visualisation de l'information urbaine

Un nombre croissant de personnes nécessitent de comprendre des données territoriales sans être contraintes de maîtriser des formules mathématiques et statistiques compliquées. Nous pensons que la visualisation est la réponse juste. Comme dit précédemment la visualisation permet de gérer une grande quantité de données de façon relativement simple et immédiate. Parmi les problématiques urbaines où sont nécessaires de nombreuses données de différents types, il existe une forte demande de pouvoir gérer et comprendre ces données, de façon à pouvoir en extraire une information. Quand la conception de la visualisation de l'information est appliquée aux données urbaines, on parle de conception d'information urbaine qui est un instrument toujours plus important dans l'analyse du comportement d'une ville. Comme on peut le voir dans la figure 1, la conception d'information urbaine (urban information design) fait partie du *cycle vertueux de l'information*.

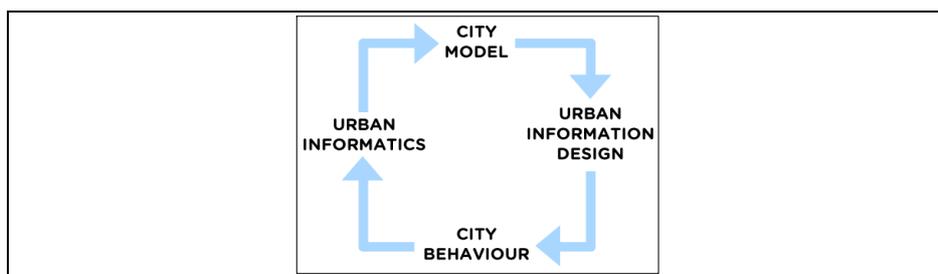


Figure 1. Cycle vertueux de l'information, (Dan Hill, www.cityofsound.com, 2008)

La conception de l'information urbaine a désormais un rôle fondamental dans le cycle vertueux de l'information urbaine.

Un des instruments les plus utilisés pour représenter les informations urbaines est l'indicateur spatial. La nécessité de représenter les indicateurs spatiaux a favorisé le

développement de diverses méthodes de représentation spatiale des données de cartes thématiques. Le choix des méthodes de représentation des données dépend de divers facteurs qui sont en même temps les éléments d'une construction correcte et de lecture de la carte :

- Le choix du modèle des données spatiales,
- La topologie,
- La classification,
- L'utilisation des couleurs,
- Le choix de l'échelle,
- Le type et le nombre d'objets géographiques qui s'utilisent pour représenter une variable spatiale (point ligne, surface),
- La typologie des attributs utilisés,
- Les modalités de traitement des attributs,
- La représentation des informations spatialement coexistants (pour une approche multi-dimensionnelle),
- La représentation d'un phénomène environnemental grâce à un indicateur intégré unique,
- L'utilisation de méthodes représentation de la variation des indicateurs dans le temps (en plus de la variation dans l'espace),
- La possibilité de représenter les données en s'inspirant des structures biologiques,
- La prise en compte des critères définis par la norme relative au processus de standardisation cartographique.

Les méthodes de représentation peuvent être appliquées soit aux données directement mesurées sur le territoire, soit aux données élaborées et intégrées sous forme d'indicateurs territoriaux qui fournissent une synthèse d'informations plus nombreuses. Diverses techniques de représentation ont été proposées ces dernières années. Mais il est nécessaire de mettre en évidence que de telles méthodes ne simplifient pas toujours l'analyse visuelle car il est parfois nécessaire de représenter plusieurs phénomènes en même temps. La représentation simultanée de plusieurs indicateurs territoriaux (ou la représentation d'un seul indicateur à deux instants différents) devient ainsi plus complexe. Ce constat ne facilite pas l'interprétation des indicateurs du côté de l'observateur. Cependant il met en évidence les interactions entre les problématiques environnementales, qui n'émergeraient pas sinon. Les modes de représentation des indicateurs environnementaux sont nombreux, aussi bien pour les indicateurs de phénomènes territoriaux à deux dimensions (surfaciens) qu'à trois dimensions (spatiaux). Si le phénomène analysé est préférentiellement mis en évidence dans une certaine dimension, il est plus aisé de le représenter avec des variables spatiales ayant la même dimension. Si, au contraire, le phénomène est plus complexe et a besoin d'être décrit par un indice qui synthétise plusieurs variables spatiales, la dimension du phénomène étudié est alors souvent perdue (en admettant qu'il soit dimensionnel).

Les figures 2, 3 et 4 présentent des exemples de représentations d'indicateurs

spatiaux en 2D et 3D.

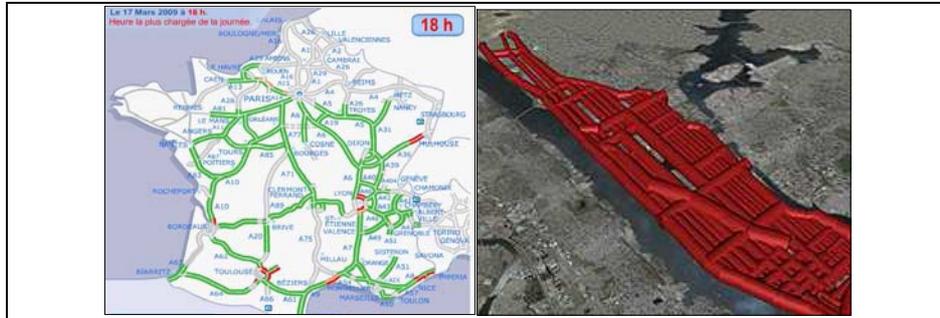


Figure 2. A gauche, indicateur de flux de véhicules en deux dimensions en zone extra-urbaine (<http://www.coralys.com>). A droite, indicateur de flux de véhicules en trois dimensions en zone urbaine (www.nadiaamoroso.com)

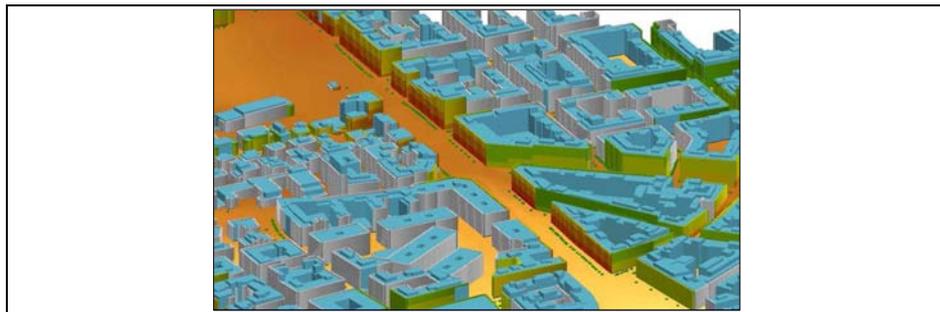


Figure 3. Indicateur de température (ou de bruit) sur façades d'immeubles en trois dimensions (M. Ioannilli, 2005)

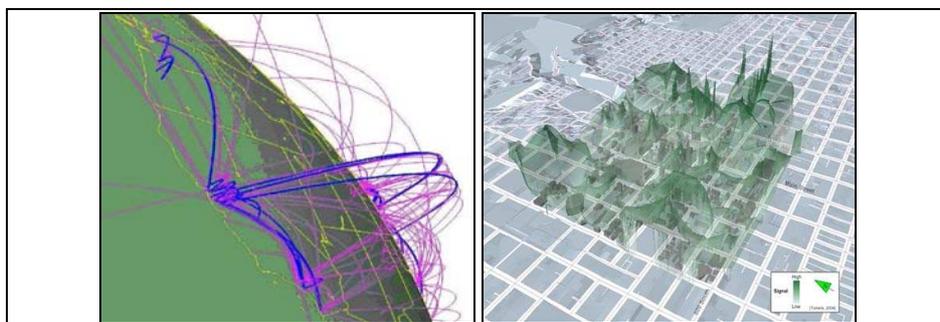


Figure 4. *A gauche, intensité d'un signal WIFI sous forme d'indicateur continu en trois dimensions. A droite, flux d'informations sur Internet ou flux de trafic routier sous forme d'indicateur continu en trois dimensions*

Pour commencer à comprendre les caractéristiques de la visualisation des données en milieu urbain, nous avons proposé des représentations différentes avec l'utilisation de différentes variables graphiques et géométriques. La limite principale dans la figure 2 à gauche qui représente le flux des véhicules est le fait d'être bidimensionnel, et de donner moins d'informations comparativement à la représentation du même indicateur (flux des véhicules) en 3D avec une seule couleur (rouge), même si cela permet une interprétation moins immédiate. La figure 3 propose une représentation en 3D à plusieurs couleurs qui permet une meilleure compréhension et interprétation de la carte du bruit d'un quartier. De plus, dans cette figure c'est aussi l'utilisation des surfaces réelles des bâtiments qui donne un air familier à l'image.

Dans la figure 4 à gauche, même si on retrouve les mêmes paramètres géométriques et graphiques (trois dimensions et utilisation des couleurs), on a moins facilité d'interprétation car les supports physiques de la représentation en figure 4 à gauche sont des arcs virtuels et non plus des bâtiments réels. La figure 4 à droite propose également une représentation tridimensionnelle très efficace des données spatiales, mais sans l'utilisation des différentes couleurs (uniquement vert) et sans le support physique d'un objet spatial réel comme les bâtiments de l'exemple de la figure 3.

3. Tests psycho-cognitifs

A partir des différentes méthodes de représentation présentées, se posent divers problèmes comme l'intelligibilité, l'interprétation et la correction de ces représentations visuelles, en bref leur performance. Une des solutions est de procéder à des tests cognitifs destinés à nombreuses personnes afin de mesurer leur efficacité.

Dans cette partie, nous détaillons la notion de tests psycho-cognitifs. Nous analysons la problématique que nous cherchons à résoudre puis nous proposons la construction d'un test afin de résoudre cette problématique en précisant la méthodologie que nous avons utilisée.

3.1. Intérêt et objectif

Les tests psycho-cognitifs semblent être un des rares instruments qui permettent de construire une « théorie empirique » sur la perception visuelle, et sur l'élaboration cognitive des données spatiales. Il est particulièrement important d'élaborer des méthodes empiriques pour obtenir une évaluation de la visualisation de l'information, [Chen00]. La méthode des tests est souvent utilisée lors d'études cognitives telles que [Mailles96, Bonhomme00, Favetta03]. La visualisation des

informations comporte un processus d'abstraction au niveau du passage du système perceptif au système cognitif. Un tel processus mental est subjectif mais présente des éléments communs parmi les différents sujets testés. Pour isoler ces éléments récurrents qui déterminent les préférences des sujets testés concernant la visualisation des données spatiales, nous avons choisi d'élaborer un test psychocognitif.

L'utilisateur *interprète* l'image de l'interface *matérialisée* par le système [Bottoni et al 95]. Pour être efficace, la représentation doit être intuitive, c'est-à-dire la plus proche possible de l'image mentale de l'utilisateur. On comprend aisément que la connaissance et la culture de l'utilisateur auront une influence fondamentale sur les réponses obtenues. Etant donné qu'il s'agit d'une transformation cognitive de l'information, il semble approprié d'utiliser une méthode empirique via un procédé de tests pour en évaluer l'efficacité.

Dans cet article, nous montrons comment nous analysons l'approche par les utilisateurs des différentes typologies de représentations de l'information territoriale afin d'en déduire les plus efficaces.

3.2. *Problématique*

Dans une première étape, notre démarche consiste à tester l'intuitivité des variables visuelles de Bertin utilisées par toute représentation, puis dans une seconde étape, nous testerons différentes représentations possibles (représentation sur façades d'immeubles et sous forme d'arcs, fig. 3 et 4). Cet article a pour objectif de décrire en détail la première étape.

Les variables visuelles sont définies dans la sémiologie graphique de Jacques Bertin [Bertin77], qui peut être considérée comme une des théories formelles les plus confirmées concernant la représentation des objets spatiaux. Dans la recherche de Bertin, on peut classer deux groupes principaux de variables : les variables « non géométriques » comme la couleur, la valeur, le grain (fig. 5) et les variables « géométriques » telles que la forme, la dimension, l'orientation, auxquelles on peut ajouter des variables de position des objets graphiques : position dans le plan (x, y) et dans l'espace (x, y, z).

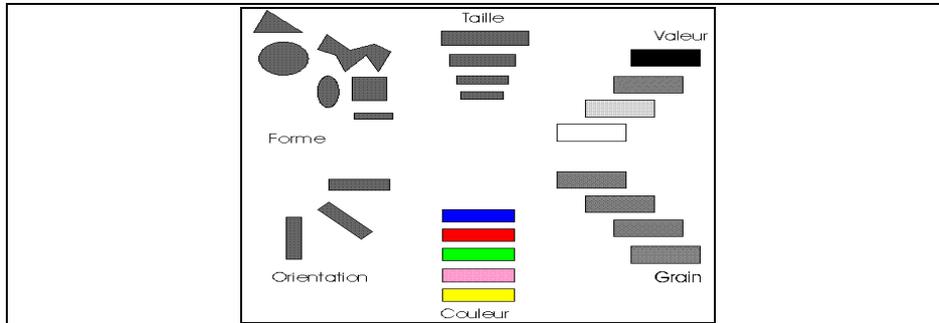


Figure 5 : Les variables visuelles de Jacques Bertin [Bertin77]

Les variables visuelles utilisées dans les représentations sont non exclusives et peuvent donc être combinées pour exprimer en un même point plusieurs informations indépendantes.

Des extensions du modèle de Bertin ont été proposées. Notamment, l'altitude (z) est une de ces extensions. Parfois, la position dans le plan et l'altitude ne sont pas considérés comme des variables et parfois comme deux variables distinctes. Le temps et le son peuvent également être considérés comme des variables supplémentaires et peuvent constituer l'animation de la représentation [Peterson95].

La première étape de notre test évalue l'efficacité des variables de Bertin utilisées dans les cartes représentant des indicateurs spatiaux continus tels que la température, le bruit ou encore un flux de véhicules ou de passagers. Nous devons d'abord définir ce que nous entendons par représentation efficace : il s'agit d'une représentation qui permet à l'utilisateur d'interpréter clairement l'ensemble des informations spatiales présentes sur la carte. Chaque variable visuelle peut se voir attribuer une information différente. Le nombre de variables étant important, un nombre élevé de combinaisons de variables visuelles est possible. On imagine intuitivement qu'au bout d'un certain nombre de variables combinées, la représentation deviendra confuse et donc source d'erreurs d'interprétation. Notre problématique est donc la suivante :

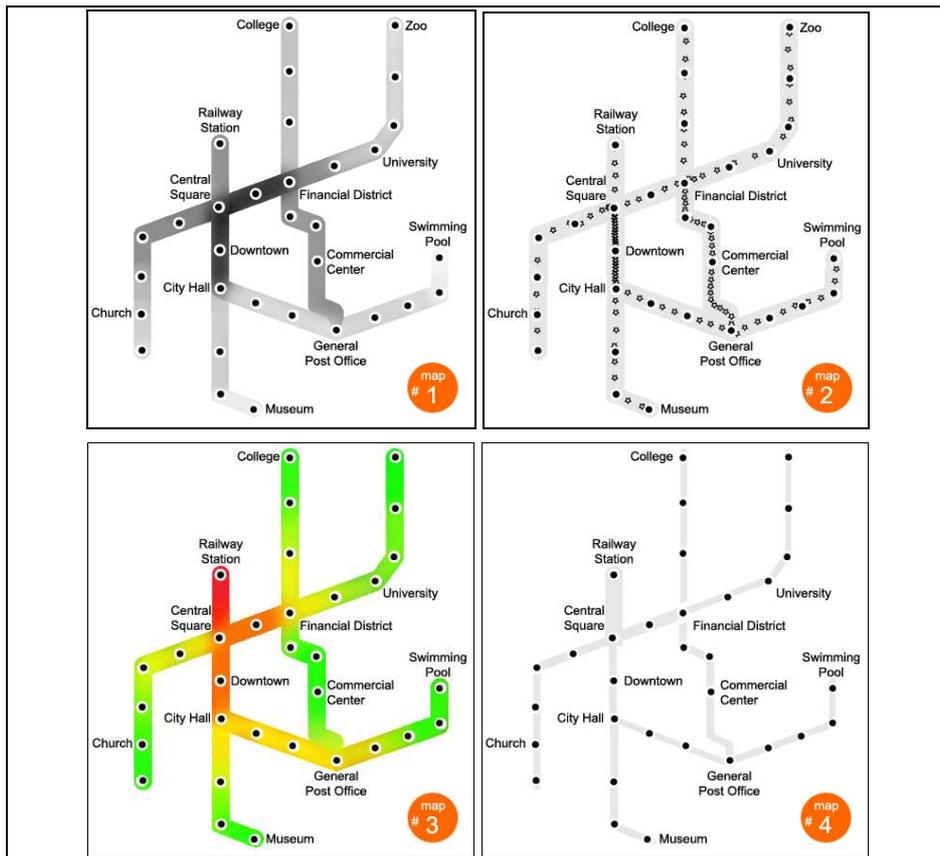
- Jusqu'à combien de variables visuelles combinées une carte reste clairement lisible ?
- Quelles combinaisons de variables visuelles sont les plus claires pour un nombre donné de variables ?

3.3. Proposition d'un test

Afin de répondre à la problématique précédemment énoncée, nous proposons une méthode « *exhaustive* » pour notre test : nous présentons à un panel d'utilisateurs potentiels toutes les combinaisons possibles des variables visuelles et nous demandons d'extraire des informations représentées par les différentes

variables en jeu. Les réponses seront confrontées aux réponses attendues et le nombre d'erreurs sera mesuré pour chaque combinaison.

Comme nous l'avons dit précédemment, le nombre de variables visuelles étant important, il conduit à un nombre de combinaisons très élevé, et ainsi à un nombre trop grand de questions possibles pour les candidats au test. En outre, notre étude portant sur le cas des indicateurs spatiaux continus (et non sur des informations ponctuelles que l'on représenterait par un ensemble d'icônes par exemple), les variables de position (x, y), d'orientation et de forme n'ont pas de pertinence ici. Pour ces raisons, nous proposons d'évaluer dans notre test la combinaison de quatre variables visuelles de Bertin : la valeur (ou luminosité ou encore intensité), le grain (ou texture), la couleur, et la taille. Nous avons choisi comme exemple de représenter le cas de l'indicateur du flux de passagers dans un métro fictif à quatre moments différents : jours ouvrables aux heures de pointe, fin de semaine, pendant les vacances et dans le moment présent. Ces quatre moments sont représentés par quatre variables visuelles, respectivement (fig. 6) : intensité, grain (ou texture), couleur, taille (épaisseur).



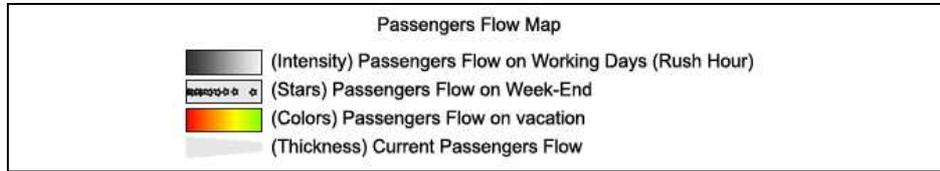
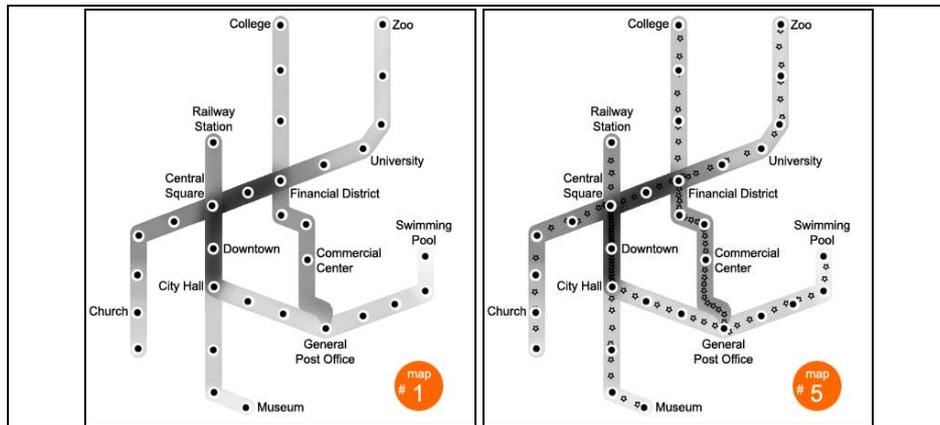


Figure 6 : Exemples de cartes de métro représentant l'indicateur « flux de passagers » à quatre moments différents au moyen de quatre variables visuelles de Bertin comme indiqué dans la légende.

Dans notre test, nous présentons au candidat toutes les combinaisons possibles de ces quatre variables visuelles :

- 4 cartes représentant chaque variable visuelle isolée (fig. 6)
- 6 cartes représentant les combinaisons de deux variables visuelles sur quatre
- 4 cartes représentant les combinaisons de trois variables visuelles sur quatre
- 1 carte représentant la combinaison des quatre variables visuelles ensemble

Au total, quinze cartes représentent toutes les combinaisons possibles des quatre variables visuelles. La figure 7 illustre quelques exemples de combinaisons des variables visuelles de Bertin : par exemple, la carte n°15 de la figure 7 combine quatre variables visuelles : intensité, couleur, taille, grain, et chaque variable représente une information sur le flux de passagers à différents moments (voir légende).



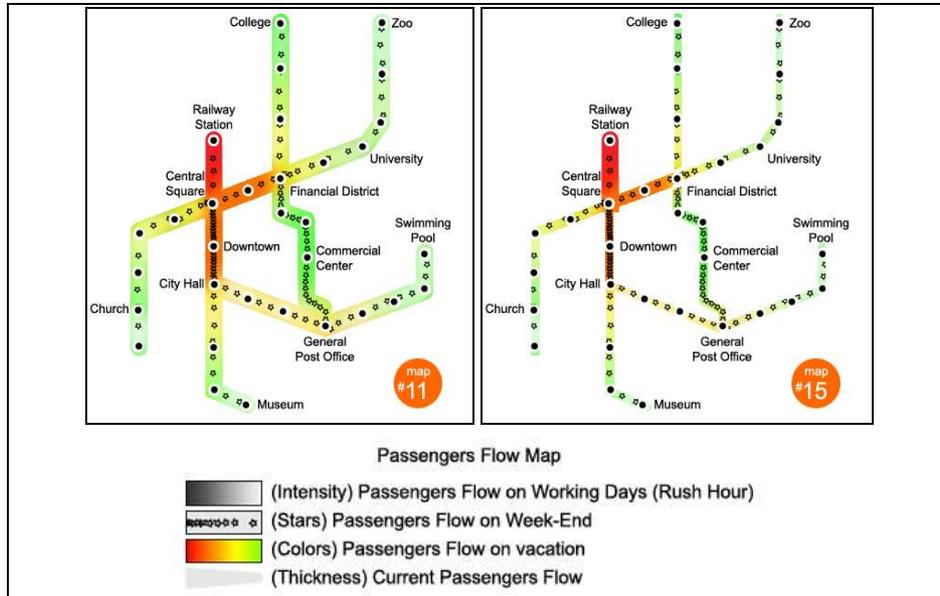


Figure 7. *Quatre exemples de cartes de métro combinant respectivement une, deux, trois et quatre variables visuelles permettant de lire sur la même carte le flux de passagers à différents moments*

Pour chaque combinaison, nous cherchons à déterminer si l'information décrite par chaque variable parmi les multiples variables est clairement interprétée par le candidat. Pour cela, nous demandons au candidat d'interpréter une information pour chacune des variables de chaque combinaison (chaque carte). Par exemple, dans la figure 8, nous demandons à l'utilisateur de nous indiquer la valeur du flux de passagers en fin de semaine entre les stations « Central Square » et « Financial District ». La carte affichée combine les quatre variables visuelles en même temps et ce flux est indiqué par la variable « texture ». Si la réponse du candidat est erronée, nous pouvons en déduire que la variable « texture » n'est pas claire dans cette combinaison des quatre variables en même temps. Cette méthode implique d'adresser au candidat :

- Une question pour chacune des quatre cartes représentant une variable visuelle
- Deux questions pour chacune des six cartes combinant deux variables visuelles
- Trois questions pour chacune des quatre cartes combinant trois variables visuelles
- Quatre questions pour la carte combinant les quatre variables visuelles ensemble

Au total, 32 questions sont adressées au candidat. Chaque question est présentée au candidat au moyen de l'interface de la figure 8.

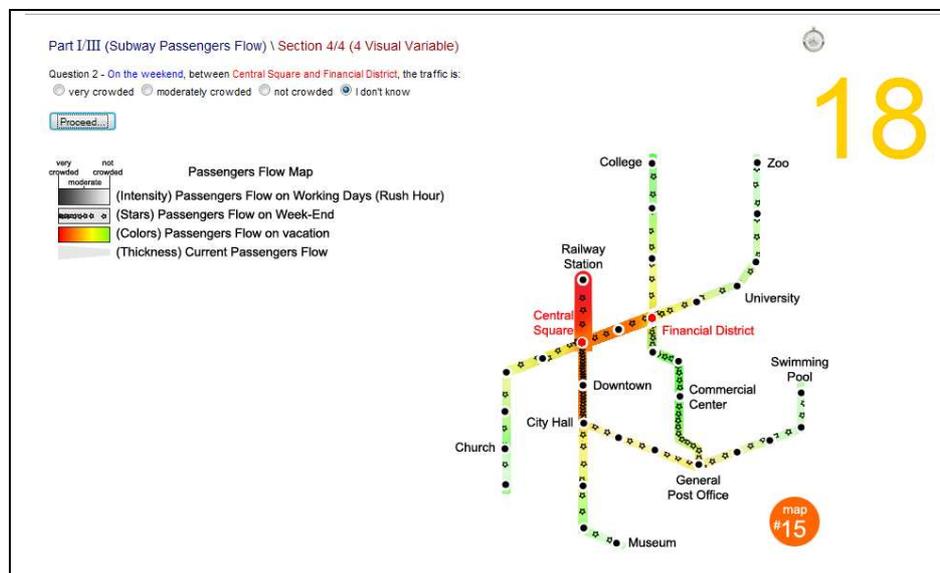


Figure 8. Interface de notre test en ligne. L'exemple demande à l'utilisateur d'indiquer le flux de passagers en fin de semaine (variable « texture ») entre les stations « Central Square » et « Financial District ». En haut à droite, un compte à rebours limite le temps de réponse

Afin d'obtenir des réponses intuitives, c'est-à-dire basées sur la perception et non sur le raisonnement, il est important que nous limitions le temps de réponse à un temps relativement court. En haut à droite de l'interface, un compte à rebours en secondes indique le temps disponible restant.

Concernant les personnes testées, nous prévoyons de soumettre le test à deux panels types de candidats : un groupe d'experts en visualisation de l'information spatiale et un groupe de non experts.

Le test et son interface sont implémentés en langages Javascript et PHP. L'implémentation de la première étape du test est en fin de réalisation et est consultable à l'adresse suivante : <http://franck.favetta.free.fr/indicators> (mot de passe : « spatial »).

4. Conclusions et perspectives

Dans le cadre de notre recherche sur le domaine des indicateurs spatiaux continus, nous avons présenté la construction d'un test en ligne qui permet de mesurer d'une part le nombre maximum de variables visuelles de Bertin qu'une représentation peut combiner pour rester claire, et d'autre part d'identifier les combinaisons les plus claires pour un nombre de variables donné. A présent, nous devons effectuer ce test auprès de candidats afin :

1. de mesurer et analyser le pourcentage de réponses erronées pour le groupe de cartes qui combinent une, deux, trois et quatre variables visuelles,
2. pour chaque groupe, d'identifier les combinaisons qui génèrent les meilleurs pourcentages de bonnes réponses.

Nous identifierons ainsi le nombre maximum de variables visuelles à ne pas dépasser dans une représentation d'indicateurs spatiaux et quelles sont les meilleures représentations.

La première étape de notre démarche de test est encore en phase d'analyse. Actuellement nous avons proposé le test à des candidats experts, et nous avons commencé à recueillir des résultats que nous permettront d'extraire des statistiques intéressantes. La seconde étape devra être mise en place afin de compléter ce test par une évaluation des meilleures représentations visuelles parmi un ensemble de représentations usuelles présentées précédemment.

Nous avons également en perspective d'étudier l'apport potentiel de l'animation et de l'interactivité pour améliorer la représentation des indicateurs spatiaux.

5. Références

- Bertin J., *La graphique et le traitement graphique de l'information*, Ed. Flammarion, 1977, 273p.
- Bertin J., *Sémiologie graphique*. Paris, Ed. Gauthier-Villars, 1967, 435 p.
- Bonavero P., Dansero E., dir., *L'Europa delle regioni e delle reti : I nuovi modelli di organizzazione territoriale nello spazio unificato europeo*, UTET libreria, 1998, 413 p.
- Bonhomme C., Un Langage Visuel dédié à l'interrogation et à la Manipulation de Bases de Données Spatio-Temporelles, Thèse de doctorat, Institut National de Sciences Appliquées de Lyon, Villeurbanne, France, 2000.
- Bottoni, P., Costabile, M.-F., Levialdi, S. & Mussio, P., *Formalising Visual Languages*, proceedings of 11th IEEE Symposium on Visual Languages, Darmstadt, Germany, September 5-9, 1995, 45-52 pp.
- Cauvin C., Reymond H., Schaub G., *Accessibilité, temps de séjour et hiérarchie urbaine : l'exemple du réseau aérien pour 53 villes d'Europe*, Sistemi Urbani 3, 1989, 297-324.
- Chen C., M.P. Czerwinski, *Empirical evaluation of information visualizations: an introduction*, Int. J. Human-Computer Studies 53, 2000, 631-635 pp.
- Chen C., *Information Visualization*, Chapter 6, *Empirical Studies of Information Visualization*, Springer London, 2006, 173-210 pp.
- Card S., *The Psychology of Human-Computer Interaction*, P. Moran and Allen Newell. 1983.
- Favetta F., Ambiguïtés dans les langages visuels d'interrogation spatiale et précision topologique des Bases de Données spatiales, Thèse de doctorat, Institut National de Sciences Appliquées de Lyon, Villeurbanne, France, 2003.
- Friedman V., *Data Visualization and Info-graphics*, Graphics, Monday Inspiration, January 14th, 2008.
- Ioannilli M., *Analisi dei sistemi urbani e territoriali*, Università di Roma Tor Vergata; 2005.

- Laurini R., *Information Systems for Urban Planning: A Hypermedia Cooperative Approach*, Taylor and Francis, February 2001, 349 p.
- Laurini R., Murgante B., *Interoperabilità semantica e geometrica nelle basi di dati geografiche nella pianificazione urbana*". L'informazione geografica a supporto della pianificazione territoriale, Beniamino Murgante editor, Franco Angeli/Urbanistica ISBN 978-88-568-0363, 2008, 229-244 pp.
- Mailles, S., Les représentations analogiques comme support de l'anticipation dans les environnements dynamiques, Thèse de doctorat, UFR de Psychologie & Laboratoire Travail et Cognition, CNRS-UMR5551, Université de Toulouse 2, 1996.
- Marr D., *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, MIT Press, 1982.
- McCormick B. H., *Visualization in Scientific Computing*. Edited with Thomas A. DeFanti and Maxine D. Brown, ACM Press, 1987.
- Peterson M.P., *Interactive and animated cartography*, New Jersey: Prentice Hall, (Prentice Hall Series in Geographic Information Science), 1995, 257 p.

Session Posters

Méthodes interactives de discrétisation de distributions statistiques et/ou/versus spatiales

Didier JOSSELIN*, Benoît PRACHE*+

*CNRS, Université d'Avignon et des Pays de Vaucluse
74, rue Louis Pasteur

+IUT STID, Agroparc
84000 Avignon Cedex

didier.josselin@univ-avignon.fr

RÉSUMÉ. Dans l'environnement de développement statistique *LispStat*, nous avons développé un ensemble d'outils, dont le *Distogramme*, méthode graphique de discrétisation de distributions spatiale et statistiques. Nous proposons différentes formes de découpages classiques, dont une approche par analyse de variance, et évaluons leurs apports respectifs en cartographie interactive.

ABSTRACT. Within *LispStat*, we developed the *Distogram*. We recently added different ways to build a partition in the data, using variance analysis, notably. We discuss the interest of such functionalities in interactive cartography.

MOTS-CLÉS : *Distogramme*, analyse interactive exploratoire, découpage, discrétisation, distribution.

KEYWORDS: *Distogram*, exploratory spatial data analysis, discretisation, distribution.

1. Le Distogramme

Pratiquement, le Distogramme est issu (i) du besoin d'une meilleure interactivité entre l'histogramme d'une variable et la représentation cartographique des individus décrits, (ii) de la difficulté de discrétiser correctement certaines variables, même transformées, (iii) de la nécessité de prendre en compte la dimension spatiale dans la construction d'un histogramme « optimal », (iv) de l'insuffisance des méthodes de discrétisation automatique fournies par les logiciels de cartographie. L'objectif du Distogramme est d'analyser dynamiquement les distributions spatiales et statistiques pour permettre une meilleure appréhension des objets géographiques et des relations qu'ils entretiennent (fonctionnelles, structurelles, statistiques, spatiales, topologiques...). Le Distogramme est basé sur la règle des D : Discrétisation, Dynamique, Discontinuité, Distribution, Dualité, notamment.

Le distogramme n'est pas autre chose qu'un véritable histogramme (figure 2). Un histogramme est un type de graphique en colonnes qui montre la répartition des données. Chaque colonne est une classe ou une catégorie. La hauteur de la colonne indique le nombre de données appartenant à cette catégorie ou la fréquence. Dans le Distogramme, la surface des colonnes est proportionnelle à l'effectif. Fondamentalement, le Distogramme ne fait que reprendre un ensemble d'outils et de concepts couramment utilisés en géographie. Il repose sur des méthodes de discrétisation éprouvées et d'animations cartographiques. Sa particularité est qu'il les regroupe en un ensemble cohérent et les met au service de l'expert dans un processus interactif. Il ne s'agit aucunement d'un outil de cartographie au sens strict, mais plutôt d'un outil d'analyse exploratoire, pour visualiser et identifier inter activement tous les individus afin de les associer ou les regrouper.

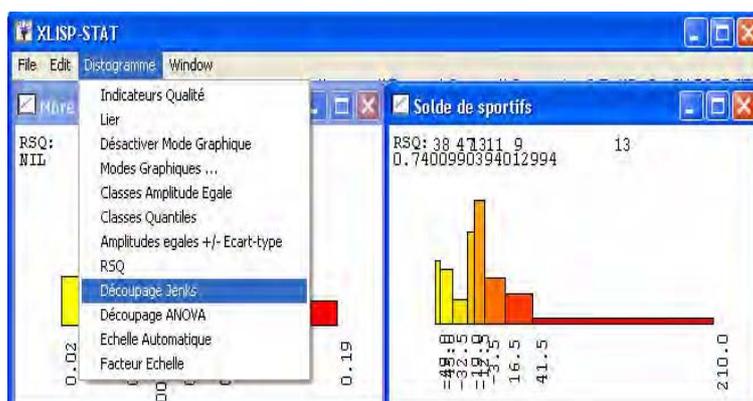


Figure 1. Fonctionnalités ajoutées au Distogramme (autour de l'analyse de Variance)

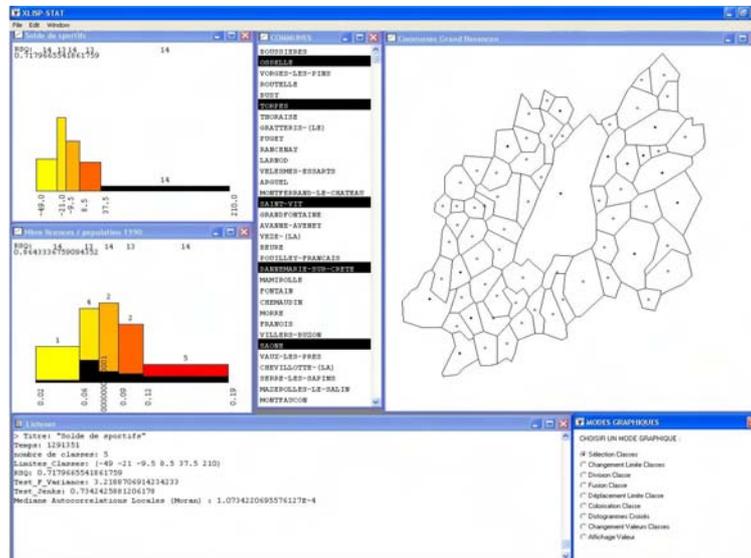


Figure 2. Le Distogramme : des cartes et des histogrammes liés

2. Analyse de variance

En plus de ses fonctionnalités usuelles, le Distogramme comporte aujourd'hui trois nouveaux outils d'analyse de variance (figure 1), le but étant, classiquement, de maximiser la variance inter-groupe et de minimiser la variance intra-groupe :

- le RSQ (R-square) : critère de classification (variance inter-classe / variance totale), utile pour obtenir le nombre de classes optimal ; il est calculé dynamiquement sur les Distogrammes et il peut aussi être calculé sous forme de graphique spécifique (figure 2).
- le découpage automatique par le critère de Jenks et le découpage automatique par analyse de variance, ces deux méthodes permettant d'obtenir une discrétisation en fonction des densités locales des points et du nombre de classes choisi.

À ces fonctionnalités strictement statistiques a été ajouté le calcul des indicateurs locaux d'autocorrélation spatiale (LISA) qui permet de reconsidérer la distribution statistique, parallèlement à la vision sur l'écran de phénomènes de regroupement et d'agrégation spatiale lors du changement du nombre ou des limites de classes.

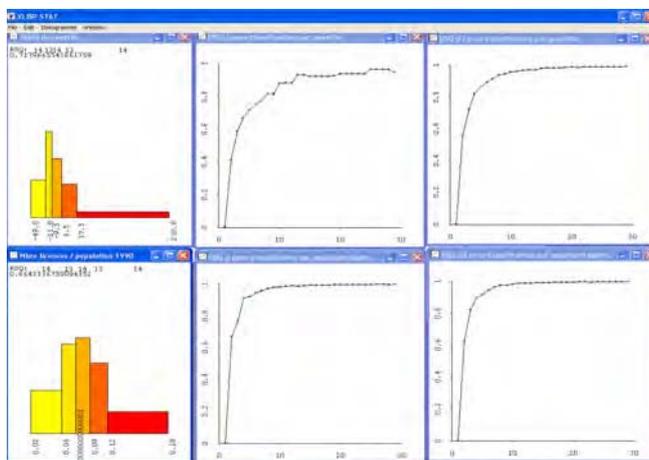


Figure 3. Courbes du RSQ en fonction du nombre de classes : le petit point vert correspond au nombre de classe optimal

La figure 3 présente des graphiques croisant les valeurs de RSQ et le nombre de classes. Ces graphiques permettent de trouver des compromis entre un nombre raisonnable de classes et une qualité suffisante de discrimination en termes de variance. L'originalité ici est qu'en cliquant sur un point de la courbe, on retrouve la distribution liée, recalculée dynamiquement.

3. Conclusion

L'objectif du poster est de présenter et discuter l'usage du distogramme et de l'ensemble de ses fonctionnalités, y compris celles qui ont été ajoutées récemment autour de l'analyse de variance, qui apportent une valeur ajoutée indéniable à la précision et à la qualité de la discrétisation de la distribution obtenue.

4. Bibliographie

- Josselin D., Le Distogramme, Actes des Journées ThéoQuant, du 21 au 23 février 2001, 10 pages.
- Josselin D., L'analyse des discontinuités spatiales avec le Distogramme, Contexte théorique, présentation, évaluation, N° spécial « cartographie animée et interactive », vol. 13, n°1/2003, pp. 29-48, *Revue Internationale de Géomatique*, Hermès, Lavoisier, Paris, 2003.

Entre lisser et raboter.

Application du filtre auto-adaptatif médien sur image satellites.

Didier JOSSELIN*, Alexandre ULMER*+

**CNRS, Université d'Avignon et des Pays de Vaucluse*

74, rue Louis Pasteur

+IUT STID, Agroparc

84000 Avignon Cedex

didier.josselin@univ-avignon.fr

RÉSUMÉ. Après avoir rappelé la construction de la médienne, combinaison linéaire de la moyenne et de la médiane, nous appliquons un filtre spatial local sur une image. Ce filtre s'adapte en fonction des distributions observées. La robustesse de la méthode est évaluée. Les résultats et les apports sont discutés avec des paramétrages différents.

ABSTRACT. We first explain how is built the MeAdian, a linear combination of the Mean and the Median. We assess the robustness of efficiency of the MeAdian. Some results are provides on satellite images filtering.

MOTS-CLÉS : Median, Mean, MeAdian, spatial filtering, remote sensing.

KEYWORDS: Médiane, Moyenne, Médiennne, filtre spatial, télédétection.

1. La médienne, pourquoi ?

La médienne est le fruit d'une réflexion à l'interface de deux disciplines : la Géographie et la Statistique. Elle a été élaborée dès 2001, puis appliquée à des distributions univariées, pour calculer des centres statistiques ou spatiaux (Josselin & Ladiray, 2002, 2004). Mathématiquement, la médienne est une méthode originale pour trouver le centre d'une distribution statistique. Le concept développé associe la moyenne et la médiane, dans le but d'améliorer la robustesse d'un estimateur.

Géographiquement, lorsqu'on souhaite analyser une image brute ou une partition spatiale à l'aide d'un calcul numérique automatique, le choix d'une méthode pertinente est crucial. Suivant la fonction choisie, les filtres lissent l'image et réduisent son « bruit » (cas de la moyenne, par exemple) ou la filtrent en éliminant les « outliers » (cas de la médiane, par exemple, qui « rabote »). Un filtre moyen permet ainsi de souligner les zones géographiques hétérogènes, un filtre médian de révéler les zones homogènes et leurs contours. Bien qu'en y regardant de plus près, la notion d'homogénéité dépend fortement de la quantité d'objets ou de pixels considérés: une grande variété, régulière dans sa composition et son organisation, peut aussi être considérée comme une forme d'homogénéité, une espèce fortement dominante, associée à quelques rares espèces différentes également.

2. Formalisme

Dès le début du XIX^e siècle, Pierre Simon de Laplace a étudié la distribution jointe de la Moyenne et de la Médiane. Il proposa une définition différente de ce que nous appelons aujourd'hui la « médienne » (ou « MeAdian » en anglais). Il y avait introduit un terme correcteur de covariance :

$$\text{Médienne de Laplace} = (1-C)*\mu - C*M$$

avec μ : la moyenne arithmétique, M : la médiane, $V(\mu)$ la variance de la moyenne, $V(M)$ la variance de la médiane.

$$C = [V(\mu) - \text{Cov}] / [V(\mu) + V(M) - 2*\text{Cov}] \quad \text{et} \quad \text{Cov} = \text{cov}(M, \mu)$$

Le formalisme de la **médienne dite "simple"** (Josselin & Ladiray, 2002) est :

$$\text{Médienne} = (1-C)*\mu - C*M \quad \text{avec} \quad C = V(\mu) / [V(\mu) + V(M)]$$

Dans ses deux formalismes, la médienne est une combinaison linéaire de la moyenne et de la médiane. Les variances des deux valeurs centrales sont estimées par un bootstrap (tirage au sort avec remise un grand nombre de fois, cf. figure 1).

3. Robustesse de la médienne face aux autres indicateurs de centralité

L'efficacité relative de l'estimateur, définie comme le rapport entre sa variance et celle du meilleur testé donne une évaluation de la robustesse de l'estimateur. Calculés sous R, les résultats montrent que pour la plupart des distributions simulées (en colonne), les médienes obtiennent des résultats relativement correctes, la moyenne et la médiane obtenant une efficacité relative nulle ou presque, respectivement pour les distributions de Cauchy et binomiale.

	Moyenne	Médiane	MeAdian	M_Laplace
Normale	100.00	61.32	91.24	98.03
Cauchy	0.00	98.35	96.54	100.00
Binomiale	51.81	0.05	41.74	100.00
Beta	75.06	24.67	56.38	100.00
Poisson	42.14	51.77	100.00	83.87
Gamma	99.90	100.00	83.01	23.15
Khi2	60.76	100.00	42.95	17.06

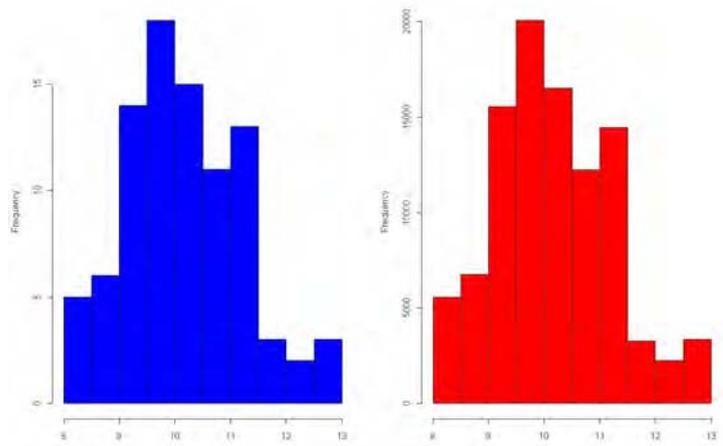


Figure 1. Distribution observée et agrégation des valeurs de 1000 distributions « bostratpées »

4. Filtrage médien d'image satellitaire

Pour la première fois, le filtre médien a été appliqué au traitement d'image, sous ses deux formes, en comparaison aux filtres moyen et médian. Ce filtrage comporte plusieurs paramètres :

4 SAGEO'2010 Toulouse

- amplitude de la fenêtre du filtre (3x3 jusqu'à 7x7) ;
- nombre d'itérations successives, qui peut influencer sur la qualité du filtrage;
- exposant des variances respectives dans la combinaison linéaire de la moyenne et de la médiane (non traité dans cette présentation).

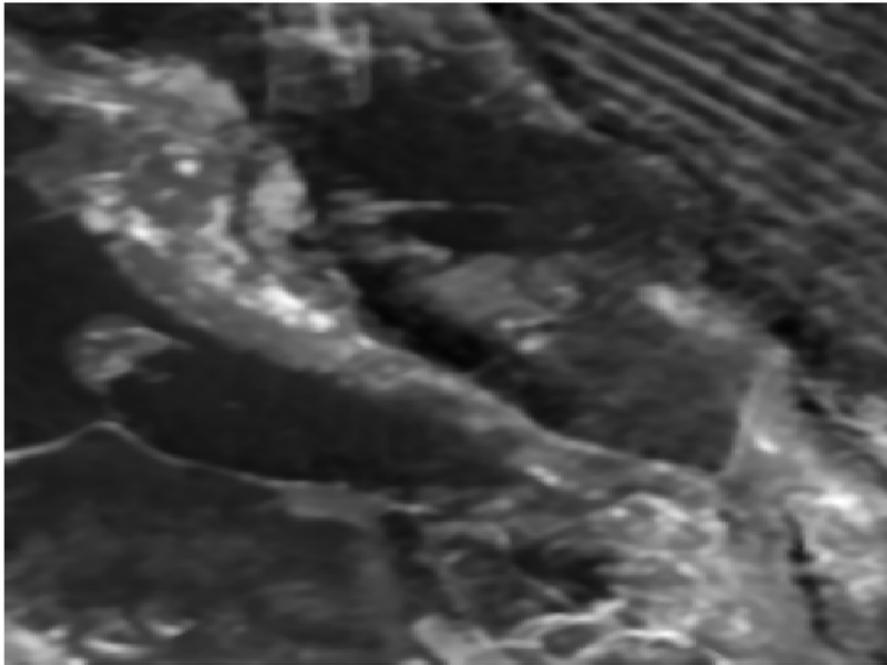


Figure 2. *Filtrage médien 3x3, une itération*

L'analyse des résultats montre la capacité d'adaptation du filtre à la distribution locale des valeurs : les zones hétérogènes sont lissées et les zones relativement sont rabotées, mettant en évidence les contrastes (figure 2). À cette résolution de filtrage (3x3), les effets sont relativement peu visibles et flous, mais l'on peut déjà voir apparaître l'effet de nettoyage dans les zones sombres où les pixels clairs sont éradiqués, et également, le sur-lignage des oppositions blanc/noir (discontinuités).

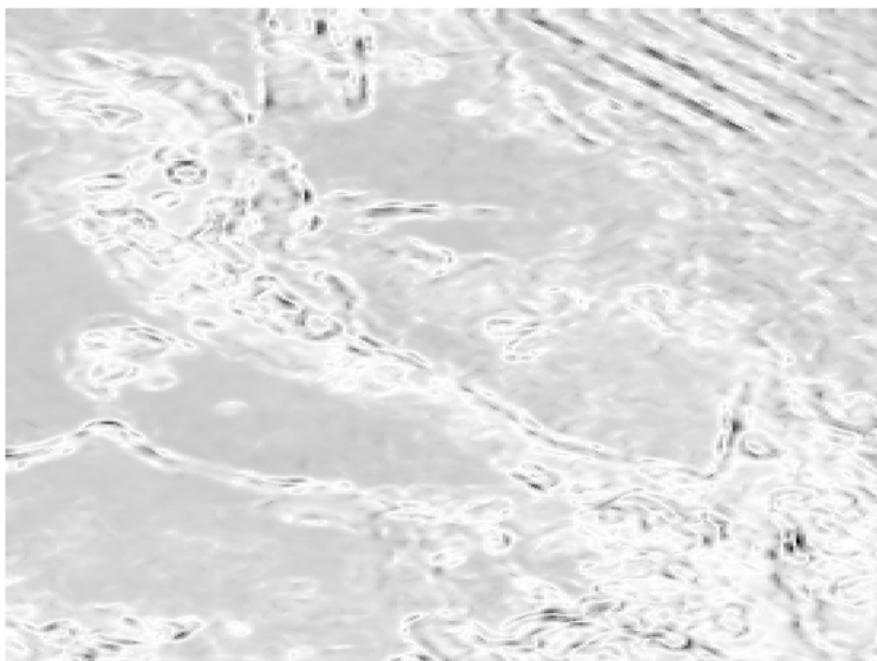


Figure 3. *Différence algébrique entre filtre moyen et filtre médien (filtre 3x3)*

À titre d'exemple, la différence entre les filtres moyen et médien indique les secteurs où les deux filtres ont convergé ou bien ont eu des comportements différents. Ils soulignent notamment les contours de zones homogènes vs hétérogènes, au niveau de résolution du filtre en question (ici 3x3, figure 3).

4. Conclusion

L'objectif est de présenter et discuter des résultats utilisables pour le filtrage d'images (lissage vs zonage, détection de contours) par le filtre médien auto-adaptatif. Il s'agit également de discuter l'effet de la résolution et du nombre d'itérations sur les résultats obtenus.

7. Bibliographie

Josselin D., Une piste pour la recherche de la "valeur centrale optimale". Discussion autour de la robustesse et du comportement de la "médienne", combinaison de normes L_p , *Cybergeo*, 2004, 18 pages.

Josselin D., Ladiray D., Combining L1 and L2 Norms, for a more Robust Spatial Analysis : the "Median Attitude", *Cybergeo*, 2002, 14 pages.

Josselin D., Ladiray D., La Médienne : une nouvelle norme pour la statistique ? Construction, propriétés et applications en analyse spatiale, Actes du colloque de la Société Francophone de Classification 2002, Toulouse, 16-18 septembre 2002, pp. 225-229.

Nonlinear source separation contribution in multispectral image classifications

Hela Elmannai^{1,2} *, Mohamed Anis Loghmari¹ **, Mohamed Saber Naceur¹ ***

Laboratoire de Télédétection et des Systèmes d'Informations à Références Spatiales

¹*Ecole Nationale d'Ingénieurs de Tunis (ENIT), Tunisia*

²*Ecole Supérieure des Communications de Tunis (SUP'COM), Tunisia*

hela.elmannai@gmail.com*, *MohamedAnis.Loghmari@isi.rnu.tn*,

****Saber.naceur@insat.rnu.tn*

ABSTRACT. Source separation is used in a wide range of applications. The most widely used algorithm is the Independent Component Analysis (ICA), which is an extension of the Principal Component Analysis (PCA). Also, in many realistic cases the process which generates the observations is nonlinear and no information is available about the mixture. In this case, the nonlinear ICA can be used for similar tasks as his linear counterpart. But, it can be expected to capture the structure of the data better if the data points lie in a nonlinear manifold instead of a linear subspace. In this paper, we try to find a model which allows a compact description of the observations in the hope of discovering some of the underlying causes or sources of the observations. Then, we will process a dimension reduction to classify the obtained sources and evaluate the performances of the proposed method.

RÉSUMÉ. La séparation des sources est utilisée dans plusieurs applications. L'algorithme le plus utilisé est l'Analyse en Composantes Indépendantes (ACI), qui est une extension de l'Analyse en Composante Principale (ACP). De plus, dans la plupart des cas réels le processus générant les observations est non linéaire et il n'y a pas d'information sur le mélange. Dans ce cas, l'ACI dans sa version non linéaire peut être utilisée d'une manière similaire à sa version linéaire. Mais, elle a de plus l'avantage de pouvoir extraire certaines structures de données plus efficacement dans un espace non linéaire. Dans cet article, on cherche à mettre en place un modèle non linéaire permettant une description compacte des observations dans le but de restituer les sources responsables de nos observations. Ensuite, on passe à l'étape de réduction de dimension pour classifier ces sources et à l'évaluation des performances de la méthode proposée.

KEYWORDS: Blind Source Separation, nonlinear mixture, Nonlinear Factor Analysis, Bayesian model, Multilayer perceptron, Clustering.

MOTS-CLÉS : Séparation aveugle des sources, mélange non linéaire, Analyse factorielle non linéaire, modèle Bayésien, Perceptron Multicouches, classification.

1. Introduction

While most natural phenomena are inherently nonlinear, the most widely used source separation methods are linear. In this work we present a nonlinear Blind Source Separation (BSS) method including Nonlinear Factor Analysis (NLFA). The proposed method estimates the nonlinear mapping without any assumption about the mixture. Then, we propose a new approach to enhance the dimension reduction.

2. Nonlinear Blind Source Separation and Nonlinear Factor Analysis model

Considering an N -sensor array that provides the vector of observed variables $X(t)$ and $S(t)$ the M unobserved sources. The signal is a nonlinear mixture of unknown statistically independent sources and $N(t)$ the noise variables. The mixture relation is $X(t) = f(S(t)) + N(t)$ (Honkela, 2005) where f is a nonlinear function. The separating system g recovers estimated sources given by $Y = g(X)$. The corresponding ICA model including noise term is called NLFA model. The nonlinear mapping is $f(S(t)) = B \tanh(A S(t) + a) + b + N(t)$ which is modelled by Multilayer Perceptron (MLP) (Lappalainen *and al.*, 2000). The model is learned by the Variational Bayesian Learning that updates iteratively the parameters of approximated distributions. Let θ the unknown parameters θ (sources, noises, mapping function parameters) and $p(\theta|X)$ the posterior probability density function. The approximation of $p(\theta|X)$ denoted $q(\theta)$ is estimated by Gaussian distributions using the Kullback-Leibler divergence.

3. General approach architecture

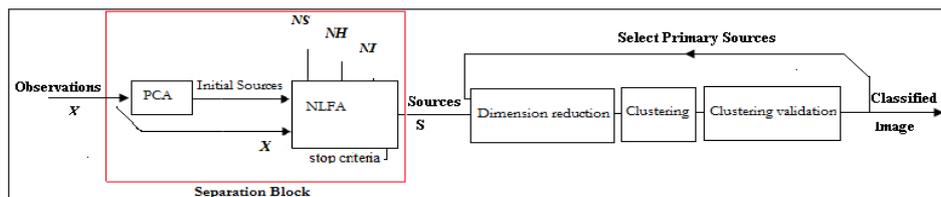


Figure 1. General approach

The general approach is presented in Figure 1. NS , NH and NI are respectively the number of sources, hidden neurons and iterations. The proposed method implements the NLFA using a PCA initialization for the sources. First, we consider that NS is equal to the number of observations. The last three steps will select only “primary” sources that have the best rate of well classification. The Dimension

reduction block generates different combinations of sources. For example for four sources, it generates $C_4^4 + C_4^3 + C_4^2 + C_4^1$ combinations. Each combination will be processed by the Clustering algorithm and Clustering validation step. The Clustering algorithms are the K-means and the Minimum Distance. The validation block measures the well classification rate comparing to the ground truth image. The MLP and the learning algorithms are susceptible to local minima. The new approach is able to avoid this problem by providing the classification criterion.

4. Experiments and analysis

We consider the case of multispectral images. The Figure 2 shows a lake environment located in central of Tunisia through four different SPOT spectral channels B1, B2, B3 and B4. The NLFA method gives four source images S1, S2, S3 and S4 (Figure 2). The last three steps show that the (S2 S3) combination provides the best well classification rate. Comparing also to other separation method such us JADE and SOBI, the proposed method have better classification rate.

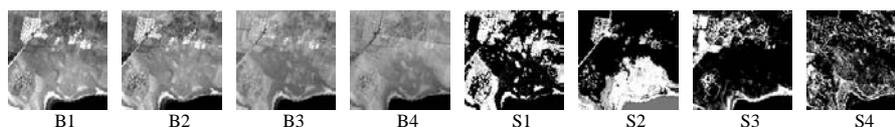


Figure 2. Channel images of SPOT-4, May 31, 1998 and sources.

5. Conclusion

This work aims to investigate the application of nonlinear BSS and the dimension reduction in multispectral image clustering. The separated sources were processed to extract the sources that contain main information about the studied area. This work could be advanced to cover hyperspectral data and secondary sources can be analysed to extract hyper-informations omitted by the given ground truth.

6. Bibliography

- Honkela A., Advances in variational Bayesian Nonlinear Blind Source Separation, Dissertation for the degree of Doctor of Science in Technology, Helsinki University of Technology, Espoo, 2005, p. 44-61.
- Lappalainen H., Honkela A., Bayesian Nonlinear Independent Component Analysis by Multi-Layer Perceptrons, Advances in Independent Component Analysis, Springer, 2000, pp. 93-121.

COMMUNICATION POSTER SaGéo 2010

Les espaces d'une consommation engagée : Une grille d'analyse des AMAP en Bouches du Rhône

MOTS-CLES : AMAP, réseau d'acteurs, analyse spatiale, territorialité, Multi échelles

Les AMAP (Association pour le Maintien de l'Agriculture Paysanne) sont un mode d'échange alimentaire en vente directe de paniers, une forme de circuit court. Un discours récurrent sur les AMAP porte sur le soutien à l'agriculture paysanne et sur la re-territorialisation du circuit alimentaire de la production à la consommation. On parle d'agriculture locale, de mise en place d'une proximité favorable au lien social. La territorialité exprimée et revendiquée est alors le garant d'un échange plus humain, basée sur la proximité, condition pour établir des liens de confiance.

L'engagement pour le maintien d'une agriculture locale pose la question de son potentiel de territorialité, à la fois dans les pratiques de consommation et pour l'activité agricole dans un espace donné. L'organisation spatiale résultante de la mise en réseau de consommateurs et de producteurs doit faire l'objet d'une observation préalable pour juger des interactions structurantes pour l'espace donné. Effectivement la spatialité du phénomène à été très peu abordée, ce qui nous apparait une nécessité pour discuter les notions de proximité, de local et de territoire.

En mettant en place ce type d'observation, nous nous faisons une idée du local que sollicite réellement le phénomène. Ensuite, le regard que nous avons sur la dynamique spatio-temporelle du phénomène, entre 2006 et 2010, nous permet une première appréciation de cette proximité revendiquée : une rotation importante des acteurs ou à l'inverse une permanence des connexions entre de mêmes espaces et populations étant des indicateurs de la pratique ou non d'une proximité.

Observer les relations spatiales d'un réseau d'échange alimentaire demande de spécifier les différents niveaux d'organisations de ce réseau, afin de mettre en valeur les types d'interactions spatiales observés et les échelles d'analyse géographique nécessaires à leur caractérisation et à leur mesure.

L'organisation d'un réseau d'acteurs détermine les lieux de l'action du celui là. Après la caractérisation structurelle, la localisation de ces espaces est donc la première étape d'une étude géographique de réseau d'acteurs. A noter que dans une approche diachronique il faut observer à la fois les évolutions structurelles et les évolutions des espaces concernés.

Notre analyse géographique veut comprendre ces localisations et cherche une explication dans les interactions entre l'organisation du réseau (structure et fonctionnement) et l'organisation de l'espace géographique. Il apparaît ici que l'observation des attributs des acteurs, le schème causal, est nécessaire pour compléter l'approche des caractéristiques interactionnelles du réseau, le schème structural. La mise en valeur de la structure spatiale du réseau d'acteur se fait par l'analyse des distances et aboutit à la construction de graphes valués.

Nous identifions donc des espaces différenciés et hiérarchisés d'après leur position dans le système d'action du réseau. Ces lieux sont les nœuds du « graphe géographique » et les liens correspondent aux distances (distance physique, distance-temps ou distance-coût selon l'analyse souhaitée) entre les lieux.

En l'occurrence pour les réseaux AMAP on identifie les lieux de production, de distribution et de consommation. Dans un second temps on mesure les distances parcourues par les acteurs du réseau. Nous ne représentons pas les déplacements des consommateurs sur le lieu de production: pour la lisibilité du graphe et pour la faiblesse de ce lien : fréquence faible et seule une partie des consommateurs – voir quasiment aucuns pour certaines AMAP – pratiquent ce déplacement.

Dans les Bouches du Rhône, nous observons à l'échelle macro géographique un semis de points qui représente les lieux clés du phénomène : la production et la distribution. A cette échelle la localisation des lieux de consommation alimentaire (lieux de résidence des acteurs-consommateurs) disparaît et se confond majoritairement avec le lieu de distribution. A l'échelle méso géographique on observe la structure du réseau AMAP, tous les lieux sont identifiés et mis en relation : c'est l'échelle privilégiée de l'observation des réseaux. A l'échelle micro géographique on observe un sous système du réseau : la localisation résidentielle des consommateurs en fonction du lieu de distribution.

Pour chacune de ces échelles nous cherchons à caractériser l'organisation spatiale des lieux par l'analyse des distances, de la dispersion des lieux par rapport à des centralités du territoire des Bouches du Rhône mais aussi en observant les corrélations des lieux avec les attributs des espaces contigus. Ainsi nous repérons des structures plus ou moins expliquées par l'hétérogénéité de l'espace. Pour chaque échelle nous identifions un système spatial différent.

De cette manière la mise en valeur de la complexité des relations à l'espace des réseaux AMAP et le constat de la souplesse actuelle de ce système d'échange alimentaire de par son caractère innovant nous amène à souligner le besoin d'une étude des effets d'une massification et d'une normalisation du phénomène AMAP pour un territoire donné.

Bibliographie

DUMOLARD P., Distances, accessibilité et diffusion spatiale, in THERIAULT M., DES ROSIERS F. (2008), *Information géographique et dynamiques urbaines - Tome 2*, Hermès – Lavoisier, 268 pages.

FONTAN J.-M., KLEIN J.-L. et TREMBLAY D.-G., (2004) *Innovation et société : pour élargir l'analyse des effets territoriaux de l'innovation*, Géographie Economie Société 2004/2, Vol. 6, p. 115-128.

LAMINE C., (2008), *Les AMAP : un nouveau pacte entre producteurs et consommateurs ?*, Gap, édition Yves Michel.

MUNDLER P. (sous la direction de) (2006), *Fonctionnement et reproductibilité des Amap en Rhône-Alpes*, ISARA.

POUZENC M., BULHER A.-E., COQUART D., GIROU S., FONTORBES J.-P., MONDY B., OLIVIER V., PILLEBOUE J., VINCQ J.-L., (2007). *Les relations de proximité agriculteurs consommateurs: Points de vente collectifs et AMAP en Midi-Pyrénées*, UMR Dynamiques Rurales, Toulouse, avril, 267 p.

PUMAIN D. et SAINT-JULIEN Th. (1997), *L'analyse spatiale. Localisation dans l'espace*. Paris, Arman Colin (Coll. « Géographie »), 167 p.

SANDERS L. (dir), (2001) *Modèles en analyse spatiale*, Paris, Hermès Lavoisier.

Les centralités commerciales de réseau dans l'unité urbaine de Paris : Hiérarchie et polarisations

Matthieu DELAGE

UMR 8504 géographie-cités, CNRS, Paris 1, Paris 7

<mailto:matthieu.delage@parisgeo.cnrs.fr>

RÉSUMÉ. Le poster proposé étudie l'organisation du commerce de réseau en Ile-de-France grâce à une méthode d'analyse spatiale. Plus particulièrement, ces commerces sont regroupés pour former des centralités grâce à une classification spatiale hiérarchique. Une série de tests quantitatifs et empiriques permet de délimiter les centralités commerciales de réseau. Ces dernières mises au jour, le poster montre la diversité morphologique et fonctionnelle des centralités commerciales de l'unité urbaine de Paris.

ABSTRACT. The poster here presented studies the retail organisation in the Paris city Region. A spatial analysis method is used (spatial cluster analysis). The poster shows the morphological and functional diversity of the retail organisation in the Paris city Region

MOTS-CLÉS : Analyse spatiale, commerces, hiérarchie, centralité, densité, Paris

KEY WORDS: Spatial analysis, retail, hierarchy, centrality, density, Paris

Introduction Générale

La localisation des commerces, concentrés ou dispersés, repose sur des logiques de complémentarité ou de concurrence, et sur des logiques d'acteurs. En France, depuis la Loi Royer de 1973, les commerces de plus de 300 m² doivent de plus demander une autorisation auprès des CDEC (Commissions Départementales d'Équipement Commercial). 37 ans après la mise en place de cette loi, qui visait un équilibre entre commerces traditionnels et centres commerciaux, où en est l'offre commerciale en Ile-de-France ? Comment se hiérarchise-t-elle ? L'objectif de ce poster est de construire une méthodologie pour décrire l'offre commerciale francilienne du point de vue des positionnements des différents commerces les uns vis-à-vis des autres. L'originalité est de s'appuyer sur une base de données regroupant 7510 commerces dits de réseau pour décrire l'offre commerciale dans sa diversité, des rues commerçantes des centres des villes de taille moyenne de l'unité urbaine de Paris, aux centres commerciaux de niveaux régional et national.

1. Concentrations commerciales dans l'espace

Du commerce en général au commerce de réseau

L'étude de l'offre commerciale dans l'unité urbaine de Paris nécessite le recours à une base de données de commerces. À l'échelle ici considérée, il n'existe pas de base exhaustive. Si la ville de Paris recense les commerces parisiens avec une nomenclature très détaillée (Base BDCOM), les départements de première et deuxième couronnes ne disposent pas de données aussi précises. Nous présentons la

base de données Point de Vente Panorama, dont les données couvrent l'espace d'étude, avec 7510 commerces recensés. Ces commerces sont des commerces dits de réseau. On définit comme appartenant à un réseau tout commerce se regroupant de manière durable avec d'autres sous une enseigne commune, par la voie de rapports contractuels : franchise, affiliation, concession, location-gérance, commerces associés en coopératives. S'y retrouvent les grands magasins parisiens, les hyper et supermarchés, mais aussi tous les autres commerces de taille plus petite avec une enseigne commune.

Méthode d'analyse

L'enjeu ici est de définir des concentrations commerciales significatives, à même de caractériser la structure de l'organisation commerciale générale de l'unité urbaine de Paris. Pour cela, une méthode d'exploration et d'analyse des données qui repose sur une analyse de la distance entre les commerces a été retenue. Ce type de méthodes permet de dégager des concentrations, et ainsi de pallier au moins partiellement au biais induit par les données, celui de la non-exhaustivité : au sein de la concentration commerciale qui apparaîtra, les interstices pourront correspondre à des commerces indépendants. Plus précisément, l'ensemble retenu se fonde sur des algorithmes de scan spatial pour étudier des variations locales dans la répartition des commerces de réseau. Nous avons retenu la classification spatiale hiérarchique, extension à l'échelle locale des méthodes du voisin le plus proche.

2. Les centralités commerciales dans l'unité urbaine de Paris : Dimensions et agrégats

Cerner la centralité commerciale : rayon de recherche et seuil d'agrégation

Deux paramètres entrent en compte pour calibrer la classification spatiale ascendante hiérarchique, le rayon de recherche de l'agrégat (ici en mètres) et la taille minimale des agrégats (en nombre de commerces pris en compte). Pour identifier les centralités, nous avons effectué une étape de calibrage sur une série de 47 réalisations en faisant varier chacun des deux paramètres. L'analyse des caractéristiques des différentes configurations obtenues (nombre d'agrégats / Taille moyenne des agrégats), a permis d'évaluer l'influence respective de chacun des paramètres. Plutôt que d'identifier une configuration optimale, l'analyse de l'évolution des structures permet de comprendre l'organisation parfois multi-scalaire des différentes centralités.

Faire des centralités imbriquées

A partir des variations des deux paramètres, en complément de cette approche quantitative, nous avons pris plusieurs zones tests, reflétant de manière thématique la diversité de l'offre commerciale et l'imbrication des centralités. C'est tout l'intérêt de la méthode retenue.

Dans Paris, trois configurations commerciales spécifiques ont été retenus et seront présentés sur le poster: l'ensemble commercial constitué par le Boulevard Haussmann et les rues adjacentes, l'ensemble constitué par la Rue de Rivoli et le Forum des Halles, et enfin la Rue du Commerce, située dans le 15^{ème} arrondissement de Paris.

Dans l'unité urbaine de Paris hors Paris, trois ensembles ont été retenus pour approcher la diversité des centralités commerciales franciliennes. Le premier exemple se penche sur une centralité commerciale organisée autour d'une avenue très commerçante (l'avenue Jean Jaurès, à Boulogne-Billancourt), le second sur un centre-ville (Saint-Germain-en-Laye) et enfin le dernier sur un centre commercial (Centre commercial Carrefour Claye-Souilly).

Une fois le nouveau découpage effectué, avec 176 centralités commerciales construites, nous avons analysé les masses commerciales, à travers une analyse des spécificités des densités.

Hiérarchie et polarisations dans l'unité urbaine de Paris

A l'échelle de l'unité urbaine, les centralités commerciales majeures (grands ensembles commerciaux du centre de Paris et centres commerciaux régionaux), participent à l'organisation générale de l'unité urbaine de Paris, de grands centres commerciaux faits pour la voiture s'étant développés en périphérie. Ces centralités polarisent les clients et les achats, et la concurrence des centralités commerciales de périphérie sur les espaces commerciaux centraux augmente, sur des biens de plus en plus rares. Ces centralités sont considérées certes comme des lieux de consommation mais aussi comme des lieux de divertissement et de tourisme. Entre les deux grandes centralités précédemment décrites apparaît un creux, correspondant à la première couronne. Des centralités plus locales s'y développent, dans une logique relevant souvent de la proximité. Ce sont ces espaces en situation d'intermédiation qui sont aujourd'hui l'objet de l'attention des acteurs du commerce.

En se réimplantant dans les interstices urbains, ces derniers participent aussi à l'enrayement du déclin des espaces commerçants.

Bibliographie:

[BER 71] BERRY B.J.L., *Géographie des marchés et du commerce de détail*, Editions Armand Colin, Paris, 1971

[FLE 07] FLEURY A., MATHIAN H., SAINT-JULIEN T., « Les centralités commerciales », *Regards sur les quartiers parisiens. Contextes spatiaux, usages politiques et pratiques citadines*, étude pour le compte de la Ville de Paris, UMR Géographie-cités, Paris, 2007

[HOT 29] HOTELLING H., « Stability in compétition », *The economic journal*, Vol.39, No 153, pp 41-57.

[LÖS 54] LÖSCH A., *The economics of location*, Yale University Press, 1954.

Fouille de données spatiales pour la caractérisation de paysages en lien avec des fonctionnalités agro-écologiques

Thomas Guyet

AGROCAMPUS-OUEST

65 rue de Saint Brieuc, CS 84215, F-35 042 Rennes Cedex, France

IRISA

Campus de Beaulieu, F-35 042 Rennes Cedex, France

thomas.guyet@agrocampus-ouest.fr

1. Introduction

De plus en plus de recherches en agronomie s'intéressent à comprendre les processus agro-écologiques en interaction avec un paysage. Mais les agronomes ne disposent généralement pas d'une variété importante de paysages réels pour identifier la façon dont les caractéristiques de paysage agissent sur un processus agro-écologique. La simulation de paysages virtuels, réalistes et paramétrables doit palier ce manque de données réelles. Ces paysages virtuels serviront de support à la simulation du processus étudié pour en comprendre les interactions. De sorte à tester leurs hypothèses, ces paysages doivent être paramétrables. Pour le reste, les agronomes cherchent à disposer de paysages réalistes, c'est-à-dire des paysages dont certaines caractéristiques sont comparables à celles rencontrées dans la nature.

Notre travail s'intéresse à l'extraction automatique de caractéristiques de paysages agricoles réels en vue de la simulation de paysages virtuels réalistes. Nous définissons ici un paysage agricole comme l'organisation spatiale d'éléments paysagés naturels (*e.g.* parcelles agricoles, haies, plans d'eau) ou artificiels (*e.g.* bâtiments, routes) avec lequel interagissent des acteurs aussi divers que des agriculteurs, des insectes, la faune, la végétation naturelle et cultivée, etc.

Nous proposons, d'une part, de représenter un paysage sous la forme d'un graphe de relations fonctionnelles qui exprime plusieurs visions de l'organisation spatiale du paysage et, d'autre part, d'utiliser des méthodes de fouille de données sur le graphe pour extraire des motifs locaux caractéristiques de l'organisation du paysage. La méthode générale de caractérisation automatique d'un paysage par des motifs locaux comprend trois étapes (*cf.* Figure 1) :

- 1) la transformation des éléments du paysage en un ensemble d'objets catégoriels,

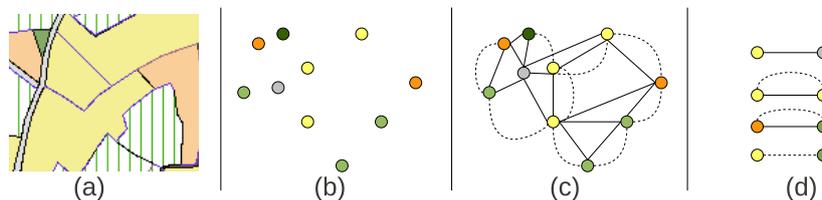


Figure 1. (a) paysage réel dans un SIG, (b) représentation des éléments paysagés (un point est une parcelle et sa couleur donne l'occupation du sol, qui est sa catégorie), (c) représentation de deux relations fonctionnelles entre les éléments paysagés : en pointillés, la proximité « à vol d'oiseaux » entre centres de parcelle et, en trait plein, l'adjacence entre parcelles, (d) motifs locaux extraits avec un support de 2.

2) la construction d'un graphe des relations fonctionnelles (spatiales) entre les éléments du paysage,

3) l'extraction de motifs locaux intéressants dans le graphe décrivant le paysage.

Là où GenExp [LEB 09] caractérise un paysage par un modèle global de statistique spatiale, *i.e.* le modèle de la distribution spatiale des centres des parcelles, nous proposons de caractériser un paysage à l'aide de motifs spatiaux locaux. Un motif spatial exprime une règle d'organisation d'éléments paysagers, par exemple « les grandes parcelles cultivées touchent une route goudronnée » ou « une parcelle de prairie est fréquemment *proche* d'une autre parcelle de prairie », etc.

2. Représentation d'un paysage par un graphe de relations fonctionnelles

Un paysage est décomposé en éléments paysagés transformés en un ensemble d'objets catégoriels. C'est-à-dire qu'un élément est un objet associé à une catégorie et à des attributs. Pour des paysages agricoles, les grandes catégories intéressantes sont les parcelles agricoles caractérisées par leur occupation du sol et les autres surfaces structurant le paysage : routes, haies, surfaces bâties, etc. Il est ensuite possible d'affiner ces catégories, par expertise ou en utilisant des méthodes automatiques de catégorisation. Il peut être intéressant de séparer les « grandes parcelles de blé » et les « petites parcelles de blés » et de créer deux catégories spécifiques. L'organisation des catégories en taxonomie permet de conserver l'information que ces deux catégories sont des sous-catégories de « parcelle de blés » et de « parcelle agricole ».

Le rôle des relations fonctionnelles est d'exprimer les liens entre les éléments du paysage, et en particulier les relations spatiales. Pour représenter l'organisation spatiale des éléments paysagés, une relation de *proximité* relie deux parcelles dans le graphe si, par exemple, la distance « à vol d'oiseau » entre leurs centres est inférieure à un seuil fixé *a priori*. La relation fonctionnelle est choisie par les agronomes en fonction des caractéristiques spatiales à mettre en avant pour caractériser un paysage. La vision de l'organisation spatiale du paysage dépend en partie des phénomènes observés : si on s'intéresse à la répartition des cultures sur une exploitation, les relations

entre parcelles doivent tenir compte de la distance par les routes et chemins. En revanche, si on s'intéresse à des flux de pollens, cette distance n'a plus d'intérêt (même si la présence de route entre deux parcelles est une information importante). De plus, des relations autres que spatiale peuvent également avoir un intérêt pour représenter le paysage (par exemple, une relation indiquant si deux parcelles appartiennent à un même exploitant).

Le représentation du paysage sous la forme d'un graphe se veut neutre. La sémantique est portée par les catégories et les types de relations qui peuvent être exprimées entre parcelles. De plus, la représentation permet d'utiliser plusieurs distances fonctionnelles dans une même représentation du paysage (*cf.* Figure 1).

3. Caractérisation automatique d'un paysage par la fouille de données spatiales

On analyse le graphe du paysage par une méthode de fouille de données inspirée de la recherche de colocalisations spatiales [HUA 04]. Cette méthode extrait des ensembles de catégories qui sont fréquemment liées deux à deux par une relation spatiale. Les motifs locaux qui sont ainsi extraits sont les ensembles maximaux de catégories dont le ratio de présence dans le paysage est supérieur à un seuil fixé. Pour plus de détails, Flouvat et *al.* [FLO 10] présentent la recherche de colocalisations appliquée à la caractérisation des zones érodées de Nouvelle-Calédonie.

Nous avons adapté la méthode pour extraire des motifs avec différentes relations fonctionnelles à partir d'un graphe en adaptant l'algorithme APriori. Cela permet d'extraire des motifs comme des ensembles maximaux de catégories liés par des relations fonctionnelles spécifiques (et pas uniquement des colocalisations).

Ce travail est en cours de réalisation dans le cadre du projet PayOTE (INRA-INRIA). Un de ses objectifs est d'explorer la caractérisation des paysages au moyen de ces motifs locaux. En particulier, nous souhaitons vérifier que l'utilisation de motifs locaux permettra de retrouver des propriétés globales d'un paysage réel dans des paysages simulés. La méthode est actuellement expérimentée pour caractériser le paysage de la zone atelier de Pleine-Fougères, en Bretagne, dans laquelle se distinguent un paysage bocagé et un paysage de grandes plaines cultivées.

4. Bibliographie

- [FLO 10] FLOUVAT F., SELMAOUI-FOLCHER N., GAY D., ROUET I., GRISON C., « Constrained colocation mining : application to soil erosion characterization », *Proceedings of the Symposium on Applied Computing*, 2010, p. 1054-1059.
- [HUA 04] HUANG Y., SHEKHAR S., XIONG H., « Discovering colocation patterns from spatial datasets : a general approach », *IEEE Transactions on Knowledge and Data Engineering*, vol. 6, n° 12, 2004, p. 1472-1485.
- [LEB 09] LE BER F., LAVIGNE C., ADAMCZYK K., ANGEVIN F., COLBACH N., MARI J.-F., MONOD H., « Neutral modelling of agricultural landscapes by tessellation methods—Application for gene flow simulation. », *Ecological Modelling*, vol. 220, 2009, p. 3536-3545.

Infrastructure de Données Spatiales et ingénierie territoriale : une approche empirique à partir du SIG Pyrénées.

LENORMAND Pauline*, — PLAINECASSAGNE Laurent***

**APEM, Assemblée Pyrénéenne d'Economie Montagnarde
22 Avenue des peupliers 31320 Castanet-Tolosan
lenormand@apem.asso.fr
lp@apem.asso.fr*

*** Laboratoire Dynamiques Rurales, Université de Toulouse II Le Mirail
5 allées Antonio Machado 31058 TOULOUSE Cedex 9*

RÉSUMÉ. L'Infrastructure de données spatiales (IDS) développée par l'APEM et support de l'observatoire socio-économique du massif pyrénéen a permis de consolider une démarche collective au service du développement économique montagnard.

MOTS-CLÉS : Infrastructure de Données Spatiales, observatoire, communauté de pratiques, ingénierie territoriale

1. Introduction

A partir d'un exemple d'observatoire basé sur une IDS, nous montrerons les modalités de construction de l'information géographique au service des politiques territoriales pyrénéennes. L'APEM, Assemblée Pyrénéenne d'Economie Montagnarde¹ est une association regroupant les interconsulaires du massif pyrénéen et a pour source de financement des fonds européens, nationaux, régionaux et des cotisations de ses membres. Le travail s'appuie sur une double démarche d'entretiens auprès de personnes ressources et d'observation participante, liée à un travail de recherche (bourse CIFRE).²

2. Une démarche collective au service du développement économique montagnard

2.1. Objectifs

La principale mission de l'APEM est de faire vivre l'observatoire socio-économique du massif pyrénéen, le SIG Pyrénées³, au service du développement montagnard. L'APEM a construit sa compétence sur des outils logiciels libres, en répondant aux besoins de ses différents partenaires composant l'ingénierie territoriale (Lardon et al., 2007) pyrénéenne. Pour cela, l'APEM a développé une méthodologie qui a permis d'assembler des outils et qui favorise l'animation de communautés d'acteurs autour du partage et de la construction de connaissances.

2.2. Difficultés

Les principaux freins rencontrés par l'association sont :

- La reconnaissance d'une échelle d'action interrégionale face à la résistance de logiques administratives.
- Une réticence à partager les données (un blocage socio-cognitif encore très présent, Noucher 2009).

2.3. Choix

Pour répondre à la politique du Massif des Pyrénées, l'observatoire est organisé en thématiques (agropastoralisme, climat, entreprise-artisanat, forêt, ...) correspondant aux principaux enjeux de la Convention de Massif 2007/2013. Ainsi nous avons créé des services informationnels dédiés à la CIMP. Pour surmonter la réticence à

¹ <http://apem.asso.fr/>

² Remarque : la communication R&D 69 centrée sur le même outil (réalisée par Thomas Portier) détaillera plus précisément les aspects techniques de l'Infrastructure de Données Spatiales.

³ <http://www.sig-pyrenees.net/>

partager les données, des interfaces personnalisées et une gestion de profils utilisateurs garantissent le respect de la confidentialité. Pour favoriser des dynamiques de travail collaboratives, un espace de travail numérique (armoire à documents, forums, annuaire en ligne) est mis à disposition de tous les acteurs pyrénéens. Et enfin, la mise en place d'une IDS transfrontalière est le dernier choix stratégique de l'APEM pour rendre cet observatoire plus attractif et plus pertinent.

2.4. Les conditions de la réussite

Il ressort de l'expérience de l'APEM que la réussite de ce type de démarche est conditionnée par une conduite conjointe de l'accompagnement technique et thématique du changement, afin notamment que les services informationnels proposés soit adaptés aux enjeux territoriaux. La séparation entre IDS et services informationnels permet son utilisation par un grand nombre d'acteurs.

3. Des innovations au service de l'ingénierie territoriale pyrénéenne

3.1. Des logiciels libres

EasySDI⁴ est un logiciel libre, prêt à l'emploi pour déployer une infrastructure de données spatiales, interopérable et basée sur les standards ISO/OGC. Son objectif, en répondant à la directive INSPIRE, est de permettre à toute organisation désireuse de diffuser de l'information géographique de le faire de manière sécurisée. Ce logiciel est utilisé par l'APEM pour illustrer l'observatoire socio-économique du massif et pour proposer des services liés à l'information géographique à ses partenaires, l'objectif étant de s'en servir comme support d'un réseau de territoires pyrénéens. L'IDS est intéressante pour l'ingénierie territoriale puisqu'elle regroupe dans un lieu unique un ensemble de données normalisées et indique comment aller les chercher de manière sécurisée.

3.2. Au service d'une co-construction de l'information

L'APEM met en place une démarche ayant pour objectif « la mise à disposition d'applications logicielles sur Internet, destinée aux acteurs et territoires souhaitant développer l'usage des TIC dans la valorisation de leurs activités »⁵. L'originalité relève de la volonté de s'inscrire uniquement dans des démarches collectives, initiées et consolidées au sein de communautés de pratique (au sens de Wenger 1991). Par cette démarche, l'APEM espère permettre aux structures territoriales de disposer d'outils informatiques performants à moindre coût, et d'assurer collectivement la maintenance évolutive des outils.

⁴ <http://www.easysdi.org/>

⁵ <http://www.boosterr.fr/>

4. Quels apports scientifiques et méthodologiques ?

4.1. *Une équipe pluridisciplinaire*

Pour mettre en place cette co-construction de l'information et ses outils supports, l'APEM s'est dotée progressivement d'une équipe composée de salariés avec des profils assez variés, permettant un agencement original de compétences (Lenormand, Barthe, 2010). On peut donc parler de construction d'une compétence collective car l'APEM a su développer une double capacité d'interprétation des enjeux du développement montagnard et de mobilisation d'outils pour rendre utilisable l'information géographique par le plus grand nombre.

4.2. *Un outil et un acteur intermédiaire*

L'observatoire du massif est un outil pour quantifier et qualifier l'espace pyrénéen, construit en s'appuyant sur les « connaissances métiers » via le réseau des acteurs économiques et par la spatialisation des enjeux. Il contribue donc à la mise en place d'une base de connaissances partagées au service du développement du Massif. L'APEM est un intermédiaire entre les outils et les territoires. L'association endosse un rôle de généraliste de la médiation (Nay O, Smith A., 2002) et de passeur, qui mobilise l'information, géographique et statistique, comme vecteur de construction d'une dynamique collective qui reste aujourd'hui encore fragmentée.

4.3 *Un croisement nécessaire / un double champ à investiguer*

Comme le montre Ertz (2009), de nombreuses évolutions techniques sont encore nécessaires pour aider la production, le partage et la diffusion de connaissances sous forme de « produits cartographiques ». Nous avons vu que l'APEM sait se positionner dans ces enjeux de recherche-développement. Les recherches actuelles sur l'ingénierie territoriale (Piveteau, 2010) gagneraient selon nous à prendre en compte les mécanismes qui précèdent à l'utilisation de ces technologies dans l'accompagnement de l'action territoriale.

5. Bibliographie

Chanal V., Communautés de pratique et management par projet : A propos de l'ouvrage de Wenger (1998) *Communities of Practice: Learning, Meaning and Identity*, *M@n@gement*, Vol. 3, No. 1, 2000, p.1-30.

Ertz O., « Collaborative Geospatial Knowledge Production », *Bulletin e-geo.ch*, 23-7/2009, p.4-6.

Lardon S, Poss Y, Moquay P., *Développement territorial et diagnostic prospectif*, Editions de l'Aube, 2007, 384p.

Lenormand P., Barthe L., "The role of development agents in territorial observatories : lessons to be learned from the "SIG Pyrénées" experience", communication au colloque ITGO, Rennes, 2010, 9p.

Noucher M., *La donnée géographique aux frontières des organisations: approche socio-cognitive et systémique de son appropriation*, thèse de doctorat soutenue le 23 avril 2009, Ecole Polytechnique Fédérale de Lausanne, 225p.

Piveteau V., « L'ingénierie territoriale : pour un parti pris géographique », Essai inédit présenté en vue de l'habilitation à diriger des recherches, ENS Lyon, 2010, 114p.

Les trajectoires des véhicules d'urgences comme outil de capitalisation d'itinéraires

Bilel Soussi* — **Jean Michel Follin**** *Laboratoire de géodésie et de géomatique (L2G) 1 boulevard Pythagore - 72000 le Mans – France*
*mohamed.soussi@esgt.cnam.fr - **jmfollin@esgt.cnam.fr

RÉSUMÉ. De nos jours, de nombreux chercheurs s'intéressent aux bases de données d'objets mobiles. Plusieurs travaux portent sur l'exploitation des données issues d'objets mobiles tout le long de leurs trajets. Nous formulons l'hypothèse que l'historique des trajectoires des véhicules d'urgence du SMUR peuvent fournir des informations exploitables lors de la recherche d'un meilleur itinéraire pour une intervention. Dans ce poster, nous proposons une méthodologie et un formalisme pour une telle approche et nous présentons nos premières réalisations.

ABSTRACT. Nowadays, many researchers are interested in the moving object databases. Several works are concerned with the utilization of data resulting from mobile objects all along their trips. We formulate the assumption that the historical trajectory of French emergency units (SMUR) can provide useful information for the research of better intervention routes. In this poster, we propose a methodology, formalism for such an approach and we present our first results.

MOTS-CLÉS : Trajectoire, appariement, algorithme de recherche d'itinéraires, base de données spatio-temporelle.

KEYWORDS: Trajectory, map matching, routing, spatiotemporal databases

La détermination du meilleur itinéraire est un facteur primordial pour le bon fonctionnement des SMUR (Service Mobile d'Urgence et de Réanimation). Pour se rendre rapidement sur une zone d'intervention, les ambulanciers du SMUR se basent généralement sur leurs connaissances du terrain. Ils utilisent leur propre expérience complétée par la consultation de cartes routières à plusieurs échelles comme outils d'aide à la détermination du meilleur itinéraire.

Dans le but de faciliter la tâche des SMUR lors de la recherche du meilleur itinéraire pour une intervention, notre travail consiste à développer une application d'aide à la navigation basée sur l'exploitation des historiques de trajectoires des véhicules.

Nous envisageons quatre étapes : la création d'une base de données de trajectoires, l'appariement des trajectoires au réseau routier, l'enrichissement du réseau routier à partir du traitement statistique d'observations réelles et la mise au point d'un algorithme temporel de recherche d'itinéraires.

1. Création d'une base de données des trajectoires brutes

Dans cette étape nous exploitons les trames GPS issues de récepteurs équipant plusieurs véhicules du SMUR du Mans. On distingue deux modèles de trajectoire :

– Modèle libre : un objet en mouvement se déplace librement dans l'espace (Triki *et al.*, 2010).

– Modèle contraint : il existe une relation entre la trajectoire et le réseau physique (par exemple une route). La proposition la plus complète d'un tel modèle est celle exposée dans (Guting *et al.*, 2006).

Nous adoptons un modèle libre historique modifié : la trajectoire est exprimée en fonction des positions absolues en 2D tout en gardant la relation avec le réseau physique par le biais de l'appariement. Par ailleurs les trajectoires sont caractérisées par des données attributaires telles que l'identifiant de l'ambulancier, le type de l'accident et le véhicule utilisé.

2. Appariement avec la base de données routière de l'IGN

Ce que l'on vise avec le processus d'appariement est de déterminer pour chaque tronçon de trajectoire le tronçon de route qui est suivi.

Pour appairer une trajectoire à des tronçons routiers nous avons adopté une approche globale plutôt qu'une approche incrémentale (Brakatsoulas *et al.*, 2005) en utilisant le plugin libre RoadMatcher¹. Il se base sur la conversion du réseau routier et de la trajectoire en graphes et la mesure de la distance de hausdorff entre chaque arc.

3. Enrichissement de la base de données routière par des données temporelles issues des trajectoires telles que la vitesse variant en fonction du moment

La notion de profil spatial de vitesse a été définie dans (Sandu-Popa *et al.*, 2009) afin d'analyser le comportement des conducteurs. Dans ce travail la vitesse est exprimée en fonction de la position relative par rapport au réseau physique (avec le constructeur gmoving). Dans nos travaux, nous visons à exprimer la vitesse moyenne variant dans le temps pour chaque tronçon routier.

En effet, on cherche à déterminer les différents profils de vitesse moyenne pour une journée type bien définie (par exemple une journée de vacances scolaires) pour une période de temps bien définie (par exemple la tranche horaire de 17h à 21h) pour chaque tronçon du réseau routier.

1. Plugin d'appariement libre pour openjump développé par l'entreprise vivid solutions.

On définit donc la notion de profil de vitesses par journée type en se basant sur le formalisme défini dans (Guting *et al.*, 2006) qui permet de prendre en compte la variabilité temporelle de valeurs (avec le constructeur moving).

La construction des profils de vitesses variant dans le temps va nécessiter un nombre suffisant d'observations afin qu'ils soient considérés comme valides. Elle est réalisable en trois étapes :

- Pour chaque observation sur un tronçon on détermine le profil de vitesse pour un intervalle $[h_e, h_s]$ avec h_e représentant l'heure où le véhicule est entré sur le tronçon et h_s celle à laquelle il en est sorti.

- On calcule la moyenne de profils de vitesses m_p pour chaque observation i , ce qui nous donne un triplet $m_p(tid, v_i, h_i)$ avec tid représente l'identifiant du tronçon, v_i désigne la vitesse moyenne du profil, et h_i renseigne sur l'heure h de l'observation (avec $h = h_e$).

- En se basant sur des approches statistiques on cherche à créer un diagramme de classes des vitesses en fonction du temps. Chaque classe regroupe un ensemble de moyennes de profils de vitesses, et est représenté par un couple $C_i ([t_i, t_{i+1}[, v_i)$ avec $[t_i, t_{i+1}[$ désigne la tranche horaire et v_i représente la vitesse moyenne de toutes les observations dans cette classe (avec élimination des valeurs aberrantes).

4. Détermination d'un meilleur itinéraire temporel

En s'inspirant des idées présentées dans (Chon *et al.*, 2003), on vise à mettre en œuvre un algorithme temporel de calcul d'itinéraires qui se basera essentiellement sur des observations réelles intégrant des spécificités liées au contexte d'urgence.

5. Bibliographie

- Brakatsoulas S., Pfoser D., Salas R., Wenk C « On Map-Matching Vehicle Tracking Data », *VLDB Conference* 2005, p.853-864.
- Chon H.D., Agrawal D., El Abbadi A., « FATES: Finding A Time dEpendent Shortest path » *Springer Berlin Heidelberg*, vol. 2574, 2003, p.165-180.
- Guting R.H., Almeida V.T., Ding Z., « modeling and quering moving objects in networks *VLDB Journal*», vol.15, 2006, p. 165-190.
- Sandu-Popa I., Kharrat A., Zeitouni K., Saint-Pierre G., « Base de données de capteurs à localisation mobile – Modèle et langage, dans *Systèmes d'Information et géolocalisation* », *RSTI – Série ISI (Ingénierie des Systèmes d'Information)*, Editions *Hermès-Lavoisier*, vol. 14, n 5, 2009.
- Triki D.Z., Frihida A., Ben Ghezala H., Claramunt C., « Modèle et langage pour la manipulation de trajectoires spatio-temporelles », *Revue internationale de géomatique*, vol 20, 2010, p 37-64.

Vers un Profil UML pour les Entrepôts de Données Intégrant les Réseaux Spatiaux

Mouna SNOUSSI¹, Sandro BIMONTE¹, Myoung-Ah KANG²
snoussi_mouna@yahoo.fr, sandro.bimonte@cemagref.fr, kang@isima.fr

¹Cemagref – Clermont-Ferrand, France

²LIMOS, UMR 6158 CNRS – Université Blaise Pascal, France

RÉSUMÉ. Les entrepôts de données spatiales et les systèmes SOLAP permettent de stocker et d'analyser les objets spatiaux au sein d'un entrepôt de données. Afin d'enrichir l'analyse SOLAP avec les méthodes d'analyse des réseaux spatiaux (i.e. réseaux routiers, etc.) nous proposons un profil UML pour représenter les relations de connectivité entre les mesures et les membres spatiaux.

ABSTRACT. Spatial data warehouses and SOLAP systems allow the storage and multidimensional analysis of huge volume of spatial data. In order to enrich SOLAP analysis with analysis methods of spatial networks (i.e. roads network, etc.), in this paper we propose a UML profile to represent connectivity relationships between spatial members and spatial measures.

MOTS-CLÉS: Entrepôt de données spatiales, Réseau, Profil UML.

KEYWORDS: Spatial Data Warehouse, Network, UML Profile.

1. Introduction et motivations

L'intégration des données spatiales dans les entrepôts a donné naissance à des nouveaux systèmes d'analyse multidimensionnelle nommés Entrepôts de Données Spatiales (EDSs) et Spatial OLAP (SOLAP). Les EDSs étendent les concepts principaux du modèle multidimensionnel pour définir dimension et mesure spatiales (Malinowski et al., 2007). Une *dimension spatiale* est une dimension qui contient au moins une composante spatiale (géométrie) dans un de ses niveaux. Une *mesure spatiale* est une collection d'objets spatiaux (géométries) et/ou le résultat d'une opération spatiale. Des modèles conceptuels pour les EDSs ont été proposés : (Malinowski et al., 2007) se base sur le modèle conceptuel MADS, et (Glorio et al., 2008) s'appuie sur le langage UML pour permettre le développement d'une architecture MDA. Un aspect important qui caractérise les objets spatiaux, et qui a été négligé par les travaux concernant la modélisation conceptuelle des EDSs, est la connectivité entre les objets (Stemliuc et al., 2009) dans un réseau des objets spatiaux (ex. réseau routier...). D'où l'intérêt de notre travail qui consiste à intégrer les concepts des réseaux spatiaux au sein des EDSs afin de représenter la connectivité entre les objets spatiaux pour permettre aux décideurs d'enrichir les requêtes SOLAP avec les méthodes d'analyses des réseaux spatiaux (i.e. le calcul du chemin le plus court, etc.). Pour ce faire nous proposons un profil UML qui intègre et adapte les modèles pour EDSs et réseaux spatiaux présentés

respectivement en (Glorio et al., 2008) et (Stempliuc et al., 2009). Pour présenter notre profil UML, nous montrons le meta-modèle, et un exemple de contrainte OCL et les cônes pour les stéréotypes. Nous avons implémenté le profile en utilisant Oracle 10g car il fournit un support natif pour les réseaux spatiaux.

2. Profil UML pour les entrepôts de réseaux spatiaux

Afin d'intégrer les réseaux spatiaux dans les EDSs en tant que dimension et mesure, nous étendons le modèle spatio-multidimensionnel pour définir les nouveaux concepts de *Dimension Réseau*, *Niveau Réseau* et *Fait Réseau*. En particulier, une *Dimension Réseau* est une dimension spatiale (Glorio et al., 2008) qui contient au moins un niveau réseau. Un *Niveau Réseau* est un niveau spatial dont les instances forment un réseau composé de membres et des éventuelles relations topologiques entre eux, comme défini en (Stempliuc et al., 2009). Un *Fait Réseau* est un fait (Glorio et al., 2008) associé à un réseau (Stempliuc et al., 2009). Nous avons implémenté ces concepts dans un profil UML (contraintes OCL¹, éléments de bases et icones). Notre profil est basé sur le méta-modèle présenté en Figure 1. Il présente les classes du modèle spatio-multidimensionnel présenté en (Glorio et al., 2008) : *Fait* , *Dimension* et *Base*  (niveau d'une dimension) avec ses extensions spatiales : *SpatialLevel* qui représente un niveau spatial, et *GeometricTypes* qui énumère les différents types de géométries. Pour la prise en compte des réseaux spatiaux nous utilisons les classes *Nœud*  et *Arc*  proposées par (Stempliuc et al., 2009) aux quelles nous avons inséré l'attribut géométrique de type *GeometricTypes*. Les nouvelles classes introduites par notre méta-modèle sont les suivantes : *Dimension_Réseau* et *Fait_réseau* . La classe *Dimension_Réseau* est associée à un niveau réseau (*Base_Réseau_nœud* /  ou *Base_Réseau_arc* / ). Les liens entre les membres d'un niveau réseau sont représentés par *Nœud_spatial* ou *Arc_spatial*. Pour la définition formelle du profil UML, nous avons défini des contraintes OCL qui décrivent le méta-modèle et un ensemble de contraintes supplémentaires. Un exemple est : "*Un Fait_Réseau ne peut être associé qu'à des classes de type Dimension ou Dimension_Réseau*"; sa définition en OCL est : "*Self.allOppositeAssociationEnds -> forAll (ae | ae.participant.oclsTypeOf (Dimension) or ae.participant.oclsTypeOf (Dimension_Réseau)* ". Un exemple de modèle pour l'analyse des accidents des réseaux routiers avec notre profil est montré en Figure 2. Ce modèle permet de répondre à des requêtes comme : "*Quel est le nombre d'accident pour chaque route (Route_Ville) entre la ville Lyon et la ville Clermont-Ferrand, et pour chaque mois ?*". Ce modèle présente un *niveau réseau* représentant le réseau routier en modélisant explicitement les liens entre les routes et les villes. Cela permet de répondre à la requête de notre exemple en retrouvant toutes les routes entre Lyon et Clermont-Ferrand, grâce à un algorithme de calcul de chemin.

¹ OCL (Object Constraint Language) est le langage formel d'expression des contraintes utilisé dans UML (<http://www.uml.org/> UML 2.0 Object Constraint Language (OCL) Object Management Group).

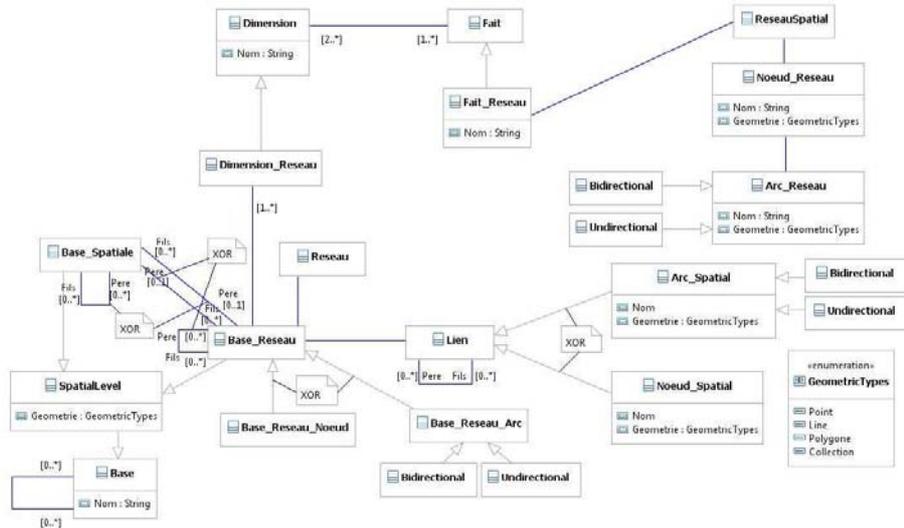


Figure 1. Meta-modèle des entrepôts de données intégrant les réseaux spatiaux.

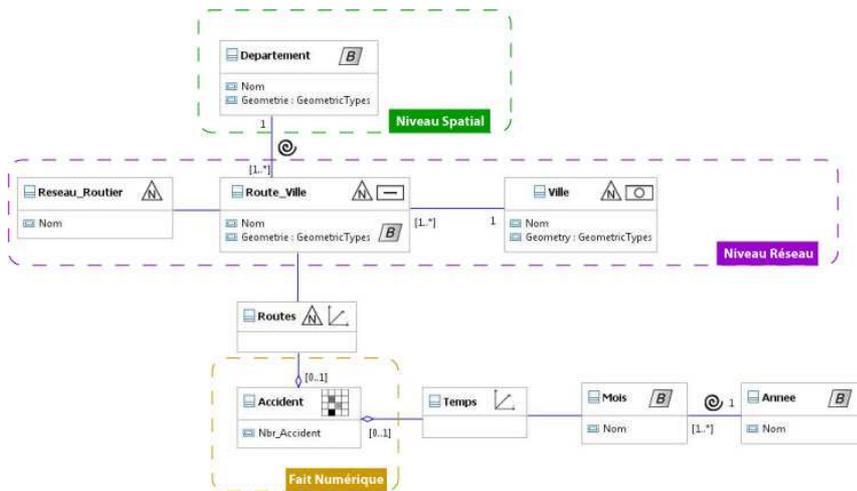


Figure 2. ED de réseaux spatiaux pour l'analyse des accidents routiers.

Références

- Glorio O., Trujillo J., « An MDA Approach for the development of Spatial Data Warehouses », *DaWaK*, Turin, Italy, 2008, p. 23-32,
- Malinowski E., Zimany E., « Logical Representation of a Conceptual Model for Spatial Data Warehouses », *GeoInformatica*, vol. 11, n° 4, 2007, p. 431-457
- Stempliuc S., Filho J., Andrade M., « Extending the UML-GEOFRAME data model for conceptual modelling of network applications », *ICEIS*, Italy, 2009, p. 164-170,

Etude de la redondance des réseaux pour évaluer leur capacité à fonctionner en mode dégradé

LHOMME Serge*** — SERRE Damien* — DIAB Youssef* — LAGANIER Richard**

**Université Paris Est - Ecole des ingénieurs de la ville de paris (EIVP)
15, rue Fénelon
75010 Paris
serge.lhomme@eivp-paris.fr*

*** Université Paris Diderot - Paris 7- UMR PRODIG
F-75205 Paris Cedex 13*

RÉSUMÉ. La résilience est un concept émergent en gestion des risques. Afin d'évaluer la résilience urbaine face aux risques d'inondation, la mise en place d'indicateurs de redondance des réseaux techniques urbains semble nécessaire. Pour cela, cette recherche se place dans la continuité des recherches concernant la vulnérabilité structurelle des réseaux.

ABSTRACT. Resilience concept is relatively new concerning flood risk management. In order to assess this concept, some redundancy indicators are needed especially for urban networks. In order to define these redundancy indicators, our study focuses on methods and tools use for assessing networks structural vulnerability.

MOTS-CLÉS: Résilience, théorie des graphes, inondation, SIG, analyse spatiale

KEYWORDS: Resilience, graph theory, flood, GIS, spatial analysis

Introduction

La tempête « Xynthia » qui a traversé la France les 27 et 28 Février 2010 a tué 53 personnes. La majorité des décès sont consécutifs à la rupture d'ouvrages de protection et cette catastrophe met en exergue les problématiques liées à l'urbanisation en zones inondables. Or, le concept de résilience, concept pluridisciplinaire et polysémique, étudié, en gestion des risques, les possibilités de concilier urbanisation et gestion des risques [LHO 2010]. C'est pourquoi, la résilience peut apparaître comme une solution alternative au tout ou rien : acceptation des constructions dans certaines zones et déclaration de zones inconstructibles dans des zones voisines. Cependant, pour que ce concept soit pleinement opérationnel, les décideurs ont besoin d'outils leur permettant d'évaluer leur politique de résilience. Pour cela, nous nous concentrons sur l'étude de la résilience des infrastructures linéaires, comme les réseaux.

Des recherches dans la continuité des études concernant la vulnérabilité structurelle des réseaux de transport

S'il n'existe pas actuellement d'outil, ni même de méthodologie, pour évaluer la résilience des réseaux, il existe une littérature importante concernant les infrastructures réseaux, dont quelques unes traitent de leur vulnérabilité. Or, il est possible d'étudier ces deux concepts dans un même continuum [PRO 2009]. Ainsi, ces recherches se placent dans la continuité des études concernant la vulnérabilité structurelle des réseaux de transport, qui ont recours à des indicateurs caractérisant la structure des réseaux [GLE 2005]. Cependant, la résilience se focalise sur la continuité possible du service suite au dysfonctionnement d'un ou de plusieurs de ses composants, autrement dit son fonctionnement en mode dégradé. C'est ce que l'on nomme les capacités d'absorption du réseau [LHO 2010]. Les problématiques de connexité sont mises en exergues, et nous cherchons alors à étudier la redondance du réseau pour évaluer ses capacités d'absorption.

La solution la plus intuitive pour étudier la redondance d'un graphe revient à étudier le nombre de chemins indépendants entre deux sommets de ce graphe. En effet, plus ce couple possède de chemins indépendants plus sa redondance sera forte. Des indicateurs comme le degré d'un point ne sont alors pas suffisant pour caractériser la redondance d'un réseau en un point. D'autres indicateurs existent comme le clustering coefficient (coefficient de regroupement ou d'agglomération), qui ne peut différencier les structures faibles, ou le redundancy ratio, qui peut donner des résultats incohérents en certains points (notamment concernant les nœuds pendants). Ces indicateurs sont donc difficilement applicables pour caractériser en un point la redondance du réseau.

C'est pourquoi deux indicateurs ont été définis. Le premier est simplement la moyenne du nombre de chemins indépendants entre ce point et les autres points du graphe, en l'occurrence les voisins des voisins de ce point (Fig. 1). Cette échelle locale, d'un point de vue topologique, permet de caractériser, en partie, la plus ou moins grande redondance d'un graphe en ce point. De plus, la redondance d'un graphe en un point doit tenir compte de la moyenne du nombre d'alternatives entre les points du graphe en cas de suppression de ce point (Fig. 1). Ainsi, ce deuxième indicateur évalue les alternatives possibles à la suppression d'un point. Pour cela, nous proposons, comme pour le premier indicateur, de s'intéresser aux voisins des voisins.

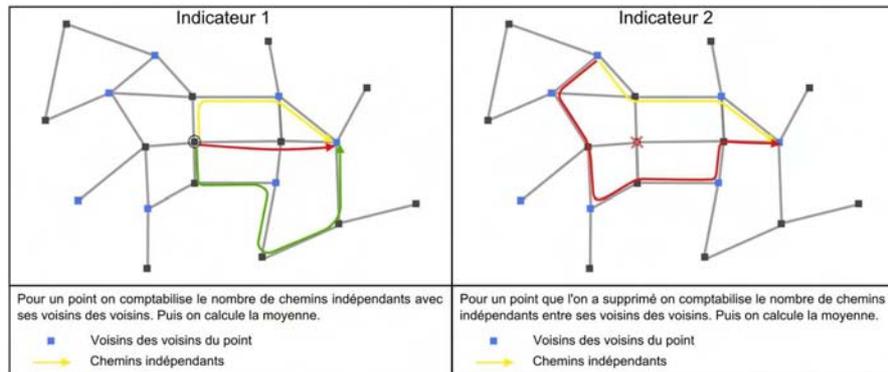


Figure 1. Illustration des deux indicateurs

Perspectives

Les deux indicateurs développés permettent d'évaluer les capacités d'un réseau à fonctionner en mode dégradé. Il est alors possible d'appliquer une méthodologie permettant d'évaluer la résilience urbaine. Pour cela, l'évaluation de la résilience d'un réseau doit tenir compte d'autres indicateurs complémentaires à ces indicateurs de redondance. Ainsi, cette recherche rentre dans un cadre plus large, aboutissant à la réalisation d'un SIG permettant l'évaluation de la résilience urbaine. En effet, les outils d'analyse spatiale développés peuvent être intégrés dans un SIG, afin de constituer un outil d'aide à la décision. Nous disposons aujourd'hui de deux outils SIG permettant l'évaluation des capacités d'absorption d'un réseau (une application locale et une application web qui rentre dans le cadre d'un projet de recherche nommé Flood Resilient City). Ces deux SIG constituent les fondements de la création d'un outil d'évaluation de la résilience urbaine.

Bibliographie

- [LHO 2010] Lhomme S., Serre D., Diab Y., Laganier R., 2010, « Les réseaux techniques face aux inondations ou comment définir des indicateurs de performance de ces réseaux pour évaluer la résilience urbaine », Bulletin de l'association des géographes français, sous presse.
- [PAS 2007] Pasche E., Geisler T. R., 2005, « New strategies of damage reduction in urban areas prone to flood », *Urban flood management*, pp 101-117.
- [PRO 2009] Provitolo D., 2009, « Vulnérabilité et résilience : géométrie variable de deux concepts », séminaire résilience urbaine de l'ENS, Paris.
- [GLE 2005] Gleyse J.-F., 2005, La vulnérabilité structurelle des réseaux de transport dans un contexte de risques, Thèse de doctorat en Analyse Théorique et Epistémologique en Science Géographique, Université Paris VII- Denis Diderot, 848p.

Evaluation de la vulnérabilité des réseaux face aux risques inondation

GRIOT Valère* — **LHOMME Serge***** — **BOURGADE Frédéric*** — **SERRE Damien*** — **DIAB Youssef*** —

**Université Paris Est - Ecole des ingénieurs de la ville de paris (EIVP)
15, rue Fénelon
75010 Paris
serge.lhomme@eivp-paris.fr*

*** Université Paris Diderot - Paris 7- UMR PRODIG
F-75205 Paris Cedex 13*

RÉSUMÉ. La vulnérabilité élémentaire a été définie comme un indicateur de référence pour caractériser la vulnérabilité structurelle d'un réseau de transport. Cependant, son temps de calcul est trop long et son estimation par d'autres indicateurs est difficile. Néanmoins, ces indicateurs permettent une bonne classification des éléments vulnérables. Il est alors possible de créer un outil SIG capable d'implémenter ces indicateurs pour réaliser une classification des éléments vulnérables d'un point de vue structurel.

ABSTRACT. An indicator has been designed in order to assess structural road network vulnerability. Yet, indicator computation process is very long and its estimation thanks to others indicators is difficult. Nevertheless it seems possible to establish vulnerable elements classification thanks to these others indicators. So a GIS solution has been designed for that classification.

MOTS-CLÉS: Vulnérabilité, théorie des graphes, inondation, SIG, analyse spatiale

KEYWORDS: Vulnerability, graph theory, flood, GIS, spatial analysis

Introduction

En France, une commune sur trois est concernée par le risque inondation, pour un coût annuel estimé à environ 250 millions d'euros, soit la première cause d'indemnisation pour les assurances en matière de risques naturels. Par ailleurs, la conjonction de nombreux facteurs comme la croissance urbaine, le réchauffement climatique, l'aversion aux risques de la société, nécessite de repenser la gestion traditionnelle des risques [LHO 2010]. Or, les réseaux techniques urbains comme les réseaux d'eau, d'énergie, de transport, de par leurs structures, leurs contraintes d'implantation et leur nécessité, peuvent être considérés comme des éléments sensibles au risque d'inondation dans la ville.

On relève aujourd'hui de plus en plus de solutions techniques pour se protéger des inondations. Les maisons flottantes, par exemple, font partie du paysage en Asie, aux Etats-Unis (San Francisco), au Canada (Vancouver) et bien sûr au Pays-Bas. Cependant, cette solution, même si elle protège l'homme, ne préserve pas la ville qui ne peut fonctionner sans ses réseaux.

L'objectif est alors de concevoir un outil permettant l'évaluation de la vulnérabilité d'un réseau. Pour cela, nous commençons par présenter les différents supports et outils d'analyse dont nous disposons. Nos algorithmes ont été testés principalement sur des échantillons des réseaux routiers des villes d'Orléans et de Bou, ainsi que sur des réseaux fictifs à mailles carrées ou triangulaires. La suite logicielle utilisée est essentiellement libre: Scilab, Vim, GRASS et Qgis.

Mise en place d'indicateurs permettant l'évaluation de la vulnérabilité des réseaux

Nous mettons en place un certain nombre d'indicateurs afin d'évaluer la vulnérabilité des réseaux. Il en résulte que l'indicateur présentant les résultats les plus satisfaisants (la vulnérabilité élémentaire) ne peut être exploité compte tenu de ses temps de calcul importants. Cet indicateur est classiquement considéré comme une référence puisqu'il compare les états normaux et perturbés d'un réseau [GLE 2005]. Une démarche utilisant la théorie des Systèmes d'Information Décisionnels permet alors de montrer qu'il est possible d'approximer la vulnérabilité élémentaire grâce aux autres indicateurs avec un taux d'erreur très faible. Ce taux étant inférieur à ce que l'on peut obtenir à l'aide de régressions linéaires. Nous dégageons ainsi les indicateurs les plus pertinents pour l'approximation de la vulnérabilité élémentaire. Ces indicateurs sont beaucoup plus rapides à calculer que la vulnérabilité élémentaire.

Malgré ces bonnes corrélations, il est encore impossible de déterminer les valeurs de vulnérabilité élémentaire à partir de ces indicateurs et d'un réseau test servant à

déterminer les valeurs des coefficients de corrélation. Cependant, les corrélations obtenues permettent de pouvoir classifier avec plus ou moins de précision les sommets en fonction de leur vulnérabilité élémentaire.

Nous montrons enfin comment les Systèmes d'Information Géographiques peuvent constituer une aide à la gestion des risques. La finalité de l'analyse menée est la création d'un outil tout SIG sous la forme d'une extension Qgis utilisant Grass de manière totalement transparente. Cet outil peut par exemple calculer les différents indicateurs et appliquer une symbologie adaptée à la classification pour créer une carte lisible.

Conclusion et perspectives

La connaissance des zones inondables ou soumises à un quelconque risque pourrait améliorer cet outil et aider les différents acteurs à localiser très simplement les éléments les plus sensibles d'un réseau. La théorie développée à l'occasion de cette étude pourrait également être étendue aux arcs d'un réseau. Les notions développées au cours de cette recherche peuvent constituer une aide précieuse pour les différents acteurs de la gestion d'un réseau. Cependant, cette approche très algorithmique doit s'accompagner d'un retour critique sur les résultats obtenus s'appuyant sur une connaissance du terrain, afin de valoriser des aspects plus qualitatifs permettant de tenir compte des spécificités du réseau.

Bibliographie

- [LHO 2010] Lhomme S., Serre D., Diab Y., Laganier R., 2010, «Les réseaux techniques face aux inondations ou comment définir des indicateurs de performance de ces réseaux pour évaluer la résilience urbaine», Bulletin de l'association des géographes français, sous presse.
- [GLE 2005] Gleyse J.-F., 2005, La vulnérabilité structurelle des réseaux de transport dans un contexte de risques, Thèse de doctorat en Analyse Théorique et Epistémologique en Science Géographique, Université Paris VII- Denis Diderot, 848p.

TOXICITY: Un modèle multi-agent pour simuler les comportements des populations lors d'un accident industriel toxique

Emmanuel Bonnet, Élise Beck, David Sheeren, Johnny Douvinet, Eric Daudé, Marion Amalric, Françoise Duraffour, Paul Salze,

Université de Caen Basse-Normandie / UMR IDEES 6266 CNRS
(emmanuel.bonnet@unicaen.fr)

Université Joseph Fourier (Grenoble) / UMR PACTE 5194 CNRS (elise.beck@ujf-grenoble.fr)

Université de Toulouse / INPT-ENSAT, UMR 1201 DYNAFOR (david.sheeren@ensat.fr)

Université de Avignon, UMR ESPACE 6012 CNRS (johnny.douvinet@univ-avignon.fr)

Université de Rouen / UMR IDEES 6266 CNRS (eric.daude@univ-rouen.fr)

Université François Rabelais (Tours) / UMR CITERES 6173 CNRS (marion.amalric@univ-tours.fr)

Université Paris 1 / PRODIG, UMR 8586 CNRS (francoise.duraffour@univ-paris1.fr)

Université de Strasbourg / ERL LIVE 7230 CNRS (paul.salze@live.cnrs.unistra.fr)

Ce poster présente le modèle TOXICITY qui vise à comprendre le niveau d'appropriation des consignes de sécurité en cas d'accident industriel. Il permet d'aider à déterminer le niveau de survie des individus en cas de diffusion d'un nuage toxique faisant suite à une explosion sur un site industriel. La présentation du modèle illustrera la démarche adoptée pour la conception et la mise en œuvre de TOXICITY en identifiant et en détaillant toutes les étapes qui permettent d'aboutir à la simulation et à l'analyse des résultats.

1. Contexte et objectif

Les risques industriels concernent de nombreuses agglomérations et sont susceptibles d'être à l'origine d'accidents auxquels les populations sont censées faire face rapidement pour se protéger. Les accidents industriels prennent différentes formes que l'on peut classer en trois catégories : les incendies, les explosions et la diffusion de nuages toxiques. Des effets combinés sont également possibles à l'image des catastrophes d'AZF à Toulouse en 2001 ou plus récemment celle de Langley West Island (UK) en 2009 qui mêlèrent explosion et diffusion d'un nuage toxique.

Dans le domaine de la modélisation multi-agents, les recherches menées sur les risques industriels ont plus souvent porté sur l'évaluation des conséquences de ces événements passés plutôt que sur l'événement lui-même (Quaranta et al., 2002; Bellasio and Tamponi, 1994 et 2005; Georgiadou et al., 2007). Les études relatives aux comportements des populations en situation de crise (Ruin, 2007 ; Creutin J.-D., 2009) montrent, de leur côté, que les chances de survie des individus sont dépendantes des connaissances qu'ils ont des risques auxquels ils sont exposés et des consignes à suivre en cas d'accident. L'exposition aux risques est d'autant plus forte que le niveau d'appropriation des consignes est faible. La prévention des risques est donc primordiale pour les populations vivant dans un territoire jugé à risque, notamment lorsqu'il contient une zone industrielle. Cependant, le niveau d'appropriation des consignes de sécurité ne suffit pas à expliquer les comportements en cas de crise, qui restent souvent imprévisibles : des mouvements de panique peuvent apparaître, aggravant alors les situations de crise (Provitolo, 2002). C'est la raison pour laquelle il nous semble nécessaire de procéder à la simulation de ces situations de crise post-accident industriel au sein du modèle TOXICITY¹. La finalité du modèle est d'aider à déterminer des niveaux de survie élevés en fonction des différents types de comportement face à la crise et des interactions entre des individus aux comportements différents.

2. TOXICITY : de la conception du modèle à la simulation

Le modèle TOXICITY proposé vise à comprendre dans quelle mesure le niveau d'appropriation des consignes de sécurité détermine la survie des individus en cas de nuage toxique, compte tenu des effets de contingences locales. La formation du nuage toxique fait suite à une explosion sur un site industriel.

Ce poster présente l'ensemble de la démarche adoptée pour la conception et la mise en œuvre de TOXICITY. Toutes les étapes pour aboutir à la création du modèle seront présentées (conceptualisation, formalisation UML, l'implémentation sous le logiciel Net Logo, analyses de sensibilité, simulations) en insistant sur les choix et les arbitrages adoptés.

Le modèle proposé s'applique à une portion d'une ville très simplifiée, constituée de bâtiments et de routes (fig. 1). Le nuage se propage à travers le réseau urbain et ne pénètre pas dans les bâtiments. Les bâtiments peuvent être ouverts ou fermés. L'utilisateur peut modifier la structure spatiale, la vitesse de propagation du nuage, le nombre d'agents, ainsi que leur proportion localisée dans les bâtiments à l'initialisation.

Les agents sont divisés en quatre catégories, suivant leur niveau d'appropriation des consignes officielles : 1/ ceux qui connaissent les consignes et cherchent à se confiner ; 2/ ceux qui ignorent les consignes et cherchent à fuir le nuage ; 3/ ceux qui ignorent les consignes et

¹TOXICITY a été développé dans le cadre de l'école thématique MAPS (modélisation multi-agents appliquée aux phénomènes spatialisés – www.csregistry.org/maps)

cherchent à imiter leurs voisins ; 4/ ceux qui ne modifient pas leur comportement et poursuivent leur activité.

L'analyse de sensibilité du modèle est proposée au travers de différents scénarios qui font varier le nombre d'agents, le taux d'ouverture ou de fermeture des bâtiments, la vitesse de propagation du nuage toxique, la position du départ du nuage. L'analyse de la configuration spatiale et des combinaisons entre les différents types de comportements permet à la fois de valider des parties de la modélisation et de pointer les améliorations à apporter. L'objectif des scénarios testés est notamment de savoir s'il est possible de déterminer le nombre minimum de personnes à informer pour assurer la survie d'un maximum de personnes.

Les analyses de sensibilité révèlent entre autres que (1) la vitesse de propagation du nuage telle que nous l'avons modélisée ne change pas fondamentalement les résultats, excepté pour les individus ne modifiant pas leur comportement ; que (2) la structure spatiale dense est plus favorable à ces mêmes individus, alors que la structure spatiale plus lâche est favorable aux « imitateurs » ; et que (3) au-delà de 70% d'individus informés, le taux de survie de ceux qui ont tendance à imiter leurs voisins ne change plus.

3. Conclusion

TOXICITY, sous sa forme actuelle, a permis d'identifier le rôle des contingences locales et de la configuration spatiale sur les taux de survie à un nuage toxique faisant suite à un accident industriel.

TOXICITY a été développé dans le cadre de l'école thématique MAPS (modélisation multi-agents appliquée aux phénomènes spatialisés – www.csregistry.org/maps) . Il est le fruit d'un travail collectif dans le cadre d'un apprentissage de la modélisation pour des thématiciens du risque. La mise en œuvre de TOXICITY dépasse maintenant le strict cadre pédagogique et s'ouvre vers l'analyse des comportements de population en situation de crise par le biais de la modélisation et de la simulation multi-agent.

4. Bibliographie

- Bellasio R., Bianconi R. (2005) – On line simulation system for industrial accidents., *Environmental Modelling & Software*, 20 (3) 329-342.
- Bellasio R., Tamponi M. (1994) - MDGP: a new numerical model for dense gas dispersion. Sensitivity analysis and first validation trials. *Air Pollution Modelling and Its Application*, 18, 439-446.
- Creutin J.-D., Borga M., Lutoff C., Scolobig A., Ruin I., Creton-Cazanave L. (2009), Catchment dynamics and social response during flash floods: The potential of radar rainfall monitoring for warning procedures *Meteorological applications* volume 16, pp 115-125
- Quaranta N., De Martini A., Bellasio R., Bianconi R. Marioni M. (2002) - A decision support system for the simulation of industrial accidents. *Environmental modeling and software*, 17, 497-504.
- Ruin I, Gaillard J.-C., Lutoff C., (2007) - How to get there ? Assessing motorists' flash flood risk perception on daily itineraries, *Environmental Hazards* 7, 3, pp 235-244
- D. Helbing, I. Farkas, P. Molnar, T. Vicsek, "Simulation of pedestrian crowds in normal and evacuation situations," In *Pedestrian and Evacuation Dynamics*, M. Schreckenberg and S. D. Sharma, (Eds.). Springer, Berlin, pp. 21-58, 2002.

Spatial information fusion

Application to expertise and management of natural risks in mountains

Jean-Marc Tacnet * — **Martin Rosalie** *,+ — **Jean Dezert**** — **Eric Travaglini** *,++

* Cemagref- *Snow Avalanche Engineering and Torrent Control Research unit (ETNA)*
2, rue de la papèterie - B.P. 76, F-38402 Saint Martin d'Herès Cedex
jean-marc.tacnet@cemagref.fr

** *The French Aerospace Lab (ONERA)*
DTIM/SIF, Bâtiment N, Chemin de la Hunière F-91761 Palaiseau Cedex
jean.dezert@onera.fr

+ *INSA Rouen - Mathematical Engineering Department*
Avenue de l'université, F-76801 Saint-Etienne du Rouvray

++ *Savoie University - CISM*
Campus scientifique F-73376 Le-Bourget-du-Lac Cedex

RÉSUMÉ. L'expertise des risques naturels en montagne, dépend de la qualité de l'information spatiale disponible. Une méthode dénommée ER-MCDA Spatiale (Aide multicritères à la décision, raisonnement crédibiliste) est proposée pour représenter l'influence de ces imperfections sur les résultats d'expertise. Elle associe l'analyse multicritères et les théories des ensembles flous, des possibilités et des fonctions de croyance en relation avec un environnement SIG.

ABSTRACT. Expertise of natural risks in mountains depends on more or less imperfect spatial information. A method called Spatial ER-MCDA is proposed to figure how these imperfections influence expertise results. It mixes multicriteria decision making, fuzzy logics and evidence theory in relation with a GIS environment.

MOTS-CLÉS : risques naturels, montagne, expertise, aide multicritères à la décision, SIG, fusion d'information, logique floue, théorie des fonctions de croyance, DST, DSMT

KEYWORDS: Mountain natural hazards, expertise, multicriteria decision making, GIS, information fusion, fuzzy logics, evidence theory, DST, DSMT

1. General context : expertise, decision and natural risks management

Natural phenomena in mountains put people and material assets at risks with dramatic consequences. Risk level is often described as a combination of hazard and vulnerability. Hazard relates to the intensity and frequency of a phenomenon while vulnerability concerns damage consequences and values assessment. One one hand, risk management implies to combine those two components to assess risk level and, on the other hand, to define and choose between the different risk reduction measures.

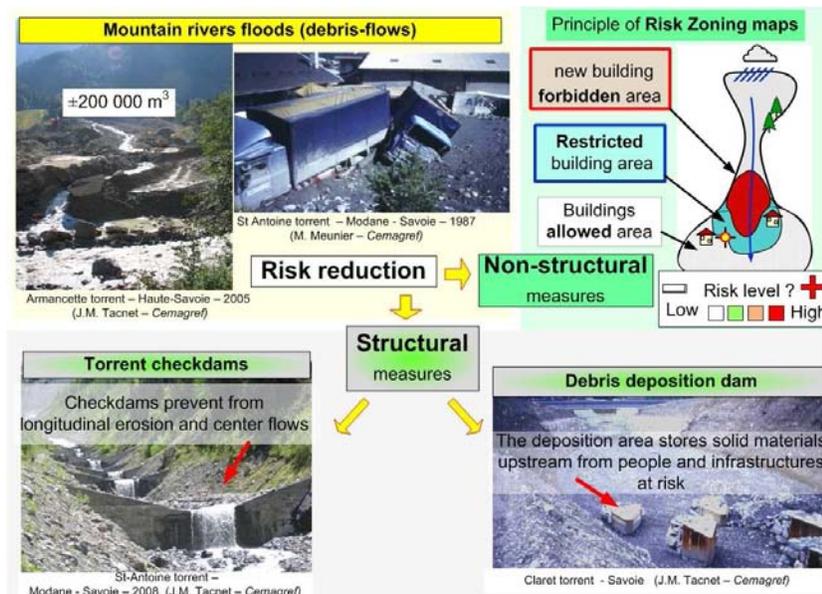


Figure 1. *Torrential flood risks and risk reduction measures*

Expertise is almost always required to take decisions related to prevention, crisis and recovery steps of the risk management process. Risk managers, local authorities, infrastructures managers have, for example to choose to implement structural measures such as civil engineering protection works and/or non structural measures such as land-use and building controls (figure 1). Expertise is based on quantitative, qualitative but also spatial description of natural phenomena. Due to partial knowledge and variable reliability of the sources, available information about natural phenomena often remain imperfect. This can lead to many imperfect experts evaluations such as : "we are certain that this avalanche or torrential flood has reached this point, area ..." or "it is possible that the flood deposit was between 1.5 m and 2.5m ...". An expert can also estimates that "the debris flows volume is up to 15000 m^3 " but we know that he is not fully reliable (a beginner, expertise conditions)... An important issue consists in determining risk limits in zoning maps applications : how far can we be

confident in expertise results and take a decision when both spatial extents, intensity and consequences of the phenomena may be assessed imprecisely ?

Expertise can be considered as a decision problem based on imperfect and heterogeneous information provided by more or less reliable and possibly conflicting sources. Ad-hoc and specific decision support systems are therefore needed to help decisions in this context of imperfect information. To be efficient, they have to take into account and trace information quality, including attribute and spatial values, in the global risk assessment decision process.

2. Information fusion to manage spatial information imperfections and heterogeneity

Fuzzy Sets (Zadeh, 1965), Possibility (Zadeh, 1978; Dubois *et al.*, 1988) and Belief Function Theories are able to consider altogether all kinds of information imperfections such as vagueness, imprecision, conflict and uncertainty. Our methodology extends the *ER-MCDA* approach (Evidential Reasoning and Multicriteria Decision Analysis) (Tacnet *et al.*, 2010) to spatial information with an application to natural risks management problems as proposed in (Tacnet *et al.*, 2009; Tacnet, 2009). The new methodology presented in this paper allows to consider in the same framework both uncertainty and imprecision of the spatial extent of information (e.g. debris-flows, avalanche extent) but also its attribute values such as quantitative values (height, speed, volume, ...) or qualitative indexes (reached, not reached ...).

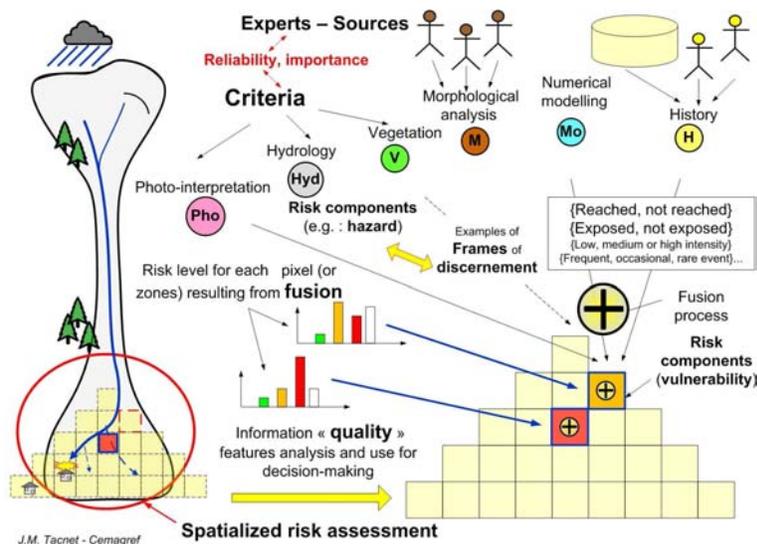


Figure 2. Risk level results from fusion of available heterogeneous information

Imperfect information (spatial extent and/or attribute values) are first represented in a G.I.S (Geographic Information System). For example, we aim to determine an hazard level using the extent and intensity of a debris-flows event. Information comes from sources such as an historical database (imprecise, not fully reliable), expert field analysis (based on an expert judgment) or numerical modeling results (whose uncertainty depend on input data quality)(figure 2).

Information is represented through vector and raster approaches. Geographic information (spatial and attribute values) are processed to be introduced in fusion calculation routines using the Dempster-Shafer theory (DST)(Shafer, 1976) and the Dezert-Smarandache Theory (DSmT)(Dezert *et al.*, 2009). Advanced fusion rules (PCR rules, partial conflict redistribution rules)(Dezert *et al.*, 2006) are used instead of the classic Dempster fusion rule to take decisions. At the end, we can spatially represent not only hazard (or risk) level but also a confidence level based on the information quality used to take decisions.

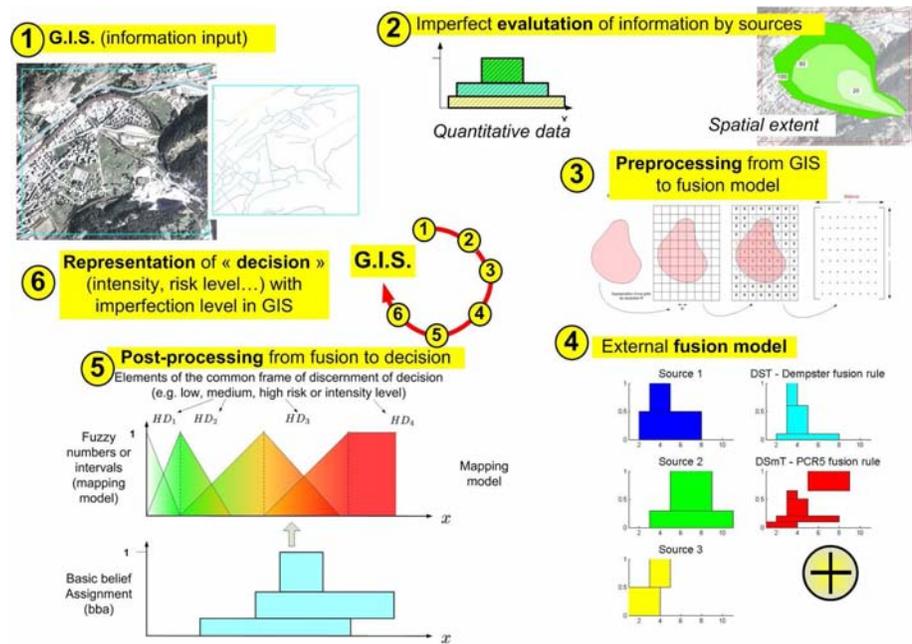


Figure 3. Principles of the Spatial ER-MCDA methodology

3. Conclusion - Discussion

Using both multicriteria decision analysis, fuzzy sets, possibility and belief function theories allows to consider any kind of information including its imperfection and

the sources heterogeneity. On one hand, the spatial *ER-MCDA* methodology allows to trace and combine spatial information imperfection (on extent and attributes) from sources to the final decision and, on the other hand, to represent the result in a G.I.S.

These developments are partially founded by the project *ADAGE* founded by the *ECCOREV* research consortium program ¹ 2010 of Provence, Alpes, Côte-d'Azur Region (PACA) and the *PARAMOUNT* Project ² of the European InterReg Alpine Space program.



4. Bibliographie

- Dezert J., Smarandache F., *Advances and applications of DSMT for Information Fusion - Collected works - Volume 2*, American Research Press, Rehoboth, USA, chapter Proportional Conflict Redistribution Rules for Information Fusion, 2006.
- Dezert J., Smarandache F., *Advances and applications of DSMT for Information Fusion - Collected works - Volume 3*, American Research Press, Rehoboth, USA, chapter 1 : An introduction to DSMT, p. 760, 2009.
- Dubois D., Prade H., *Possibility Theory : An approach to Computerized Processing of Uncertainty*, Plenum Press, New York (U.S.A), 1988.
- Shafer G., *A mathematical theory of Evidence*, Princeton University Press, 1976.
- Tacnet J.-M., *Prise en compte de l'incertitude dans l'expertise des risques naturels en montagne par analyse multicritères et fusion d'information*, Thèse de doctorat en sciences et génie de l'environnement, Ecole Nationale Supérieure des Mines de Saint-Etienne, 2009.
- Tacnet J.-M., Batton-Hubert M., Dezert J., *Advances and applications of DSMT for Information Fusion - Collected works - Volume 3*, American Research Press, Rehoboth, USA, chapter Information fusion for natural hazards in mountains, p. 565-659, 2009.
- Tacnet J.-M., Batton-Hubert M., Dezert J., « A two-step fusion process for multi-criteria decision applied to natural hazards in mountains », *Belief 2010, Proceedings of International Workshop on the Theory of Belief Functions, Brest (France), 1-2 april 2010*, 2010.
- Zadeh L., « Fuzzy sets », *Information and Control*, vol. 8, p. 338-353, 1965.
- Zadeh L. A., « Fuzzy Sets as a basis for a theory of possibility », *Fuzzy Sets and Systems*, vol. 1, p. 3-28, 1978.

1. <http://eccorev.cerege.fr/>

2. <http://www.paramount-project.eu/>

Gravitation et étalement urbain ; un essai de mécanique du couple vitesse de circulation routière et densité de population

RÉSUMÉ. La question de la dynamique urbaine a largement été explorée du côté des économistes sous l'aspect des densités de population mais également du côté des géographes. Les premiers travaux sur cette question datent des années 80 avec les ouvrages portant sur les dynamiques générales et les systèmes chaotiques. L'approche incrémentale s'est aujourd'hui affinée pour évoluer vers les systèmes multi agents ou les automates cellulaires. Nous proposons dans ce poster de définir un système dynamique qui ne serait cette fois-ci pas basé sur des principes par contagions. Aussi, nous nous proposons de construire un couple vitesse routière-densité de population reposant sur des hypothèses issues des théories du trafic mais également du modèle gravitaire ou des grandes lois du mouvement. Ce papier propose d'appliquer le modèle défini dans le cadre d'une aire d'étude du Nord Est de la France : Dijon et son aire urbaine.

ABSTRACT. The question of the urban dynamics was widely investigated as for the economists under the aspect of the densities of population but also the highly-rated of the geographers. The first works on this question date the 80s with the works concerning the general dynamics and the chaotic systems. The incremental approach became refined today to evolve towards the multi-systems agents or the cellular automata. We suggest in this paper defining a dynamic system which would be this time this based on principles by contagions. So, we suggest building a couple road speed density of population basing on stemming hypotheses theories of the traffic but also the model of the gravity or big laws of the movement. This paper suggests applying the model defined within the framework of an area of study of the northeast of France: Dijon and its urban area.

MOTS-CLÉS : gravitation, densité de population, dynamique urbaine, théorie du trafic, vitesse
KEYWORDS: gravitation, density of population, urban dynamic, traffic theory, speed

Dans un contexte de plus en plus tendu en matière d'énergie et de gestion des coûts urbains, la question de la maîtrise l'étalement urbain apparaît de plus en plus fondamentale pour le décideur public. Longtemps méconnue dans ses principes, la dynamique urbaine est aujourd'hui mieux appréhendée avec le développement des nouveaux outils de simulation. Au-delà d'une simple connaissance théorique ou d'une recherche de nouvelles méthodologies, l'approche prospective de l'étalement urbain, à une échelle suffisamment fine, peut se révéler être particulièrement efficace

dans le cadre d'une aide à la décision pour les politiques publics. Employée essentiellement, dans les premiers temps, comme outils de prévision, la modélisation urbaine s'appuie, aujourd'hui, d'avantage sur une approche probabiliste, particulièrement adaptée pour expérimenter la ville de demain. L'aménagiste devient ainsi acteur du développement urbain en mettant en scène le profil le plus proche de la ville durable. A l'échelle du modélisateur, l'objectif est de fiabiliser ses simulations. Cela passe aussi bien par le développement de nouveaux outils informatiques que par une réflexion plus en amont sur les formalismes à intégrer. Ce papier se positionne plus particulièrement sur ces aspects théoriques de la dynamique urbaine en mettant en place une modélisation du couple vitesse routière-densité urbaine. Pour ce faire, le modèle envisagé se base sur deux grandes hypothèses que sont la gravitation universelle et la mécanique des fluides. A partir de ces éléments, il élabore un système dynamique théorique. Quelle est la nature de ce système final ? Comment l'utiliser en interaction avec un Système d'Information Géographique. Pour répondre à ces interrogations, nous évoquerons, tout d'abord, les différentes hypothèses employées, puis le modèle en tant que tel pour enfin finir par un test de validité à partir de données réelles (aire urbaine dijonnaise) ainsi que des projections de la dynamique en 2030 (2 scenarii de la croissance urbaine).

Trois hypothèses principales sous tendent notre modélisation : le modèle gravitaire dans sa forme généralisée, le modèle LWR pour la congestion routière ainsi que la seconde loi de Newton. Pour commencer, comme généralement en économie urbaine, l'argumentaire de base consiste en un espace vide, sans contrainte, lisse assimilable à un plan avec une parfaite capillarité. La vitesse et la densité sont agrégées à l'échelle communale. Sur ce territoire, la première hypothèse présuppose que la densité de population est reliée mécaniquement au flux. Conformément à la gravitation universelle, deux masses s'attirent alors et forment entre elle un flux. La généralisation de cette loi conduit à une émission multicentre selon la loi suivante :

$$Q_i = G \sum_j M_j e^{-\alpha_i x_{ij}} \quad [1]$$

Où Q_i représente le flux en i , G une constante, M_j la densité de population en j , α_i des paramètres à évaluer et x_{ij} les distances au point i .

Dans la seconde hypothèse, on envisage que l'espace se comporte comme un fluide et donc qu'il existe en tout point i du territoire une circulation selon les lois du trafic. Trois grandeurs sont définies dans les modèles macroscopiques du trafic : la vitesse de circulation du flux, la densité du flux et enfin le flux. Les modèles de premier ordre sont les plus connus. On utilise alors généralement les modèles LWR. Plusieurs hypothèses sont admissibles quant à la forme du diagramme fondamental, le poster décrira le modèle le plus approprié dans le cadre d'une étude urbaine à une grande échelle. Enfin la troisième hypothèse suppose que les flux ont une orientation et décrivent des trajectoires de type radiales en direction ou en provenance de la totalité des lieux i . Cette hypothèse met en relation la dynamique de la vitesse dans le temps et la masse de population. Cela signifie plus clairement que la masse de

population agit directement sur la dynamique de la vitesse en construisant peu ou prou « des puits » de vitesse. La forte densité attire mais réduit également la facilité de circulation.

Après résolution mathématique du problème, on parvient à établir un système différentiel reliant la vitesse V_i avec la densité de population que nous noterons g_i .

$$\begin{aligned}\frac{dV_i}{dt} &= -\lambda g_i - \Gamma & [2] \\ \frac{dg_i}{dt} &= \text{aléa} \tau_i \xi V_i g_i + \text{aléa} \Phi \tau_i V_i\end{aligned}$$

V_i vitesse routière, g_i densité de population, λ , Φ , Γ , ξ des paramètres à évaluer, τ_i le taux de variation annuel du lieu i et *aléa* une fonction aléatoire pour fixer la parité convergente ou divergente à partir ou des points i^l . Le système [2] peut être rapproché d'un champ de vecteur si l'on considère qu'il existe une infinité de localisation i . Au regard du modèle gravitaire qui tente d'approcher la question des flux routiers, notre modélisation simule d'autres phénomènes (densité, vitesse) en ajoutant de plus une dynamique de nature systémique.

Au-delà de l'intérêt purement méthodologique de cette approche, le poster présente également un aspect plus thématique avec une application sur l'aire urbaine dijonnaise. Le calage sur les données réelles donne alors de bons résultats de corrélation ($R^2 = 0.93$ entre les données réelles et les données simulées pour 1999 en partant du recensement de 1954). Partant de ce jeu de paramètres (en faisant toutefois varier Φ , ce qui ne change pas notre R^2), il est alors possible de projeter plusieurs scénarii de la ville à l'échelle des communes à l'horizon 2030. Le poster présente une cartographie des densités mais également une visualisation de l'espace-vitesse sous tendant la dynamique. La visualisation des évolutions met clairement en évidence le processus d'étalement urbain au sein de l'aire d'étude sans toutefois fixer les limites de cette croissance. Deux scénarii se font alors face, un premier supposant une évolution modérée et un second avec un étalement massif.

Bibliographie

- Anas A., Arnott R., Small K.A. (1998) « Urban Spatial Structure », *Journal of Economics Literature*, n°36, p.1426-1464
- Lighthill M.J., Whitham G.B. (1955) « On kinematic waves : a theory of traffic flow on long crowded roads », *Proceedings of Royal Society*, n°1178, p. 145-317
- Pumain D. (1982) *La dynamique des villes*, Paris, Economica
- Richards P.I. (1956), « Shockwaves on the highway », *Operations Research*, Vol 4, pp. 42-51
- Isard W., Bramhall D.F. (1960), « Gravity, Potential and Spatial Interaction models », in *Methods of Regional Analysis*, W. Isard, MIT Press, Cambridge, Massachusetts, pp. 493-568.

¹ *aléa* = $(-1)^{\text{random}(0)+1+f}$ f un paramètre compris entre 0 et 1.

Cartomouv' : informer et partager sur la cartographie animée

Groupe Cartomouv¹

« Cartographie du changement, cartographies en mouvement »

UMR Géographie-cités

13, rue du four

75 006 Paris

RÉSUMÉ. Ce poster est l'occasion de présenter le contenu d'un site réalisé par un groupe de recherche qui travaille sur les productions de cartes animées. Cette production est analysée selon une grille de lecture qui vise à formaliser le processus de productions des cartes : depuis les caractéristiques des données mises en jeu jusqu'à leurs perceptions .

Conçu comme un outil de travail interne, le site souhaite aujourd'hui s'ouvrir pour partager les ressources créées. Il concentre des ressources bibliographiques, ainsi que des expérimentations graphiques et cartographiques émanant du groupe. Il est voué à évoluer pour mener d'autres expérimentations allant jusqu'à des enquêtes de perception.

ABSTRACT. This poster aims to present the contents of a website implemented by a research group. This group works on the production of animated maps through a grid of formalisms ranging from data themselves to the users' perceptions. This website was originally designed as a tool for internal work but it is expected to become a way to share bibliographic resources, as well as graphics and mapping experiments. Afterwards it would evolve to conduct other experiments such as perception surveys.

MOTS-CLÉS: Cartographie animée, représentation spatio-temporelle, sémiologie graphique

KEYWORDS: Animated maps, cartography of spatial dynamics, graphic semiology

¹Le groupe Cartomouv constitue un axe rattaché à l'UMR Géographie-cités, composé de cartographes et géographes de différentes spécialités thématiques, ingénieurs et chercheurs, Françoise Bahoken (1), Maher Ben Rebah, Annabelle Boffet Mas (2), Claire Cuntz (3), Antoine Fleury, Nicolas Lambert (4), Liliane Lizzi, , Hélène Mathian, Antonine Ribardièrre(5), Christine Zanin, UMR Géographie-cités, Paris, (1) UMR Géographie-cités, Paris et UMR Migrinter, Poitiers, (2) CEDETE, Orléans, (3) UMR Environnement, Ville, Société, Lyon, (4) UMS RIATE, Paris, (5) UMR PRODIG, Paris

1. Introduction

S'intéressant à la représentation de phénomènes temporels et constatant la prolifération des cartes animées sur Internet, le groupe de travail « cartographie du changement, cartographies en mouvement » s'est constitué autour d'une réflexion sur les méthodes et techniques cartographiques pour la représentation des dynamiques spatio-temporelles. L'objectif est d'identifier et analyser les liens entre les différentes formalisations pour la représentation de ces dynamiques qui vont - de la représentation informatique de l'information géographique - à leur représentation cartographique, en s'attachant particulièrement à identifier les potentiels et les problèmes de sémiologie graphique posés par les nouvelles technologies.

Ces questions ont été simultanément abordées sur le plan théorique et technologique, et le groupe s'est donné pour objectif de réaliser à la fois un travail d'analyse, d'expérimentation et de diffusion autour de ces questions. Le poster que nous proposons a pour but de présenter la démarche et les objectifs du travail du groupe, via l'environnement commun sur lequel il s'appuie et qui consiste en un site internet. Même si le caractère « statique » d'un poster et la présentation de représentations « dynamiques » semblent antinomiques, c'est ici l'occasion de présenter notre approche très analytique et d'ouvrir le site à un public plus large.

La dynamique spatiale est le support du questionnement. Dans cette première version, ce site vise à rendre compte des différents aspects de la démarche :

- proposer une bibliographie ouverte, triée selon un degré d'intérêt par rapport à des catégories méthodologiques ;
- accueillir des cartographies de phénomène-spatio-temporels commentées, discutant les choix qui ont été faits par rapport aux résultats attendus ;
- donner à voir des expérimentations pour une sémiologie animée sous la forme de prototypes conçus comme des éléments de base d'une grammaire visuelle animée ;
- recenser et informer sur les solutions technologiques.

Nous présenterons dans ce poster les deux volets les plus avancés de ce site, à savoir les catégories établies pour structurer la recherche bibliographique et les prototypes de cartographie animée.

2. Une bibliographie commentée

Une bibliographie sélective est proposée regroupant, actuellement, environ 70 références: articles, revues et ouvrages portant sur les questions que posent la représentation du temps, du changement et du mouvement. Ces textes sont analysés à différents niveaux de conception et formalisation qui sont au cœur des questionnements présentés : modèles informatiques associés au changement et au

mouvement, représentation cartographique de dynamiques. Cette bibliographie présente l'originalité de décrire chaque référence par un profil de spécialisation selon les catégories suivantes :

- Modélisation des données et des processus
- Gestion et représentation du temps
- Sémiologie, communication, cognition (théorie)
- Design et interface cartographique
- Technologies informatiques
- Application thématique
- Etat de l'art.

Chaque notice bibliographique comprend les références complètes du texte, un résumé ainsi que le niveau de pertinence dans chacune des sept catégories, selon la manière dont l'aspect de la catégorie est abordé (« pas du tout », « évoqué », « central »).

Les catégories ont été définies sur la base d'une première analyse sémantique de la production scientifique du domaine. Outre l'intérêt évident du recensement de ces références spécialisées, cette bibliographie catégorisée pourrait donner lieu à d'autres utilisations, comme celle d'une évaluation épistémologique de la production du domaine et de sa dynamique. En effet, la lecture des catégories au prisme du temps donne d'ores et déjà à voir des déplacements de centres d'intérêts.

3. Un laboratoire d'expérimentation sémiologique animée

L'animation et l'interactivité sont les deux composantes technologiques des ces nouvelles représentations cartographiques, donnant à l'observateur/utilisateur le loisir d'être spectateur et acteur de sa carte.

Un « laboratoire » propose d'explorer les potentialités de l'intégration du déroulement du temps dans une représentation cartographique, en partant de l'animation d'objets graphiques simples et génériques : le point, la ligne et la surface. Il s'agit de mettre au point et de tester des « briques élémentaires » d'une « grammaire » possible de la cartographie animée [DIB 92, PET 94, MAC 95]. L'expérimentation porte sur la mise en regard d'une formalisation numérique de l'information et ce qu'elle implique du point de vue de sa représentation cartographique. On part ainsi des changements qui peuvent affecter l'une de leurs trois composantes génériques des objets géographiques : la géométrie, la sémantique et l'identité [CHE 07] De la même manière, deux types de formalisation du temps sont spécifiquement considérés, dans un but de comparaison des résultats en terme de représentation : enregistrements à des dates régulières (ex : phénomène connu tous les mois) et enregistrements irréguliers (ex : lorsqu'un événement survient).

Pour chaque combinaison distincte (objet / changement / temps), un prototype est proposé et permet de visualiser la différence entre une représentation discontinue

4 SAGEO'2010

ou continue (i.e. avec interpolation entre deux états), ainsi qu'en testant des effets visuels permis par l'animation (trace, empreinte de la situation précédente, etc.).

En combinant de manière systématique les critères énoncés ci-dessus, ces différents prototypes permettent d'illustrer et de comparer l'efficacité de chacune des solutions proposées. Ces premières expérimentations sont spécifiques à certaines combinaisons [formalisation de l'espace- formalisation du temps – représentation graphique]. Elles sont là pour discuter de l'intérêt de la démarche dans une perspective de prolongement à d'autres combinaisons plus complexes de formalisations de l'espace (surface continues par exemple) ou de structures du temps (enregistrements non datés avec certitude, datations relatives et/ou incertaines...).

Conclusion

Ce site internet, conçu comme un outil de travail et de partage des connaissances, est destiné à être complété, en intégrant les avancées du groupe de recherche et aussi en s'appuyant sur la participation des utilisateurs du site, que ce soit via la base de données bibliographique ou le laboratoire sémiologique.

Enfin ce site est destiné à être le support et le vecteur d'un autre volet des recherches du groupe de travail, portant plus spécifiquement sur la dimension cognitives de ces nouvelles représentations cartographiques.

Bibliographie

- Andrienko N., G. Andrienko and P. Gatalaky, 2000. " Supporting visual exploration of object movement. In: Proceedings of the Working Conference on Advanced Visual Interfaces AVI, Palermo, Italy, ACM Press, pp. 217-220, 315.
- Andrienko N., G. Andrienko and P. Gatalaky, 2000, Visualization of spatio-temporal information in the Internet. In: Proceedings of the DEXA'2000 Workshop on Web-Based Information Visualization, Greenwich, IEEE CS Press
- Cheyran J.P., 2007 « Les processus spatio-temporels : quelques notions et concepts préalables à leur représentation », *Mappemonde*, n° 87 (2007.3)
- DiBiase D., MacEachren A.M., Krygier J.B., Reeves C., 1992. « Animation and the role of map design in scientific visualisation », *Cartography and Geographic Information Systems*, 19 (4), pp.201-214.
- Kraak M.-J., Edsall R. et MacEachren A. M., 1997, Cartographic animation and legends for temporal maps: exploration and or interaction, 7 p. <http://www.itc.nl/~kraak/legends/>

Cartomouv' : Un lieu de ressource pour la cartographie animée⁵

- MacEachren A. M., 1994, « Time as a cartographic variable », in Unwin D.J. Hearnshaw H.M. (Eds.), *Visualization in Geographic Information Systems*, Wiley, pp.115-130
- Peterson M.P., 1994. « Spatial Visualization through Cartographic Animation: Theory and Practice », *Proceedings of Geographic Information Systems / Land Information Systems GIS/LIS 1994*, pp. 250-258.
- Peuquet D., 2002. *Representation of Space and Time*. The Guilford Press; illustrated edition 380 p.
- Thomas Ott Frank Swiaczny, 2001, *Time-integrative Geographic Information Systems - Management and Analysis of Spatio-Temporal Data*. Springer edition. 234 p.

Session R & D

Visualisation interactive de résultats de simulations numériques sous Google Earth

Application au risque d'inondation

RESUME. La simulation numérique occupe un vaste espace, tant dans la recherche que dans l'industrie. Quels que soient les modèles utilisés, on constate qu'il n'existe encore que très peu de solutions techniques simples permettant une visualisation et/ou une évaluation rapide et ergonomique des résultats obtenus en sortie de modèle.

L'étude du risque d'inondation, en ce qu'elle manipule des données spatio-temporelles, peut rendre les post-traitements relativement fastidieux. Une solution consistera donc à proposer une application générant de façon automatisée un document au format KML, contenant les données spatialisées sous une forme vectorielle, et reproduisant la dimension temporelle par une animation chronologique. La méthode proposée s'attache à résoudre un certain nombre de difficultés techniques découlant de la volonté de représenter, non des données SIG, mais des résultats bruts de simulations numériques, et ce par l'intermédiaire d'un processus complètement intégré.

Si le gain de temps en ingénierie est alors indéniable, l'apport du logiciel grand-public qu'est Google Earth, par opposition aux solutions orientées métier, permet surtout d'ouvrir de nouvelles perspectives en termes de communication. Si la solution intéressera prioritairement les professionnels, ingénieurs et techniciens, ayant à manipuler des résultats de modélisations, en l'occurrence hydrologiques, elle s'adresse également à des acteurs dont le profil est moins technique. En effet, cela constitue une opportunité d'accès simplifié à l'information sous une forme de représentation plus interactive, à même de susciter des réactions accrues des preneurs de décisions ou du public, par exemple en ce qui concerne la perception du risque d'inondation.

La présentation s'attachera à (1) préciser les enjeux techniques, (2) expliquer les différentes étapes clés du processus, (3) détailler les aspects innovants, (4) présenter des exemples de visualisations interactives, et enfin (5) exposer les évolutions possibles et/ou à venir.

MOTS-CLES : *Simulation, modèle numérique, visualisation interactive, Google Earth, KML, données spatio-temporelles, risque inondation*

LiveFire, dispositif de géolocalisation automatisée et temps-réel de prises de vues aéroportées dans le cadre de la lutte contre les feux de forêt

L'incendie de forêt, naturel ou provoqué par l'homme, est considéré comme un risque naturel majeur en France. Sa gestion se révèle de plus en plus complexe en raison d'une évolution des territoires, essentiellement caractérisée par l'augmentation des interfaces habitat-forêt ou la déprise agricole. Face à ce risque, le Service Départemental d'Incendie et de Secours (SDIS) des Pyrénées-Orientales, acteur de la Défense de la Forêt Contre l'Incendie, se veut novateur par la recherche et le développement de nouvelles technologies. Cette démarche s'inscrit dans le cadre de partenariats privés et publics, notamment avec plusieurs laboratoires de recherche universitaire, et suit les directives préconisées par le Ministère de l'Intérieur : amélioration des actions de prévention, de la lutte et du Retour d'Expérience.

Le projet LiveFire répond au besoin de disposer d'outils d'aide à la décision facilitant l'interprétation du comportement des incendies et leur développement en situation opérationnelle. En effet, face à un incendie de grande ampleur, la lutte peut s'avérer complexe en raison de conditions de propagation défavorables et souvent difficiles à estimer dans un contexte spatio-temporel très particulier (fumées importantes, propagation rapide, sautes de feu, fronts de feu multiples...). L'apport d'un dispositif aéroporté, en l'occurrence embarqué à bord d'un avion départemental de surveillance et reconnaissance, représente une source accrue d'information. Le dispositif LiveFire de cartographie et d'acquisition d'images embarqué, adapté aux besoins des sapeurs pompiers, permet de géolocaliser précisément le sinistre et de réaliser des prises de vue verticales et obliques à l'aide de deux caméras (MIR, pour moyen infrarouge, et visible) et d'un appareil photo numérique. Les photos aériennes et les images traitées en temps-réel (extraction des contours de feu et géoréférencement) sont transmises vers le Poste de Commandement des Opérations de Secours, au sol. Cette chaîne de traitement répond alors à un véritable besoin opérationnel puisqu'en localisant en temps-réel les fronts de feux par géoréférencement des captures vidéo MIR, le dispositif contribue de manière importante à l'anticipation des moyens à engager ou des actions à mettre en œuvre en fonction des enjeux. Des systèmes similaires existent mais ce projet est novateur dans la mesure où il se révèle être une solution pragmatique, conçue pour répondre à une demande spécifique, dans des conditions particulières et économiquement exploitable dans un contexte opérationnel.

La présentation s'attachera à (1) préciser les objectifs de cet outil d'aide à la décision, (2) développer l'architecture du système et sa mise en œuvre technique, (3) exposer les perspectives d'évolution et de généralisation du système développé sous licence libre, facilitant sa diffusion et son évolutivité.

MOTS-CLÉS : *Incendie de forêt, Aide à la lutte, Vidéographie MIR aéroportée, Géoréférencement, Traitement automatisé et temps-réel d'image.*

Mariage du SIG et de la 3D au service du grand paysage

Etude de cas sur le Mont St Michel

Hélène DURAND *

** Alisé géomatique*

Helene.durand@wanadoo.fr

1. Le contexte

La 3d se généralise comme un outil au service de la concertation, de la visualisation et de la prospective paysagère. Les représentations en 3D, vaste domaine, recoupe en fait des supports, des niveaux de rendus et des finalités très variées (du bloc diagramme, au drapage d'une ortho photographie sur un Modèle numérique de terrain en passant par un Kml sous Google earth ou des maquettes photo-réalistes).

Alors que de telles applications sont maintenant courantes en milieu urbain, nous en sommes encore au démarrage de telles applications en périphérie urbaine, voire en milieu rural ou naturel. Pourtant les enjeux sont majeurs : corridors écologiques, incidence des infrastructures terrestres, ouvrage ponctuel et incidence paysagère (éolien, lignes à haute tension) , mitage péri-urbain et maîtrise de l'étalement urbain, fermeture des milieux par abandon de pratiques agricoles, sont autant de sujet auxquelles ces maquettes 3D peuvent apporter des réponses.

2. Le développement d'une maquette temps réel sur la Baie du Mont St Michel

Pour répondre à de multiples enjeux sur ce territoire emblématique (mitage urbain, disparition de la trame bocagère, deux DREAL ont souhaité développer une maquette 3D photo réaliste, temps réel. Cette réalisation nous a amené à faire des choix de représentations et des traitements spécifiques sur certaines composantes du paysage, en particulier sur les Haies. Un travail sur la typologie de ces haies a été engagé avec une équipe de paysagiste. Cette typologie en 14 postes a ensuite été traduite par des bases de données levées sur le terrain.

Un effort de terrain considérable a été mené à cette occasion: par des méthodes de levés innovantes (Laser et Sig nomade), 6.500 arbres isolés, et 25.000 haies ont été levés. Un observatoire photographique de plus de 1400 clichés est venu alimenter les bibliothèques de textures et façades, il autorise une comparaison des premières simulations à la réalité du terrain. Un travail tout aussi original a été mené pour rendre compte des ambiances variées à travers la baie (paysage littoral, herbus, polder et bocage).

3. Perspectives et améliorations

De nombreuses questions méthodologiques ont été soulevées lors de cette réalisation :

- ✓ Comment passer des BD géoréférencées à une Maquette 3D temps réel.
- ✓ Quelles sont les nouveaux enjeux de cette imbrication étroite entre Système d'information et Modélisation 3D
- ✓ Comment mettre à jour ou enrichir une telle maquette ?
- ✓ Comment rendre compte de la diversité des paysages rencontrés ?
- ✓ Quel niveau de détail représenter ?
- ✓ Quelles sont les spécifications des bases de données nécessaires ?.

En marge des recherches en cours ou à venir, sur ces nouvelles formes de représentation 3D, cette réalisation aura permis de lister les principaux bénéfices, mais également les améliorations nécessaires pour que de telles maquettes deviennent de véritables outils d'aide à la concertation autour de projets d'aménagement.

Apports d'un gazetier pour l'exploration de connaissances hétérogènes

Olivier Bedel * — Erwan Quesseveur** — Ludovic Dupuis ***

* *Alkante SAS*
1 rue du Chêne Morand
35510 Cesson-Sévigné
o.bedel@alkante.com

** *ESO Rennes, Université Rennes 2, UEB*
Place Henri Le Moal, CS 24307
35043 Rennes
erwan.quesseveur@univ-rennes2.fr

*** *University of Greenwich, School of Science*
Medway Campus - Kent - ME4 4TB
L.A.Dupuis@greenwich.ac.uk

RÉSUMÉ. Le projet CHARM 3 (Interreg IV.A) piloté par le centre IFREMER de Boulogne, vise à accroître la connaissance de l'espace maritime de la Manche dans le but de contribuer à la gestion intégrée de cet espace frontalier. Il rassemble 17 partenaires scientifiques anglais et français aux compétences très complémentaires. Le projet comporte ainsi de nombreux volets thématiques portant sur :

- l'inventaire des données physiques, environnementales, biologiques et d'activité humaine disponibles en Manche
- l'inventaire des taxons planctoniques
- l'étude des invertébrés benthiques et de leurs habitats
- la classification des habitats marins
- l'étude de l'exploitation commerciale des espèces halieutiques
- l'analyse de la législation en vigueur
- la cartographie et la modélisation de l'habitat
- l'analyse de la dynamique des communautés halieutiques et la viabilité des pêcheries
- l'analyse du changement climatique sur la distribution des espèces et des communautés halieutiques
- la modélisation des réseaux trophiques
- la mise en œuvre d'un plan de gestion spatialisé de la Manche orientale

Un volet destiné à la diffusion des connaissances rassemblées doit permettre de les porter à la connaissance des services de l'Etat, des collectivités territoriales et des acteurs socio-économiques. Dans cette logique un web-atlas interactif sera mis en œuvre.

Pour améliorer les possibilités d'exploration de ces ressources hétérogènes, une nouvelle forme de gazetier va être expérimentée.

Notre présentation détaillera les apports de ce gazetier qui s'appuie sur une organisation flexible de l'information inspirée de la structure des systèmes d'information logique.

La connaissance produite de manière indépendante par les partenaires CHARM 3 est hétérogène, complexe et massive. Le constat que l'utilisateur ne doit pas connaître l'organisation de la donnée pour pouvoir l'explorer (structuration des couches, thématiques, logique de modélisation...) est le cœur de cette recherche. Un index de lieux s'appuyant sur une organisation flexible, facilite une appréhension non experte de cette production.

Ce type d'index doit permettre de prendre en compte des données comme les photos, les coupures de presse, les entretiens, et de les associer à des données géographiques classiques. Par exemple, grâce à ce dispositif, on pourra identifier les thématiques pour lesquelles on dispose d'information pour une zone précise.

La structuration du gazetier s'appuie sur les propositions de normes Gazetier de l'Open Geospatial Consortium (OGC, WFS profile Gazetteer). Cette organisation associée à des services web adéquats doit garantir l'interopérabilité du gazetier avec d'autres systèmes d'organisation des données géographiques, notamment ceux du projet CHARM 3. Dans ce contexte, l'index de lieu du gazetier pourra servir de point d'entrée pour interroger différentes bases de données géographiques et visualiser les données disponibles sur une zone précise.

La présentation permettra de comparer le gazetier avec un outil d'interrogation de métadonnées, approche plus classique que le gazetier cherche à enrichir.

MOTS CLES : gazetier, exploration de données, structuration logique, approche spatiale.



Exemple d'utilisation dans l'IDS SIG-Pyrénées d'EasySDI :

Solution open-source d'Infrastructure de Données Spatiales configurable par Joomla!

EasySDI est une solution simple et prête à l'emploi d'Infrastructure de Données Spatiales. Elle est interopérable et respecte les standards ISO et OGC. Ces objectifs consistent à répondre aux exigences de la directive INSPIRE ainsi qu'aux besoins d'organismes souhaitant partager leurs données géographiques. La solution a été conçue pour proposer des services de découverte, de visualisation et de téléchargement de données géographiques dans un environnement sécurisé et multilingue.

Visant à être le moins intrusif possible, EasySDI peut être intégré dans un environnement existant respectant les norme ISO et OGC indépendamment de la technologie ou le type de licence utilisée. En outre, EasySDI n'a pas vocation à remplacer les produits existants. Pour se faire, il s'appuie sur les bibliothèques open-source et des logiciels choisis pour leur dynamisme comme OpenLayers, GeoTools, ExtJS, Joomla!, Geonetwork, Geoserver.

EasySDI s'appuie sur l'expérience d'organismes actifs dans toute l'Europe travaillant dans le domaine de l'information géographique, en veillant à la pertinence et la durabilité des concepts du projet. L'objectif du comité de pilotage est de concevoir les évolutions des fonctionnalités du logiciel en fonction des besoins exprimés par la communauté, qui est actuellement formé de 6 structures actives en France, en Suisse et au Luxembourg.

EasySDI est utilisée par l'APEM en France dans le cadre de l'observatoire <http://www.sig-pyrenees.net>.

www.easysdi.org

www.sig-pyrenees.net

www.apem.asso.fr

Atelier Ontologies géographiques

OntoGéo 2010

Atelier « *Ontologies Géographiques* », dans le cadre de
Sagéo'10 « *Spatial Analysis and GEomatics* »

Toulouse, 18 novembre 2010

Atelier organisé dans le
cadre du projet ANR
GeOnto

Actes électroniques

Objectifs et thématiques

Comité Scientifique

Programme

Résumés des articles

Résumés des démonstrations

Liste des articles et des résumés de démonstrations

Objectifs et thématiques

La description et l'intégration cohérente des données géographiques se révèle très difficile car ces données sont très nombreuses et diverses. Une approche de plus en plus privilégiée pour réaliser cette intégration, autant dans le monde des bases de données que dans celui de la recherche d'information et des systèmes d'information géographiques, est d'appuyer l'intégration des données sur une ontologie du domaine concerné.

Dans cet atelier, nous proposons ainsi de faire un panorama des recherches en cours sur les ontologies géographiques et de présenter les résultats obtenus dans le cadre de projets de recherche menés dans cette thématique. Cet atelier se veut un lieu de rencontre et d'échanges de façon à favoriser le partage d'expérience et d'expertise concernant la construction et l'exploitation d'ontologies géographiques et initier d'éventuelles coopérations.

Quatre présentations orales, suivies de questions et discussions ont été programmées pour cet atelier d'une demi-journée : un exposé d'un conférencier invité et trois exposés d'articles acceptés. Une session démonstrations de résultats de recherche fait suite à ces présentations.

Le processus de sélection a été réalisé sur la base de soumissions de papiers de 12 pages maximum pour les articles et de résumés pour les démonstrations. Chaque article a fait l'objet de deux relectures.

L'atelier est organisé dans le cadre du projet ANR GeOnto, qui a financé les frais de transport d'un des auteurs de chaque papier et démonstration et qui a permis l'invitation d'Oscar Corcho-Garcia de l'Université Polytechnique de Madrid.

Nous adressons tous nos remerciements aux auteurs des articles et résumés soumis ainsi qu'à tous les membres du comité de programme.

Liste (non exhaustive) des thèmes :

- Construction d'ontologies géographiques
- Enrichissement, évolution d'ontologies géographiques
- Ontologies géographiques et interopérabilité
- Exploitation d'ontologies géographiques pour la recherche d'information, l'indexation de documents, dans le cadre de l'apprentissage humain, ...

Comité d'organisation

- Chantal Reynaud (LRI Univ. Paris-Sud 11 & CNRS, INRIA Saclay Île-de-France) : Présidente
- Nathalie Abadie (Lab. COGIT, IGN)
- Fayçal Hamdi (LRI Université Paris-Sud 11 & CNRS, INRIA Saclay Île-de-France)
- Ammar Mechouche (Lab. COGIT, IGN)
- Yassine Mrabet (LRI Univ. Paris-Sud 11 & CNRS, INRIA Saclay Île-de-France)

Comité de programme

- Nathalie Abadie (Lab. COGIT, IGN)
- Nathalie Aussenac-Gilles (IRIT, Université Paul Sabatier)
- Alain Bouju (Lab. Informatique, Image, Interaction, Université de La Rochelle)
- Mauro Gaio (LIUPPA, Université de Pau et des Pays de l'Adour)
- Jérôme Gensel (LIG, Université Pierre Mendès France, Grenoble)
- Fayçal Hamdi (LRI Université Paris-Sud 11 & CNRS, INRIA Saclay Île-de-France)
- Eric Kergosien (LIUPPA, Université de Pau et des Pays de l'Adour)
- Marion Laignelet (IRIT, Université Paul Sabatier)
- Thérèse Libourel (Laboratoire d'Informatique, Université de Montpellier II)
- Ammar Mechouche (Lab. COGIT, IGN)
- Yassine Mrabet (LRI Univ. Paris-Sud 11 & CNRS, INRIA Saclay Île-de-France)
- Sébastien Mustière (Lab. COGIT, IGN)
- Van Tien Nguyen (LIUPPA, Université de Pau et des Pays de l'Adour)
- Chantal Reynaud (LRI Univ. Paris-Sud 11 & CNRS, INRIA Saclay Île-de-France)
- Raphaël Troncy (EURECOM, Sophia Antipolis)
- Christelle Vangenot (Université de Genève, Suisse)

Programme

18 novembre, 2010

14h30-15h30	Conférence invitée, <i>Oscar Corcho</i>	
	Ontology-based data integration in the geospatial domain: a case study in Spain. <i>Oscar Corcho, Universidad Politécnica de Madrid</i>	p. 8
15h30-17h00	Présentations d'articles	
15h30-16h00	Recherche automatique de concepts et de relations sémantiques dans des documents géographiques. <i>Marion Laignelet, Mouna Kamel, Nathalie Aussenac- Gilles</i>	p.9
16h00-16h30	Une approche basée sur les ontologies et les agents pour l'interopérabilité des SIG <i>Alnafie Emdjed, Hamadou Djamila, Libourel Thérèse</i>	p. 24
16h30-17h00	AOC, une ontologie formelle pour la modélisation de systèmes complexes en géographie <i>Patrice Langlois, Eric Daudé, Baptiste Blanpain, Emmanuel Sapin</i>	p.36
17h00-17h30	Session démonstrations	
17h00-17h15	TaxoMap Framework appliqué à l'alignement d'ontologies géographiques au sein du projet Geonto <i>Fayçal Hamdi, Chantal Reynaud, Brigitte Safar</i>	p. 51
17h15-17h30	Vers un système en ligne d'aide à la découverte du contenu des bases de données géographiques <i>Ammar Mechouche, Nathalie Abadie, Emeric Prouteau, Sébastien Mustière.</i>	p. 54

Résumés des articles

Ontology-based data integration in the geospatial domain: a case study in Spain.

Oscar Corso, Universidad Politécnica de Madrid, Madrid, Espagne

In this talk we will present the work that is being carried out under the geo.linkeddata.es initiative, where geographical data provided by the Spanish Instituto Geográfico Nacional is being migrated to the Web of Linked Data. Methodological and technological details about the migration process will be provided, together with a discussion on guidelines for naming geographical features, generation and management of feature geometries, integration of heterogeneous geographical and statistical databases, identification of duplicated individuals, and best practices on the usage of these data in applications. Finally a list of open research topics will be identified and discussed, including aspects related to the maintenance of database to ontology mappings, to the management of multilinguality and to provenance recording.

Recherche de concepts et de relations sémantiques pour la géographie. Traiter automatiquement des documents de spécification de bases de données.

Marion Laignelet^{}, Mouna Kamel^{**}, Nathalie Aussenac-Gilles^{**}*

^{}Maison de la recherche – Université de Toulouse le Mirail, Toulouse, France*

*^{**}Institut de Recherche en informatique de Toulouse (IRIT) – CNRS, Toulouse, France*

Basée sur une méthodologie exploitant la notion de marqueur, nous proposons de repérer des marques linguistiques de concepts et de relations sémantiques en utilisant non seulement des informations de type syntaxique mais également la structure hiérarchique des documents. Nous travaillons sur des documents de spécification de bases de données géographiques dont la structure et l'organisation, du fait notamment de leur fonction, est sémantiquement explicite. Matérialisée par un balisage XM, cette structure permet de construire un noyau d'ontologie qui est ensuite enrichi par des connaissances extraites à l'aide de patrons combinant analyse lexico-syntaxique et analyse structurelle.

Une approche basée sur les ontologies et les agents pour l'interopérabilité des SIG

Emdjed Alnafie^{}, Djamila Hamdadou^{*}, Thérèse Libourel^{**}*

^{}Université d'Oran Es-senia, laboratoire LIO, Algérie*

*^{**}Université de Montpellier, laboratoire LIRMM, France*

Le domaine de l'information a connu un progrès extraordinaire permettant à de nombreuses spécialités et disciplines, de partager des sources d'information hétérogènes et de grandes masses. Cette communication, en particulier entre les systèmes d'information géographique (SIG), est malheureusement confrontée au problème de l'hétérogénéité des sources d'information. L'interopérabilité des SIGs est la solution qui permet de créer un unique et homogène système « virtuel » avec un accès simple, transparent et efficace à l'information nécessaire. Cet accès est assuré quelle que soit la localisation des SIGs ou de leurs méthodes de stockage. L'objectif principal de cette étude, et dont les contributions seront en mesure de faciliter la prise de décision des gestionnaires concernés, est de proposer une approche fondée

sur l'utilisation combinée des géo-ontologies et des agents. Cette approche adopte une politique de traitement permettant de trouver rapidement des réponses aux requêtes tout en garantissant un accès direct à ces réponses dans le cas d'une prochaine éventuelle requête de même nature.

AOC, une ontologie formelle pour la modélisation de systèmes complexes en géographie

*Patrice Langlois, Eric Daudé, Baptiste Blanpain, Emmanuel Sapin
Université de Rouen, Mont-Saint-Aignan, France*

Cet article présente un méta-modèle nommé AOC (Agents-Organisations-Comportements), qui s'exprime par une ontologie formelle et sert de cadre conceptuel à une plateforme de simulation de systèmes complexes spatialisés, compatible avec les problématiques de dynamiques géographiques les plus larges. Le modélisateur géographe doit pouvoir concevoir son modèle et le formaliser dans un langage cohérent relativement à son cadre de pensée. AOC offre ainsi la possibilité de structurer les relations entre agents via le concept d'organisation, qui permet de créer des modèles multi-niveaux. De plus, une formalisation générale de la notion de comportement permet de modéliser la dynamique d'une classe très large de systèmes.

Résumés des démonstrations

TaxoMap Framework appliqué à l'alignement d'ontologies géographiques dans le projet GéOnto.

Fayçal Hamdi, Chantal Reynaud, Brigitte Safar

Les outils actuels d'alignement d'ontologies ne sont pas performants dans tous les domaines ni quelles que soient les ontologies. La qualité des mappings obtenus peut être améliorée si le processus d'alignement tient compte des spécificités des ontologies alignées. L'environnement TaxoMap Framework a été conçu pour aider un expert du domaine à spécifier des traitements sur les mappings produits par l'outil d'alignement TaxoMap, afin de les raffiner ou bien de fusionner, restructurer ou enrichir des ontologies. Nous proposons une démonstration de deux modules implémentés dans cet environnement : le module d'alignement et le module de raffinement.

Vers un système en ligne d'aide à la découverte du contenu de bases de données géographiques.

Ammar Mechouche, Nathalie Abadie, Emeric Prouteau, Sébastien Mustière

Nous présentons ici un prototype d'une application web proposant à un utilisateur quelconque de pouvoir découvrir le contenu de bases de données géographiques, en lui permettant de savoir comment et où sont représentées les données qui répondent à son besoin. Cette application s'appuie sur une ontologie du domaine décrivant les entités topographiques du monde réel et permettant de guider l'utilisateur dans l'expression de son besoin. Elle s'appuie également sur des ontologies formalisant les spécifications des bases de données géographiques, et inclut une solution de cartographie en ligne pour un aperçu des données.

Ontology-based data integration in the geospatial domain: a case study in Spain.

Oscar Corcho

*Universidad Politécnica de Madrid
28660 Boadilla del Monte, Madrid, Spain*

ABSTRACT. In this talk we will present the work that is being carried out under the geo.linkeddata.es initiative, where geographical data provided by the Spanish Instituto Geográfico Nacional is being migrated to the Web of Linked Data. Methodological and technological details about the migration process will be provided, together with a discussion on guidelines for naming geographical features, generation and management of feature geometries, integration of heterogeneous geographical and statistical databases, identification of duplicated individuals, and best practices on the usage of these data in applications. Finally a list of open research topics will be identified and discussed, including aspects related to the maintenance of database to ontology mappings, to the management of multilinguality and to provenance recording.

Recherche de concepts et de relations sémantiques pour la géographie

Traiter automatiquement des documents de spécification de bases de données

Marion Laignelet*, Mouna Kamel**, Nathalie Aussenac-Gilles**

* CLLE-ERSS - UMR 5610

Maison de la Recherche - Université de Toulouse-Le Mirail

5, allée Antonio-Machado, F-31058 Toulouse Cedex 9

marion.laignelet@univ-tlse2.fr

** Institut de Recherche en Informatique de Toulouse (IRIT) - CNRS

UPS, 118, route de Narbonne, F-31062 Toulouse Cedex

mouna.kamel@irit.fr

aussenac@irit.fr

RÉSUMÉ. Basée sur une méthodologie exploitant la notion de marqueur, nous proposons de repérer des marques linguistiques de concepts et de relations sémantiques en utilisant non seulement des informations de type syntaxique mais également la structure hiérarchique des documents. Nous travaillons sur des documents de spécification de base de données géographiques dont la structure et l'organisation, du fait notamment de leur fonction, est sémantiquement explicite. Matérialisée par un balisage XML, cette structure permet de construire un noyau d'ontologie qui est ensuite enrichi par des connaissances extraites à l'aide de patrons combinant analyse lexico-syntaxique et analyse structurelle.

ABSTRACT. Based on the notion of pattern, our methodology aims at tracking linguistic evidences of concepts and semantic relations by using not only syntactic information but also the hierarchical structure of documents. The documents that we analyse are geographical and cartographic databases. Due to their role, their structure is semantically meaningful. This structure is implemented with XML tags, which makes it possible to automatically build an ontology kernel. This kernel is enriched with pieces of knowledge that are extracted thanks to patterns that combine lexico-syntactic and structural features.

MOTS-CLÉS : ontologie, géographie, TAL, patrons lexico-syntaxiques, structure de document, repérage automatique de concepts et relations

KEYWORDS: ontology, geography, NLP, lexicale and syntactical patterns, document structure, automatic tracking of concepts and relations

1. Introduction

Nous proposons de trouver en corpus des traces linguistiques de concepts et de relations sémantiques dans le but de construire une ontologie géographique à partir de documents de spécifications de bases de données. La méthode que nous proposons exploite non seulement des patrons lexico-syntaxiques [BOU 03, BOU 04, AUS 09, AUS 08] mais également des patrons basés sur la structure hiérarchique des documents et matérialisée dans nos textes par des balises XML. La mise en forme des documents de spécification de base de données géographiques de l'IGN, du fait notamment de leur fonction descriptive et explicative, est sémantiquement explicite et permet de construire un noyau d'ontologie. Ce noyau est ensuite enrichi par des connaissances analysées et extraites à l'aide de patrons plus locaux, combinant analyse lexicale, syntaxique et structurelle.

Ce travail s'inscrit dans le cadre d'un projet de recherche, le projet Géonto, dont l'objectif est de capitaliser des connaissances hétérogènes à partir de documents géographiques diversifiés. D'un point de vue méthodologique, l'objectif du projet est d'éprouver et d'adapter des outils et méthodes existants pour créer des ontologies à partir de données et de connaissances imprécises et hétérogènes. Du point de vue applicatif, les ontologies créées doivent permettre un accès facilité à de grandes masses de données géographiques. Parallèlement, nous espérons par cette méthode appréhender les différences de point de vue sous-jacentes au domaine.

Dans cet article, nous présentons la méthode que nous avons définie pour exploiter au mieux les documents de spécification de bases de données géographiques de manière à constituer une ontologie reflétant la structuration des connaissances définie par ces spécifications. L'originalité de cette méthode est de rechercher des connaissances en exploitant non seulement le langage naturel présent dans ces textes, mais aussi les indications explicites de structure qui encadrent les zones en langage naturel. Dans la section 2, nous présentons un panorama des différentes méthodes de construction d'ontologies à partir de texte, que ce soit pour la recherche des concepts ou celle des relations entre les concepts. La section 3 présente le projet Géonto et la problématique d'acquisition de connaissances pour la géographie et la cartographie. La section 4 décrit l'approche mise en oeuvre ainsi que les outils développés pour extraire les concepts potentiellement pertinents pour le domaine de la géographie et les relations sémantiques susceptibles de relier ces concepts. Enfin, dans la dernière section, nous évaluons quantitativement et qualitativement la méthode mise en oeuvre.

2. Construire des ontologies à partir de textes en langue naturelle

La construction d'ontologies, et plus généralement de ressources termino-ontologiques (RTO), à partir de textes nécessite la mise en place de méthodologies complexes exploitant massivement des outils du traitement automatique des langues [BOU 04, JAC 06, SCH 05, MAY 09, ROL 06]. L'utilisation de techniques de TAL permet notamment d'extraire les éléments qui vont constituer la RTO, ou qui vont servir d'indices à l'ontographe pour définir des éléments de RTO. Selon le type de RTO visé (hiérarchie de termes, thé-

saurs, hiérarchie de concepts, ontologie plus ou moins riche en relations) les besoins sont variables. Mais dans tous les cas, il est nécessaire d'extraire des termes, lesquels sont potentiellement aptes à être considérés par la suite comme des étiquettes de concepts pour le domaine en question.

2.1. Méthodologies basées sur la notion de marqueur ou patrons

L'utilisation de patrons pour extraire des connaissances est largement utilisée. [HEA 92] fait figure de référence en ce qui concerne les techniques de linguistique computationnelle. Meyer (2001) introduit la notion de patrons de connaissances (*knowledge rich contexts*) pour faire référence à patrons linguistiques exprimant diverses relations sémantiques liées à la définition, telles l'hyponymie, la synonymie ou la méronymie. [HAD 02] définit la notion de marqueur comme une « *forme linguistique faisant partie de catégories prédéfinies (grammaticales, lexicales, syntaxiques ou sémantiques) dont l'interprétation définit régulièrement le même rapport de sens entre les termes.* » Et, dans le contexte de la recherche de relations sémantiques en corpus, [GRA 04] définissent le patron lexicosyntaxique de la manière suivante : « *à la différence des marqueurs, les patrons identifient la relation recherchée plus précisément en définissant également des contraintes syntaxiques ou typographiques sur le contexte des termes.* » Dans tous les cas, un patron définit des contextes suffisamment fins et précis, relevant de niveaux linguistiques variables et révélateurs de connaissances ou de comportements linguistiques particuliers.

Dans le cas de la recherche de relations sémantiques, un patron permet d'indiquer l'existence d'une relation sémantique particulière entre deux entités (au moins). Appliquée à la construction d'ontologies, une telle approche fait l'hypothèse que les relations lexicales peuvent fournir des indices pour définir des relations conceptuelles et, avec elles, de nouveaux concepts et termes associés. [BOU 04] mettent en place une méthodologie fondée sur l'utilisation de patrons. Une première phase d'extraction de termes¹ construit un réseau de mots et de syntagmes (le « réseau terminologique ») dans lequel chaque syntagme est relié à sa tête et à ses expansions. À ce niveau, les éléments du réseau sont des candidats-termes potentiellement aptes à devenir des concepts. La seconde étape est une analyse distributionnelle² : à partir du réseau terminologique, un calcul des proximités distributionnelles entre les unités est effectué. La dernière phase extrait les relations³. Des outils de modélisation⁴ permettent *in fine* de définir les concepts à retenir. Dans cette approche, il est important de noter que les termes sont les occurrences des mots du corpus et les relations sont extraites à l'aide de patrons lexico-syntaxiques. D'autres travaux mettent en oeuvre cette méthodologie ([JAC 06]).

1. À l'aide du logiciel Syntex, analyseur syntaxique.

2. À l'aide du logiciel Upery.

3. À l'aide de Yakwa (concordancier) et Caméléon (logiciel de recherche de relations lexicales à partir de marqueurs linguistiques).

4. TermOnto, Terminae

La question des patrons lexico-syntaxiques pour la création et la population d'ontologies est également traitée dans [MAY 09]. Ces auteurs proposent la notion de patrons lexico-syntaxiques (*lexico-syntactic ontology design patterns*, ODPs) qui génèrent de l'information ontologique à partir de textes non-structurés dans le but de créer une nouvelle ontologie ou enrichir des ontologies existantes.

2.2. Les verbes pour exprimer les relations

Au coeur d'un patron lexico-syntaxique traduisant une relation lexicale, se trouve souvent un verbe comme indice fort de cette relation. Par exemple, *comporter*, *contenir*, *se composer de* peuvent marquer la relation d'un tout vers ses parties. [SCH 05] insiste sur le rôle des verbes comme éléments centraux pour déterminer une relation entre deux concepts d'une ontologie. Pour les auteurs, les verbes spécifient l'interaction entre les participants d'une action ou d'un événement. Leur système identifie des triplets (*i.e.* des paires de concepts reliés par une relation) qui pourront être intégrés à une ontologie existante. Cet outil extrait les termes pertinents et les verbes des textes à l'aide d'une approche combinant des techniques linguistiques et statistiques. L'approche mise en place est de type linguistique. Les structures de dépendance sont reconnues : fonction grammaticale, structure des phrases, étiquetage morpho-syntaxique et lemmatisation, normalisation des syntagmes nominaux complexes sur la base de leur tête nominale. Une reconnaissance des entités nommées et des concepts propres au domaine (le foot) est également menée (utilisation de l'ontologie du domaine) ainsi que la reconnaissance des synonymes. Enfin, un processus statistique fournit des mesures de pertinence et des mesures de cooccurrence. L'approche se veut robuste et adaptable à d'autres domaines.

2.3. Utilisation de la structure des documents

L'idée d'exploiter la structure des documents pour améliorer le processus d'acquisition des connaissances est proposée par [ROL 06]. Les auteurs partent du constat suivant lequel les documents ont la plupart du temps une structure logique cohérente et porteuse de sens. C'est le cas, par exemple, des manuels techniques, des dictionnaires, des codes juridiques, etc. Généralement, cette structure se matérialise en XML. Leur objectif est de construire une ontologie des plantes tropicales en s'appuyant sur la structure du document⁵ à deux niveaux, d'un côté pour obtenir une hiérarchie de classes reflétant la taxinomie botanique traditionnelle (famille, genre, espèce), d'un autre pour cibler plus précisément les traitements linguistiques à effectuer pour compléter la hiérarchie de classes par des indications méronymiques. Les auteurs insistent sur le fait que la botanique est un domaine qui se prête naturellement à la représentation de taxinomies : le découpage en genres et en espèces est représenté dans la structure du document et permet ainsi d'initialiser la hiérarchie de classes de l'ontologie. Le résultat produit consiste en

5. La *Flore du Cameroun* qui représente 40 volumes publiés entre 1963 et 2001. La structure des différents volumes est assez régulière, soit une fiche par espèce.

une première ontologie intermédiaire⁶ qui facilite les traitements linguistiques ultérieurs plus complexes pour créer une ontologie du domaine. D'autres travaux comme [DES 00] utilisent également la structure des documents pour construire des ontologies du domaine.

Dans tous les cas, construire des RTO à partir de textes n'est pas une tâche triviale et l'importance du domaine, de la qualité du corpus ou encore de la finalité de la ressource sont essentiels.

3. Acquisition de connaissances dans le domaine de la géographie-cartographie

Les données géographiques sont à la fois nombreuses et diverses. Par exemple, l'INSEE propose une ontologie en français⁷ qui décrit les événements liés à la construction ou la modification des territoires ; le projet TOWNTOLOGY⁸ développe une ontologie dédiée à l'enseignement de l'urbanisme. A notre connaissance, il n'existe pas d'ontologie dédiée aux objets cartographiques, en français du moins⁹. Dans ce contexte, le projet Géonto, dans lequel cet article s'inscrit, propose des méthodes et des outils permettant la description et l'intégration cohérente des données géographiques [MUS 09, KER 09]. Pour répondre à cet objectif, nous proposons la construction d'une ontologie du domaine qui doit permettre l'intégration de sources d'information multiples et hétérogènes. La méthodologie mise en place suit les étapes suivantes :

- création d'une première ontologie (le noyau),
- enrichissement de cette ontologie à partir des spécifications de bases de données topographiques de l'IGN.

Ce qui nous intéresse plus particulièrement concerne la conception et l'enrichissement du noyau d'ontologie. Le noyau de l'ontologie est obtenu à partir de l'exploitation de la structure des documents de spécification des bases de données. L'enrichissement se fait à partir de l'analyse de la structure et de l'analyse linguistique de zones textuelles spécifiques, les zones de définition.

Dans les figures suivantes, un extrait du corpus est présenté, d'abord sous leur format original puis en XML.

D'une manière générale et comme cela a été dit dans la section 2, l'extraction de connaissances à partir de textes nécessite la constitution d'un corpus. Il s'agit d'une étape importante car, dans une approche automatisée, le corpus est la source essentielle d'information : la taille du corpus doit être suffisamment importante pour fournir une couverture large du domaine et il doit être suffisamment homogène pour permettre le repérage de régularités. Par ailleurs, [BOU 04] affirment qu'il est impossible de définir *a priori* des instructions méthodologiques précises car le processus de construction reste lié à l'application et aux exigences des spécialistes du domaine.

6. Dans notre travail, nous parlons de noyau d'ontologie.

7. <http://www.insee.fr/fr/methodes/default.asp?page=xml/xml.htm>

8. <http://liris.cnrs.fr/townto/>

9. La situation pour les ontologies en anglais est différente.

A – Voies de Communication Routière

Tronçon de Chemin

Définition : Voie de communication terrestre non ferrée destinée aux piétons, aux cycles ou aux animaux...	
Regroupement : Voir les différentes valeurs de l'attribut <nature>.	
Sélection : Voir les différentes valeurs de l'attribut <nature>.	
Modélisation géométrique : A l'axe, au sol.	
<hr/>	
Attribut : Nature	
Définition :	Permet de distinguer plusieurs types de voies de communication terrestres.
Type :	Énuméré
Valeurs :	Chemin empierré / Chemin / Sentier / Escalier / Piste cyclable
<hr/>	
Nature = « Chemin empierré »	
Définition : Route sommairement revêtue ou chemin empierré (pas de revêtement de surface ou revêtement très dégradé), mais permettant la circulation de véhicules automobiles de tourisme par tous temps.	
Regroupement : Allée (carrossable) Piste Route empierrée	
Sélection : Toutes les routes empierrées sont incluses.	
...	
<hr/>	
Attribut : Franchissement	
Définition :	Attribut indiquant la présence d'un obstacle physique dans le tracé d'une route et la manière dont il est franchissable.
Type :	Énuméré
Valeurs :	Bar piéton / Gué ou radier / Pont / Tunnel / Sans objet
<hr/>	
Franchissement = « Gué ou radier »	
Définition : Passage naturel ou aménagé permettant de traverser un cours d'eau sans avoir recours à un pont ou un bateau.	
Regroupement : Gué Radier	
...	
<hr/>	
Attribut : Nom	
Définition :	Nom du chemin.
Type :	Caractères
Valeur nulle :	Le champ contient la chaîne de caractères "Valeur non renseignée" pour tous les chemins n'appartenant pas à un grand itinéraire routier nommé (référence = BDCart0).

Figure 1. Un extrait du corpus : sa mise en forme

```

<?xml version="1.0" encoding="ISO-8859-1" ?>
<document> <domaine>
  <nom_domaine>A - Voies de communication routière</nom_domaine>
  <classe>
    <nom_classe>Tronçon de chemin</nom_classe>
    <definition>Voie de communication terrestre ... </definition>
    ...
    <regroupement> Voir les différentes valeurs de l'attribut &lt;nature&gt;
  </regroupement>
  ...
  <attributs>
    <attribut>
      <nom_attribut> Nature </nom_attribut>
      <definition> Permet de distinguer plusieurs... </definition>
      <type> Énuméré</type>
      <valeurs> Chemin empierré/Chemin/Sentier/Escalier ... </valeurs>
      <attribut_valeur>
        <valeur> Chemin empierré </valeur>
        <definition> Route sommairement revêtue ... </definition>
        <regroupement> Allée (carrossable)|Piste|Route empierrée
      </regroupement>
      </attribut_valeur>
    </attribut>
  </attributs>
</classe></domaine> </document>

```

Figure 2. Un extrait de la version XML du corpus

La construction d'une ontologie (dite de référence) dans le projet Géonto a pour but de faire collaborer des données géographiques hétérogènes. Dans ce contexte, le sens des termes et des concepts associés est susceptible de varier selon les sous-domaines de la géographie. Par exemple, le terme de grotte peut être relié à des concepts différents selon que l'on prend le point de vue d'une carte routière ou celui d'une carte de randonnée : dans un cas, "grotte" est relié au concept "curiosité touristique" par une relation d'hyponymie alors que dans un contexte de randonnée, une relation d'hyponymie avec le concept "cavité souterraine" serait sans doute plus appropriée, peut-être même en association avec l'idée de danger. Dans notre corpus, l'exemple du pont est plus concret : le "pont" peut être une sorte de "franchissement" lorsqu'un objet passe par dessus un autre ou une sorte d'"obstacle" lorsqu'il passe par dessous. Dans ce contexte, il est nécessaire de recourir à un corpus suffisamment large et diversifié pour être capable de mettre au jour l'ensemble des points de vue possibles pour un même objet.

Les documents sur lesquels nous travaillons sont des documents de spécification de bases de données géographiques fournies par le COGIT (IGN). Nous disposons de deux textes différents, BDCarto et BDTopo. L'ensemble de ces deux corpus constitue un corpus de 23 884 mots (17 069 pour BDCarto et 6 815 pour BDTopo). Ces textes sont en XML. Les connaissances exprimées dans ces documents est cependant relativement variée puisqu'ils décrivent les objets nécessaires à la création de cartes, qu'elles soient touristiques, routières, pédestres, etc. Des points de vue différents sur les objets manipulés dans le domaine sont ainsi mis en valeur, notamment dans les zones de définition¹⁰. En plus d'un travail sur la structure logique de ces textes (notamment à travers l'organisation des titres et des sous-titres) qui fournit des indications de relation d'hyponymie, nous proposons une étude approfondie des zones de définition. À ce propos, [AGU 09] constatent que dans un contexte définitoire, les relations sémantiques à considérer sont la synonymie, la méronymie, la causalité et le but.

4. Présentation de la méthode

Dans cette section, nous présentons brièvement le module de construction du noyau d'ontologie. Cette partie du travail a été décrit avec précision dans [KAM 09a] et [KAM 09b]. Nous focalisons notre attention sur la chaîne de traitement destinée à l'enrichissement du noyau d'ontologie. Cette chaîne met au jour des concepts et des relations sémantiques entre ces concepts à l'aide de patrons exploitant à la fois des informations lexicales, syntaxiques et structurelles.

4.1. Construction du noyau d'ontologie

Partant du constat que les éléments de structure servent à organiser, subdiviser, hiérarchiser le contenu d'un document, alors cette structure est porteuse de sémantique, soit des relations hiérarchiques entre unités textuelles. Partant de là, il est possible de construire

10. Matérialisées par des balises XML spécifiques.

un noyau d'ontologie valide pour un corpus donné. L'idée de patron structurel proposé et mis en oeuvre par [KAM 09b] permet d'associer une propriété sémantique à un élément structurel. Le patron lexico-syntaxique, quant à lui, prend généralement en compte des éléments lexicaux et syntaxiques présents dans le texte non structuré. Plus précisément, les titres (emboîtement ou parallélisme de titres et sous-titres) ou encore les structures énumératives peuvent permettre de mettre en lumière des relations de subordination ou de juxtaposition entre les termes contenus dans ces mêmes éléments structurels. Les auteurs insistent sur la nécessité de disposer d'un document au format qui rend explicite sa structure et son caractère hiérarchique, par exemple, le XML. Enfin, les termes doivent désigner sinon des concepts, au moins des termes valides pour le domaine visé. La démarche proposée se fait en quatre étapes : *(i)* associer une sémantique aux balises et aux relations qui les relient ; *(ii)* ré-annoter (automatiquement) le document pour mettre en évidence les propriétés des balises et pour attribuer des catégories grammaticales ou sémantiques aux unités textuelles ; *(iii)* définir les patrons structurels qui caractérisent la sémantique portée par un élément de structure et produisent un fragment d'ontologie ; et *(iv)* projeter les patrons. Cette méthode¹¹ décrite dans [KAM 09a] met en évidence 1183 concepts extraits à partir de la structure.

4.2. Chaîne de traitement pour l'enrichissement du noyau d'ontologie

Cette chaîne de traitement a été produite avec le logiciel LinguaStream¹². Son objectif est d'extraire automatiquement dans des textes en langage naturel des concepts et des relations entre ces concepts pour produire des fragments d'ontologie (de la forme : terme/concept-relation-terme/concept). La chaîne de traitement se décompose en cinq modules principaux : *(i)* la préparation du corpus en XML, *(ii)* les pré-traitements linguistiques et le repérage des syntagmes nominaux, *(iii)* la recherche de relation, *(iv)* la prise en compte de la structure (les titres exclusivement), *(v)* l'export en XML.

4.3. La préparation du corpus en XML

Ce module permet d'abord de sélectionner les textes du corpus à analyser ainsi que, à l'intérieur des textes, les zones textuelles délimitées par des balises particulières. Pour ce qui nous concerne, nous exploitons les zones de titres et les zones de définition (cf. 1 et 2.

4.4. Les pré-traitements linguistiques et le repérage des syntagmes nominaux

Cette étape est constituée de plusieurs modules. Tout d'abord, un segmenteur, fourni nativement par LinguaStream, permet de découper le texte en mots. Puis l'analyseur morpho-syntaxique TreeTagger¹³ fournit un étiquetage morpho-syntaxique des mots du

11. Développée à l'aide de Gate (<http://gate.ac.uk>)

12. <http://www.linguastream.org/>

13. <http://www.ims.uni-stuttgart.de/projekte/complex/TreeTagger/>

corpus. Un troisième module, basé sur des macro-expressions régulières, permet le découpage du corpus en phrase. 23 884 mots et 1 978 phrases sont annotés dans le corpus. Le repérage des syntagmes nominaux est réalisé à l'aide d'une grammaire EDCG. Il permet de repérer et d'annoter :

- des SN syntaxiquement simples (*un gué*)
- des SN avec des noms composés, [N de N] (*l'hôtel de police*)
- des SN plus complexes, constitués soit d'adjectifs (*un chemin étroit, un bâtiment industriel et commercial*), soit de compléments du nom "spéciaux" où l'un des éléments introduit une relation sémantique particulière comme *partie, tas, ensemble, etc. (un tronçon de route)*

En plus du repérage des syntagmes nominaux, une structure de traits sémantiques est associée automatiquement à chacune des expressions. Les informations suivantes sont associées : le type de marqueur éventuellement présent (méronyme, holonyme, hypéronyme, mot composé, mot simple), le lemme court (dans la plupart des cas le lemme du nom tête du SN), le lemme long (totalité de l'expression nominale) et des informations de type syntaxique. Ce module annote 6 244 syntagmes nominaux. Par exemple, l'occurrence *Ligne de métro* sera annotée comme cela :

```

<lss :sem type="sn" id="296">
<lss :text>Ligne de métro</lss :text>
<lss :value>
<marqueur>compose</marqueur>
<lemmasCourt>ligne</lemmasCourt>
<lemmasLong>ligne de métro</lemmasLong>
<syntaxe>null</syntaxe>
</lss :value>
</lss :sem>
```

Figure 3. Exemple de repérage et d'annotation d'un syntagme nominal

Sur la base de l'ensemble de ces traitements, l'étape suivante consiste à mettre au jour des relations sémantiques entre les syntagmes repérés, sans préfigurer, à ce niveau de l'étude, de leur qualité de concept ou non.

4.5. La recherche de relations sémantiques

Cette étape se concentre sur le repérage des relations sémantiques. Sont repérées les relations traditionnelles telles que l'hyponymie et la méronymie. Les deux grammaires correspondantes ont été créées et adaptées à partir du travail accompli avec Caméléon [AUS 08]. Il s'agit à la fois d'une adaptation et d'un enrichissement de ces travaux relativement nos propres problématiques et nos corpus. Concernant les relations d'artefact et de fonction, elles ont été créées pour le projet Géonto, sur la base d'une étude manuelle du corpus. Dans chacune des grammaires, la relation s'établit entre un X et un Y. Dans certains cas, lorsque que le X ou le Y est absent de la relation intra-phrastique, c'est le titre qui s'avère jouer, la plupart du temps, le rôle du X ou du Y.

La relation de méronymie est transitive et anti-symétrique. Elle est repérée dans notre grammaire à l'aide de règles du type « X est une partie de Y » ou « Y est composé de X ». Le X n'est donc pas forcément le premier élément de l'expression repérée. La grammaire comporte 68 types de règles, chacune étant divisée en plusieurs configurations (présence du verbe être ou de virgules, absence explicite du X, ...). Dans le corpus, 18 cas de méronymie sont repérés¹⁴. Dans l'exemple suivant, la règle r12c s'applique à la phrase *Voie qui fait partie du domaine public* et extrait une relation de méronymie entre "voie" (X) et "partie du domaine public" (Y).

```
<lss :sem type="meronymie" id="1">
<lss :text>Voie qui fait partie du domaine public</lss :text>
<lss :value>
<regleMero>r12c</regleMero>
<lemmasCourtX>voie</lemmasCourtX>
<lemmasLongX>voie</lemmasLongX>
<snX>Voie</snX>
<lemmasCourtY>domaine</lemmasCourtY>
<lemmasLongY>partie domaine public</lemmasLongY>
<snY>partie du domaine public</snY>
</lss :value>
</lss :sem>
```

Figure 4. Exemple de relation de méronymie

La relation d'hyperonymie est une relation transitive, anti-symétrique et hiérarchique du type « X est un Y », « X est une sorte de Y ». Dans l'exemple suivant, la règle r20a trouve une relation d'hyperonymie entre *autoroute* et *route*.

```
<lss :sem type="hyperonymieInSentence" id="1">
<lss :text>Les autoroutes sont des routes</lss :text>
<lss :value>
<regleHyperoS>r20a</regleHyperoS>
<lemmasCourtX>autoroute</lemmasCourtX>
<lemmasLongX>autoroute</lemmasLongX>
<snX>Les autoroutes</snX>
<lemmasCourtY>route</lemmasCourtY>
<lemmasLongY>route</lemmasLongY>
<snY>routes</snY>
</lss :value>
</lss :sem>
```

Figure 5. Exemple de relation d'hyperonymie

La grammaire dédiée à la relation d'hyperonymie comporte 20 règles. Dans notre corpus, 48 relations d'hyperonymie sont instanciées (seulement 2 dans BDTopo). Dans la littérature, cette relation est souvent citée comme étant la plus riche [AGU 09] alors que dans notre corpus, il semble que ce soit le contraire. Cette différence est sans doute liée au type de corpus qui, dans notre cas, est très spécialisé.

14. Il y a une différence entre le nombre de règles et le nombre d'instances : un certain nombre de règles/ patrons ne s'instancient pas sur ce corpus, les patrons ayant été développés et évalués sur des textes diversifiés et pas uniquement sur les textes de spécification de bases de données. L'intérêt de cette méthode est de se donner les moyens de traiter efficacement de nouveaux textes.

La relation d'artefact correspond à une relation du type « X est représenté par » ou « X est employé pour ». C'est une relation symétrique. Ce type de relation se caractérise par le fait que deux concepts sont mis en relation de quasi-synonymie dans un domaine spécifique, dans notre cas, les bases de données de l'IGN. Dans le monde *réel* (*i.e.* général), ces concepts ne seraient en aucun cas considérés comme des synonymes. Cette relation est à mettre en parallèle avec la distinction entre monde réel et monde carto proposé par le Cogit¹⁵ : un artefact est utilisé pour désigner de manière plus ou moins détournée des objets ou des ensembles d'objets du monde réel [KAS 09]. Dans l'exemple ci-dessous, le concept X n'est pas explicite dans la phrase mais doit être repris du titre, d'où la valeur « fromTitre » de « lemmasCourtX », « lemmasLongX » et « snX ».

```
<lss :sem type="artefact" id="5">
<lss :text>représentant un danger potentiel</lss :text>
<lss :value>
<regleArtef>r4aSansX</regleArtef>
<lemmasCourtX>fromTitre</lemmasCourtX>
<lemmasLongX>fromTitre</lemmasLongX>
<snX>fromTitre</snX>
<lemmasCourtY>danger</lemmasCourtY>
<lemmasLongY>danger potentiel</lemmasLongY>
<snY>un danger potentiel</snY>
</lss :value>
</lss :sem>
```

Figure 6. Exemple de relation d'artefact

Cette grammaire est constituée de 6 règles principales, chacune étant divisée en plusieurs configurations possibles (présence du verbe être ou de virgules, absence explicite du X dans la zone de définition, ...). Ce module repère 14 relations d'artefact dans notre corpus.

La relation de fonction indique à quoi un objet sert. C'est une relation anti-symétrique qui introduit un Y composé soit d'un SN seul (*le transport des marchandises*) ou un syntagme verbal qui pourrait être nominalisé (*transporter les marchandises, le transport des marchandises*). Dans l'exemple qui suit, une relation fonctionnelle de type *faire passer des véhicules* est associée à *trajet de bateau*.

La grammaire consacrée aux artefacts est constituée de 6 règles principales, chacune étant divisée en plusieurs configurations (présence du verbe être ou de virgules, absence explicite du X, ...). Dans notre corpus, elle repère 82 relations de fonction.

4.6. La prise en compte de la structure

Ce dernier module repère deux types différents de relations nécessitant la prise en compte de la structure du document. Il est basé sur les observations faites dans [KAM 09b] concernant la notion de patron structurel. Trois types de relations sont mises en avant :

15. Laboratoire de l'IGN, partenaire du projet Géonto.

```

<lss :sem type="fonction" id="4">
<lss :text>Trajet du bateau servant à passer des véhi-
cules</lss :text>
<lss :value>
<regleFonct>r2b</regleFonct>
<lemmasCourtX>trajet</lemmasCourtX>
<lemmasLongX>trajet de bateau</lemmasLongX>
<snX>Trajet du bateau</snX>
<lemmasCourtY>véhicule</lemmasCourtY>
<lemmasLongY>passer véhicule</lemmasLongY>
<snY>passer véhicules</snY>
</lss :value>
</lss :sem>

```

Figure 7. Exemple de relation de fonction

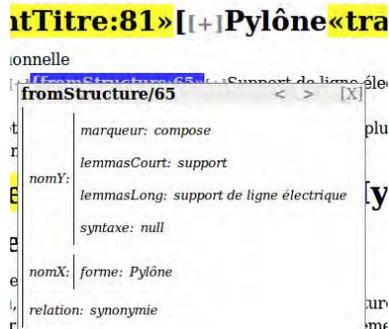


Figure 8. Une relation de synonymie entre deux concepts extrait à l'aide de la structure et d'informations lexico-grammaticales

- le type « synonymie stricte » : un terme-concept est défini par un SN simple (un nom/terme seul). Par exemple : *cascade* / *définition* : *chute d'eau*.
- le type « hyperonymie » : un terme-concept est défini par un un nom/terme auquel sont adjoints des adjectifs/propriétés. Par exemple : *gorge* / *définition* : *vallée étroite et encaissée*
- le type « méronymie » : un terme-concept est défini par un sn introduit par un marqueur de méronymie. Par exemple : *tronçon de route* / *définition* : *portion de voie de communication*

Dans le module, le premier traitement concerne les SN constitués d'un marqueur de méronymie, d'holonymie ou d'hyperonymie. Dans ce cas, le titre est nécessaire pour la création de l'un des deux concepts de la relation (X ou Y selon les cas). Le second cas concerne les SN simples qui suivent immédiatement un titre. Deux possibilités sont envisagées : soit le SN est seul et nous concluons sur une relation de synonymie ; soit le SN ne constitue que le début de la définition et nous concluons à une relation d'hyperonymie. Ce module repère 406 relations. La figure 8 montre un exemple pour la relation de synonymie.

5. Résultats et discussion

L'évaluation des ontologies en termes de performance est un problème reconnu mais délicat [SCH 05]. Dans l'idéal, disposer d'un gold-standart permettrait de fournir une évaluation quantitative intéressante. Ce type de processus nécessite un investissement humain coûteux essentiellement en termes d'annotation manuelle. Nous n'avons malheureusement pas pu mettre un tel système en place.

Nous avons mené une évaluation qualitative et manuelle des concepts et relations sémantiques associées aux concepts. Pour chaque relation annotée nous avons jugé de la pertinence de la relation elle-même ainsi que celle des concepts X et Y. Cinq valeurs sont possibles :

- la valeur est **valide** si la relation est jugée valide ainsi que les deux concepts X et Y reliés par cette relation,
- la valeur est **inverse** si la relation est valide mais est inversée (c'est le cas des relations anti-symétriques comme l'hyperonymie et la méronymie),
- la valeur est **approximative** lorsque la relation est valide mais les concepts sont approximatifs soit parce qu'ils ont été mal extraits par les programmes soit parce qu'il y a une relation de pronominalisation avec le titre et que c'est ce dernier qu'il faudrait prendre en compte,
- la valeur est **incertaine** lorsqu'une expertise particulière est requise (nécessité de recourir à des experts en cartographie),
- enfin, la valeur est **fausse** quand la relation est éronnée (dans ce cas, les concepts sont également souvent non valides).

Les résultats de cette évaluation sont indiqués dans le tableau suivant.

Relation	Artefact	Fonction	Hyperonymie	Synonymie	Méronymie	Holonymie	Total
Valide	16,7 %	14,3 %	38,8 %	25 %	19 %	58,3 %	33,1
Inverse	0 %	0 %	1 %	0 %	43 %	0 %	2,7
Approximative	16,7 %	23,8 %	12 %	25 %	19 %	8,3 %	15,4
Incertaine	66,6 %	55,5 %	19,6 %	0 %	0 %	16,7 %	22,2
Fausse	0 %	6,4 %	28,5 %	50 %	19 %	16,7 %	26,5

Ces résultats mettent en avant de grandes disparités qualitatives selon les types de relations sémantiques. Tout d'abord, les relations d'artefact et de fonction sont les relations présentant le plus haut pourcentage d'incertitude sur la pertinence de la connaissance extraite même si d'un point de vue linguistique, elles sont tout à fait acceptables. Cette incertitude est liée à leur spécificité pour le domaine et au recours nécessaire à des experts géographes pour leur validation. La relation d'hyperonymie présente des résultats corrects mais insuffisants. Presque 30 % des cas sont incorrects. Ces résultats vont dans le sens de ceux de [MAY 09] qui constatent, sur leurs données, une situation de surgénération de leurs patrons d'hyperonymie. Notre proposition pour le traitement de la synonymie n'est clairement pas convainquant : exploiter la structure et la syntaxe n'est pas suffisant pour distinguer l'hyperonymie de la synonymie. Enfin, les résultats concernant la méronymie et l'holonymie sont encourageants même si les patrons de méronymie doivent être revus pour ne pas générer des relations sémantiques inversant les concepts.

Une évaluation quantitative est en cours sur ces données. Nous mettons en place une procédure de comparaison des concepts existants dans le noyau ontologique avec les concepts (introduit à la section 4.4) et les relations fournis par la chaîne de traitement décrite dans cet article. Cette évaluation devrait fournir des résultats quantitatifs à même de rendre compte de l'intérêt de notre méthode d'enrichissement d'ontologie.

6. Conclusion

Cet article présente une méthodologie d'extraction de connaissances à partir de textes mise en place dans le cadre d'un projet de recherche visant l'appariement de connaissances géographiques hétérogènes. Il s'agit plus particulièrement de repérer de manière automatique des termes susceptibles d'être des concepts du domaine et des relations sémantiques particulières entre ces termes.

Dans la tradition des travaux exploitant les notions de marqueur et de patron lexicosyntaxique, notre approche utilise également la structure même des documents pour mettre en évidence des relations sémantiques entre les termes. Les résultats confirment l'idée que la structure logique permet de mettre au jour des relations d'hyponymie entre les termes du titre et les termes contenus dans les zones de définition. Mais la situation est moins convaincante pour ce qui est de la relation de synonymie.

L'évaluation des résultats a uniquement porté sur un jugement humain de la validité des relations entre termes extraits des documents. Une seconde évaluation, en cours, va permettre de mesurer dans quelle mesure les concepts et relations extraites enrichissent le noyau d'ontologie initial. Enfin, une évaluation par les experts est nécessaire pour notamment valider le statut de concept de chacun des termes repérés.

7. Bibliographie

- [AGU 09] AGUADODECEA G., ALVAREZDEMON I., MONTIEL PONSODA E., « From linguistic pattern to ontology structures », *TIA'2009*, 2009.
- [AUS 08] AUSSENAC-GILLES N., JACQUES M.-P., « Designing and evaluating patterns for relation acquisition from texts with Caméléon », *Terminology*, vol. 14, n° 1, 2008, p. 45-73.
- [AUS 09] AUSSENAC-GILLES N., HERNANDEZ N., « Du linguistique au conceptuel : identification de relations conceptuelles à partir de textes », *TIA*, 2009.
- [BOU 03] BOURIGAUT D., AUSSENAC-GILLES N., « Construction d'ontologies à partir de textes », *Conférence TALN 2003*, 2003.
- [BOU 04] BOURIGAUT D., AUSSENAC-GILLES N., CHARLET J., « Construction de ressources terminologiques ou ontologiques à partir de textes : un cadre unificateur pour trois études de cas », *Revue d'intelligence artificielle*, vol. 18, n° 1, 2004, p. 87-110.
- [DES 00] DESMOULINS C., GRANDBASTIEN M., « Des ontologies pour indexer des documents techniques pour la formation professionnelle », *IC'2000 : ingénierie des connaissances (Toulouse, 10-12 mai 2000)*, 2000.

14 Nom de la revue ou conférence (à définir par \submitted ou \toappear)

- [GRA 04] GRABAR N., HAMON T., « Les relations dans les terminologies structurées : de la théorie à la pratique », *Revue d'intelligence artificielle*, vol. 18, 2004.
- [HAD 02] HADDAD M., « Extraction et impact des connaissances sur les performances des systèmes de recherche d'information », PhD thesis, Université de Grenoble, 2002.
- [HEA 92] HEARST M., « Automatic Acquisition of Hyponyms From Large Text Corpora », *Proceedings of the 14th International Conference on Computational Linguistics*, 1992, p. 539-545.
- [JAC 06] JACQUES M.-P., AUSSENAC-GILLES N., « Variabilité des performances des outils de TAL et genre textuel », *TAL*, vol. 47, n° 1, 2006, p. 11-32.
- [KAM 09a] KAMEL M., AUSSENAC-GILLES N., « Construction automatique d'ontologies à partir de spécifications de bases de données », *Conférence IC*, 2009.
- [KAM 09b] KAMEL M., AUSSENAC-GILLES N., « Utiliser la Structure du Document dans le Processus de Construction d'Ontologies (regular paper) », L'HOMME M.-C., SZULMAN S., Eds., *Conférence Internationale sur la Terminologie et l'Intelligence Artificielle (TIA)*, <http://www.irit.fr/>, novembre 2009, IRIT, page (on line).
- [KAS 09] KASSEL G., « Vers une ontologie formelle des artefacts », *20es Journées Francophones en Ingénierie des Connaissances, Hammamet, Tunisie*, 2009.
- [KER 09] KERGOSIEN E., KAMEL M., SALLABERRY C., BESSAGNET M.-N., AUSSENAC-GILLES N., GAIO M., « Construction et enrichissement automatique d'ontologies à partir de ressources externes », *Conférence JFO 2009*, 2009.
- [MAY 09] MAYNARD D., FUNK A., PETERS W., « Using Lexico-Syntactic Ontology Design Patterns for Ontology creation and population », EVA BLOMQUIST KURT SANDKUHL F. S. V. S., Ed., *Proceedings of the Workshop on Ontology Patterns (WOP 2009)*, Washington DC, USA, vol. 516, 2009.
- [MUS 09] MUSTIÈRE S., ABADIE N., AUSSENAC-GILLES N., BESSAGNET M.-N., KAMEL M., KERGOSIEN E., REYNAUD C., SAFAR B., « GéOnto : Enrichissement d'une taxonomie de concepts topographiques », *SAGEO'2009*, Marne-la-Vallée, 2009.
- [ROL 06] ROLE F., ROUSSE G., « Construction incrémentale d'une ontologie par analyse du texte et de la structure du document », *Document numérique*, vol. 9, n° 1, 2006, p. 77-91.
- [SCH 05] SCHUTZ A., BUITELAAR P., « RelExt : A Tool for relation Extraction from Text in Ontology Extension », *Proceedings of the 4th International Semantic Web Conference (ISWC)*, 2005.

Une approche basée sur les ontologies et les agents pour l'interopérabilité des SIG

ALNAFIE Emdjed* — **HAMDADOU Djamilia**** — **LIBOUREL Thérèse*****

* Université d'Oran Es-senia, Laboratoire LIO, ALGÉRIE amazad@hotmail.fr

** Université d'Oran Es-senia, Laboratoire LIO, ALGÉRIE dzhammadoud@yahoo.fr

*** Université de Montpellier, Laboratoire LIRMM, FRANCE libourel@lirmm.fr

RÉSUMÉ. Le domaine de l'information a connu un progrès extraordinaire permettant à de nombreuses spécialités et disciplines, de partager des sources d'information hétérogènes et de grandes masses. Cette communication, en particulier entre les systèmes d'information géographique (SIG), est malheureusement confrontée au problème de l'hétérogénéité des sources d'information. L'interopérabilité des SIGs est la solution qui permet de créer un unique et homogène système "virtuel" avec un accès simple, transparent et efficace à l'information nécessaire. Cet accès est assuré quelle que soit la localisation des SIGs ou leurs méthodes de stockage. L'objectif principal de cette étude, et dont les contributions seront en mesure de faciliter la prise de décision des gestionnaires concernés, est de proposer une approche fondée sur l'utilisation combinée des géo-ontologies et des agents. Cette approche adopte une politique de traitement permettant de trouver rapidement des réponses aux requêtes tout en garantissant un accès direct à ces réponses dans le cas d'une prochaine éventuelle requête de même nature.

ABSTRACT. The domain of information has known an amazing progress which allowed many specialities and disciplines, to share at the same time large and heterogeneous sources of information. This communication, in particular between the geographical information systems (GIS), is unfortunately, confronted to the problem of heterogeneity of the information resources. The GIS's Interoperability is the solution which makes possible to create a unique and homogeneous "virtual" system with a simple, transparent and effective access to required information, whatever their localization or their storage methods. The principal aim of this study, and whose contributions will be able to facilitate the decision-making of the managers concerned, is to propose an approach based on geo-ontologies and agents. This approach adopts a requests treatment policy making possible to find request's answers rapidly and guarantee a direct access to these desired answers in case of next possible requests of same nature.

MOTS-CLÉS : Système d'Information Géographique (SIG), Interopérabilité, Hétérogénéité, Géo-ontologies, Agent.

KEYWORDS: Geographic Information System(GIS), Interoperability, heterogeneity, Geo-ontologies, Agent

1. Introduction

Le domaine de l'information et de la communication a connu une avancée remarquable qui ne cesse d'engendrer de nouvelles techniques. Cette évolution exponentielle a permis à un nombre important de personnes, de différentes spécialités et domaines, de partager des sources d'information à la fois hétérogènes et d'une grande masse.

Il est donc clair que la recherche d'information qu'effectue un utilisateur va s'élargir pour toucher aux ressources en possession d'autres utilisateurs. Cette communication va malheureusement se heurter au problème d'hétérogénéité des ressources d'information, chose qui nécessite une standardisation de l'univers de discours, et c'est là que l'interopérabilité des systèmes joue un rôle primordial dans le partage des informations et des services.

Notre objectif à travers cette étude, est de proposer une approche pour l'interopérabilité des SIGs, fondée sur l'utilisation combinée des ontologies et des agents. Cette combinaison vise à unir la flexibilité des ontologies à l'efficacité des agents pour remédier au mieux du problème de l'hétérogénéité. L'approche permet donc d'échanger de manière cohérente les informations entre les SIGs et de trouver des réponses aux requêtes afin de répondre aux besoins des utilisateurs.

L'article présente en section 2 les concepts de base de l'interopérabilité des SIGs. La section 3 donne un rapide aperçu préliminaire sur les ontologies et présente les spécificités des géo-ontologies en particulier. L'approche proposée est décrite dans sa globalité en section 4 et nous concluons notre propos en section 5.

2. L'interopérabilité des SIGs : État de l'art

L'étude des phénomènes et des objets géographiques englobe des quantités de plus en plus grandes d'informations dispersées géographiquement et en possession d'utilisateurs qui ne partagent pas les mêmes contextes et ne suivent pas les mêmes contraintes. Cela dit, un problème sérieux et crucial n'est autre que l'interopérabilité des systèmes d'informations géographiques qui ont besoin énormément et fréquemment de coopérer et partager leurs ressources, pour répondre à leurs besoins.

L'interopérabilité est l'une des meilleures solutions dans le sens où elle permet non seulement de fournir aux utilisateurs la vision et l'impression que le système "virtuel" sur lequel ils travaillent est unique et homogène, alors qu'en réalité l'information se trouve distribuée dans des environnements hétérogènes, mais elle leur donne également la possibilité d'accéder d'une manière simple, transparente et efficace aux informations demandées, quelles que soient leur localisation ou leurs modalités de stockage, mais il ne faut pas oublier que cette partageabilité doit être réalisée à travers des mécanismes assurant le bon fonctionnement des systèmes en question.

Cependant, avant de se réjouir des avantages de l'interopérabilité, il est essentiel d'étudier les techniques permettant de la réaliser et parmi lesquelles "**l'utilisation des ontologies**" se montre très intéressante.

3. Les ontologies : concepts et définitions

Tout d'abord, une ontologie au sens global possède plusieurs définitions. La première, mais aussi la plus fréquemment utilisée, est apparue pour le projet ARPA Knowledge sharing, elle définit l'ontologie comme "une spécification explicite d'une conceptualisation" [GRU 93]. Elle peut être aussi définie comme une formalisation à un certain niveau de complexité d'informations structurées [SMI 01], mais il est possible également de proposer une version simplifiée et plus facile à comprendre qui présente l'ontologie comme "une hiérarchie de termes reliés par des relations générique/spécifiques [FLE et Al 09]", ou encore un ensemble de concepts (classes) ainsi que les relations existant entre ces concepts.

L'ontologie qui a pour rôle de prendre en compte et gérer l'information géographique est appelée "géo-ontologie" ou ontologie d'un domaine relevant de la géographie. Elle est définie par les différentes spécificités auxquelles est liée la multidimensionnalité des objets géographiques.

Cependant il faut bien faire la différence entre les géo-ontologies relatives à une thématique (agriculture, hydrologie, l'urbain...) et qui doivent leur appellation au domaine qu'elles traitent, de celles purement spatiales (d'où l'appellation "ontologies spatiales") qui se focalisent sur la spécification de concepts spatiaux dans un but de représentation cartographique.

Il faut aussi prendre en compte, avant de construire une géo-ontologie, l'identification des utilisateurs concernés ainsi que les différents contextes qui rentrent en jeu. Cela dit, une géo-ontologie se présente sous forme de vues multi-utilisateurs pour des besoins multi-contextes.

Le contexte peut être défini comme étant "un ensemble de conditions et influences environnantes qui permettent de comprendre la situation en faisant d'elle une situation unique". En termes de méthodologie de construction des géo-ontologies, il est possible d'en distinguer deux types, le premier basé sur la structure de l'objet géographique et le deuxième sur son fonctionnement.

4. L'approche proposée (INSIGO)

L'objectif principal de notre modèle "**INSIGO**" est de résoudre le problème de l'interopérabilité des SIGs. Il s'agit donc de lever toute sorte d'ambiguïté qui peut être due à l'hétérogénéité qui existe entre les différents SIGs communiquant entre eux.

Le modèle que nous proposons dans la présente étude est non seulement basé sur les ontologies mais aussi sur les SMAs (Système Multi-Agents) dans le but d'unir les forces de ces deux grands concepts afin d'aboutir à une approche à la fois concise et consistante.

4.1. Présentation de l'approche

INSIGO est une approche basée sur un modèle cylindrique réalisé à travers la connexion de plusieurs cellules de forme hexagonale de façon circulaire. Cela signifie qu'en partant d'une cellule donnée et en parcourant ses voisines¹, il est possible d'y retourner. La figure 1 illustre la composition de chaque cellule :

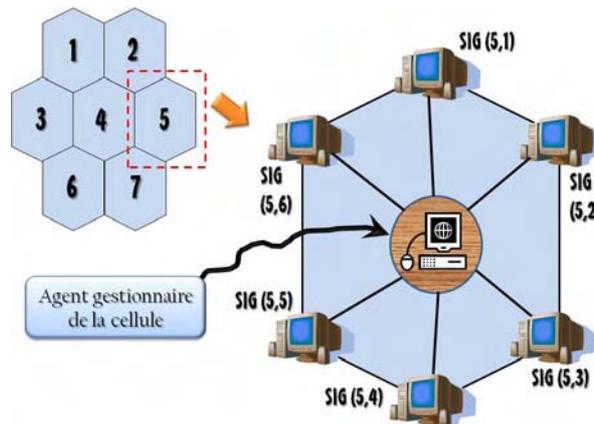


Figure 1. La composition des cellules

Sur chaque tête de l'hexagone se trouve un SIG muni d'une adresse composée du numéro de la cellule à laquelle il appartient ainsi que son numéro dans la cellule (ex : SIG(4,3) est le SIG numéro 3 de la cellule 4). Chaque SIG possède trois voisins directs et appartient pratiquement à trois cellules voisines et est doté, par conséquent, de trois adresses différentes avec lesquelles il est reconnu dans chaque cellule à laquelle il appartient(adresse locale).

Dans le but de réaliser l'interopérabilité, chaque SIG sera doté d'un module appelé "module d'interopérabilité" composé des éléments suivants :

- Une **ontologie de domaine** modifiée à chaque mise à jour du SIG, celle-ci

1. Le nombre de voisins est six exactement

contiendra l'ensemble des concepts qui relèvent du domaine que vise le SIG (agriculture, météorologie, urbanisme, etc.) en question ainsi que les rapports qui les relient. Cette ontologie subira des révisions chaque fois que le SIG subit des mises à jour influentes, c'est-à-dire des mises à jour qui ajoutent, suppriment ou modifient des concepts ou des relations entre ces concepts.

– Une **base de données** pour enregistrer les adresses des SIGs ayant déjà répondu à des requête antérieure afin d'y accéder directement. Cette base de données contient aussi les adresses locales du SIG pour chaque cellule à laquelle il appartient.

La connexion des SIGs de la cellule est assurée, comme le montre la figure (1), grâce à un agent appelé **gestionnaire de cellule** ou encore **GC**. Du moment que chaque cellule possède son GC, les gestionnaires sont à leur tours connectés entre eux. Chaque gestionnaire de cellule est connecté au GCs qui lui sont directement voisins soit six GC au nombre des cellules voisines comme le montre la figure (2).

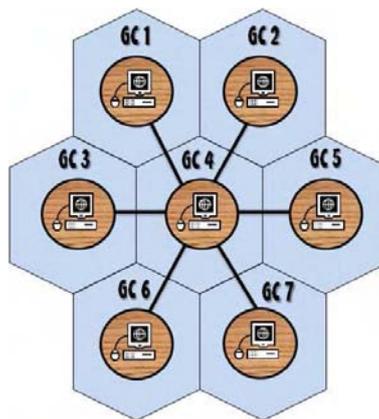


Figure 2. Connexion d'un GC avec ses voisins directs

Un gestionnaire de cellule possède également un "module d'interopérabilité" composé des éléments suivants :

- Un module de traduction des requêtes en langage ontologique
- Une ontologie d'application de la cellule qui assurera l'analyse syntaxique et l'orientation des requêtes vers les SIG de la cellules (plus précisément vers les ontologies de domaines)
- Une base de données qui contiendra :
 - La liste des cellules voisines et les SIGs partagés avec ces cellules ;
 - Les adresses des SIGs qui ont répondu à des anciennes requêtes ;

Les gestionnaires sont connectés à un agent appelé **gestionnaire global de cellule** ou encore **GGC**.

Le GGC est un agent désigné pour jouer le rôle d'intermédiaire entre les différents GCs en particulier ceux qui ne sont pas voisins directs. En dépit de son appellation qui donne l'impression que sa structure est plus complexe que celle des GCs, le GGC est en réalité une version simplifiée des gestionnaire des cellules et il ne fait que transmettre les requêtes qui lui arrivent vers le destinataire désiré. L'utilisation d'un tel agent permet de réduire les coûts de connexion entre les GCs surtout lors de l'ajout d'une nouvelle cellule au réseau. La figure (3) illustre la connexion entre les gestionnaires des cellules et le gestionnaire global.

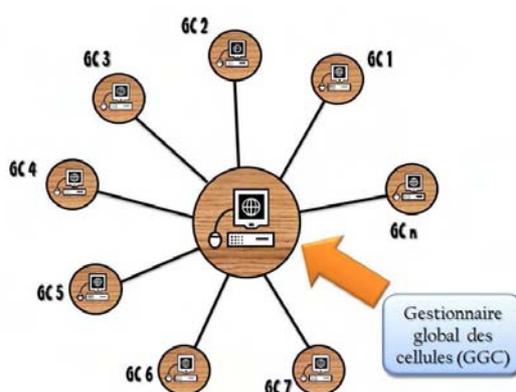


Figure 3. Connexion des GC avec le gestionnaire globale

La présence de agents gestionnaire de cellules crée deux niveau dans notre approche : le niveau SIGs et le niveau gestionnaires. les deux niveaux sont construits sur la même base, c-à-d une connexion avec les voisins directs et une connexion avec un agent assurant le rôle de gestion.

4.2. Le traitement des requêtes

Le traitement des requêtes se fait à base des deux niveaux de "INSIGO" : SIGs et gestionnaires.

4.2.1. Le niveau des SIGs

Le SIG interrogeant tente, en premier lieu, de répondre à la requête localement, et renvoie la réponse à l'utilisateur si elle existe et sinon il transmet la requête aux SIGs

qui lui sont directement voisins.

Chacun de ces derniers tente un traitement local de la requête à condition qu'il soit homogène avec le SIG interrogeant. La requête doit être écrite dans le même langage que celui du SIG voisin. Dans le cas contraire, un rejet de la requête est signifié au SIG envoyant la requête (les rejets sont mentionnés par les points rouges dans le diagramme de séquence niveau SIG).

Si la réponse existe, elle sera transmise dans le chemin inverse qu'a parcouru la requête munie de l'adresse du SIG ayant répondu à la requête pour l'enregistrer dans le module BDD du SIG interrogeant, sinon, les SIGs voisins transmettrons à leur tour la requête à leurs voisins et le cycle continue tant qu'il n'y a pas de rejet ou que la requête retourne au SIG interrogeant sans réponse(modèle cylindrique circulaire). Le diagramme de séquence correspondant au traitement des requêtes au niveau des SIGs est illustré par la figure (4).

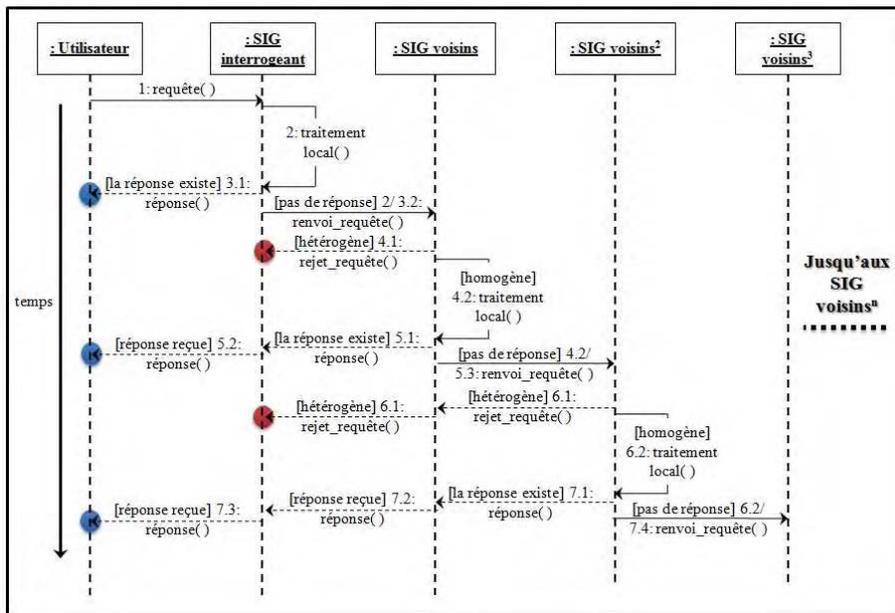


Figure 4. Diagramme de séquence, niveau SIG

La figure (5) montre l'organigramme du fonctionnement de l'approche au niveau des SIGs.

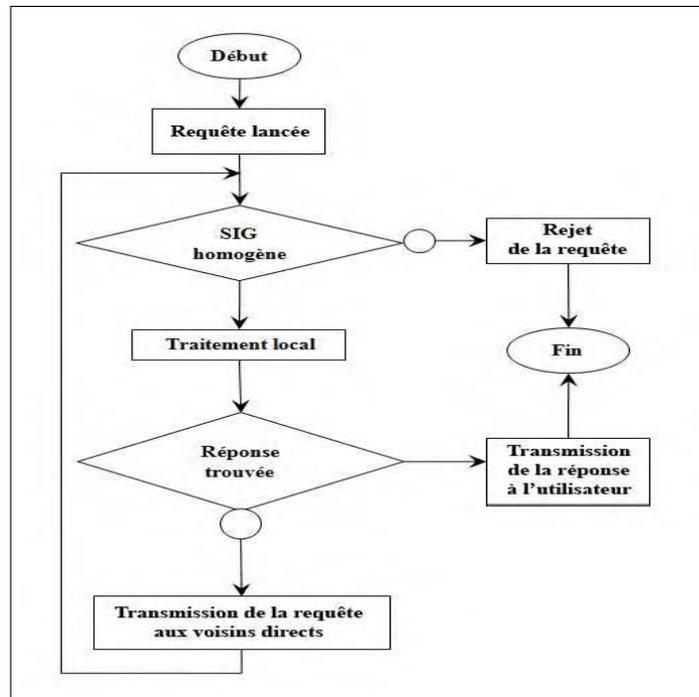


Figure 5. L'organigramme du fonctionnement du niveau des SIGs

4.2.2. Le niveau des gestionnaires

Le deuxième diagramme illustre le passage au deuxième niveau (niveau gestionnaire) à partir des points rouges sur le premier diagramme tel que ces points représentent les rejets de la requête. Une fois la requête rejetée ou revenue au SIG interrogeant sans réponse, ce dernier la transmet au gestionnaire de la cellule qui utilisera son module de traduction afin de transformer la requête en langage ontologique dans le but d'enlever toute sorte d'ambiguïté due à l'hétérogénéité entre les SIGs.

Une fois la traduction terminée, la nouvelle requête est vérifiée syntaxiquement par l'ontologie d'application pour éviter toutes sortes d'erreurs, ensuite le GC accédera aux ontologies de domaine de chaque SIG de sa cellule sauf l'interrogeant dans une tentative de trouver un SIG qui peut répondre à la requête et transmettre la réponse au SIG interrogeant. Si cette tentative échoue, le GC consulte ses voisins qui vont à leurs tours consulter les SIGs de leurs cellules et ainsi de suite jusqu'à ce que la réponse soit trouvée.

A ce moment, le GGC intervient pour assurer une communication directe entre le GC qui détient la réponse et le GC interrogeant afin de lui transmettre cette réponse

qui sera munie de l'adresse du SIG lui ayant répondu. cette adresse est enregistrée à la fois dans le module BDD du SIG interrogeant ainsi que dans le module BDD du GC. La figure (6) montre le diagramme de séquence correspondant au traitement des requêtes au niveau des gestionnaires.

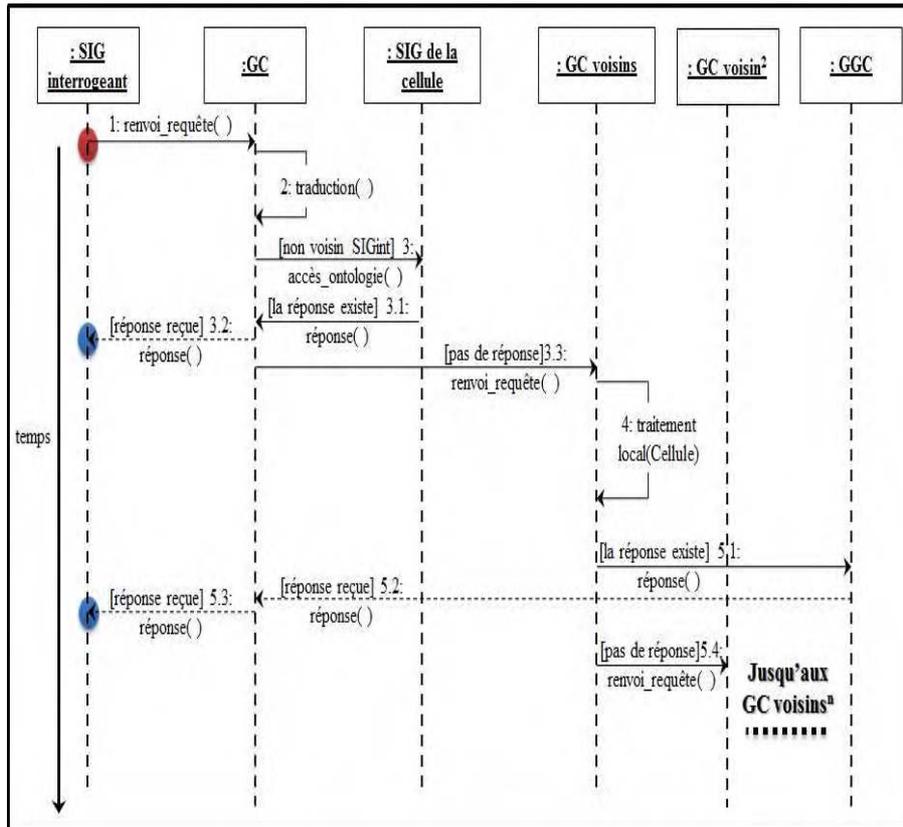


Figure 6. Diagramme de séquence niveau gestionnaire

La figure (7) montre l'organigramme du fonctionnement de l'approche au niveau des gestionnaires.

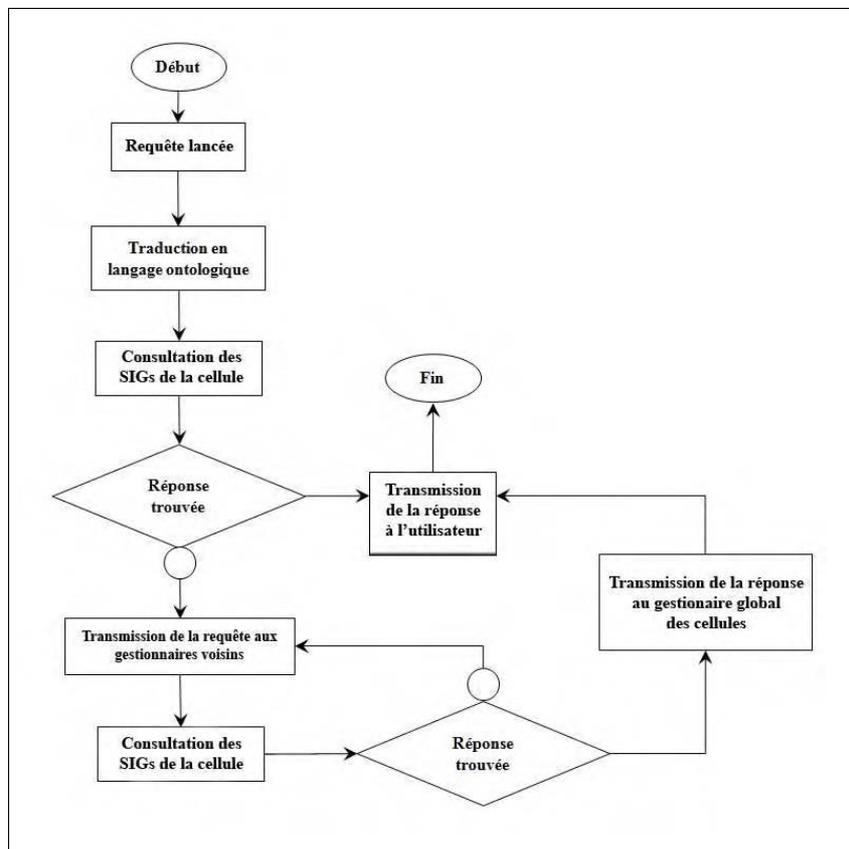


Figure 7. L'organigramme du fonctionnement du niveau des GCs

4.3. Avantages de l'approche

"INSIGO" adopte un traitement basé sur le même principe, traitement local suivi d'une consultation des voisins jusqu'à ce que la réponse soit trouvée ; cette uniformité du principe facilite la réalisation de cette approche. D'autre part, l'utilisation des agents pour la gestion des cellules au lieu de SIGs, permet au SIG de se consacrer à répondre aux requêtes et au GC d'en gérer le trafic. Il faut aussi noter le plus important à savoir l'utilisation des ontologies (la traduction de la requête en langage ontologique) permet d'assurer une homogénéité entre les SIGs sans pour autant être obligé d'uniformiser leurs langages internes.

5. Conclusion

La communication entre les systèmes d'informations géographiques est devenue une nécessité incontournable (catastrophes naturelles, météo, recherche scientifique ...). Cependant, partager les informations est une tâche variée et complexe en même temps. Cela est dû non seulement à la spécificité de chaque système, voire son organisation et le langage qu'il utilise, mais aussi à son type, le mode de transfert des données échangée et le coût des technologies utilisées pour répondre au besoin des utilisateurs.

L'interopérabilité pourra sans doute relever tous ces défis et en même temps rassembler toutes les préoccupations qui y sont attachées. En connaissance de tous cela, le besoin de techniques assurant cette interopérabilité se montre plus que nécessaire. Parmi ces techniques nous avons cité l'utilisation des applications et services Web, ainsi que les modèles basés sur des logiques terminologiques ou encore les modèles sémantiques.

En dépit des avantages de ces techniques, nous avons ciblé l'utilisation des ontologies comme technique qui se montre capable de résoudre le problème dans plusieurs domaines et disciplines de l'informatique pure à la médecine en passant par la biologie et la géographie.

Cette efficacité de l'ontologie est exprimée par sa définition simplifiée. L'association des ontologies et des agents informatiques pourra donner naissance à des structures flexibles et facile à implémenter, ce qui permet une large panoplie d'idées et d'architectures parmi lesquelles se situe notre approche "**INSIGO**" basée sur une connexion de cellules hexagonales ainsi que les agents gestionnaires de cellules.

Le traitement des requêtes se fait à deux niveaux en se basant sur un traitement local suivi d'une consultation des voisins directs (au niveau SIG tout comme au niveau GC), sachant que le modèle cylindrique construit de plusieurs cellules hexagonales permet un voisinage circulaire important. Enfin, bien que cette approche soit structurée en deux niveaux, ces derniers utilisent le même principe de base.

6. Bibliographie

- [KHO 07] *Conception d'une nouvelle ontologie pour l'interopérabilité des systèmes d'informations géographique*. Hacene BELHADEF et Dr Mohamed-Khireddine KHOLLADI, 4th International Conference : Sciences of Electronic, Technologies of Information and Telecommunications March 25-29, 2007 - TUNISIA
- [GRU 93] Gruber T.R., 1993, *A translation Approach to Portable Ontology Specifications*, Technical report KSL 92-71, Knowledge systems laboratory, Stanford University, California, 27p
- [SMI 01] Smith B., 2001, *Geographical categories : an ontological investigation*, in international journal of geographical information science, Vol.15 No. 7, pp. 591-612.
- [FLE et AI 09] Yann FLETY, Marie-Hélène DE SEDE, Vers une géo-ontologie pour les Systèmes Energétiques Territoriaux, XVI rencontres de Rochebrune, "Ontologies et dynamique des systèmes complexes", Janvier 2009, Megève.
- [GAN 02] Gandon F., 2002, *Distributed artificial intelligence and knowledge mangement : Ontologies and multi-agent systems for a coperate semantic web*, Thèse de Doctorat, University of Nice and INRIA, 486p.
- [SMI 04] Smith B., 2004, Beyond concepts : ontology as reality representation, in Proceedings of the international Conference on Formal Ontology and Information Systems FOIS 2004, Turin, Italy, 4-6 November 2004, 12p.
- [LEC et AI 98] Eric Leclercq, Djamel Benslimane, Kokou Yetongnon, *"ISIS : une architecture multi- agents pour l'interopérabilité des SIG"*, colloque SMAGETn Cemagref-Engref, Clermont-Ferrand, France 1998.
- [NOY et AI 00] *"Développement d'une ontologie 101 : Guide pour la création de votre première ontologie"* Natalya F. Noy et Deborah L. McGuinness Université de Stanford, Stanford, CA, 94305
- [MOR et AI 97] *Le Robert Collège* Daniel MORVAN, François GÉRARDIN, Annick DEHAIS, Brigitte VIENNE, Marie-José BROCHARD, Alette LUCOT, Alain REY, Dictionnaires IE ROBERT, 1997
- [PHI et AI 05] *De l'interopérabilité des SIG* Claude Philipona, Alexandre Fellay, bulletin HEC 71, Entreprise Camptocamp SA, FRANCE, 2005
- [DEG et AI 97] *Les enjeux de l'interopérabilité sémantique dans les systèmes d'information de santé* Patrice Degoulet, Marius Fieschi, Christophe Attali, Informatique et Gestion Médicalisée, Volume 9 Springer-Verlag France, Paris, 1997
- [PSY et AI 03] *Apport de l'ingénierie ontologique aux environnements de formation à distance*, Valéry PSYCHÉ, Olavo MENDES et Jacqueline BOURDEAU. Revue STICEF, 10, 2003
- [MEL 07] *Réalisation de l'interopérabilité sémantique des systèmes, basée sur les ontologies et les flux d'information* Thèse pour obtenir le grade de docteur de l'université de Savoie, Naçima MELLAL, Laboratoire d'Informatique, Systèmes, Traitement de l'Information et de la Connaissance, Université de Savoie, 2007
- [GUA et AL 95] *Ontologies and knowledge bases : Towards a terminological clarification*, Nicola GUARINO et Pierdaniele GIARETTA. In N MARS, réd., Towards Very Large Knowledge Bases : Knowledge Building and Knowledge Sharing, pages 25-32. IOS Press, 1995.

AOC, une ontologie formelle pour la modélisation de systèmes complexes en géographie

Patrice Langlois — Eric Daudé — Baptiste Blanpain — Emmanuel Sapin

UMR CNRS 6228 "IDEES", équipe MTG

Université de Rouen

1 rue Thomas Becket, bât. IRED, 76821 Mont-Saint-Aignan Cedex

Patrice.langlois@univ-rouen.fr

RÉSUMÉ. Cet article présente un méta-modèle nommé AOC (Agents-Organisations-Comportements), qui s'exprime par une ontologie formelle et sert de cadre conceptuel à une plateforme de simulation de systèmes complexes spatialisés, compatible avec les problématiques de dynamiques géographiques les plus larges. Le modélisateur géographe doit pouvoir concevoir son modèle et le formaliser dans un langage cohérent relativement à son cadre de pensée. AOC offre ainsi la possibilité de structurer les relations entre agents via le concept d'organisation, qui permet de créer des modèles multi-niveaux. De plus, une formalisation générale de la notion de comportement permet de modéliser la dynamique d'une classe très large de systèmes.

MOTS-CLÉS : Système complexe, système multi-agents, ontologie géographique, méta-modèle, dynamique spatiale, plateforme de modélisation.

ABSTRACT. This paper presents a meta-model named AOC (Agents-Organisation-Behaviour) that is described by a formal ontology. This meta-model is the theoretical concept of a simulation platform of spatial complex systems that are compatible with the largest dynamic geographic problems. The geographer should be able to conceive his model and to formalise it in a coherent language according to his way of thinking. Thanks to the concept of 'organisation', AOC offers the possibilities to structure the relationships between agents and to create multi-level models. Moreover, a general formalisation of the notion of behaviour is used to simulate the dynamic of very large types of systems.

KEYWORDS: Complex system, multi-agent system, geographic ontology, meta-model, spatial dynamic, simulation platform.

1. Introduction

Une problématique fondamentale en informatique moderne est l'étude des systèmes dits complexes, dans lesquels des agents indépendants peuvent interagir pour s'influencer entre eux et faire évoluer le système. Parmi les systèmes complexes, les systèmes multi-agents, dans lesquels des agents artificiels opèrent collectivement et de façon décentralisée pour accomplir une tâche ou simuler un phénomène, sont étudiés dans de nombreuses disciplines (informatique, géographie, sociologie, économie, biologie, etc.), par des informaticiens comme par des non informaticiens (Phan et al., 2007)

La nécessité d'un cadre théorique dans lequel le modélisateur peut penser son modèle et le formaliser est grandissante. Dans cet article, nous présentons l'ontologie d'un paradigme, c'est-à-dire un cadre conceptuel associé à un moyen de produire une certaine classe de modèles (Ramat, 2006). Ce paradigme, résumé par l'acronyme AOC (Agents-Organisations-Comportements), sert de base à une plateforme intégrée permettant la conception, le codage et la simulation de systèmes complexes en géographie (programme ANR et FEDER « MAGEO »).

Notre approche s'est inspirée initialement de l'ontologie AGR (Agent-Groupe-Rôle) de Ferber (Ferber, 1995), (Ferber et al., 2005), en y ajoutant le concept d'organisation multi-échelles, qui s'applique aussi bien à la structuration spatiale que sociale. Par ailleurs, le concept d'agent, dans le contexte AOC, s'applique aussi bien à des individus sociaux qu'à des entités structurantes de l'environnement. En effet, l'environnement est, pour le géographe, souvent autant le sujet d'étude que les individus qui s'y meuvent, même si la temporalité de sa dynamique est différente de celles des individus qui s'y déplacent.

AOC est une manière de représenter le monde à travers des modèles, par combinaison de quelques concepts simples à travers une certaine formalisation. Cet article présente le paradigme AOC sous forme d'une ontologie formelle orientée utilisateur, indépendante de toute application informatique.

On peut dire que l'ontologie présentée est formelle car ses entités sont abstraites (universelles) et permettent de générer des objets concrets (particuliers) dans un modèle. Le paradigme AOC fournit un cadre théorique au modélisateur qui est indépendant de l'informatique et dans lequel il peut penser son modèle, pour le décrire ensuite dans un langage utilisateur cohérent par rapport à ce cadre théorique. Ce cadre doit donc être facilement transposable à sa discipline, à ses thématiques, à ses méthodes et à son langage ; il est donc proche de sa manière de penser, à l'instar du langage à base de règles de SpaCelle (Dubos-Paillard et al., 2003). Ce paradigme offre un ensemble de concepts à la fois suffisamment abstraits pour que le modélisateur puisse y puiser une grande variété de réalisations, mais aussi suffisamment simples afin qu'il puisse les utiliser facilement. Enfin, le contexte théorique de la plateforme ne doit pas être alourdi et obscurci par des aspects

techniques purement informatiques n'ayant rien à voir avec la problématique du modélisateur. Dans cet article, nous ne décrivons donc pas une ontologie de la plateforme elle-même, mais l'ontologie de la classe des modèles qu'elle peut générer.

2. Définition des concepts de AOC

Après avoir exposé les concepts préalables d'espace et de temps, nous décrivons les concepts spécifiques à AOC que sont les agents, les organisations et les comportements.

2.1. *L'espace et le temps*

Toute modélisation géographique se situe à une certaine échelle d'espace et de temps. Cette échelle se caractérise par une extension (un domaine d'espace et de temps) et une granularité (discrétisation). L'espace géographique est constitué d'un référentiel géométrique paramétrable (nombre de dimensions, origine géographique, orientation des axes, unités, pas élémentaires, extension du domaine et métrique). De même le temps est représenté par un référentiel monodimensionnel qui définit une temporalité (par une origine, une unité de temps, un pas de temps élémentaire, une extension). A un référentiel d'espace on associe un référentiel de temps qui lui est compatible en termes de discrétisation et d'extension, aussi nous parlerons de référentiel d'espace-temps.

Les problématiques géographiques utilisent souvent plusieurs échelles de représentation des phénomènes. A chaque échelle est associé un référentiel d'espace-temps. Néanmoins il est nécessaire de définir une correspondance entre ces référentiels, en prenant le référentiel le plus fin (dit référentiel général) comme base de tous les autres.

Dans l'ontologie AOC, les objets du monde se localisent dans des référentiels d'espace-temps, associés à des organisations spatiales qui ont pour rôle de gérer les contraintes de déplacement et les interactions avec les autres objets. Ces organisations spatiales sont abordées dans la partie 2.3.

2.2. *Agent*

Dans le paradigme des systèmes multi-agents, un agent est une entité qui possède une certaine individualité, une indépendance, une certaine autonomie d'action, et une connaissance partielle de son environnement. Un agent possède une mémoire (dont la valeur à un instant donné représente son état) lui donnant la possibilité de conserver, à travers le temps, certaines de ses caractéristiques matérielles, énergétiques ou informationnelles. La dynamique d'un agent s'exprime par des changements internes de son état et des changements externes, par ses

actions qui modifient l'état des objets de son environnement. Elle est décrite dans la partie 2.4.

Un agent représente donc une entité du monde réel ayant un comportement. Cette entité peut être aussi bien une maison qu'un être humain, voire une entité abstraite.

A un instant donné, un agent peut être à la fois « positionné » socialement et localisé dans l'espace géographique.

Agent géographique

Un agent géographique est un agent localisé (pays, ville, individu) qui peut avoir une dynamique. Il appartient souvent à une couche d'information géographique, en général importée d'un SIG (Système d'Information Géographique). Il possède une emprise spatiale (ponctuelle, linéaire, ou surfacique), appelée aussi son support. Cette localisation peut être exprimée dans le repère de l'espace général ou être relative à un composant sur lequel il se localise (par exemple, un voyageur peut avoir des coordonnées terrestres absolues, ou avoir seulement une localisation relative dans le train dont il est passager). Un agent géographique possède donc des attributs spatiaux qui servent à décrire sa forme et sa localisation dans un référentiel donné.

2.3. Organisation

Si l'agent modélise ce qui est de l'ordre de l'individuel ou de l'élémentaire, dans AOC, une organisation modélise ce qui est de l'ordre du collectif. Mais ne nous trompons pas, une organisation n'est pas une collection d'objets ou un ensemble d'éléments ; une organisation est la structure qui permet à ces objets ou à ces éléments de se relier, de se séparer, de se combiner, de communiquer, d'interagir entre eux. En somme, l'organisation est ce qui fait la différence entre un ensemble d'objets et un système d'objets.

Définition : une organisation O définie sur un ensemble A d'agents, appelés ses composants, est une structure S munie d'un ensemble de règles d'évolution et de gestion des interactions avec les autres agents du système.

Cette définition, très générale, recouvre des façons très diverses de structurer des collectivités d'agents. Ces différentes approches peuvent néanmoins être classées en plusieurs catégories, comme nous allons le voir.

Si l'organisation structure les agents dans le référentiel spatial, on parle d'organisation spatiale, sinon on parle d'organisation sociale. L'adjectif social est pris ici dans un sens très large, qui va au-delà des exemples précédents.

Organisations spatiales

Les organisations spatiales disposent d'une position et d'un volume dans le référentiel spatial du modèle ; elles y organisent les agents les uns par rapport aux autres. Parmi les organisations spatiales, on peut distinguer :

- Les organisations formées de composants définissant un maillage de l'espace. Ces maillages peuvent être réguliers (rectangles, hexagones, triangles), ou irréguliers (découpages administratifs).
- Les organisations qui portent sur des composants matériels dissociés les uns des autres, donc ne représentant pas une partition de l'espace. Ces composants peuvent occuper des volumes nuls (semis de points), ou plus ou moins étendus dans une ou plusieurs dimensions (réseaux routiers faits de tronçons et de nœuds, ou les agglomérations dans un espace national).
- Les organisations qui portent sur des composants situés, mais non matériels, ce qui leur confère la propriété d'être interpénétrables. Les composants ne décrivent donc pas nécessairement une partition du volume, puisqu'ils peuvent couvrir des volumes d'intersection non nulle (zones d'influence religieuse, zones d'influence commerciale, etc.).

Organisations sociales

Les organisations dites sociales sont les organisations qui structurent leurs composants les uns par rapport aux autres sur d'autres critères que leur position spatiale. Les deux catégories remarquables sont :

- Les organisations qui structurent un ensemble d'agents abstraits (c'est-à-dire sans réalité matérielle) les uns par rapport aux autres. Un exemple peut être l'ensemble des catégories socioprofessionnelles (CSP), où chaque catégorie prend le statut d'agent ; un autre exemple est celui des différentes fonctions d'une entreprise, à savoir les fonctions d'administration, de production, de recherche, de direction, etc. Ces différents composants abstraits (qui, rappelons le, possèdent le statut d'agents), trouvent leur intérêt en permettant à d'autres agents de s'en déclarer membre (nous parlons de « passagers »), ce qui confère à ces derniers des compétences associées au composant qui les accueille. Ce point concerne également toutes les organisations qui agencent des composants les uns par rapport aux autres ; il sera développé dans la partie 3.1.
- Les organisations qui structurent socialement des agents (par exemple des individus sociaux), par des liens qui expriment des relations pouvant être de natures variées : liens mono ou bidirectionnels, liens étiquetés ou non (texte, pondération), etc.

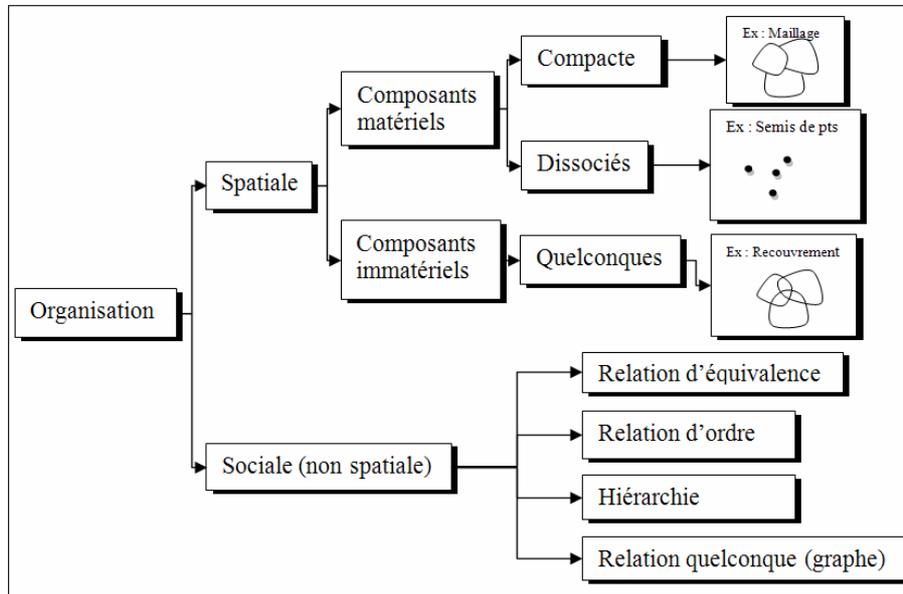


Figure 1. Hiérarchie des organisations

2.4. Compétence, comportement et dynamique

La dynamique du système résulte du comportement de ses agents qui agissent et interagissent entre eux et avec leur environnement.

La dynamique d'un agent est l'expression de ses compétences au cours du temps, (c'est-à-dire le processus d'exécution de ses compétences). Cette dynamique sera appelée comportement de l'agent.

Le comportement d'un agent peut se manifester de différentes manières (non exclusives entre elles) :

- Affecter son support spatial : si celui-ci subit une transformation isométrique qui, à chaque pas de temps, effectue une translation suivie d'une rotation, on dit que l'agent est mobile (son support reste invariant par rapport à un repère local attaché à l'agent). Si ses attributs spatiaux subissent un autre type de transformation géométrique, on dit que l'agent est déformable.
- Lorsque la transformation s'opère sur des attributs non spatiaux de l'agent (internes ou relationnels) on parle de dynamique évolutive (interne ou relationnelle) de l'agent.
- Lorsque des agents ou des objets (matière, flux, énergie, etc.) se transforment au cours du temps en d'autres agents ou d'autres objets,

comme des réactions chimiques, des dynamiques foncières (remembrement), ou des possibilités de reproduction ou de réplication, on parlera de dynamique transformationnelle.

Une compétence peut être innée (permanente), lorsqu'elle est constitutive de l'agent, mais peut également être acquise (temporaire), lorsqu'elle lui est offerte par une organisation (ou un de ses composants) dont il est passager à un instant donné ; l'agent peut ainsi adapter son comportement à son environnement social et spatial. Par exemple, dans une organisation qui décrit la « structure familiale d'un ménage », le composant « père » dispose de compétences spécifiques à ce rôle, ce qui permet à un agent passager de ce composant, d'acquérir, en plus de ses compétences innées, ces compétences temporaires. Mais ensuite, ce même agent humain partant travailler, endossera le rôle d'automobiliste et adoptera ainsi les compétences liées à la conduite automobile, afin de se déplacer sur l'organisation spatiale de type réseau routier.

Le comportement global d'un agent est en général la combinaison de processus plus simples, que l'on appelle comportements élémentaires, et qui sont l'expression (l'exécution en informatique) de compétences élémentaires, celles-ci étant des transformations de l'agent en fonction de son état présent et de celui de son environnement.

Formalisation de compétences simples

Notre modèle de compétence part du principe qu'à un instant donné, il peut, ou non, être pertinent d'exécuter une compétence donnée. Tout agent possède des compétences innées ou acquises, qu'il exécute ou pas en fonction du contexte. Selon le contexte social, spatial ou temporel, il sera pertinent ou non d'exécuter une *action* en particulier. Dans le modèle comportemental proposé, la pertinence liée à l'environnement présent sera exprimée par une clause *pertinence* associée à chaque compétence. Cette clause est une fonction qui renvoie une valeur réelle entre 0 et 1 à partir d'une évaluation de l'état de l'agent et de son environnement.

Les bases de cette approche ont été développées dans le modèle d'agents cellulaires décrit dans (Dubos-Paillard et al., 2003), (Langlois, 2010), basé sur le couple *compétence* = \langle *pertinence, action* \rangle . Nous y avons ajouté un troisième terme nommé *satisfaction*, pour obtenir une modélisation de compétence sous forme d'un triplet :

$$\textit{compétence} = \langle \textit{pertinence}, \textit{action}, \textit{satisfaction} \rangle.$$

La notion de pertinence d'une compétence permet de dépasser le simple conditionnement de l'exécution d'une action à une évaluation logique (de type SI pertinent ALORS action). En effet, la pertinence d'une compétence a pour but d'être évaluée et comparée à celles des autres compétences qu'un agent pourrait utiliser à un instant donné. Si plusieurs compétences sont en concurrence à un instant donné, seule la plus pertinente verra son action exécutée par l'agent.

L'action décrit les opérations exécutées par l'agent, autant sur ses attributs que sur son environnement.

La satisfaction est un nombre réel entre 0 et 1 représentant l'évaluation de la bonne exécution de l'action associée. Elle vaut zéro si celle-ci n'a pas été exécutée. Elle peut servir de pertinence pour une action suivante, lors de l'enchaînement séquentiel de deux comportements ; elle peut également permettre de quantifier l'état d'avancement vers la réalisation d'un objectif.

Formalisation de compétences complexes

La clause pertinence offre donc la possibilité de choisir, parmi une liste de compétences, celle qu'il convient d'exécuter à un instant donné. Le choix le plus naturel consiste à choisir la plus pertinente en cet instant (on note cette opération C_1 / C_2), mais on peut également choisir une compétence tirée au hasard ($C_1 ? C_2$), avec éventuellement une pondération par la pertinence.

D'autres opérations permettent de faire se succéder l'exécution de plusieurs compétences dans le temps : la plus classique est l'exécution séquentielle d'une série de compétences élémentaires. Dans le cas de deux compétences C_1 et C_2 , leur séquence s'écrit « $C_1 ; C_2$ » et signifie simplement qu'à l'exécution de C_1 succèdera celle de C_2 .

D'autres possibilités sont l'exécution en parallèle de plusieurs compétences (la marche met en jeu de nombreux mouvements élémentaires simultanés), ou encore l'exécution séquentielle avec un *GO/NOGO* après l'exécution de chaque compétence : c'est la séquence conditionnelle « $C_1 ; > C_2$ », qui signifie qu'après l'exécution de C_1 , la compétence C_2 ne sera examinée que si la satisfaction de C_1 n'est pas nulle ; C_2 ne sera donc jamais exécutée si l'action de C_1 ne l'a pas été.

Il est possible d'ordonner les compétences de plusieurs autres façons, par exemple en séquence dans l'ordre décroissant des pertinences (exécution de plusieurs tâches avec un ordre de priorité), ou encore en exécutant une compétence autant de fois qu'il est nécessaire, soit jusqu'à ce qu'une autre compétence devienne pertinente (un agent se déplace jusqu'à ce qu'il ait trouvé un objet à ramasser), soit jusqu'à ce qu'un événement se produise (un agent se déplace jusqu'à ce qu'on lui dise « stop »).

3. Relations entre les concepts de AOC

3.1. La relation de positionnement / localisation

Pour modéliser le fait qu'un ensemble A d'agents qui sont les composants d'une organisation, se localisent et se déplacent dans un environnement lui-même formé d'un ensemble d'agents B composants d'une autre organisation, on utilise une fonction de positionnement de A vers B , qui associe à tout agent de A un agent au

plus dans B . Les liens de cette relation sont des couples (passager, hôte) de la forme $(a, b) \in A \times B$, qui s'expriment par une phrase du type « le passager a est sur l'hôte b ». Si de plus le composant b est un « conteneur spatial » le passager a est alors localisé relativement à ce composant (qu'on appelle alors son localisateur). Par exemple, si b est un tronçon de route (élément linéaire), a peut être localisé par son abscisse curviligne relative à b . Si b est un carreau (élément surfacique) a peut être localisé par deux coordonnées relativement à b .

Par exemple, un agent « wagon de voyageurs » peut être modélisé en lui adjoignant une organisation spatiale composée de sièges et d'un couloir qui sont les composants de l'organisation du wagon, par abus de langage on dira simplement que ce sont les composants du wagon. Les voyageurs sont positionnés dans le wagon, comme passagers, chacun étant localisé sur un siège ou sur le couloir, qui sont les hôtes. Les passagers peuvent éventuellement se déplacer dans le wagon, selon certaines règles. A une autre échelle, un wagon peut être composant d'un train, si l'on considère qu'il est fixe dans la composition de ce train, même si le train est mobile (avec ses wagons et ses voyageurs) sur la ligne qu'il emprunte (il est alors passager sur un tronçon de ligne du réseau ferré). La notion de composant et de passager est donc relative à une organisation.

Les passagers sont, via une relation de localisation, les usagers de l'organisation. Ils interagissent avec elle et, éventuellement, la font évoluer. Un agent peut être l'hôte de plusieurs passagers, mais un agent ne peut être passager que d'un unique composant hôte d'une couche d'organisation. Un agent peut en revanche être passager de plusieurs organisations à la fois. Ainsi, un homme sera passager d'un service d'une entreprise, d'une CSP, et d'une maille spatiale. Un agent peut également être passager de plusieurs couches d'une même organisation, par exemple une maison est à la fois dans une commune, un département et une région de l'organisation administrative du territoire français.

3.2. Mobilité, processus d'interaction

L'organisation possède des méthodes et des données de gestion collective. Elle est capable d'autoriser ou refuser l'arrivée, le départ, la création ou la suppression d'un nouvel agent comme passager sur un hôte. Elle gère la mobilité des passagers sur ses composants-hôtes et d'une manière générale les interactions.

La dynamique de la relation de localisation définit le processus de mobilité des agents à travers les compétences offertes aux passagers par l'organisation-hôte. Ainsi, la mobilité apparaît comme un exemple simple (et fréquent) de processus d'interaction, dans lequel un agent a passé d'un lieu i à un lieu j au cours d'un pas de temps élémentaire. Il y a donc interaction entre les lieux i et j qui échangent simultanément leur contenu.

D'une manière générale un processus d'interaction est défini sur un ensemble d'agents A (appelés hôtes dans (Langlois et al., 2007) sous le contrôle d'un ensemble M de médiateurs. Par exemple la collision entre deux boules de billard, la

prédation d'un loup sur un mouton, la transmission d'une maladie d'une personne infectée à une personne saine, sont des interactions élémentaires qui entrent dans des processus d'interaction. Ainsi, un processus d'interaction est composé de trois entités, $P = \langle A, M, f \rangle$. A est l'ensemble des agents (A peut être hétérogène : les loups et les moutons, ou homogène : les boules de billard). M est l'ensemble des médiateurs de l'interaction à travers lesquels s'effectuent les interactions élémentaires (M est appelé parfois le vecteur). La fonction f représente ici de manière théorique le processus, c'est-à-dire le mécanisme d'évolution des états des agents, lorsque les médiateurs opèrent sur eux. Les médiateurs peuvent être représentés par les « particules » (vision discrète) qui sont échangées ou par le flux ou le champ (vision continue) qui s'applique lors d'une interaction élémentaire entre deux agents de A . Par exemple, dans l'interaction des boules de billard, le médiateur est la quantité de mouvement échangée sur le lieu de la collision. Dans la prédation, la proie disparaît au profit de l'énergie gagnée par le (ou les) prédateur(s). L'épidémie se propage entre les individus par transmission et multiplication du vecteur (bactérie ou virus). Comme pour la mobilité, c'est encore l'organisation des agents de A qui gère collectivement l'interaction.

3.3. Modèles multi-niveaux : Emboîtement agent – organisation – agents

L'organisation apparaît toujours comme un intermédiaire entre un agent global et ses agents internes. Un agent n'est jamais considéré comme contenant directement une collection d'agents, mais comme contenant une ou plusieurs organisations, qui gèrent des composants internes et des passagers.

Un agent est dit terminal (ou simple ou élémentaire), s'il ne contient pas d'organisation interne, et par conséquent pas de composant interne. Dans le cas contraire il est non-terminal (ou composé ou complexe). L'agent est alors un système composé d'organisations d'agents plus élémentaires en interaction. Il possède une double description, externe en tant qu'agent, et interne en tant que système formé de ses composants internes.

Cette approche permet une modélisation multi-échelles d'un modèle, par raffinements successifs à partir d'un agent qui englobe tous les autres. On peut ainsi envisager de modéliser des systèmes multicouches, tels que les maillages géographiques sur plusieurs échelles (par exemple, pays-région-département-canton).

L'agent qui englobe tous les autres peut être appelé agent-système. Son organisation interne décrit les entités et la dynamique du modèle. Ses entrées sont connectées à l'interface homme-machine, permettant à l'utilisateur de paramétrer le modèle et d'interagir en cours de simulation. Ses sorties permettent de visualiser et/ou de mémoriser les observations opérées sur le système au cours de la simulation.

4. Schéma ontologique

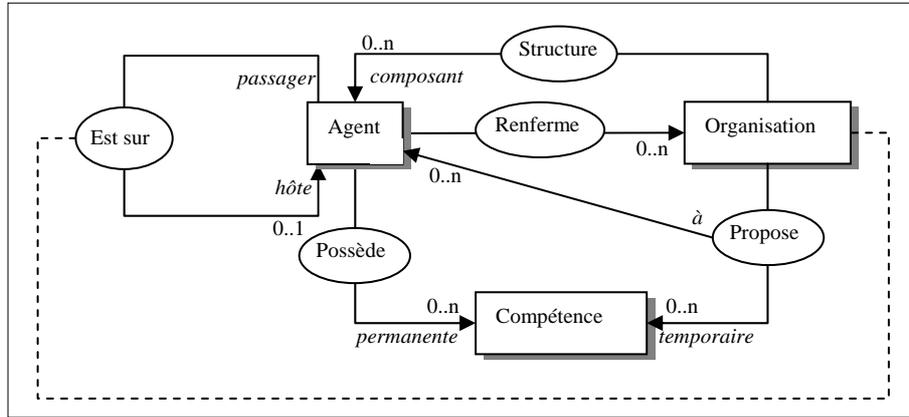


Figure 2. Schéma simplifié du paradigme AOC

Le schéma de la Figure 2 résume les principales relations entre les concepts d'agent, d'organisation et de comportement : une organisation structure un ensemble d'agents qui sont alors composants de l'organisation. Mais un agent complexe peut aussi renfermer en son sein des organisations qui structurent ses constituants internes. Une compétence peut être utilisée de manière permanente par un agent, ou bien être proposée par une organisation de manière temporaire aux passagers. Une relation de positionnement/localisation « est sur » est une fonction définie entre deux ensembles d'agents, les passagers et les hôtes. Les contraintes de localisation et de mobilité entre les passagers sont gérées par l'organisation des hôtes.

5. Application

Le modèle des termites (Wilensky, 1997) a été choisi pour être modélisé avec l'approche AOC. Ce modèle peut être décrit de la manière suivante : des termites se déplacent sur un espace discret formé d'un carroyage, une case ne pouvant contenir plus d'un termite à la fois. Des grains sont disposés dans les cases (un grain au plus par case). Les termites ramassent les grains qu'elles trouvent, les portent au cours de leur déplacement et les déposent à proximité d'autres grains. Un grain est donc, à un instant donné, soit dans une case, soit porté par un termite. A l'initialisation tous les grains sont dans des cases.

Montrons comment ce modèle peut être formalisé dans le paradigme AOC.

Agents

- Agent *Case* : possèdera un attribut booléen, *YaGrain*, indiquant s'il y a un grain dans la case ;
- Agent *Termite* : possèdera un attribut booléen, *PorteGrain*, indiquant si le termite porte un grain.

Ici les Termites, ainsi que les Cases, sont des agents terminaux.

Organisations

Terrain : Carroyage de *Cases* (par exemple un tore avec voisinage de Moore).

Relation de localisation des *Termites* vers les *Cases* : un termite est passager d'une case et une seule, et il ne peut y avoir plus d'un termite sur une case.

Compétences

Les seuls agents ayant une dynamique propre sont les termites. Cette dynamique se résume aux trois compétences suivantes :

- *SeDéplacer* : le termite se déplace vers une case voisine choisie au hasard ;
- *Prendre* : dans le cas où sa case contient un grain, le termite le ramasse ;
- *Déposer* : dans le cas où le termite porte un grain et se trouve sur une case vide, s'il y a au moins un grain dans son voisinage, alors il dépose le grain qu'il porte.

Traduisons ces compétences sous forme de triplets $\langle \textit{Pertinence}, \textit{Action}, \textit{Satisfaction} \rangle$:

$\textit{SeDéplacer} = \langle P_1, A_1, S_1 \rangle$

$P_1 = 1$ s'il y a au moins une case voisine sans termite, 0 sinon.

$A_1 =$ Se déplacer vers une case sans termite du voisinage choisie au hasard.

$S_1 = 1$ lorsque l'on s'est déplacé, 0 sinon.

$\textit{Prendre} = \langle P_2, A_2, S_2 \rangle$

$P_2 = 1$ si le termite ne porte aucun grain et que sa case en contient un, 0 sinon.

$A_2 =$ Le grain passe de la case au termite.

$S_2 = 1$ lorsque un grain est passé au termite, 0 sinon.

$Déposer = \langle P_3, A_3, S_3 \rangle$

P_3 = Le termite porte un grain, il n'y en a pas dans sa case, et il y en a au moins un dans une de ses cases voisines.

A_3 = Le grain passe du termite à la case.

S_3 = 1 lorsque le grain est déposé, 0 sinon.

Ces compétences doivent maintenant être ordonnancées dans le temps, à l'aide des opérations décrites dans la partie 2.4. Nous proposons deux façons de procéder, mais d'autres approches sont évidemment possibles.

Première approche

Lors d'un pas de temps, un termite peut soit prendre, soit déposer (s'il le peut). Cette première action est nécessairement suivie d'un déplacement (si celui-ci est possible). Le comportement d'un termite peut donc être formulé de la façon suivante :

$(Prendre \mid Déposer) ; SeDéplacer$

Il se peut qu'un termite ne puisse ni prendre ni déposer, dans ce cas le termite va juste se déplacer, si son environnement le permet. S'il ne peut pas se déplacer, il restera bloqué tant que les termites de son voisinage ne lui auront pas laissé de place.

On vérifie facilement que la succession de plusieurs pas de temps va permettre à un termite de réaliser l'ensemble d'un cycle ($SeDéplacer, Prendre, SeDéplacer, Déposer$).

Seconde approche

Une autre façon de formaliser ce processus consiste à réaliser l'ensemble du cycle ($SeDéplacer, Prendre, SeDéplacer, Déposer$) au cours d'un unique pas de temps : le termite (à vide) commence par se déplacer jusqu'à arriver sur une case où il pourra ramasser un grain, puis il se déplace jusqu'à arriver sur une case où il pourra déposer son grain :

JUSQUA $Prendre$ REPETER $SeDéplacer$;

JUSQUA $Déposer$ REPETER $SeDéplacer$;

La compétence *SeDéplacer* va ainsi, par exemple dans la première ligne, être exécutée plusieurs fois, jusqu'à ce que la compétence *Prendre* devienne pertinente, et soit exécutée.

Lorsqu'il vient de déposer, à la fin d'un cycle, le termite recommence le cycle suivant en se déplaçant au moins une fois avec de prendre. Il ne peut donc pas ramasser à nouveau le grain qu'il vient de déposer, à moins que son environnement l'ait empêché de se déplacer.

6. Conclusion

La vision systémique et complexe du monde conduit les chercheurs à investir l'univers des méthodologies individus-centrées, tels que les automates cellulaires et les systèmes multi-agents. Dans ces univers complexes, une grande variété de facteurs sociaux, économiques, naturels et géographiques interagissent et affectent la dynamique des objets étudiés. Que ce soit en biologie, en économie ou en géographie, cette prise en compte de la complexité des relations entre des entités d'un même niveau et de niveaux différents sert de levier au besoin de modèles de simulation informatique, afin de comprendre et prévoir les évolutions possibles de tels systèmes.

Le paradigme AOC permet au chercheur de penser et construire des modèles multi-niveaux et dynamiques. Cette possibilité est offerte grâce à la récursivité de la notion d'agent qui autorise l'emboîtement des échelles sociales et géographiques. Les dynamiques qui s'exercent au sein de ces modèles peuvent être aussi bien pensées au travers d'une modélisation de compétences propres aux organisations et à leurs composants qu'au niveau des agents eux-mêmes. A l'emboîtement des échelles géographiques et des niveaux d'organisation répond l'emboîtement des compétences. Ainsi des compétences complexes, évolutives, peuvent être construites à partir de compétences élémentaires.

Le paradigme AOC est formalisé dans un schéma XML qui permet d'écrire des modèles sous forme de documents XML, indépendamment de toute plateforme de simulation. Cette méthode permet ainsi la communication et le partage de modèles au sein d'une communauté de modélisateurs. La plateforme MAGéo, qui intègre ce schéma XML, permet une modélisation relativement simple basée sur de l'interaction graphique et des règles non algorithmiques, adaptée à des utilisateurs non informaticiens.

7. Références bibliographiques

Dubos-Paillard E., Guermond Y., Langlois P., 2003. Analyse de l'évolution urbaine par automate cellulaire : le modèle SpaCelle. *L'espace géographique*, vol 4, p. 357-378.

- Ferber J., 1995. *Les systèmes multi-agents. Vers une intelligence collective*. InterEditions, Paris.
- Ferber J., Michel F., Baez J., 2005. AGRE: Integrating Environments with organizations. In *E4MAS'04*, Australie, p. 127-134.
- Langlois P., Daudé E., 2007, Concepts et modélisation de la diffusion, *Cybergeo*, n°. 364, <http://cybergeo.revues.org/index2898.html>, 23p.
- Langlois P., 2010. *Simulation des systèmes complexes en géographie, fondements théoriques et applications*. Hermès-Lavoisier, Paris.
- Phan D., Amblard F., 2007. *Agent-based modelling and simulation in the social and human sciences*.
- Ramat E., 2006. Introduction à la modélisation et à la simulation à événements discrets. In Phan D.& Amblard F., 2007, *Agent-based modelling and simulation in the social and human sciences*, p. 87-103.
- Wilensky, U., 1997. NetLogo Termites model. <http://ccl.northwestern.edu/netlogo/models/Termites>. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.

TaxoMap Framework appliqué à l’alignement d’ontologies géographiques dans le projet *GéOnto*

Fayçal Hamdi, Chantal Reynaud, Brigitte Safar

CNRS - Université de Paris-Sud 11 (LRI)
INRIA Saclay Ile-de-France (LEO)
Parc Orsay Université 4 rue Jacques Monod
91893 Orsay France

Faycal.Hamdi, Chantal.Reynaud, Brigitte.Safar@lri.fr

RÉSUMÉ. Les outils actuels d’alignement d’ontologies ne sont pas performants sur tous les domaines ni quelles que soient les ontologies. La qualité des mappings obtenus peut être améliorée si le processus d’alignement tient compte des spécificités des ontologies alignées. L’environnement TaxoMap Framework a été conçu pour aider un expert du domaine à spécifier des traitements sur les mappings produits par l’outil d’alignement TaxoMap, afin de les raffiner ou bien de fusionner, restructurer ou enrichir des ontologies. Nous proposons une démonstration de deux modules implémentés dans cet environnement : le module d’alignement et le module de raffinement.

ABSTRACT. The current ontology alignment tools are not efficient on all domains and whatever the ontologies to be aligned. The quality of the results can be improved if the alignment process take into account the specificities of the ontologies. The environment TaxoMap Framework was conceived to help a domain expert to specify treatments, based on mappings generated by the alignment tool TaxoMap, in order to refine or merge, restructure, or enrich ontologies. We propose a demonstration of two modules implemented in this environment: the alignment module and the refinement module.

MOTS-CLÉS : Alignement d’ontologies, raffinement d’ontologies

KEYWORDS: Ontology alignment, ontology refinement

1. Le module "Alignement"

TaxoMap est un logiciel d'alignement de taxonomies développé au sein de l'équipe IASI du LRI (Université Paris-Sud). L'outil a été conçu pour découvrir des alignements entre des ontologies $O = (C, H)$ exprimées en OWL et supposées décrire le même domaine d'application. C est un ensemble de concepts où chaque concept est caractérisé par un ensemble de labels et H est une hiérarchie de subsomption qui contient un ensemble de relations *isA* établis entre les nœuds correspondants aux concepts. Le processus d'alignement est un processus orienté qui cherche à relier par un *mapping* chaque concept d'une ontologie source à un unique concept de l'ontologie cible.

La découverte des mappings repose sur des techniques variées, terminologiques ou structurelles, qui s'appuient sur l'utilisation de l'analyseur morpho-syntaxique Tree-Tagger [SCH 94]. Celui-ci permet un paramétrage en fonction de la langue, une lemmatisation et une catégorisation des mots qui composent un label. Les sorties de l'analyseur sont utilisées par une mesure de similarité appliquée aux labels des concepts vus comme des ensembles de tri-grammes [LIN 98]. Pour chaque concept c_s de l'ontologie source O_S , la mesure de similarité identifie les concepts de l'ontologie cible O_T , qui ont les meilleures similarités et sont candidats au mappings avec c_s . Les techniques d'alignement appliquées en séquence permettent ensuite de déterminer le concept le plus pertinent parmi les candidats et quelle type de relation (relations d'équivalence, de subsomption ou de proximité) doit être établie entre c_s et ce concept.

2. Le module "Raffinement"

Les tests effectués sur les taxonomies mises à disposition par notre partenaire le COGIT de l'IGN dans le projet *GéOnto*, ont montré que le module d'alignement fournissait de très bons résultats (précision de 92,3%) mais que ces derniers pouvaient encore être améliorés.

Une étude des traitements d'améliorations souhaitées par les experts a montré que celles-ci étaient souvent spécifiques aux ontologies alignées. Pour ne pas faire de *TaxoMap* un outil uniquement dédié à l'alignement de taxonomies topographiques, nous avons développé un environnement permettant aux experts du domaine de spécifier et de mettre en oeuvre des traitements supplémentaires sur des alignements préalablement obtenus. Dans cette démonstration, l'environnement est utilisé pour raffiner l'alignement construit par *TaxoMap*. Il peut aussi être utilisé pour tout autre traitement prenant appui sur les résultats d'un alignement entre ontologies, comme la fusion ou l'enrichissement d'ontologies.

L'environnement permet une spécification déclarative des traitements basés sur les résultats d'un alignement particulier et concernant des ontologies particulières, en utilisant un ensemble de primitives génériques et prédéfinies. Les traitements sont spécifiés sous la forme de pattern composés d'une partie condition et d'une partie action. La partie condition s'exprime au travers d'un ensemble de primitives (prédicats)

permettant d'expliciter l'ensemble des caractéristiques qui doivent être vérifiées pour l'application du pattern. Les caractéristiques peuvent être des relations structurelles entre des concepts ou des relations terminologiques sur leur labels ou d'autres relations.

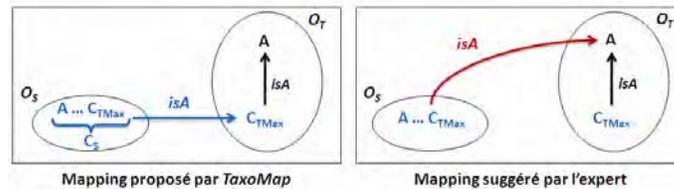


Figure 1. Illustration d'un Pattern

La partie contexte du Pattern :

$$\begin{aligned} & \exists x \exists y (isAStrictInclusion(x, y) \\ & \wedge \exists z (isSubClassOf(y, z, O_T) \wedge strictInclusionLabel(z, x))) \end{aligned}$$

La partie solution du Pattern :

$$Delete_Mapping(x, y, _) \wedge Add_Mapping(x, z, isA)$$

Les traitements qui peuvent être spécifiés dépendent des caractéristiques des ontologies concernées et de la tâche à exécuter (raffinement d'alignement ou enrichissement d'ontologies par exemple). Ces traitements sont associés à des modules de spécification indépendants, un pour chaque tâche, chacun ayant leur propre vocabulaire. Le module est extensible et a priori applicable à n'importe quel traitement basé sur des résultats d'alignement.

3. Bibliographie

- [LIN 98] LIN D., « An Information-Theoretic Definition of Similarity », *In Proceedings of the 15th International Conference on Machine Learning*, 1998, p. 296–304.
- [SCH 94] SCHMID H., « Probabilistic Part-of-Speech Tagging Using Decision Trees », *Proceedings of the International Conference on New Methods in Language Processing*, 1994, p. 44–49.

Vers un système en ligne d'aide à la découverte du contenu de bases de données géographiques

Mechouche Ammar, Abadie Nathalie, Prouteau Emeric, Mustière Sébastien

Laboratoire COGIT, Institut Géographique National

73 Avenue de Paris

94160 Saint-Mandé, FRANCE

ammar.mechouche@ign.fr, nathalie-f.abadie@ign.fr, emeric.prouteau@ign.fr, sebastien.mustiere@ign.fr

RESUME. Nous présentons un prototype d'application Web proposant à un utilisateur de découvrir le contenu de bases de données géographiques. Cette application repose sur des ontologies pour décrire et annoter les bases de données décrites par le système. Cette application inclut une solution de cartographie en ligne pour afficher les données.

ABSTRACT. We present a Web application prototype enabling any user to discover geographic databases' content. It uses ontologies to describe and annotate the different geo-databases considered by the system. A Web mapping application is added to display geo-data samples.

MOTS-CLÉS: Bases de données géographiques, spécifications, ontologie, IHM, cartographie en ligne.

KEYWORDS: Geographic databases, specifications, ontology, user interface, Web mapping.

1. Introduction

Il peut être difficile d'appréhender la complexité des diverses sources de données géographiques désormais disponibles (Craglia et al., 2008). Nous présentons ici un prototype d'application Web proposant à un utilisateur de découvrir le contenu des bases de données de l'IGN en lui permettant de savoir comment et où sont représentées les données qui répondent à son besoin.

2. Architecture du système

Notre système est composé de trois modules principaux. Les choix techniques effectués pour l'implémentation du système sont présentés en figure 1.

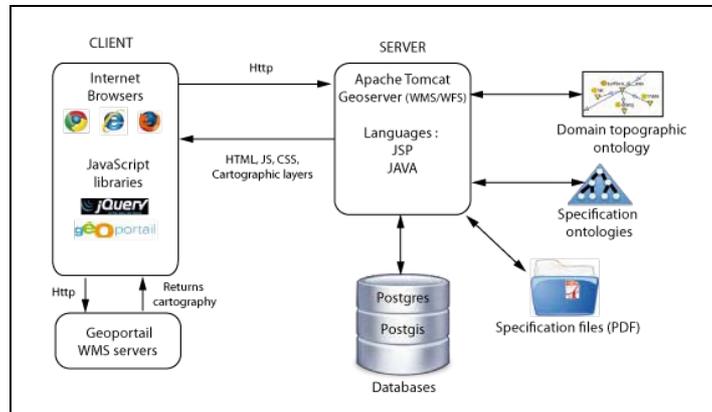


Figure 1: Implémentation du système.

Le module de recherche aide l'utilisateur à exprimer sa requête en lui proposant, à l'aide d'une application d'auto-complétion, d'interroger le système à l'aide d'un vocabulaire unifié fourni par une ontologie globale du domaine de la topographie. Nous utilisons une ontologie bilingue (Français et Anglais), de plus de 700 concepts, créée à partir des textes de spécifications de bases de données géographiques à l'aide d'outils de traitement automatique du langage naturel (Abadie et al., 2008).

Le module d'extraction d'informations recherche les données disponibles correspondant aux termes de la requête de l'utilisateur, en utilisant des ontologies locales de spécifications qui formalisent les spécifications de contenu de chacune des bases de données considérées (Abadie et al., 2010). Chaque classe du schéma d'une base de données y est traduite en OWL et annotée à l'aide des concepts issus de l'ontologie du domaine de la topographie. Des connaissances supplémentaires tirées de ses spécifications, comme des critères de sélection du type « *La classe 'Rivière' comprend tous les cours d'eau permanents de plus de 10 mètres de large* », sont formalisées et ajoutées dans l'axiome utilisé pour l'annoter. Toutes ces connaissances concernant des classes annotées par le concept de l'ontologie globale recherché par l'utilisateur sont ainsi récupérées et renvoyées à l'utilisateur.

Le module cartographique affiche sous forme de cartes les données identifiées par le module d'extraction d'informations afin de permettre à l'utilisateur de visualiser simplement les diverses données disponibles correspondant au thème qui l'intéresse.

3. Conclusion

L'interface Web, composée de trois parties, est présentée en figure 3 : Une zone de texte (en haut) pour saisir les requêtes des utilisateurs, des onglets (à gauche),

correspondant chacun à une base de données et servant à afficher les informations envoyées par le système sur les données issues de cette base et qui correspondent à la requête de l'utilisateur. La troisième, à droite de la page, comporte l'affichage cartographique des données.

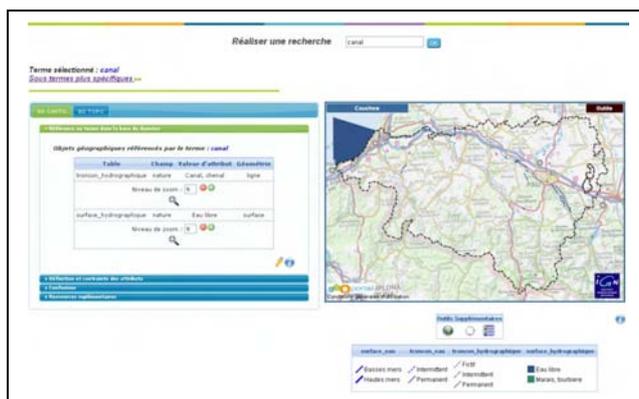


Figure 1: Interface Web du prototype implémenté.

4. Remerciements

Cette recherche a été réalisée dans le cadre du projet GeOnto, financé par l'Agence Nationale de la Recherche (ANR-O7-MDCO-005).

5. Bibliographie

- Abadie N. and Mustière S., « Constitution d'une taxonomie géographique à partir des spécifications de bases de données ». *Colloque International de Géomatique et d'Analyse Spatiale, SAGEO'08*, Montpellier, France, 2008.
- Abadie N., Mechouche A., Mustière M., « OWL-based formalization of geographic databases specifications ». *17th International Conference on Knowledge Engineering and Knowledge Management, EKAW'10*, Lisbonne, Portugal, 11-15 octobre 2010.
- Craglia M, Goodchild M, Annoni A, Camara G, Gould M, Kuhn W, Mark D, Masser I, Maguire D, Liang S, Parsons E. « Next-Generation Digital Earth - A Position Paper from the Vespucci Initiative for the Advancement of Geographic Information Science ». *International Journal of Spatial Data Infrastructures Research*. 3; 2008. p. 146-167. JRC47746 1.5 Article contribution to other periodicals.