

# Estimation de l'aléa pluvial en France métropolitaine

Patrick Arnaud, Jacques Lavabre





# Estimation de l'aléa pluvial en France métropolitaine

Patrick Arnaud, Jacques Lavabre

## Collection Guide pratique

Les requins. Identifier les nageoires

Pascal Deynat

2010, 380 p.

Cactus et plantes succulentes du monde

Francis Bugaret

2010, 240 p.

Mieux intégrer la biodiversité dans la gestion forestière

Marion Gosselin, Yoan Paillet

2010, 156 p.

Les Lamproies en Europe de l'Ouest

Catherine Taverny, Pierre Élie

2010, 112p.

Le potager familial méditerranéen

Charles-Marie Messiaen, Fabienne Messiaen-Pagotto

2009, 192 p.

Utilisation des bois de Guyane pour la construction

Sylvie Mouras, Michel Vernay

2009, 2<sup>e</sup> édition, 160 p.

Les filets maillants

Gérard Deschamps, coordinateur

2009, 272 p.

Durabilité naturelle et préservation des bois tropicaux

Daniel Fouquet

2009, 128 p.

Forêts de protection contre les aléas naturels. Diagnostics et stratégies

Freddy Rey, Jean Ladier, Antoine Hurand, Frédéric Berger, Guy Calès,

Sylvie Simon-Teissier

2009, 112 p.

Les termites dans le monde

Alba Zaremski, Daniel Fouquet, Dominique Louppe

2009, 96 p.

Les orchidées sauvages de Paris

Coordination éditoriale de Sébastien Lesné

2009, 136 p.

© Éditions Quæ 2010 ; ISBN : 978-2-7592-0862-3 ; ISSN 1952-2770

Le code de la propriété intellectuelle interdit la photocopie à usage collectif sans autorisation des ayants droit. Le non-respect de cette disposition met en danger l'édition, notamment scientifique, et est sanctionné pénalement. Toute reproduction, même partielle du présent ouvrage est interdite sans autorisation du Centre français d'exploitation du droit de copie (CFC), 20, rue des Grands-Augustins, Paris 6<sup>e</sup>.

# Remerciements

Cet ouvrage a été réalisé à l'initiative du ministère de l'Écologie, de l'Énergie, du Développement durable et de l'Aménagement du Territoire (direction de la Prévention des pollutions et des risques, bureau des Risques naturels), avec le suivi de l'avancement du travail par Marie Renne. Il a été rédigé pour décrire de façon détaillée la méthode SHYREG – Pluie, développée par le Cemagref ; elle a conduit à l'élaboration de la base de données de quantiles de pluies de durée d'une heure à 72 heures pour les périodes de retour de deux à 100 ans. Ce guide accompagne la mise à disposition de ces données *via* la Climathèque de Météo-France.

La rédaction a été coordonnée par Patrick Arnaud et Jacques Lavabre du Cemagref à Aix-en-Provence, avec la participation de Luc Neppel (partie « Comparaison à une approche statistique régionale ») et Aurélie Muller (partie « Intervalles de confiance »).

Les auteurs remercient vivement les membres du comité de relecture qui ont contribué à la lisibilité de cet ouvrage et à sa validation. Ce comité de validation est composé d'experts en hydrométéorologie :

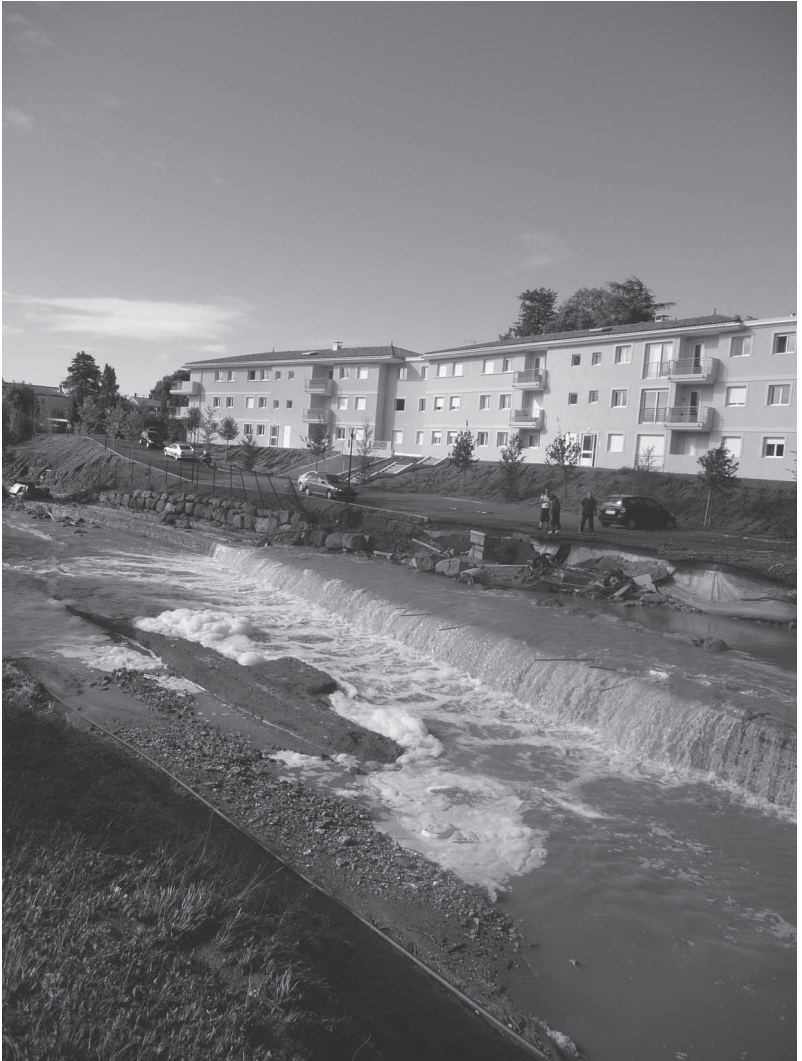
Jacky Astier	BRL Ingénierie Nîmes
Rémi Garçon	EDF – DTG Grenoble
Jean-Pierre Laborde	Université de Nice
Luc Neppel	Université Montpellier II – Hydrosociences
Jean-Michel Veysseire	Météo-France Toulouse

Qu'ils en soient tous ici remerciés, de même que ceux qui ont contribué de près ou de loin à l'émergence de la méthode SHYREG.



# Sommaire

<b>Problématique</b>	7
Introduction	7
Difficultés des statistiques classiques	10
Autres approches	18
<b>Simulation de chroniques de pluies horaires</b>	23
Les générateurs de pluie	23
Principe du modèle de pluie horaire de SHYPRE	27
Particularités du générateur	39
Application	56
<b>Paramétrisation par l'information journalière</b>	67
Les paramètres fixés	68
Paramètres « journalisés »	71
Corrections systématiques	78
Application sur 2 812 postes journaliers	84
<b>Régionalisation des paramètres journaliers</b>	97
Descriptif de la méthode utilisée	97
Cartes résultantes	100
Contrôle de la régionalisation des paramètres	107
Restitution des quantiles journaliers	111
Analyse des pluies horaires	118
<b>Analyse des résultats. Discussion</b>	125
Les cartes induites	125
Robustesse	127
Stabilité spatiale	133
Comportement à l'infini	135
Intervalle de confiance (Muller, 2006)	138
Comparaisons à une approche statistique régionale d'après Neppel <i>et al.</i> , 2007	143
Conclusion	151
<b>Références bibliographiques</b>	153



Crue du Réal aux Arcs, © P. Cantet, 2010.



# Problématique

## Introduction

L'évaluation des risques hydrologiques est l'une des préoccupations essentielles des hydrologues. Elle est nécessaire pour la gestion de l'aménagement du territoire rural ou urbain, pour le dimensionnement d'ouvrages hydrauliques, de ponts, de voies de communication à proximité de cours d'eau, etc.

Parmi les risques hydrologiques, le risque de crue est le plus souvent étudié. Classiquement, les études sur le risque hydrologique sont effectuées sur les bassins versants concernés, et passent essentiellement par l'analyse statistique des événements passés pour estimer la probabilité d'apparition d'un événement hydrologique. C'est ce que l'on appelle couramment la prédétermination des débits.

La prédétermination d'un risque se résume en général à l'association d'une fréquence à une caractéristique de ce risque. Cela revient à associer une valeur de cette variable à sa probabilité de dépassement, généralement ramenée à l'année (notion de période de retour). Une distribution empirique est établie à partir d'un échantillon de valeurs observées. Cette distribution guidera alors le choix d'un modèle théorique qui permettra d'extrapoler vers des fréquences plus rares pour estimer différents niveaux de l'aléa en fonction des enjeux (figure 1).

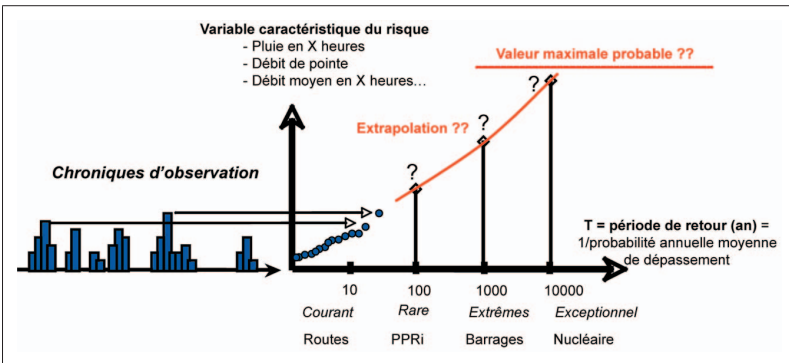


Figure 1. Principe général de la prédétermination.

Si la prédétermination des débits peut être étudiée uniquement à partir de l'information des débits observés (lorsque la chronique d'information est suffisamment longue, et pour les événements de fréquence rare à extrême), les hydrologues préféreront l'utilisation de l'information pluviométrique pour la connaissance des écoulements de fréquence encore plus rare (Lang, Lavabre, 2007). Ainsi, l'information pluviométrique est largement utilisée pour l'estimation des débits de crue sur l'ensemble de la plage des fréquences, du courant à l'exceptionnel. Généralement, cette étude amène à la détermination

d'une pluie de projet liée à une probabilité d'apparition. La pluie de projet peut être appréhendée de deux façons (CIGB, 1992) :

– **par une approche maximaliste ou d'événement maximal probable** : la pluie de projet est alors définie comme la plus forte précipitation physiquement possible sur une surface donnée, pour un site et une époque donnés (Hershfield, 1961), (Collier, Hardaker, 1996), (Koutsoyiannis, 1999). Cette approche est cependant peu utilisée en France ;

– **par une approche probabiliste** : la pluie de projet est liée à une probabilité de dépassement ou une période de retour. Son élaboration repose sur l'analyse statistique de divers éléments (hauteur, durée, forme) des événements pluvieux (Sighomnou, Desbordes, 1988).

En France, c'est l'approche probabiliste qui est utilisée. Dans l'approche probabiliste, la prise en compte de la pluviométrie se résume généralement à l'ajustement d'une loi de probabilité sur la distribution de fréquences empirique des pluies maximales observées. Les pluies de fréquences rares se déduisent de l'extrapolation de la loi statistique ajustée. Les quantiles de pluie ainsi estimés sont utilisés ensuite pour la construction de pluies de projet, pour le calcul de quantiles de débits, pour l'extrapolation des distributions de fréquences des débits. C'est le principe notamment de la méthode du Gradex, initiée par EDF et actuellement largement utilisée en ingénierie en France (Guillot, Duband, 1967).

On mesure alors l'importance de l'extrapolation des pluies vers les fréquences rares, qui conditionne largement les estimations des débits de crues. Or, la prédétermination des quantiles de pluies extrêmes reste très délicate pour différentes raisons : le problème de l'identification d'une loi de probabilité sur un échantillon réduit de valeurs observées qui reste sensible à la période d'observation, le poids des valeurs extrêmes, l'influence du choix de la loi de probabilité qui conditionne son comportement en extrapolation vers les fréquences rares, l'impossibilité de valider pour les fréquences rares...

À cela se rajoute la difficulté d'associer une « pluie de projet » à une fréquence donnée. En effet, un événement pluvieux se caractérise par différents critères, et en analyse multicritère, un événement ne peut avoir qu'une seule fréquence. La notion même de « pluie de projet » est critiquable puisqu'elle ne prend pas en compte les différentes formes de pluie, qui peuvent être associées à une même période de retour à travers une de ses caractéristiques. De plus, l'attribution de la même fréquence à la « pluie de projet » et à la « crue de projet » qui en résulte, n'est pas non plus évidente.

L'utilisation de l'information pluviométrique améliore la connaissance des quantiles de crues rares. Cependant, l'absence de chroniques suffisamment longues, les problèmes de métrologie lors de l'observation de forts événements (pluviomètres qui débordent, stations limnimétriques noyées ou emportées, extrapolation hasardeuse de la courbe d'étalonnage...) et la difficulté d'attribuer une fréquence aux valeurs exceptionnelles observées, rendent les résultats de ces méthodes statistiques extrêmement sensibles à l'échantillonnage des valeurs fortes.

Les recherches en milieu non jaugé ont conduit certains chercheurs à introduire, dans l'estimation des quantiles de crues, l'utilisation de générateurs de pluie couplés à une transformation de la pluie en débit. Cette nouvelle approche (Eagleson, 1972) est utilisée dans de nombreuses autres études. La méthode est basée sur deux éléments principaux (Cadavid *et al.*, 1991) :

- un modèle probabiliste qui introduit une composante stochastique à travers un modèle de génération de pluie ;
- un modèle de transformation de la pluie en débit qui fournit un hydrogramme pour une pluie efficace donnée.

Par la combinaison de ces deux éléments, on calcule la distribution de probabilité des débits de pointe. Différents modèles hydrologiques ont été utilisés : ceux basés sur la théorie de l'onde cinématique (Cadavid *et al.*, 1991 ; Eagleson, 1972 ; Shen *et al.*, 1990), comme ceux basés sur l'hydrogramme unitaire (Hebson, Wood, 1982 ; Rodriguez-Iturbe, Valdes, 1979). Peu d'études ont cependant comparé les distributions générées aux distributions observées. C'est le cas de certains travaux qui présentent le couplage d'un modèle stochastique de pluie avec le modèle hydrologique TOPMODEL (Beven, 1987 ; Beven, Kirkby, 1979 ; Blazkova, Beven, 1997).

C'est à ce type d'approche que s'apparente la méthode SHYPRE pour laquelle un modèle stochastique de génération de hyétogrammes au pas de temps horaire a été développé. Une fois couplé à une modélisation de la transformation de la pluie en débit, ce modèle permet la génération de multiples scénarios de crues de diverses formes qui remplacent les crues de projet classiques et fournissent une information temporelle complète à l'aménageur pour dimensionner un ouvrage. Outre son utilisation pour la gestion des crues, l'étude de la pluie au pas de temps horaire et sa modélisation en scénarios de crues sur de longues périodes de simulation permet une autre approche de la prédétermination des pluies et des débits, y compris sous forme d'hydrogramme de crue, ce qui a son importance pour le dimensionnement d'un ouvrage (barrage, bassin de rétention) où la notion de volume intervient.

### Étude de l'aléa pluviométrique

L'étude d'aléa pluviométrique est une étape préalable essentielle à l'évaluation des risques hydrologiques. Dans ce contexte, le développement de simulateurs de pluies permet de s'affranchir de la difficulté d'associer une probabilité d'apparition à une caractéristique unique des pluies. En analysant et modélisant l'ensemble de la structure temporelle des pluies, les générateurs de pluies extraient un maximum d'informations à partir des observations. On peut alors penser qu'ils puissent être moins sensibles à l'échantillonnage des observations, point sensible des méthodes statistiques classiques.

L'étude des crues nécessite de descendre en dessous du pas de temps journalier. C'est pourquoi le générateur de pluies a été développé au pas de temps horaire.

Enfin, si au pas de temps journalier l'information pluviométrique ponctuelle (pluviomètres) couvre correctement le territoire, l'information *infra-journalière* est plus difficilement accessible, rendant difficile les approches développées à ces pas de temps. La recherche d'approches permettant une régionalisation plus aisée de ces paramètres est alors aussi un atout pour pouvoir espérer l'appliquer sur l'ensemble d'un territoire avec le moins d'erreur possible.

# Difficultés des statistiques classiques

L'estimation des quantiles de pluie repose généralement sur le traitement statistique d'une chronique d'observation. Le traitement statistique comporte trois phases :

- **L'échantillonnage**, à savoir l'extraction de variables caractéristiques de la chronique d'observation ;
- **le choix d'un modèle** qui représente « au mieux » l'échantillon. C'est sur ce modèle que reposera l'extrapolation pour l'estimation des quantiles exceptionnels ;
- **le calage des paramètres du modèle** (ou ajustement du modèle).

Nous n'aborderons pas ici les problèmes métrologiques liés à l'acquisition des chroniques observées, sur lesquelles portent tous les calculs. Pourtant, ces problèmes peuvent entraîner des incertitudes notables. Différents problèmes sont associés à la métrologie, tels que les erreurs liées à la mesure physique du phénomène (précision, robustesse, homogénéité de la chronique), la gestion des lacunes (qui affectent souvent la connaissance des événements les plus forts), la représentativité des observations locales par rapport à l'environnement régional, etc.

*Remarque. Par la suite, toutes les fréquences empiriques seront calculées par la formule de Hazen(1).*

## L'échantillonnage

### CHOIX DE LA MÉTHODE D'ÉCHANTILLONNAGE

Deux méthodes d'échantillonnage sont couramment utilisées :

- **la méthode des maxima annuels** (ou saisonniers) qui consiste à retenir une seule valeur (la plus forte) sur l'année (ou la saison). L'échantillon est constitué de N valeurs, pour une durée d'observation de N années. Cette méthode est simple, mais présente le désavantage d'éliminer certaines valeurs fortes si elles sont observées la même année ;
- **la méthode des valeurs supérieures à un seuil** : un seuil de détection est préalablement fixé et on échantillonne toutes les valeurs qui excèdent ce seuil. L'échantillon est alors formé de K valeurs pour N années d'observation (avec K généralement supérieur à N). Cette méthode reste cependant tributaire du choix du seuil qui doit garantir l'indépendance des observations.

On voit sur le graphique de la figure 2 présenté en exemple, que suivant la méthode d'échantillonnage choisie, on ne donne pas les mêmes fréquences empiriques aux valeurs observées, en particulier pour les valeurs les plus faibles. Dans cet exemple, trois échantillonnages ont été effectués à partir de la même série pluviométrique observée :

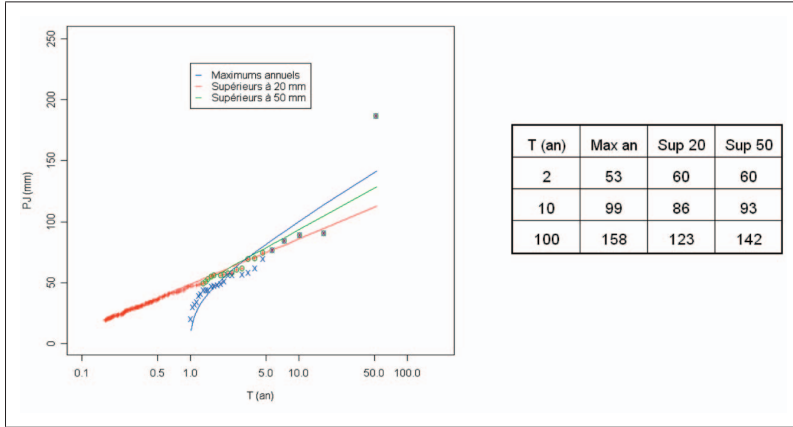
- un échantillonnage des pluies journalières maximales annuelles : Max an ;
- un échantillonnage des pluies journalières supérieures à 20 mm : Sup 20 ;
- un échantillonnage des pluies journalières supérieures à 50 mm : Sup 50.

---

1. La probabilité de non-dépassement des valeurs  $x_i$  est calculée par la formule de Hazen  $= \frac{i-0,5}{n}$ , où  $i$  est le rang des valeurs  $x_i$  classées par ordre croissant et  $n$  le nombre de valeurs de l'échantillon.

Si on ajuste une loi de type exponentiel sur ces différents échantillons provenant de la même chronique de pluie observée, respectivement Gumbel pour Max an, et exponentielle pour Sup20 et Sup50, on obtient des estimations de quantiles très différentes.

Sur cet exemple, alors que l'on part de la même information, la méthode d'échantillonnage conduit à des écarts de près de 30 % sur l'estimation du quantile centennal.



**Figure 2.** Effet de la méthode d'échantillonnage sur la distribution de fréquences empiriques des valeurs observées et sur l'estimation des quantiles (poste d'Aix-en-Provence, 1961-1994).

On notera que la présence de valeurs faibles dans un échantillonnage par la méthode des maximums annuels peut conduire à une surestimation de l'écart-type de la variable, qui constitue souvent l'un des estimateurs d'un paramètre ayant le plus de poids dans la loi de probabilité que l'on va ajuster sur l'échantillon.

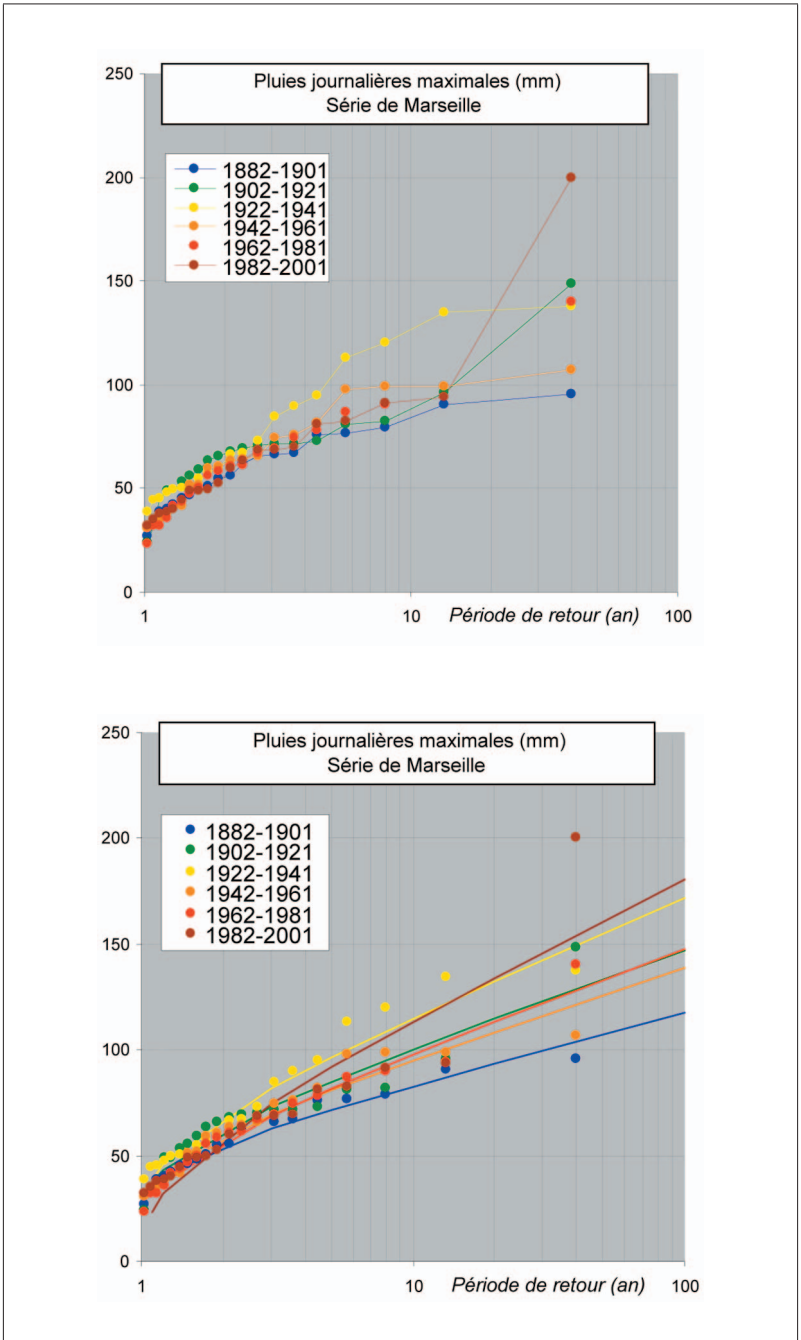
#### INFLUENCE DE LA PÉRIODE D'OBSERVATION

Au-delà du strict problème de la méthode d'échantillonnage, il est bien clair que l'échantillon est tributaire des caractéristiques de la période d'observation, c'est-à-dire :

- de sa durée : plus la durée d'observation sera grande et plus l'échantillon sera représentatif ;
- de sa position par rapport à des successions d'années sèches et humides (bien que l'on fasse généralement l'hypothèse de stationnarité du phénomène et de l'indépendance des années successives).

Pour illustrer ce propos, le graphique de la figure 3 présente les distributions de fréquences empiriques des pluies journalières de Marseille observées sur différentes périodes successives de 20 ans, entre 1882 et 2001.

On observe la variabilité liée à l'échantillonnage, avec des distributions de fréquences empiriques extrêmement variables en fonction de la période d'observation choisie. Si l'on peut étudier cet effet lorsque l'on dispose de très longues séries d'observation, ce n'est pas le cas pour la plupart des chroniques d'observation, beaucoup plus courtes. Si pour la série de Marseille on estime



**Figure 3.** Influence de la période d'observation sur la distribution de fréquences empiriques des valeurs observées et sur l'ajustement de la loi de Gumbel (méthode des moments).

la valeur de la pluie journalière décennale à 100 mm, on n'a observé aucune valeur décennale sur les 20 ans de la période (1882-1901) et on a observé quatre valeurs décennales sur les 20 ans de la période 1922-1941. Du coup, avec 20 ans d'observation on aurait estimé la pluie journalière décennale autour de 80 mm à partir des observations de la période 1882 et 1901, et autour de 120 mm à partir des observations de la période 1922-1941. Cela pose notamment problème lors de la comparaison des résultats de stations géographiquement proches mais ayant des séries de profondeur différentes, et soulève également la question de la stationnarité.

#### IMPACT DE LA PRÉSENCE DE VALEURS EXTRÊMES

La présence ou l'absence de valeurs « extrêmes » (quelquefois dénommées « horsains » ou « outliers ») dans l'échantillon pose aussi un réel problème. En effet, l'influence de ce genre de valeur est très forte sur l'estimation des quantiles, notamment en extrapolation. Par exemple, sur le premier graphique de la figure 4, on observe en 1993 une valeur de pluie journalière deux fois plus forte que la plus forte valeur observée en 30 ans de mesure. La présence de cette nouvelle valeur dans l'échantillon va alors fortement influencer l'ajustement d'une loi de probabilité. De même, sur la longue série de Marseille, on a observé en quatre ans deux valeurs nettement supérieures à la plus forte valeur observée sur les 120 années précédentes. Là encore, l'ajustement d'une loi de probabilité et son extrapolation vers les fréquences rares vont être fortement perturbées par l'observation de quatre années supplémentaires.

On admet généralement que l'estimation des quantiles associés à des périodes de retour inférieures à la durée d'observation est correcte, et que de grandes incertitudes apparaissent pour les quantiles de fréquences plus rares (période de retour supérieur à 2 ou 3 fois la durée d'observation). Cependant sur les exemples précédents, on voit que les quantiles proches de la durée d'observation peuvent aussi être fortement perturbés par la présence ou l'absence d'une valeur forte.

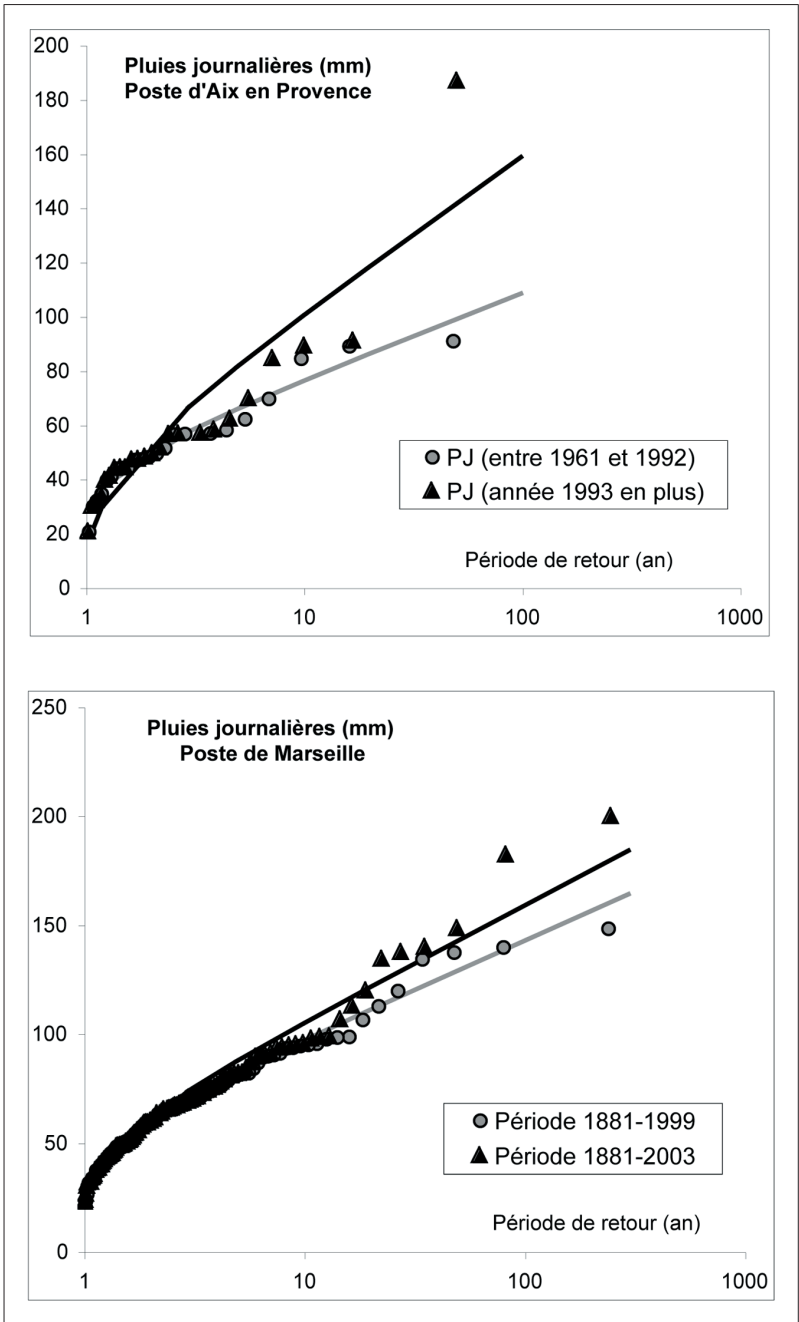
#### Sensibilité de l'ajustement statistique

Il est très sensible à la présence, ou à l'absence, de valeurs « exceptionnelles » dans l'échantillon des observations. Le poids de ces valeurs, souvent entachées d'une forte incertitude métrologique, est extrêmement important dans la détermination des quantiles rares, mais parfois aussi pour des quantiles plus courants.

## Choix d'une loi de probabilité théorique

Bien entendu le choix de la loi de probabilité influe sur les estimations, et notamment lors d'une utilisation en extrapolation. Le choix des modèles probabilistes utilisés varie avec le type d'échantillonnage adopté.

**Pour un échantillonnage de valeurs maximales annuelles**, la théorie des valeurs extrêmes conduit, sous les hypothèses les plus couramment utilisées, au choix de la **loi GEV** (*Generalized Extreme Value*) (Fisher, Tippett, 1928 ; Gnedenko, 1943). On évite en général d'utiliser ce type d'échantillonnage lorsque la série est courte (car on ne retient qu'une seule valeur par an).



**Figure 4.** Influence de la présence de valeurs « extrêmes » dans l'échantillon, sur la distribution de fréquences empiriques des valeurs observées et sur l'ajustement de la loi de Gumbel (méthode des moments).



$$- P(X \leq x) = \exp \left\{ - \left[ 1 - \gamma \frac{x - \alpha}{\beta} \right]^{1/\gamma} \right\} \quad \text{si } \gamma \neq 0$$

$$- P(X \leq x) = \exp \left\{ - \exp \left[ - \frac{x - \alpha}{\beta} \right] \right\} \quad \text{si } \gamma = 0$$

Où  $\alpha$  est le paramètre de position,  $\beta$  est le paramètre d'échelle et  $\gamma$  est le paramètre de forme.

Le paramètre de forme traduit le comportement à l'infini de la loi (figure 5) :

- si  $\gamma = 0$ , le comportement asymptotique est exponentiel (loi de Gumbel ou EV1) ;
- si  $\gamma < 0$ , le comportement asymptotique est plus qu'exponentiel (loi de Fréchet ou EV2) ;
- si  $\gamma > 0$ , le comportement asymptotique est moins qu'exponentiel (loi de Weibull ou EV3).

Pour un échantillonnage de valeurs supérieures à un seuil (POT : Peak Over Threshold), on choisit généralement une **loi GPD** (Generalized Pareto Distribution) (Pickands, 1975) On préfère cet échantillonnage pour traiter les séries courtes.

$$- P(X \leq x) = 1 - \left[ 1 - \gamma \frac{x - \alpha}{\beta} \right]^{1/\gamma} \quad \text{si } \gamma \neq 0$$

$$- P(X \leq x) = 1 - \exp \left[ - \frac{x - \alpha}{\beta} \right] \quad \text{si } \gamma = 0$$

C'est aussi une loi de probabilité à trois paramètres.

La loi exponentielle simple est un cas particulier du modèle de Pareto ( $\gamma = 0$ ).

En règle générale, un modèle s'ajuste d'autant mieux à la distribution empirique que son degré de paramétrisation augmente. À l'inverse, le modèle est moins robuste car le poids de l'échantillon affecte d'au « représentatif » : chronique de courte durée, présence de valeurs extrêmes et/ou de lacunes...

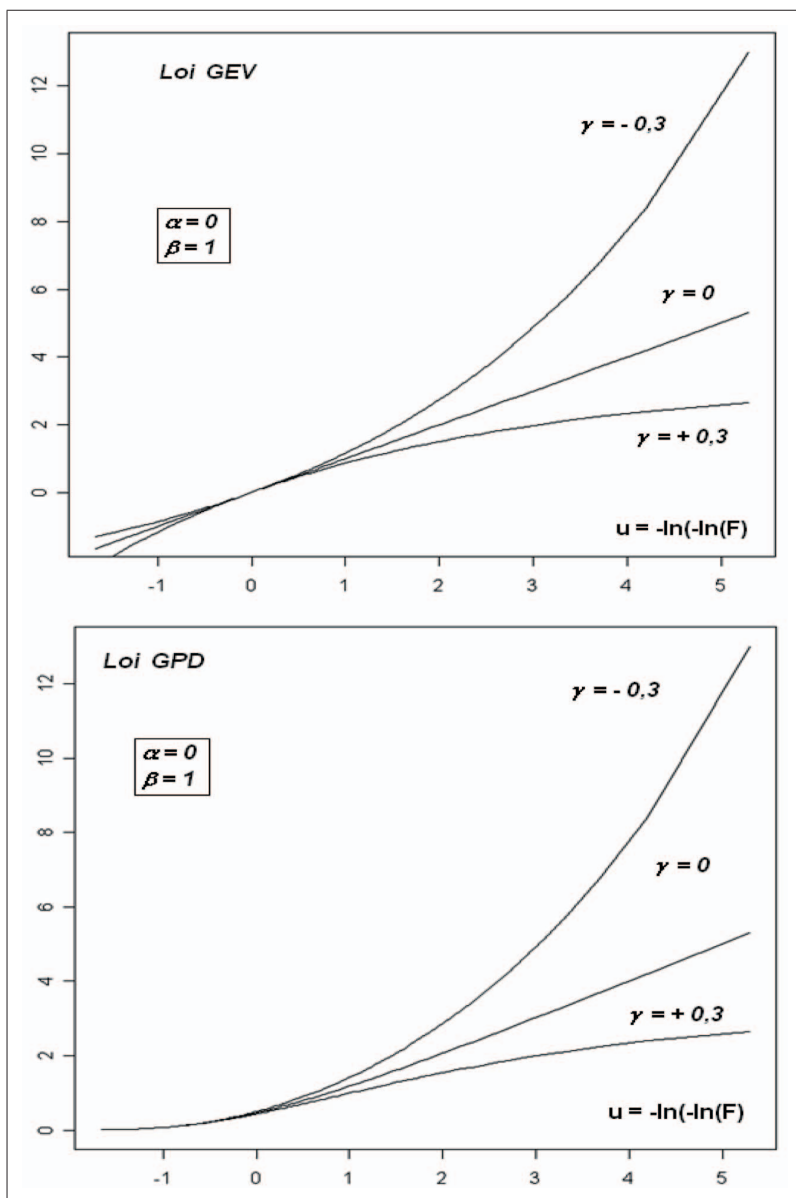
## Estimation des paramètres des lois de probabilité

L'ajustement de ces modèles consiste à déterminer leurs paramètres à partir de l'échantillon des valeurs observées. Il existe pour cela de nombreuses méthodes classées en deux catégories :

- les méthodes paramétriques comme la méthode du maximum de vraisemblance, la méthode des moments, la méthode des moments pondérés, les méthodes bayésiennes ;
- les méthodes non-paramétriques.

Les méthodes les plus utilisées de façon opérationnelle, sont la méthode des moments et la méthode du maximum de vraisemblance. Le problème est que chaque méthode peut conduire à des estimations différentes des paramètres et donc à des estimations différentes des quantiles.

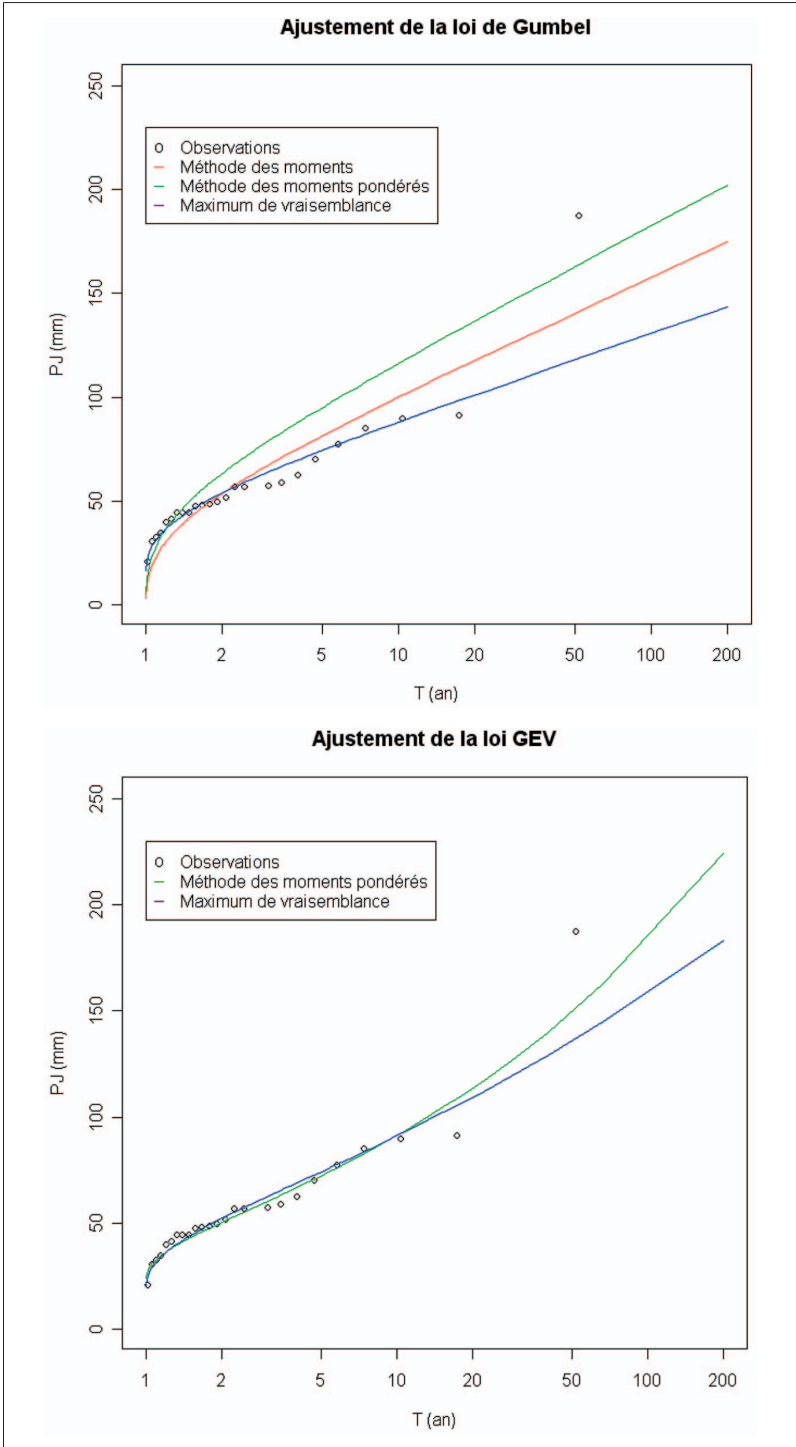
On voit bien sur l'exemple du poste d'Aix-en-Provence (figure 6), présentant une valeur extrême dans son échantillon, que la même loi peut conduire à des estimations très différentes en fonction de la méthode d'ajustement des



**Figure 5.** Distribution de fréquences des lois GEV et GPD pour différentes valeurs du coefficient de forme  $\gamma$ .

paramètres que l'on choisit. Les méthodes des moments, des moments pondérés et du maximum de vraisemblance conduisent respectivement pour la loi de Gumbel aux quantiles suivants : valeurs décennales 100, 116 et 88 mm et valeurs centennales 161, 187 et 133 mm, soit jusqu'à plus de 30 % d'écart en fonction de la méthode d'ajustement retenue pour la loi de Gumbel.

Le problème reste le même avec l'utilisation de la loi exponentielle. Il se rajoute en plus la difficulté du choix du seuil de sélection des événements,



**Figure 6.** Ajustement d'une loi de Gumbel et d'une loi GEV sur la série d'Aix-en-Provence, suivant différentes méthodes : méthode des moments, méthode des moments pondérés, méthode du maximum de vraisemblance.

devant garantir l'indépendance des valeurs de l'échantillon. Plus le seuil sera bas, plus on donnera du poids aux valeurs faibles avec une tendance à sous-estimer les quantiles forts. À l'inverse, si le seuil est trop élevé, la réduction de l'effectif de l'échantillon, le rend plus sensible à l'échantillonnage.

En général, la méthode des moments donne des quantiles plus forts que la méthode du maximum de vraisemblance, en particulier si la distribution de fréquences a une allure plus qu'exponentielle ou s'il y a la présence d'une valeur extrême dans l'échantillon. Ces méthodes sont cependant comparables s'il n'y a pas de valeur singulière dans l'échantillon.

### Les difficultés des statistiques

Les approches basées sur les statistiques classiques sont sensibles aux choix de nombreuses hypothèses : le choix de la méthode d'échantillonnage, le choix du modèle statistique, le choix de la méthode d'ajustement. La multiplication de ces hypothèses peut conduire à des estimations très variables des quantiles de pluies, à partir de l'information obtenue sur une même série pluviométrique. Les incertitudes qu'elles entraînent sur les estimations des quantiles de pluie ne sont pas négligeables et excèdent largement les intervalles de confiance classiques généralement affichés, qui ne représentent eux que l'incertitude d'échantillonnage et ne tiennent pas compte de l'incertitude due à l'ensemble des autres hypothèses.

## Autres approches

Pour traiter les difficultés des méthodes statistiques classiques d'autres approches ont été développées. L'objectif principal est de réduire l'effet de l'échantillonnage, très dommageable pour l'estimation des valeurs rares. Afin de mieux prendre en compte le poids des valeurs extrêmes, une des solutions est d'enrichir l'échantillon des valeurs observées, pour augmenter sa représentativité.

Pour cela, on peut chercher de l'information sur d'autres sites de mesure (approche régionale), sur différentes variables caractéristiques (approche multi-durées), sur l'information historique (approche historique), sur la nature même du signal de pluies (approche par modélisation). Un bref aperçu du principe de ces approches est donné ici.

## Approche régionale

L'approche régionale, aussi appelée « méthode des années-stations », prend en compte de l'information régionale pour l'estimation des pluies en un lieu donné. À ne pas confondre avec une approche « régionalisée » qui reste une approche locale pour laquelle les paramètres ont été régionalisés.

L'approche régionale vise à augmenter le nombre d'observations en considérant le phénomène non plus localement mais sur l'ensemble d'une zone donnée. On utilise l'information de plusieurs stations pour réduire l'influence de l'échantillonnage sur les données locales. On augmente ainsi l'échantillon des valeurs exceptionnelles. En respectant l'hypothèse d'indépendance des différentes observations, on peut mélanger les données observées sur différents

sites supposés homogènes entre eux. L'homogénéité des sites peut nécessiter une « normalisation » des variables étudiées. Après cette « homogénéisation », les données de différentes stations de mesure sont agglomérées et étudiées de façon régionale. L'accroissement de l'effectif permet aussi l'utilisation de lois à trois paramètres comme la loi GEV (Onibon *et al.*, 2004), voire à quatre paramètres comme la loi TCEV (Two Component Extreme Values)(Rossi *et al.*, 1984), avec plus de robustesse que pour une approche locale.

Notons que de nombreuses applications de méthodes régionales conduisent à des valeurs négatives pour les coefficients de forme de la loi GEV ou GPD utilisées (– 0.15 en Espagne (Ferrer, 1992) ; – 0.15 en Grèce (Koutsoyiannis, 2004) ; – 0.13 sur la région Languedoc (Dominguez *et al.*, 2005)), sur des données journalières.

## Approche multi-durées

Le risque pluvial peut être étudié à partir de différentes caractéristiques de la pluie, par exemple les pluies maximales de différentes durées.

L'analyse classique de la pluie sur différentes durées consiste en général à étudier une distribution de probabilité associée à chaque durée (appelées lois marginales). Ce point peut conduire à des contradictions, avec des lois de probabilité qui se croisent en extrapolation vers les fréquences rares. Par exemple, la présence d'une valeur extrême de pluies en 6 heures dans l'échantillon peut conduire à une estimation de la pluie maximale centennale en 6 heures supérieure à la pluie maximale centennale en 12 heures.

Pour s'affranchir de ce genre d'incohérence, une analyse conjointe des pluies de différentes durées est préférable.

En particulier, l'analyse des courbes Intensité-Durée-Fréquences (courbes I-D-F.) empêche l'intersection des distributions des quantiles des différentes durées.

De façon classique, cette approche multi-durées passe par trois étapes :

- ajustement des lois marginales (une distribution de probabilité par durée de pluies) ;
- calcul des quantiles des pluies de différentes durées associés à différentes périodes de retour ;
- ajustement d'équations paramétriques pour chaque période de retour.

Plusieurs modèles empiriques peuvent être proposés de la forme :

$$I(d, T) = \frac{\delta \times T^\alpha}{(d + \lambda)^\beta} \text{ avec } I(d) = \text{l'intensité de la pluie de durée } d, \text{ en mm/h}$$

T = la période de retour, en année

d = la durée de la pluie, en h

$\alpha$ ,  $\beta$ ,  $\delta$  et  $\lambda$  des coefficients.

La plus utilisée actuellement est une formule simple proposée par Montana :

$$I(d) = \alpha \times d^\beta$$

où  $\alpha$  et  $\beta$  sont des coefficients fonction de la période de retour et de certaines plages de la durée d.

L'inconvénient de ces modèles est qu'ils nécessitent un grand nombre de paramètres (coefficients pouvant être variables avec la période de retour) et qu'ils sont basés sur des régressions de valeurs dépendantes.

À partir d'approches issues de la théorie des multifractales, exploitant l'hypothèse de l'invariance d'échelle, on peut réduire considérablement le nombre de paramètres (Burlando, Rosso, 1996).

Les modèles multi-durées présentent un grand intérêt en ingénierie puisque des données journalières peuvent être utilisées pour estimer des quantiles à des pas de temps plus fins. Ces données journalières, plus largement disponibles que les données à pas de temps plus fins peuvent alors améliorer l'estimation des pluies à pas de temps fins. Par exemple, Muller *et al.* (2008b) montrent, sur une longue chronique de pluies à Marseille, que les enregistrements journaliers sont de meilleure qualité et présentent moins de lacunes que les enregistrements horaires, en particulier lors des événements extrêmes. Dans ce cas, les estimations des quantiles de pluies de courte durée par une approche multi-durées sont plus robustes et plus fiables que celles obtenues par une approche classique par durée de pluie.

## Approches historiques

Ces approches sont surtout développées pour l'analyse statistique des crues. L'objectif est d'utiliser l'information historique pour une meilleure connaissance du risque inondation, en consolidant la distribution des valeurs observées vers les probabilités de dépassement faibles (événements rares).

L'utilisation de l'information historique dans une analyse fréquentielle permet de mieux intégrer l'information réellement disponible et devrait donc permettre d'améliorer l'estimation des quantiles de grandes périodes de retour. Pour les débits, cette information historique est relative à des grandes crues qui se sont produites avant le début de la période de mesures systématiques. La prise en compte statistique de l'information historique peut être réalisée par des méthodes bayésiennes (Miquel, 1983).

De façon générale, l'utilisation de l'information historique conduit à une diminution de l'impact des valeurs singulières dans les séries d'enregistrements systématiques et à une diminution de l'écart-type des estimations (Ouarda *et al.*, 1998). L'incertitude d'échantillonnage sur les quantiles estimés est donc réduite lorsque l'information historique est exploitée.

Une des difficultés de cette approche est peut être l'estimation des débits de crues historiques, et en particulier le passage des hauteurs d'eau (niveaux historiques atteints) aux débits (Naulet, 2002 ; Naulet *et al.*, 2005). De plus, si l'information historique sur les crues peut être relativement bien recensée, surtout pour les grands cours d'eau traversant les grandes villes, il n'en est pas de même pour l'information pluviométrique. Il existe toutefois des séries pluviométriques anciennes souvent plus longues que les séries continues de hauteurs d'eau.

## Approche par modélisation du signal de pluie

L'objectif de ces approches est de modéliser le signal de pluies pour recréer des chroniques de pluies statistiquement équivalentes aux chroniques de pluies

simulées. L'extrapolation vers les fréquences rares se fait par augmentation de la durée de simulation.

Généralement, ces méthodes reconstituent des chroniques de pluies, fournissant toute l'information temporelle liée au pas de temps du modèle. Dans ce cas, ces méthodes prennent en compte implicitement toutes les durées de pluies.

Ces approches permettent surtout de s'affranchir de la notion de pluie de projet unique en proposant de multiples scénarios de pluie associés à une caractéristique dont on peut déterminer la fréquence. La notion même de crue de projet n'est plus nécessaire puisque l'on dispose de multiples scénarios de crues directement utilisables pour tester le fonctionnement hydraulique des ouvrages sollicités.

C'est à cette famille d'approche qu'appartient la méthode SHYPRE, détaillée dans ce document.

### **Lien entre les différentes approches**

Le lien essentiel entre les approches citées ici est l'augmentation de la quantité d'informations utilisées pour étudier le phénomène. Elles améliorent la représentativité des échantillons de valeurs étudiés, réduisant les problèmes d'échantillonnage.

Pour certaines, elles permettent aussi une appréhension plus cohérente du phénomène dans sa dimension temporelle (approches multi-durées et par simulation).



La Nartuby en crue à Trans-en-Provence, © P. Cantet, 2010.



La Nartuby en crue à Trans-en-Provence, © P. Cantet, 2010.



# Simulation de chroniques de pluies horaires

L'étude du risque pluvial nécessite la connaissance de la pluie à différents pas de temps. L'exploitation des chroniques de pluies observées permet l'étude statistique de ces caractéristiques. Cependant, pour l'analyse des événements extrêmes, la longueur des séries d'observation reste le facteur déterminant sur l'incertitude des quantiles estimés. Si l'étude des pluies extrêmes nécessite une information large des pluies à pas de temps fins, l'étude du risque hydrologique peut aussi nécessiter l'utilisation de chroniques de pluie.

Face au problème de la disponibilité de l'information pluviométrique à pas de temps fins (chroniques trop courtes, lacunes, faible couverture spatiale...), des générateurs de pluies ont été développés de façon à simuler des chroniques de pluies statistiquement équivalentes aux chroniques de pluies observées. La modélisation de ces pluies permet alors de générer des scénarios de pluies probables directement utilisables pour l'étude du risque pluvial mais aussi comme variables d'entrée pour différentes modélisations (hydrologie, érosion, transfert de polluants...).

*L'intérêt initial de ces modèles est de s'affranchir de la notion de pluie de projet unique et de fournir une information temporelle complète sur les événements pluvieux.*

## Les générateurs de pluie

La pluie résulte de processus physiques naturels dont on connaît les principes fondamentaux. Les phénomènes énergétiques et mécaniques concernant la physique atmosphérique sont à la base des modèles de prévision météorologique. Mais ces modèles concernent des échelles trop grandes pour être utilisés en hydrologie. Pour traiter des phénomènes à l'échelle du bassin versant, des modèles mécanistes tentent de modéliser la physique atmosphérique et de caractériser la pluviométrie à une échelle plus petite que celle utilisée en météorologie. Dans ces modèles, le système concerné est la masse nuageuse. Les nuages sont considérés comme des réservoirs d'eau condensée dont l'évolution est définie par les équations de la physique atmosphérique. Ces modèles sont basés sur la conservation de la masse de l'eau condensée contenue dans une colonne nuageuse verticale (Georgakakos, Bras, 1984).

Les phénomènes naturels, tels que la pluie, sont complexes et il est actuellement difficile de reconstituer les mécanismes successifs reliant les causes aux effets sans un ajustement constant du modèle au cours d'un événement. En effet, cette complexité rend la modélisation très sensible aux erreurs sur les variables initiales. C'est pourquoi on préfère considérer ces phénomènes comme des variables aléatoires et les étudier de façon statistique.

Les modèles de génération de pluies sont plus fréquemment associés à un pas de temps journalier. La modélisation des pluies au pas de temps journalier est en effet plus simple car le phénomène est moins complexe qu'aux pas de temps fins. On dispose aussi d'une information bien plus importante au pas de temps journalier, ce qui facilite leur mise en application. L'étude des crues nécessite cependant d'avoir une information à pas de temps plus fins. Bien que leur développement soit plus réduit qu'au pas de temps journalier, les modélisations de chroniques de pluies à pas de temps fins ont fait l'objet de nombreuses recherches à travers différentes approches.

## Modèles basés sur des processus d'agrégation

Initiés par les travaux de LeCam (1961), certains modèles sont basés sur des processus d'agrégation, par la combinaison de deux processus aléatoires : l'occurrence et la position de cellules pluvieuses et la génération de la durée et de l'intensité de ces cellules.

Le processus d'agrégation est un processus de groupement stochastique basé sur l'hypothèse que l'espace est organisé en groupes hiérarchiques dont l'occurrence des éléments dans le temps est décrite par une loi de Poisson. Les auteurs de ce processus (Neyman, Scott, 1958) l'élaborèrent pour traiter de façon probabiliste des phénomènes de cosmologie (représentation de galaxies). C'est Waymire et Gupta (1981) qui furent les premiers à utiliser ce processus pour décrire l'occurrence des épisodes pluvieux, ainsi que des cellules pluvieuses qui les composent. Ces cellules pluvieuses ainsi générées et positionnées, sont affectées d'une durée et d'une intensité. La pluie sur chaque pas de temps est alors déterminée par la somme des intensités des cellules présentes sur le pas de temps (figure 7).

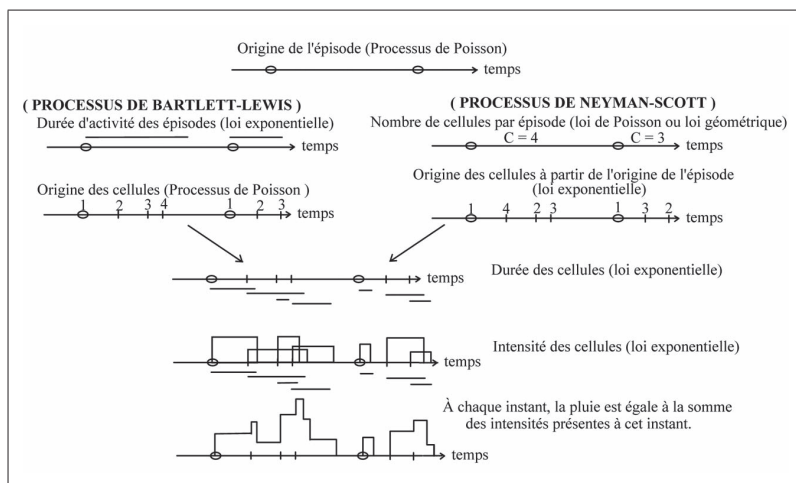


Figure 7. Principe des modèles basés sur un processus d'agrégation.

Deux approches sont principalement utilisées pour positionner les cellules pluvieuses : le processus de Neyman-Scott et le processus de Bartlett-Lewis. Dans le processus de Neyman-Scott, le nombre de cellules pluvieuses dans l'épisode est une variable aléatoire et les cellules pluvieuses sont positionnées

de façon indépendante à partir du début de l'épisode considéré. Dans le processus de Bartlett-Lewis, c'est la durée d'activité de l'épisode qui est une variable aléatoire, et les cellules pluvieuses sont positionnées en fonction de la position de la cellule précédente par un processus de Poisson tant que l'on ne dépasse pas la durée d'activité de l'épisode.

De façon générale, les étapes successives jusqu'à l'obtention d'une chronique pluvieuse sont :

- la génération de l'origine des divers épisodes pluvieux par un processus de Poisson ;
- pour chaque épisode, la génération et le positionnement des cellules par un processus de Neyman-Scott ou de Bartlett-Lewis ;
- la génération de la durée de chaque cellule dans une loi exponentielle ;
- la génération de l'intensité de chaque cellule dans une loi exponentielle ;
- calcul des pluies à chaque pas de temps par la somme des intensités des cellules présentes sur le pas de temps.

Ces modèles fonctionnent en continu. Ils nécessitent un nombre relativement faible de paramètres pour générer des épisodes pluvieux (5 ou 6), mais qui doivent être déterminés pour chaque mois (d'où plus de 60 valeurs).

Les différents travaux sur ces modèles (Acreman, 1990 ; Cowpertwait, 1991 ; Rodriguez-Iturbe *et al.*, 1987) montrent que cette approche donne des résultats relativement bons. Ces modèles sont capables de reproduire correctement la distribution des pluies à différents pas de temps. Ils présentent toutefois une faiblesse dans la reproduction de la proportion des périodes sèches. Rodriguez-Iturbe *et al.* (1988) proposent alors de considérer le paramètre de la loi des durées des cellules comme une variable aléatoire. La valeur de ce paramètre est alors générée dans une loi Gamma pour chaque épisode. Cette modification améliore la restitution des périodes sèches.

Il reste que ces modèles ont quelques difficultés à reproduire des valeurs extrêmes (Onof *et al.*, 1995). Les développements actuels de ces modèles portent sur une régionalisation de leurs paramètres. Leur variation saisonnière est assez marquée et leur signification physique facilite l'interprétation de la variation spatiale. Certains travaux notent aussi une dépendance entre les durées et les intensités des cellules, caractéristique à prendre en compte (Onof *et al.*, 1995).

Ces modèles sont les plus couramment utilisés par les anglo-saxons. En France on utilise plutôt des modèles basés sur une simulation directe de hyétoigrammes.

## Modèles basés sur la simulation directe de hyétoigrammes

Ces modèles partent du principe que la pluie peut être assimilée à un processus aléatoire et intermittent (succession d'états secs et pluvieux) dont l'évolution est décrite par des lois de nature stochastique. En outre, ces modèles reposent sur l'hypothèse d'indépendance des variables décrivant les hyétoigrammes, ainsi que sur l'hypothèse de stationnarité du phénomène étudié (tout comme les modèles basés sur les processus d'agrégation).

La génération du signal de pluie comporte deux étapes (figure 8) :

- une première étape constitue l'étude descriptive du phénomène. Elle amène au choix des variables qui vont caractériser le phénomène étudié. Par hypothèse, et dans un premier temps, chaque variable descriptive devra être indépendante des autres, et sera représentée par une loi de probabilité théorique ajustée sur les valeurs observées. Cette hypothèse d'indépendance sera discutée par la suite ;
- la seconde étape est la construction de la chronique de pluie par des variables descriptives générées de façon aléatoire à partir de leur loi de répartition. De manière générale, si une variable aléatoire  $X$  a pour fonction de répartition  $F(x)$ , la génération de variables suivant la loi  $F(x)$  se fait en tirant de façon aléatoire une valeur de  $F(x)$  dans une loi uniforme entre 0 et 1, puis en cherchant la valeur correspondante de  $x$ .

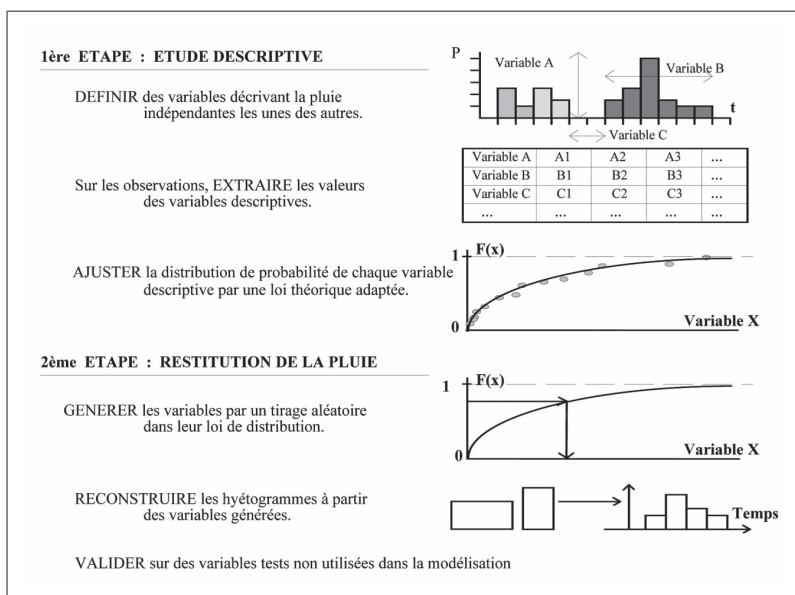


Figure 8. Principe d'un modèle stochastique de simulation des hyétoigrammes.

### Principe général

L'élaboration d'un modèle stochastique de simulation de hyétoigrammes se résume donc à trouver les bonnes variables aléatoires indépendantes décrivant le processus de pluie, ainsi que les lois de probabilité qui les représentent le mieux.

La reconstruction du signal s'effectue grâce à une hiérarchisation logique du tirage des différentes variables, afin de parvenir à une représentation la plus fidèle possible du signal de départ, tout en respectant la définition des variables utilisées. La validation du modèle consiste à étudier sa capacité à reproduire des variables représentatives du signal simulé, **non utilisées lors de la modélisation**.

Sur cette base, différents modèles peuvent être élaborés, les différences provenant en particulier du choix des variables et des lois de probabilité. Ce type de

modèle est souvent retrouvé au pas de temps horaire (Arnaud, Lavabre, 1999 ; Cernesson *et al.*, 1996 ; Lebel, 1984 ; Tourasse, 1981), mais aussi au pas de temps journalier (Buishand, 1978).

## Les modèles basés sur la désagrégation de la pluie journalière

On trouve de nombreux modèles basés sur la désagrégation de la pluie journalière. **L'élaboration de tels modèles trouve son intérêt dans la disponibilité plus large des données journalières par rapport aux données horaires.**

Glasbey *et al.* (1995) (Glasbey, 1995) rappellent quelques travaux initiaux portant sur cette approche. Hershendorff et Woolhiser (1987) ont modélisé les épisodes pluvieux d'une journée (nombre de périodes pluvieuses dans un jour, pluie tombée, durée et point de départ de chaque période), à partir de la pluie totale du jour considéré, et du jour antérieur et postérieur. Econopouly *et al.* (1990) ont testé ce modèle sur d'autres sites. Les résultats obtenus sont corrects mais le passage des périodes pluvieuses aux chroniques de pluies horaires n'a pas été réalisé.

L'approche que présentent Glasbey *et al.* (1995) pour désagréger des données journalières est la suivante. Des séries de pluies horaires sont simulées à partir d'un modèle horaire, qui n'est pas forcément calé, et que l'on contraint à respecter les totaux journaliers observés. Cela n'est pas évident de façon directe, car on ne connaît pas l'expression de la distribution des pluies horaires conditionnées sur les totaux journaliers. Une solution alternative est de générer des pluies horaires inconditionnellement à partir du modèle horaire, jusqu'à obtenir une séquence ayant des totaux horaires satisfaisants.

La méthode présentée agrège des pluies horaires avant de désagréger la pluie journalière. Les résultats sont indépendants du modèle utilisé car les simulations sont contraintes à respecter les pluies journalières. Les modèles de pluies horaires utilisés peuvent être du type de ceux présentés précédemment.

Notons que si cette approche permet de désagréger des données de pluies journalières observées, pour une étude des pas de temps plus fins, elle peut aussi être utilisée pour désagréger des pluies journalières modélisées. Les modèles de génération de chroniques de pluies journalières ne manquant pas (toute la famille des modèles auto-régressifs ou autres...), leur couplage avec cette approche de désagrégation fournit une nouvelle méthode d'étude des distributions des pluies à pas de temps fins.

## Principe du modèle de pluie horaire de SHYPRE

Le générateur de pluies horaires développé dans la méthode SHYPRE est un modèle basé sur la simulation directe de hyétoigrammes (voir p. 25). Il a été développé dans le cadre de deux thèses (Arnaud, 1997 ; Cernesson, 1993). C'est un modèle événementiel, c'est-à-dire qu'il crée de façon indépendante et discontinue des événements pluvieux au pas de temps de l'heure. Seul le nombre moyen d'événements dans l'année est respecté, sans situer les événements dans une chronique quelconque.

La mise en œuvre du générateur de pluies horaires implique trois étapes, présentées sur le schéma de la figure 9 (les différents paramètres présentés sur le schéma seront définis dans le paragraphe ci-dessous *Analyse descriptive des pluies observées*).

- **La première étape est le calage du générateur** : elle est issue d'une étude descriptive du phénomène. Elle conduit au choix des variables qui définissent le phénomène et au choix des lois de probabilité qui vont décrire ces variables. On vérifie l'hypothèse d'indépendance des variables descriptives les unes par rapport aux autres. On estime ensuite les paramètres des lois de probabilité de chaque variable à partir d'un ajustement statistique des distributions de fréquences des valeurs observées.
- **La seconde étape est la simulation d'événements pluvieux** : différentes valeurs des variables définies précédemment sont générées par une méthode de Monte Carlo, c'est-à-dire par un tirage aléatoire dans leur loi de probabilité théorique. La génération des variables est réalisée suivant un ordre dicté par la construction des événements pluvieux. L'hypothèse initiale d'indépendance des différentes variables descriptives permet le tirage des valeurs des variables indépendamment les unes des autres. Cette hypothèse d'indépendance doit être vérifiée. Si ce n'est pas le cas, la prise en compte de la dépendance des variables est alors nécessaire.
- Enfin **la dernière étape est la validation des résultats obtenus** : à partir de la simulation d'une multitude d'événements, associée à une longue période de simulation, on extrait des caractéristiques pluviographiques non directement utilisées par le modèle. Les distributions de fréquences de ces variables de contrôle (les pluies maximales de 1 à 72 heures) permettent de rendre compte statistiquement de la cohérence temporelle des événements simulés, par comparaison avec les distributions empiriques des valeurs observées de ces mêmes variables.

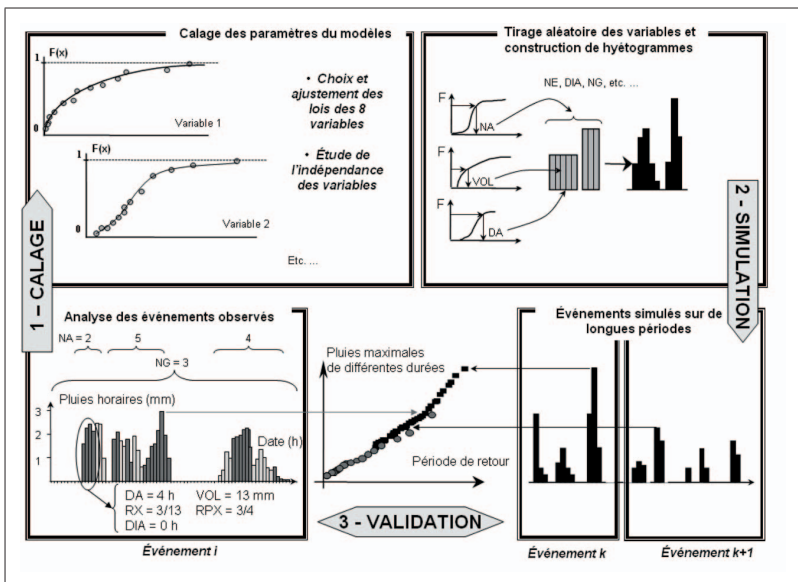


Figure 9. Principe du modèle de pluie horaire de SHYPRE.

## Analyse descriptive des pluies horaires observées

Les événements pluvieux servant au calage du modèle, sont sélectionnés à partir d'un critère journalier. Ces événements sont définis comme étant une succession de pluies journalières supérieures à 4 mm et comprenant au moins un cumul journalier supérieur à 20 mm (figure 10).

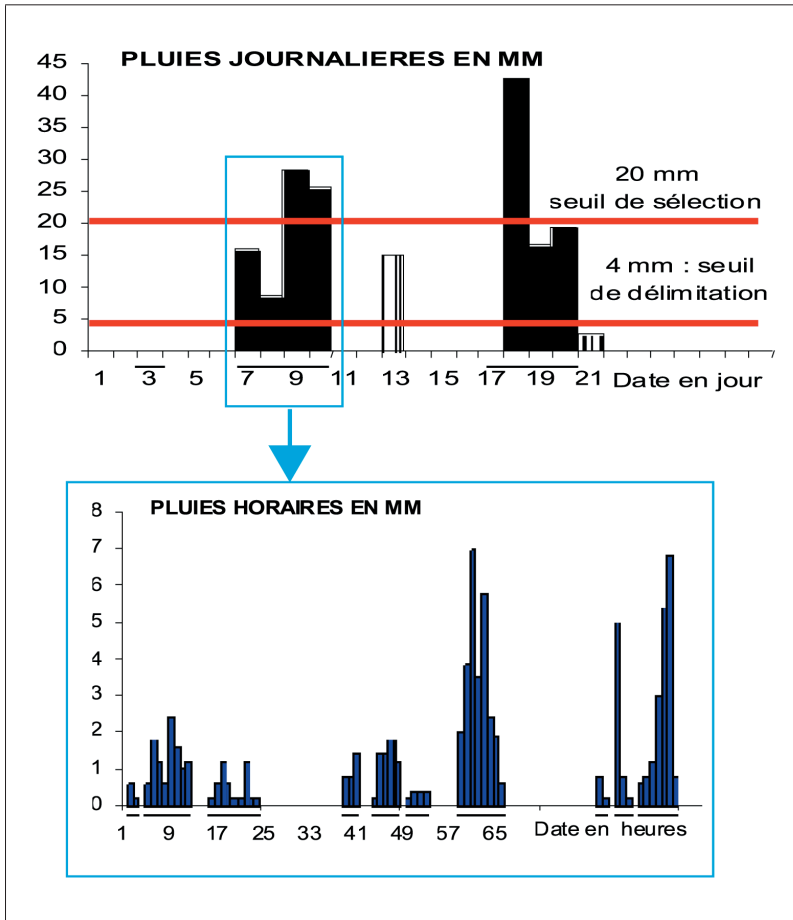


Figure 10. Échantillonnage des données analysées.

### a) Sélection des événements pluvieux à partir de l'information journalière.

Un événement est caractérisé par :

- présence d'une pluie journalière > 20 mm ;
- toutes les pluies journalières sont > 4 mm.

Ici on a sélectionné deux événements de durées respectives quatre et trois jours.

## b) Analyse de la structure des pluies horaires sur la durée des événements pluvieux sélectionnés.

Chaque événement pluvieux sélectionné au pas de temps journalier, est analysé au pas de temps horaire comme une succession d'averses. Une averse est définie comme une succession de pluies horaires ne présentant qu'un maximum local.

L'échantillonnage des événements pluvieux sur la base d'un critère journalier conduit à la définition de la première variable du modèle : la **variable NE** qui est le nombre d'événement pluvieux par an.

L'analyse des pluies au pas de temps horaire est ensuite réalisée événement par événement (cf. exemple figure 11).

Un événement pluvieux comportera une ou plusieurs périodes pluvieuses. Les périodes pluvieuses sont séparées par un ou plusieurs pas de temps horaires sans pluie. On définit alors la **variable NG** qui est le nombre de périodes pluvieuses par événement pluvieux et la **variable DIA** (en heure) qui est la durée sèche entre deux périodes pluvieuses.

On analyse alors les périodes pluvieuses sur la base d'averses. Une averse est une succession de pluies horaires ne présentant qu'un maximum, avec une décroissance des pluies autour de ce maximum local. Une période pluvieuse comporte donc une ou plusieurs averses successives séparées par un minimum relatif affecté arbitrairement à l'averse qui le précède. On définit alors la **variable NA** qui est le nombre d'averses par période pluvieuse.

Chaque averse est alors définie par différentes variables qui sont :

- la durée de l'averse (en heure) : **variable DA** ;
- le volume de l'averse (en mm) : **variable VOL** ;
- le rapport entre la pluie horaire maximale de l'averse et son volume : **variable RX** ;
- et la position relative de la pluie horaire maximale dans l'averse : **variable RPX**.

### Effectif de l'échantillon

L'analyse descriptive des événements pluvieux enregistrés sur une station pluviographique conduit à étudier plusieurs événements par an, composés de plusieurs périodes pluvieuses ayant elles-mêmes plusieurs averses. L'effectif des variables étudiées est alors bien supérieur à une valeur par an, ce qui participera à la robustesse de l'approche.

*Remarque. Le choix d'analyse des hyétogrammes sur la base des averses (successions de pluies horaires présentant un seul maximum) permet de prendre en compte implicitement l'auto-corrélation forte qu'il y a entre les pluies horaires successives d'un événement pluvieux. La sensibilité de ce mode d'analyse à la précision des enregistrements sera discutée plus loin.*



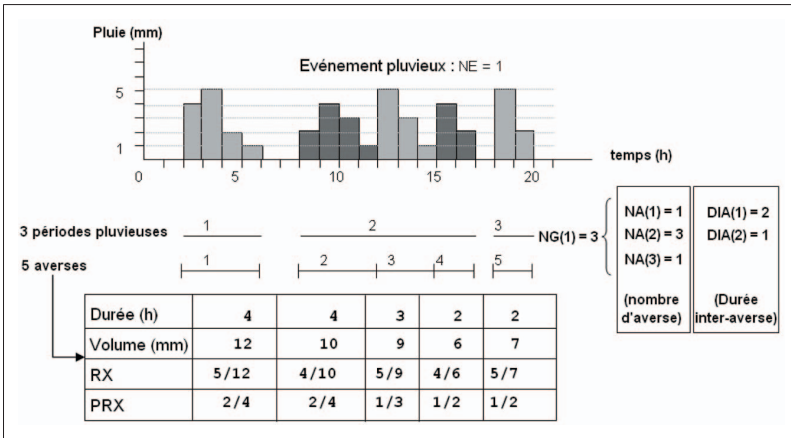


Figure 11. Détail de l'analyse d'un événement pluvieux : calcul des valeurs des différentes variables.

## Typologie des averses

L'entité de base dans l'analyse des hyétogrammes horaires est l'averse. Rappelons qu'une averse est une succession de pluies horaires ne présentant qu'un maximum relatif.

Dans un premier temps, toutes les averses ont été analysées ensemble, considérant qu'elles appartenaient à une même population. Cependant, l'étude quantitative des intensités des averses et de leur durée, a montré la nécessité de distinguer deux types d'averses. Une typologie des averses est donc nécessaire. Différents essais ont été faits sur l'analyse des averses.

La séparation des averses en fonction de leur nombre dans la période pluvieuse conduit à des distributions de fréquences des intensités moyennes relativement différentes. Deux populations distinctes d'intensités d'averses ont alors été considérées : les intensités des averses isolées et celles des averses groupées (Cernesson, 1993).

Cependant l'écart entre les caractéristiques statistiques de ces deux types d'averses restait faible sur la globalité des postes étudiés. Les travaux effectués sur le modèle en climat méditerranéen ont conduit à la définition d'averses dites « principales » et d'averses dites « ordinaires ». On définissait alors une seule averse principale par événement pluvieux comme étant l'averse qui apporte la plus grande quantité d'eau lors de l'événement. Toutes les autres averses étaient alors définies comme étant des averses ordinaires (Cernesson *et al.*, 1996). Cette fois-ci la différenciation entre deux types d'averses s'applique à la fois pour les intensités moyennes et pour les durées.

Cette typologie s'est avérée insuffisante pour caractériser les pluies extrêmes observées sous d'autres climats. En effet, l'analyse des hyétogrammes observés sur l'île de la Réunion, a montré la nécessité de considérer plus d'une averse « principale » par événement (Fine, Lavabre, 2002).

Afin de ne pas alourdir la paramétrisation du modèle, dans l'objectif de sa régionalisation, la solution retenue a été de définir **une typologie des averses**

**sur la base d'un critère journalier.** Ce critère doit permettre de distinguer les averses responsables des forts cumuls de pluie. On choisit alors de retenir au moins une averse principale par événement. **On rajoute** ensuite autant d'averses principales que de jours où la pluie journalière a dépassé 50 mm. Le tableau 1 montre quelques exemples de calcul du nombre d'averses principales à analyser dans un événement, à partir du critère journalier.

	Durée (jour)	Valeur des pluies journalières				NAVP
Évènement 1	1	35				1 (1 + 0)
Évènement 2	1	59				2 (1 + 1)
Évènement 3	2	65	35			2 (1 + 1)
Évènement 4	2	57	125			3 (1 + 2)
Évènement 5	3	24	44	30		1 (1 + 0)
Évènement 6	4	16	104	41	51	3 (1 + 2)

**Tableau 1.** Exemples de calculs du nombre d'averses principales à considérer dans un événement.

Une fois le nombre d'averses principales déterminé sur ce critère journalier (NAVP), on analyse l'information au pas de temps horaire. On analyse alors les NAVP averses les plus fortes de l'événement comme étant des averses principales, indépendamment de leur position dans l'événement. Les autres sont analysées comme étant des averses ordinaires. On s'assure cependant que le nombre d'averses principales ne dépasse pas le nombre total d'averses de l'événement (cas pouvant se produire si l'événement est par exemple composé d'une seule averse de plus de 50 mm).

Le seuil de 50 mm a été choisi de façon empirique et permet de ne pas donner trop d'importance à des événements pluvieux longs mais peu intenses. Par exemple, l'événement n° 5 de 3 jours présenté dans le tableau 1 n'aura qu'une averse principale.

D'autres valeurs du seuil ont aussi été testées. Des valeurs trop faibles conduisaient à mélanger trop d'averses faibles avec la population des averses fortes. On augmente alors le nombre d'averses principales mais elles sont en moyenne plus faibles et moins concernées par le phénomène de persistance dont on parlera par la suite. Des valeurs du seuil trop fortes conduisent à réduire le nombre d'averses principales à un effectif trop faible pour être traité statistiquement. Des essais ont aussi été réalisés pour faire varier ce seuil en fonction de la pluviométrie du poste. Les résultats obtenus n'étaient pas non plus satisfaisants car l'observation du phénomène de persistance (voir plus loin) devenait très variable et nécessitait une paramétrisation plus importante pour être prise en compte.

Ce seuil de 50 mm, fixe quelque soit le climat, conduit à la détermination d'une neuvième variable qui est le nombre d'averses principales de l'événement (NAVP). On en déduit le paramètre  $\mu_{NAVP}$  qui est le nombre moyen d'averses principales par événement. Ce paramètre est toujours supérieur à 1.

La typologie choisie pour les averses permet de sélectionner des averses dites « principales » ayant des caractéristiques bien différentes des autres averses. On trouve en moyenne que ces averses « principales » sont deux à trois fois plus intenses et deux à trois fois plus longues que les averses « ordinaires ». La typologie retenue permet aussi de résoudre le problème du choix de la loi de probabilité à retenir pour décrire la variable « volume » des averses. En effet, le mélange de deux types d'averses conduisait à la définition d'une loi basée sur la combinaison de deux lois exponentielles. La distinction des deux types d'averses conduit à l'utilisation d'une loi exponentielle simple pour décrire la variable « volume » de chacune d'entre elles.

À travers la nouvelle variable NAVP liée à la typologie effectuée, on pourra alors imposer lors de la génération de chaque événement pluvieux, d'avoir la présence d'une ou plusieurs averses plus fortes que les autres.

La typologie des averses concerne les variables VOL, DA et RX, qui seront caractérisées par des lois de probabilité ajustées sur les deux échantillons distincts d'averses. On notera alors ces variables VOLP, DAP et RXP si elles concernent les averses principales, et VOLO, DAO et RXO si elles concernent les averses ordinaires.

Pour la variable RPX, la distinction des types d'averses n'est pas nécessaire.

## Choix de deux saisons

La typologie retenue pour distinguer les averses ne tient pas compte des types de perturbations générant les cumuls de pluies. On traite les observations sans distinguer les événements convectifs, frontaux ou cycloniques (en climat tropical). Pour traiter ce point, une information historique sur les types de temps serait nécessaire, sans pour autant apporter une information indispensable à notre approche, qui reste strictement statistique.

Un traitement par saison permet en partie de séparer les événements plutôt orageux, des événements moins intenses et de durées plus longues. En effet, lors de l'analyse des événements pluvieux, deux saisons sont étudiées de façon distincte :

- la saison « été » concerne tous les événements des mois de juin à novembre ;
- la saison « hiver » concerne tous les événements des mois de décembre à mai.

Notons que le traitement par saison est surtout pertinent pour le passage aux débits. Il permet en effet de distinguer deux types de conditions initiales dans la modélisation de la transformation de la pluie en débit. Sous nos latitudes, des conditions hydriques plutôt « sèches » en été et à l'automne, et des conditions hydriques plutôt « humides » en hiver et au printemps. **La distinction de deux types de conditions hydriques nous semble plus importante qu'une distinction par type de temps, puisque l'objectif de la méthode est avant tout l'étude des crues.**

## Les lois de probabilité caractérisant les différentes variables

Le calage des lois de probabilité des variables du modèle est effectué sur les échantillons des valeurs extraites des hyétogrammes observés.

Les neuf variables définies sont utilisées pour décrire les événements pluvieux observés. L'étude d'une cinquantaine de postes pluviographiques, situés dans les départements du pourtour méditerranéen français, a permis de déterminer les lois de probabilité théoriques qui reproduisent au mieux les distributions de fréquences empiriques des différentes variables : la loi de Poisson est utilisée pour les variables NE, DA ; la loi géométrique est utilisée pour les variables NG, NA, et DIA ; la loi normale, tronquée entre 0 et 1, est utilisée pour la variable RPX (Arnaud, 1997 ; Cernesson, 1993 ; Cernesson *et al.*, 1996). Les variables RX, VOL et NAVP sont définies par des lois particulières qui seront présentées dans les paragraphes suivants.

L'application de ces lois sur de nouvelles séries de mesures en climat tempéré et tropical a aussi montré leur pertinence, conduisant à les garder quelle que soit la zone d'étude.

On retrouve dans le tableau 2 les différentes lois de probabilité et le nombre de paramètres associés à chaque variable du modèle.

Le calage du générateur nécessite la détermination d'une vingtaine de paramètres pour chacune des deux saisons considérées. Cependant, une étude de sensibilité du modèle à ses paramètres (Arnaud, 1997) a montré qu'il est peu sensible à certain d'entre eux. Ceux-ci permettent la construction des hyétogrammes de formes « réalistes », sans influence sur les quantités d'eau générées. D'autres sont peu variables, et peuvent même être fixés à une valeur régionale (voir p. 68 « Les paramètres fixés »).

À partir de ces lois théoriques calées, on peut générer de façon aléatoire des valeurs de chacune des variables afin de simuler des événements.

## Simulation d'un événement

Des hyétogrammes horaires sont construits averse après averse à partir des variables descriptives générées par un tirage aléatoire dans leur loi de probabilité par une méthode de Monte-Carlo. Pour cela on inverse la loi de probabilité, c'est-à-dire que l'on exprime la valeur de la variable en fonction de sa probabilité cumulée. On génère ensuite un nombre aléatoire suivant une loi uniforme entre 0 et 1, que l'on assimile à une probabilité cumulée et on en déduit la valeur de la variable.

La génération des variables se fait indépendamment au fur et à mesure des besoins et suivant un ordre précis. Le modèle simule ainsi une multitude des scénarios de pluies horaires (hyétogrammes).

L'organigramme de la figure 12, p. 36, présente l'ordre des tirages aléatoires effectués pour la génération des variables nécessaires à la simulation d'un événement pluvieux au pas de temps horaire.

Si l'on veut générer un échantillon d'événements équivalant à 100 ans de données, on génère 100 valeurs de la variable NE suivant sa loi calée sur les observations de la saison « été » et 100 valeurs de cette même variable suivant

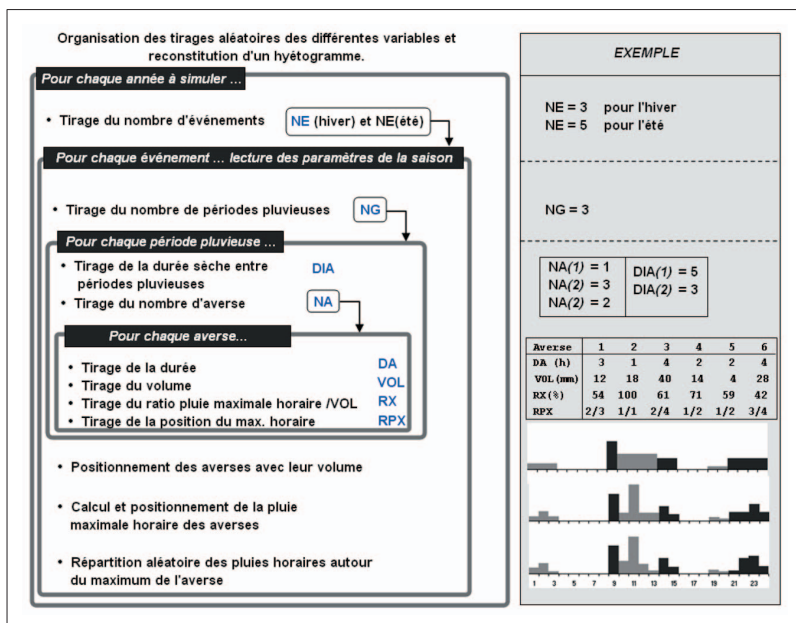
Variable	Loi de probabilité	Paramètres de détermination	Nombre de paramètres	Total
NE	• Loi de Poisson $P(k) = \sum_{j=0}^k e^{-p} \times \frac{p^j}{j}$	$p = \text{moyenne}$	1 par saison	2
NG	• Loi Géométrique $P(x \leq k) = 1 - (1 - p)^k$	$p = 1/\text{moyenne}$	1 par saison	2
DIA	• DIA < 12 h : loi Géométrique • DIA > 12 h : loi Uniforme	$1/\text{moyenne}$ et $\text{Seuil de troncature}$	2 par saison	4
NA	• Géométrique	$1/\text{moyenne}$	1 par saison	2
DA	• Averse principale, saison Hiver : – Si DA < 10 h : loi de Poisson – DA de 10 à 20 h : loi Uniforme • Averse principale, saison Été : – Si DA < 8 h : loi de Poisson – DA de 8 à 14 h : loi Uniforme • Averse ordinaire : – Si DA < 6 h : loi de Poisson – DA de 6 à 12 h : loi Uniforme	$\text{Moyenne}$ et $\text{seuil de troncature}$	2 par saison et par type d'averse	8
RPX	• Normale $F(x) = \frac{1}{\sigma_x \times \sqrt{2\pi}} \times e^{-\frac{1}{2} \left  \frac{x-\mu}{\sigma_x} \right ^2}$	$\text{Moyenne}$ et $\text{écart-type}$	2 par saison	4
VOL	• Type Exponentielle	Voir § 2.3.1	3 par saison et par type d'averse	12
NAV	• Type Exponentielle	Voir § 2.3.2	1 par saison	2
RX	• Type Uniforme	Voir § 2.3.4	1 par saison et par type d'averse	4

**Tableau 2.** Loi de probabilité et nombre de paramètres associés à chaque variable.

sa loi calée sur les observations de la saison « hiver ». On a alors le nombre total d'événements à générer.

Puis pour chacun des événements à générer, on procède au tirage suivant, en prenant les lois de probabilité des variables ajustées sur la saison appropriée :

- tirage du nombre de périodes pluvieuses : NG ;
- pour chaque période pluvieuse : tirage du nombre d'averses qui la compose et de la durée sèche qui la sépare de la période pluvieuse suivante. On génère donc NG valeurs de NA et (NG-1) valeurs de DIA. On peut alors calculer le nombre total d'averses composant l'événement ;
- tirage du nombre d'averses principales à considérer. Ce nombre sera borné par le nombre d'averses total déterminé précédemment. On génère ensuite la position des averses principales dans l'événement ;
- pour chaque averse, on génère les valeurs des variables DA, VOL en distinguant le cas où l'on a affaire à une averse principale ou à une averse ordinaire. La détermination de la durée de chaque averse permet alors de positionner



**Figure 12.** Organisation des tirages aléatoires des différentes variables pour la construction de hyétoGrammes simulés.

les différentes averses dans le temps, et de calculer la durée totale de l'événement. On dispose aussi de la hauteur de pluie totale de l'événement, par sommation de toutes les valeurs de la variable VOL ;

- pour chaque averse, on génère les valeurs des variables RX (en distinguant la typologie de l'averse) et RPX. On peut alors calculer le maximum horaire de chaque averse et le positionner ;
- reste alors à répartir la quantité restante,  $(1-RX)*VOL$ , de façon aléatoire de part et d'autre du maximum horaire, en respectant la décroissance des pluies lorsque l'on s'éloigne du maximum. On respecte alors la notion d'averse qui est la succession de pluies horaires ne présentant qu'un maximum relatif.

On génère ainsi un échantillon d'événements pluvieux au pas de temps horaire, d'effectif directement proportionnel au nombre d'années simulées. L'échantillon de ces événements pluvieux est statistiquement équivalent à l'échantillon des événements observés. Pour vérifier cela, on va comparer les courbes IDF (intensité durée fréquence) des pluies observées à celles des pluies simulées.

## Les variables de contrôle

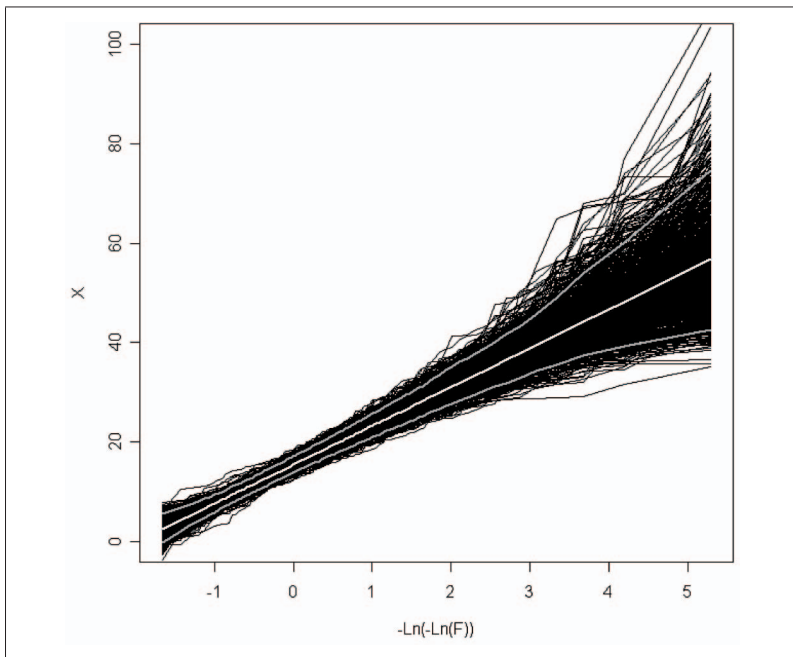
L'échantillon des événements pluvieux simulés est comparé à l'échantillon des événements pluvieux observés ayant servi au calage du générateur. Cette comparaison ne peut être que statistique. Pour juger de la pertinence de la structure temporelle des événements simulés, on doit prendre en compte différentes durées d'analyse. Le contrôle de la cohérence statistique et de la structure temporelle des événements simulés est basé sur l'analyse des distributions de fréquences des pluies de différentes durées (courbes IDF). Ces

variables ne sont pas utilisées lors de la modélisation. Elles sont déduites des hyétoigrammes simulés, de la même façon que pour les hyétoigrammes observés.

Les variables de contrôle sont les pluies maximales de 1, 2, 3, 4, 6, 12, 24, 48 et 72 heures, extraites des hyétoigrammes, qu'ils soient observés ou simulés.

À partir de la même paramétrisation du modèle, différentes simulations de  $N$  années peuvent être réalisées. Par le biais du hasard, ces différentes simulations donneront des échantillons d'événements différents. Les résultats de la simulation peuvent donc conduire à différentes distributions de fréquences uniquement tributaires de l'échantillonnage des événements simulés.

À titre d'exemple, on étudie la variable  $X$ , de moyenne 20 et d'écart-type 10 et qui suit une loi de Gumbel, pour laquelle on génère de façon aléatoire des échantillons de taille  $n$ . On trouve par exemple sur le graphique de la figure 13, 1 000 distributions de fréquences empiriques issues du tirage aléatoire de 100 valeurs de la variable  $X$ . On représente en blanc la distribution théorique de la variable  $X$  et en gris les bornes entre lesquelles sont contenues 90 % des valeurs.



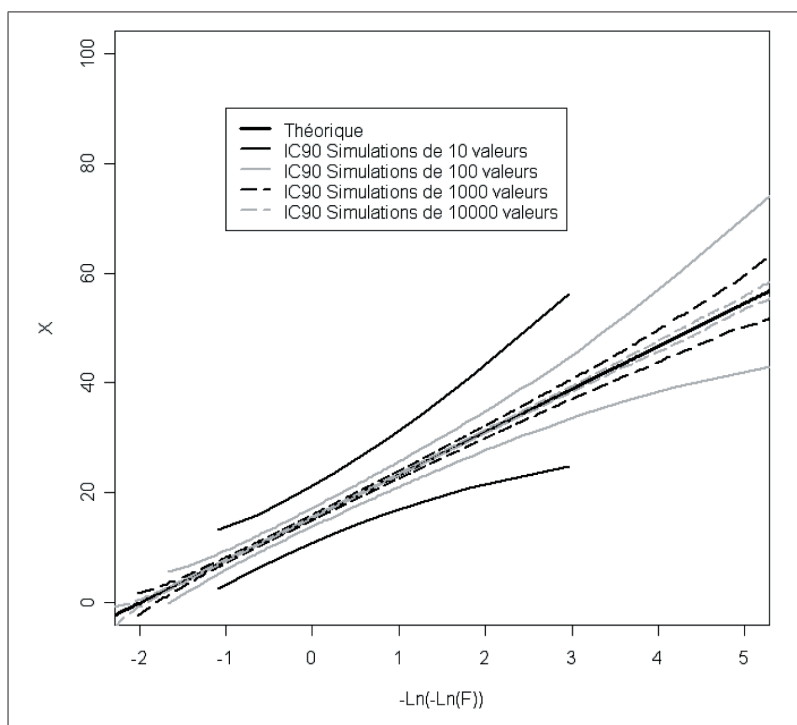
**Figure 13.** Distributions de fréquences de 1 000 échantillons de 100 valeurs tirées au hasard dans une loi de Gumbel caractérisant la variable  $X$  de moyenne 20 et d'écart-type 10 (courbes noires), intervalle de confiance à 90 % (courbes grises) et loi théorique (courbe blanche).

On observe ici la variabilité de l'estimation des quantiles liée à l'échantillonnage des valeurs simulées. Sur les observations, on a la même incertitude. Elle représente plus une incertitude sur l'estimation des périodes de retour à attribuer aux événements (observés ou simulés), qu'une incertitude sur les quantiles. Les écarts d'un échantillon à l'autre sont d'autant plus marqués que l'on

s'intéresse aux valeurs de rang proche de 1 (si le rang est associé aux valeurs classées par ordre décroissant) dont on a du mal à estimer la fréquence.

À partir de ces différents échantillons de simulation, on peut calculer des intervalles de confiance empiriques, associés au rôle de l'échantillonnage des simulations. **Cet intervalle n'a cependant rien à voir avec des incertitudes sur l'estimation des quantiles de pluies. En effet, il suffit d'augmenter la période de simulation pour stabiliser les valeurs des quantiles estimés, comme on peut le voir sur la figure suivante.**

Sur la figure 14 on représente cette fois les intervalles de confiance à 90 % des distributions de fréquences issues de 1 000 échantillons d'effectif variable : 10, 100, et 1 000 valeurs (pouvant être assimilées à un nombre d'années pour des échantillons de valeurs maximales annuelles). On observe toujours la même tendance, à savoir une forte variabilité du quantile associé au rang 1, et une décroissance progressive de cette variabilité lorsque le rang augmente. Cependant, si on prend par exemple le quantile associé à la fréquence 0,9 (valeur décennale pour un échantillonnage de valeurs maximales annuelles), sa variabilité est grande si on simule seulement 10 valeurs (valeur de rang 1). Elle diminue lorsque l'on augmente l'effectif des échantillons. Pour des simulations de 100 valeurs, la fréquence 0,9 est associée au rang 10, et pour des simulations de 1 000 valeurs, la fréquence 0,9 est associée au rang 100. On s'aperçoit alors de la forte diminution de l'intervalle de confiance du quantile associé à la fréquence 0,9 lorsque l'on augmente la taille de l'échantillon simulé. On raisonne de la même façon avec le générateur de pluie.



**Figure 14.** Intervalles de confiance à 90 % issus de 1 000 simulations dans une loi de Gumbel de la variable X de moyenne 20 et d'écart-type 10, et d'effectif variable.



Afin de stabiliser les estimations des quantiles et retenir une distribution de fréquences indépendante de l'échantillonnage des événements simulés, il faut générer pratiquement un échantillon équivalent à 1 000 fois la plus grande période de retour désirée. Par exemple, pour obtenir une distribution de fréquences stabilisée à 100 ans, on doit simuler 100 000 ans et ne retenir que les valeurs de rang inférieur à 1 000. Une autre procédure équivalente est de simuler 1 000 échantillons de 50 ans (la période de retour maximale calculée par la formule de Hazen est alors de 100 ans) et de calculer la distribution moyenne (Arnaud *et al.*, 1998).

### Présentation des résultats

Les résultats de SHYPRE, à travers les courbes IDF déduites des hyétogrammes simulés, sont issus de simulations 1 000 fois plus longues que la plus grande période de retour désirée, afin de s'affranchir du problème de représentativité des échantillons aléatoires de pluies simulées. On ne s'intéresse alors qu'à des valeurs de rang inférieur à 1 000, faiblement influencées par l'échantillonnage.

## Particularités du générateur

Une des hypothèses de base des générateurs stochastiques est l'indépendance des variables du modèle. Ces variables peuvent ainsi être générées de façon indépendante lors de la modélisation. Le choix des variables du modèle va d'ailleurs dans ce sens lors de sa conception. Une étude de la dépendance des variables entre elles est cependant nécessaire pour vérifier cette hypothèse.

Les réalisations successives d'une même variable sont aussi censées être indépendantes, ce qui permet de générer une même variable de façon aléatoire par tirage dans sa loi de probabilité.

### Spécificités du générateur

L'une des particularités du générateur de la méthode SHYPRE concerne l'étude et la prise en compte de la dépendance observée entre certaines variables du modèle et de la dépendance observée entre certaines réalisations d'une même variable. Ce travail permet de respecter au mieux les observations à travers la génération des variables aléatoires.

La seconde particularité est liée à sa robustesse face à l'échantillonnage des valeurs extrêmes, à travers une étude adimensionnelle des lois sensibles du modèle.

Le générateur de pluie est basé sur l'analyse de variables aléatoires descriptives permettant de générer des chroniques horaires d'événements pluvieux. Ces variables aléatoires sont définies par leur distribution de probabilité ajustée sur l'échantillon des valeurs observées. On pourrait alors penser être soumis aux mêmes difficultés que celles présentées en introduction sur les statistiques classiques. Ce n'est pas le cas pour deux raisons :

– **d'une part l'effectif des valeurs observées est nettement plus important** que lorsque l'on s'intéresse à une seule observation par événement (la pluie maximale en 24 heures de l'événement par exemple), voire à une seule observation

par an (échantillonnage des valeurs maximales annuelles). En effet, on a plusieurs événements par an qui sont composés de plusieurs averses. L'analyse des lois caractéristiques des averses repose généralement sur des échantillons d'effectif supérieur à 100, voire 1 000 ;

– **d'autre part, un effort particulier a été fait dans le choix des lois** de probabilité utilisées, en choisissant des lois **robustes**, c'est-à-dire le plus faiblement paramétrées. Les lois les plus sensibles sont étudiées de façon adimensionnelle afin d'augmenter la représentativité de l'échantillon par agglomération des variables.

L'étude de la loi du volume des averses et de la loi du nombre des averses a été réalisée de façon adimensionnelle et est présentée ici. On présente aussi la dépendance observée entre certaines variables et leur modélisation.

## Loi du volume des averses

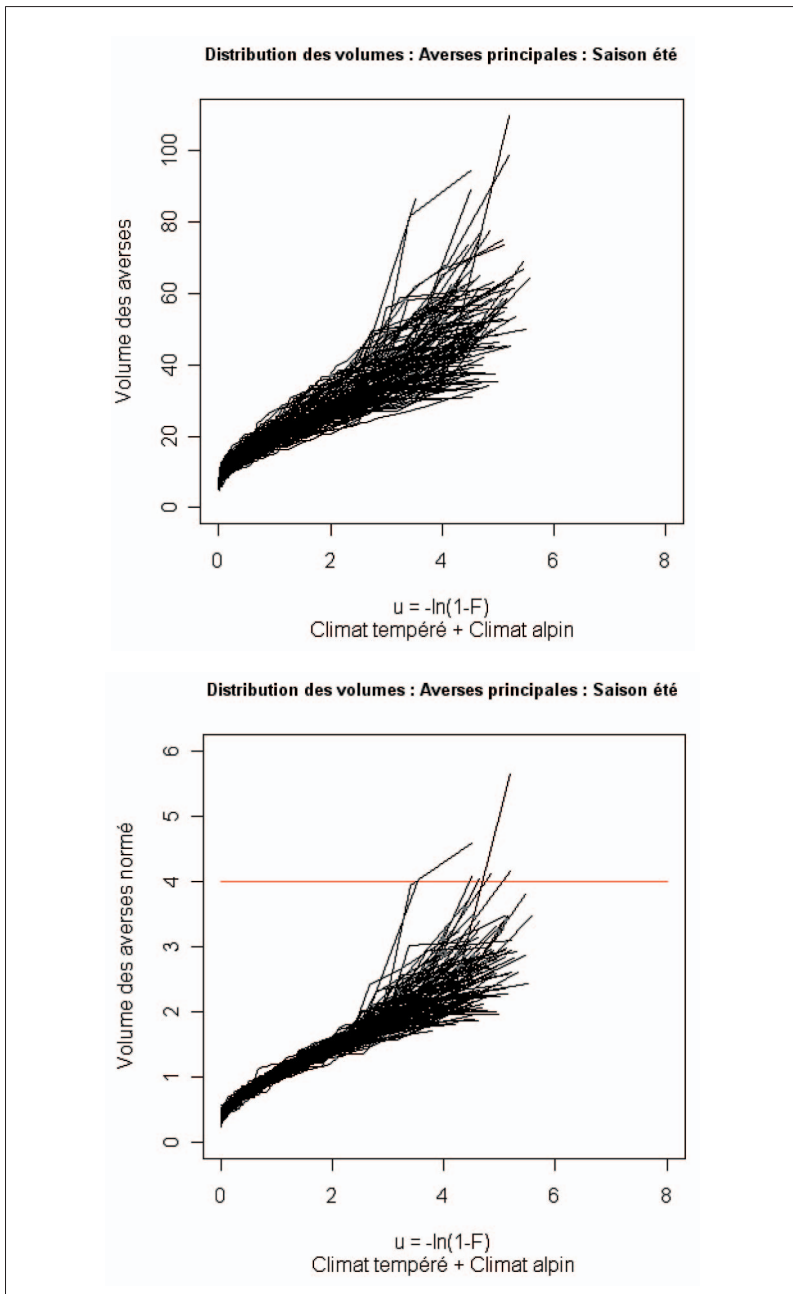
Le choix de la loi des volumes des averses est délicat car il peut fortement conditionner le comportement à l'infini du modèle (Arnaud, 1997). Dans un premier temps nous n'avons pas voulu faire d'hypothèse forte sur l'extrapolation de la loi des volumes d'averse, **notre objectif étant en priorité la recherche d'une loi peu sensible à l'échantillonnage.**

Cette recherche nous a conduit à choisir une loi exponentielle simple calée sur les valeurs inférieures à quatre fois la moyenne de la variable (Arnaud, 1997).

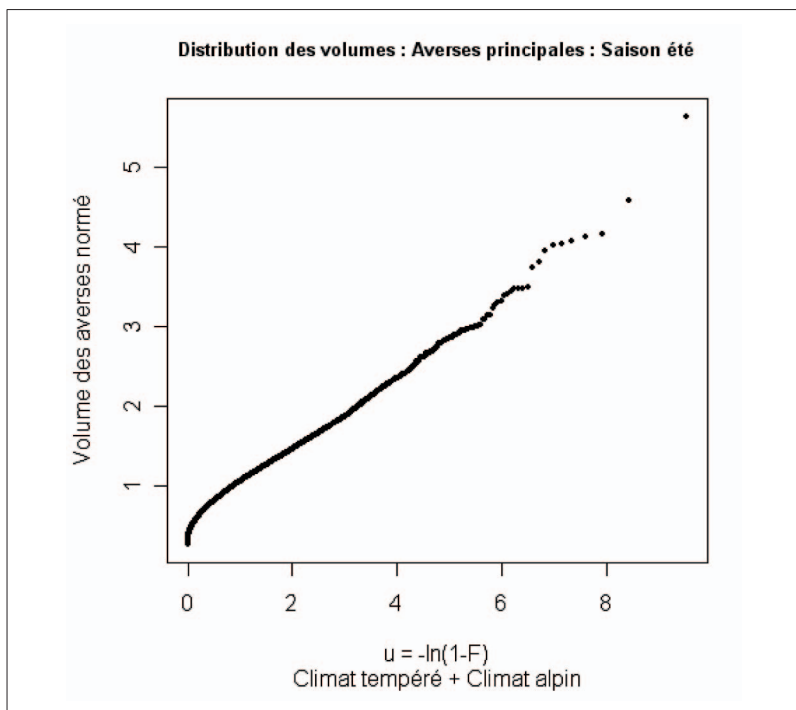
Les graphiques de la figure 15 montrent les distributions de fréquences de la variable VOLP, observées sur une centaine de postes pluviographiques français (en climat « tempéré » et « alpin »). On observe que la loi exponentielle (représentée par une droite sur ce système d'axes) est adaptée si l'on ne tient pas compte de certaines valeurs, par exemple les valeurs supérieures à quatre fois leur moyenne. On peut noter que ce seuil conduit à exclure de l'analyse un nombre très faible de valeurs, et pour la plupart des postes aucune valeur.

On choisit donc de modéliser la distribution de fréquence des volumes des averses principales par une loi exponentielle calée sur des valeurs de volumes inférieures à quatre fois la moyenne locale de la variable. L'ajustement des paramètres de la loi est effectué par une régression linéaire entre les valeurs observées de la variable et leur fréquence de non-dépassement exprimée en logarithme. Dans ce cas les valeurs de la variable supérieure à quatre fois la moyenne sont ignorées dans la régression, mais leur présence dans l'échantillon est prise en compte dans le calcul des fréquences empiriques des autres valeurs. Par exemple, si on a 100 valeurs observées dont deux dépassent le seuil de quatre fois la moyenne, la fréquence empirique des 98 valeurs prises en compte pour l'ajustement est associée à un effectif de 100 valeurs (les rangs 1 et 2 n'étant pas attribués).

Si la présence de valeurs fortes dans les distributions locales pourrait conduire à rejeter la loi exponentielle simple, la distribution « régionale » (figure 16), établie par l'agglomération des valeurs normées de l'ensemble des postes appartenant à une même zone climatique et une même saison, semble confirmer le choix de la loi exponentielle pour décrire la variable volume des averses.



**Figure 15.** Distribution de fréquences des volumes des averses principales pour l'ensemble des postes du climat « tempéré + alpin » : valeurs brutes et valeurs normées par la moyenne locale.



**Figure 16.** Distributions de fréquences « régionales » des volumes des averses principales, réduits par leur moyenne locale, sur l'ensemble des postes du climat « tempéré + alpin » (plus de 7 000 points).

Pour les averses ordinaires la même étude a été faite et la loi exponentielle est aussi adaptée, pour les valeurs inférieures à six fois la moyenne locale.

La loi des volumes d'averses est donc calée suivant une loi exponentielle simple, en ne prenant pas en compte les observations particulièrement extrêmes. Un seuil de sélection de la variable est établi à quatre fois la moyenne locale pour les averses principales et six fois la moyenne locale pour les averses ordinaires. Bien entendu, lors de la simulation des volumes d'averse dans ces lois exponentielles, des valeurs supérieures à ce seuil pourront être générées.

Ces valeurs extrêmes, exclues de l'ajustement de la loi de la variable, peuvent être associées à différents problèmes d'échantillonnage. Elles peuvent être simplement dues au découpage des averses. Par exemple, une des deux averses conduisant à un volume supérieur à 4 % la moyenne sur le poste de Versailles, est liée à une averse unique tombée le 3 septembre 1987 et dont l'enregistrement est caractérisé par une intensité constante de 7,6 mm/h pendant 11 heures. Dans ce cas on peut penser qu'une précision dans le dépouillement aurait pu conduire à séparer cette averse en deux ou trois, séparée par un minimum relatif.

Nous avons calculé la durée moyenne des averses concernées par le dépassement de seuil, en distinguant la climatologie des postes, la saison et le type d'averse. Il s'avère que dans tous les cas la durée de ces averses est en moyenne **deux fois** plus forte que la durée moyenne des autres averses.

Cet écart, plus marqué sur les averses ordinaires, peut aussi révéler un problème de typologie des averses, mais qui ne concerne qu'une faible proportion d'averses (moins de 1 %). Par exemple, pour les distributions de la variable VOLO en « été » pour les postes « méditerranéens », certaines averses ordinaires extrêmes auraient dû être analysées comme des averses principales. Ce cas peut se produire, si trois fortes averses successives se trouvent dans un événement de deux jours, entraînant une seule pluie journalière dépassant les 50 mm, on n'en choisira que deux comme « principales ».

Le choix effectué de ne pas prendre en compte les valeurs extrêmes des volumes d'averses, repose sur le souci de ne pas mélanger une information relevant de l'observation des phénomènes extrêmes et une information induite pas l'échantillonnage et/ou le mode d'analyse des données.

### La loi des volumes

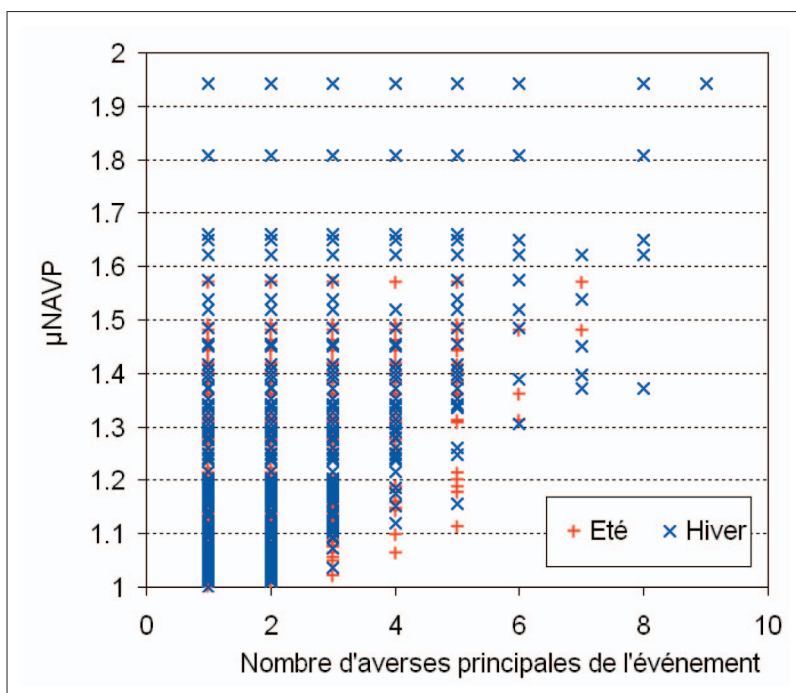
La typologie des averses reste une étape délicate qui peut influencer le calage de certaines lois de probabilité telles que la loi des volumes d'averses. Afin de ne pas être trop influencé par l'échantillonnage de cette variable, il a été choisi de ne pas tenir compte des valeurs extrêmes prises par cette variable. La loi de probabilité utilisée pour les « volumes des averses » est une loi exponentielle calée sur les valeurs inférieures à quatre fois la moyenne de la variable pour les averses « principales », et sur les valeurs inférieures à six fois la moyenne pour les averses « ordinaires ».

## Loi du nombre d'averses principales

Le nombre d'averses principales est égal à 1 + le nombre de jours où la pluie a dépassé 50 mm. Cette définition vient de la typologie des averses que l'on a choisie. Le graphique de la figure 17 montre, pour l'ensemble des postes retenus pour cette étude, le nombre d'averses principales observées pour chaque événement, en fonction de sa valeur moyenne (paramètre  $\mu$ NAVVP).

On déduit sur ce graphique, que la distribution de fréquences de la variable NAVVP doit avoir une dissymétrie positive. Les valeurs maximales observées du nombre d'averses principales sont fortes comparées à leur moyenne. Pour les postes présentant des valeurs moyennes de NAVVP proches de 2 (fortes valeurs), on observe des événements présentant jusqu'à huit cumuls journaliers ayant dépassé les 50 mm (NAVVP = 9). Pour les postes présentant des valeurs faibles de  $\mu$ NAVVP (proche de 1) on observe des événements présentant jusqu'à trois cumuls journaliers ayant dépassé les 50 mm. Ce graphique nous donne donc une idée de la distribution de la variable NAVVP, caractérisant la persistance pluviométrique des événements.

La distribution de fréquence du nombre d'averses principales étant plutôt dissymétrique, on choisit une loi de probabilité du type exponentielle pour cette variable. Si une loi du type exponentielle semble préférable pour modéliser la distribution de fréquences de la variable NAVVP, la loi exponentielle simple paramétrée par la moyenne de la variable ne s'ajuste pas toujours correctement sur les distributions observées. De plus on ne dispose pas d'assez de valeurs pour ajuster une loi de probabilité sur chaque poste, en particulier lorsque la variable  $\mu$ NAVVP est proche de 1.



**Figure 17.** Nombre d'averses principales (d'un événement) en fonction de sa valeur moyenne.

Pour ces raisons et afin de ne pas introduire des paramètres supplémentaires au modèle, on choisit d'ajuster une loi de probabilité « adimensionnelle » paramétrée par la variable  $\mu_{NAVP}$ . Pour cela, on mélange toutes les valeurs de NAVP observées pour des postes présentant des valeurs de  $\mu_{NAVP}$  comparables, afin d'ajuster une loi de probabilité unique pour la variable NAVP.

On détermine alors une loi de type exponentiel rendue discrète par l'expression :

$$NAVP = 1 + ENT \left[ - (2 \times \mu_{NAVP} - 1,7) \times \text{Ln} (1 - F^{\mu_{NAVP}}) \right]$$

Où ENT() est la fonction partie entière, F est la fréquence cumulée.

Cette loi est paramétrée uniquement par la valeur de  $\mu_{NAVP}$  qui représente le nombre moyen d'averses principales par événement. L'utilisation de cette loi « unique » peut, cependant, introduire un biais dans la valeur moyenne de NAVP simulée. C'est pourquoi, un point essentiel dans la modélisation effectuée pour la génération de la variable NAVP est de respecter sa valeur moyenne observée (c'est-à-dire la valeur  $\mu_{NAVP}$ ) par le biais de retraiter.

Les graphiques de la figure 18 montrent les distributions « régionales » observées pour différentes classes de  $\mu_{NAVP}$ , ainsi que la loi de probabilité théorique que l'on va utiliser.

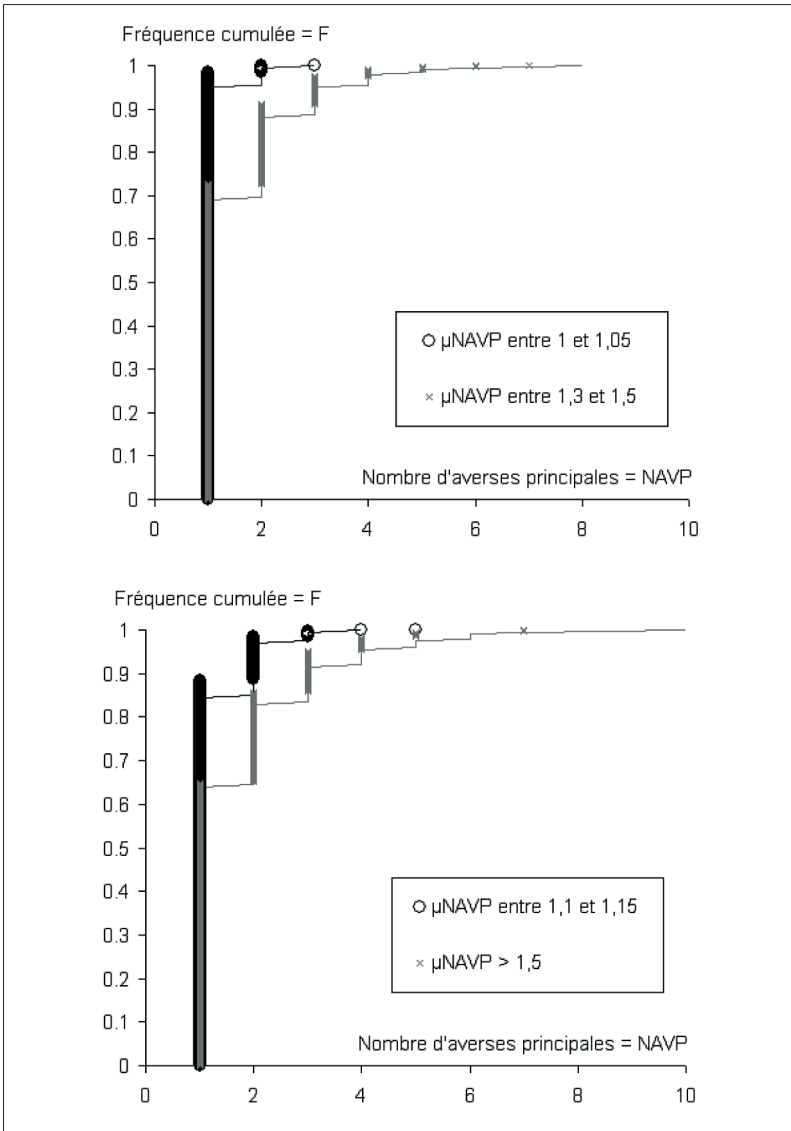


Figure 18. Distributions du nombre d'averses principales par classe de  $\mu_{NAVP}$  : valeurs observées (points) et valeurs théoriques (traits).

### La loi du nombre d'averses principales

Ce nombre est donc généré de façon aléatoire à partir d'une loi de type exponentiel, uniquement paramétrée par la valeur  $\mu_{NAVP}$ , et de façon à respecter cette valeur moyenne du nombre d'averses principales observées.

## Persistance des averses principales

Le phénomène de persistance correspond à la propagation dans le temps des caractéristiques des averses. Ce phénomène observé sur la plupart des événements présentant de forts cumuls de pluies (en particulier pour les postes d'altitude) se traduit par la succession de fortes averses. Ainsi, un très fort cumul de pluies sur plusieurs heures, peut être consécutif à une agglomération d'averses qui sont importantes sans être exceptionnelles.

La persistance des averses signifie que la succession des différentes averses observées n'est pas complètement aléatoire. Lorsque l'on observe un événement présentant une averse « forte » en termes de volume, il y a généralement présence d'autres averses « fortes » dans l'événement, du moins en proportion plus importante que le hasard seul ne peut expliquer.

Afin de mieux modéliser ce phénomène de persistance, on a cherché à mieux le caractériser, par l'intermédiaire d'une analyse des observations, pouvant conduire à une modélisation objective du phénomène. Le phénomène de persistance des averses conduit généralement à l'apparition des événements pluvieux extrêmes. Il peut être difficile de le paramétrer du fait du faible nombre d'observations de ces événements extrêmes. C'est pourquoi, il est intéressant d'agglomérer les observations de différents postes pluviographiques, afin de consolider l'échantillon des valeurs observées, en particulier vers les fréquences rares.

### Intérêt des fréquences empiriques

Pour consolider l'échantillon des valeurs fortes par agglomération des observations, on étudie non pas les valeurs prises par la variable, mais la fréquence empirique des valeurs observées.

L'étude des fréquences empiriques de la variable permet de s'affranchir de la distribution de fréquence de la variable étudiée, et peut faciliter la modélisation à effectuer dans des modèles stochastiques. C'est pourquoi l'on travaille sur la fréquence des variables plutôt que sur les valeurs de la variable. Par la suite, lorsque l'on parlera de la fréquence d'une variable, il s'agira de la fréquence de non-dépassement, calculée par exemple par la formule de Hazen, donnant une fréquence proche de 1 pour la plus forte valeur observée.

### MISE EN ÉVIDENCE DE LA PERSISTANCE

L'événement représenté sur la figure 19 illustre le phénomène de persistance des averses, et montre les difficultés que l'on peut avoir pour générer de tels événements à partir de modèles basés sur l'indépendance des variables.

Cet événement, formé de 10 averses dont trois principales, correspond au plus fort événement pluvieux observé sur ce poste, avec une pluie totale de 234 mm dont 152 sont tombés en 24 heures.

Les trois averses principales de cet événement, sont aussi les trois plus fortes averses observées sur la chronique entière, pour la saison hiver. On a donc les trois plus fortes averses observées de la saison qui se retrouvent dans le même événement. De plus, si on étudie le rang des averses ordinaires qui composent cet événement, on retrouve les deux plus fortes averses ordinaires observées sur l'ensemble des événements de la saison. Ces deux plus fortes



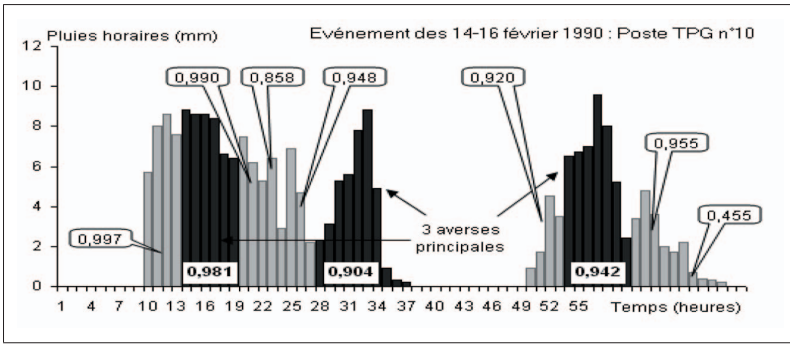


Figure 19. Événement des 14-15-16 février 1990, sur le poste n° 10 du TPG (Isère).

averses ordinaires se trouvent en plus placées à côté de la plus forte averse principale.

Cet exemple montre les difficultés que peuvent rencontrer les modèles stochastiques pour générer des événements extrêmes, si l'on reste dans un schéma d'indépendance des tirages aléatoires.

La dépendance mise en évidence reste indépendante de la loi de probabilité théorique choisie pour décrire le volume des averses. En effet, on juge de la dépendance des fréquences empiriques, qui sont indépendantes de la loi qui sert à décrire la variable.

La dépendance n'est pas non plus liée au mode d'analyse des averses. Par exemple, on pourrait penser que la présence d'un minimum relatif peut séparer une succession de pluies horaires en deux averses, alors qu'elle aurait pu être analysée comme appartenant à une même période pluvieuse. Une étude de l'auto-corrélation des averses successives a été menée par Muller (2006) sur l'ensemble des averses (sans distinction liée à la typologie). Elle a montré que l'auto-corrélation des averses est maximale à l'ordre 4. C'est-à-dire qu'il y a plus de dépendance entre deux averses séparées par deux ou trois averses, qu'entre deux averses successives. En fait, on revient à une auto-corrélation d'ordre 1 si l'on supprime les « petites » averses séparant les plus fortes. Ce point confirme la présence d'une « persistance » sur les averses les plus fortes (les averses « principales » pour nous).

Les diverses observations présentées ci-dessus nous conduisent à poser cette question fondamentale. **Dans ce cas, avec quelle probabilité peut-on simuler un tel événement ?**

On retrouve ce cas de figure pour de nombreux événements extrêmes observés. Pour les postes du climat tempéré où le nombre d'averses principales par événement est généralement de 1, la persistance des averses principales est moins visible mais reste présente. Par exemple, l'événement des 5 et 6 août 1997, sur le poste de Trappes (78) est composé de deux fortes averses principales ( $F = 0,943$  et  $0,898$ ) ainsi que d'une forte averse ordinaire ( $F = 0,975$ ).

Cette persistance est plus marquée pour les postes du climat méditerranéen où l'on trouve fréquemment les deux plus fortes averses principales de la série dans un même événement. La persistance est encore plus marquée pour les

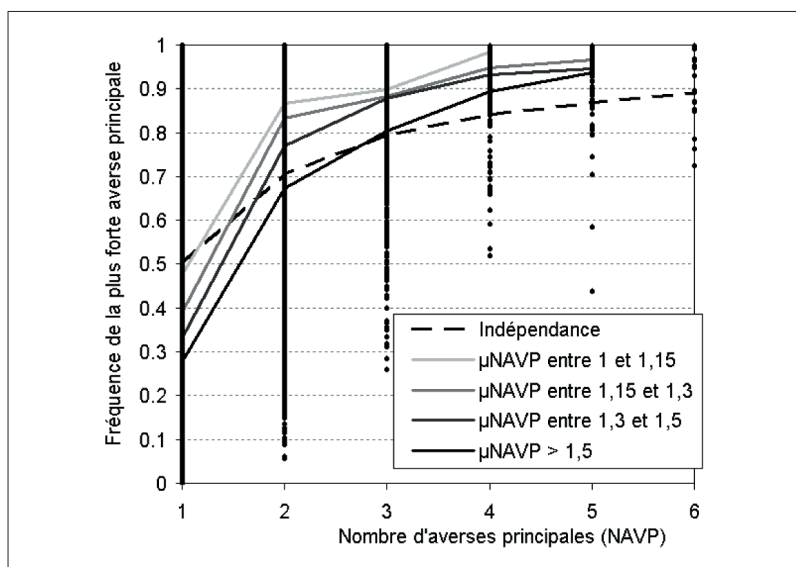
postes du climat tropical (par exemple île de la Réunion) où il arrive que les trois ou quatre plus fortes averses principales de la série se retrouvent dans un même événement pluvieux.

### ANALYSE DE LA PERSISTANCE

Ce phénomène de persistance a été étudié sur la variable VOLP (volume des averses principales) d'une part parce qu'il est essentiellement marqué pour cette variable, et d'autre part parce que c'est la variable la plus importante du modèle.

L'analyse de la fréquence de non-dépassement des averses « principales » a été généralisée pour l'ensemble des postes étudiés.

On s'intéresse tout d'abord à la plus forte averse principale de l'événement. La figure 20 présente l'évolution de la fréquence de la plus forte averse principale de l'événement, en fonction du nombre d'averses principales de l'événement. Les points correspondent aux valeurs obtenues à partir des postes étudiés (soit près de 40 000 valeurs). Chaque point correspond donc à un événement pluvieux analysé. Les courbes en trait plein représentent la fréquence médiane en fonction du nombre d'averses, calculée pour les postes appartenant à une même classe de valeur du paramètre  $\mu\text{NAVVP}$ . Enfin, la courbe en pointillés correspond à la courbe théorique que l'on devrait obtenir en cas d'indépendance des valeurs du volume des averses principales d'un événement.

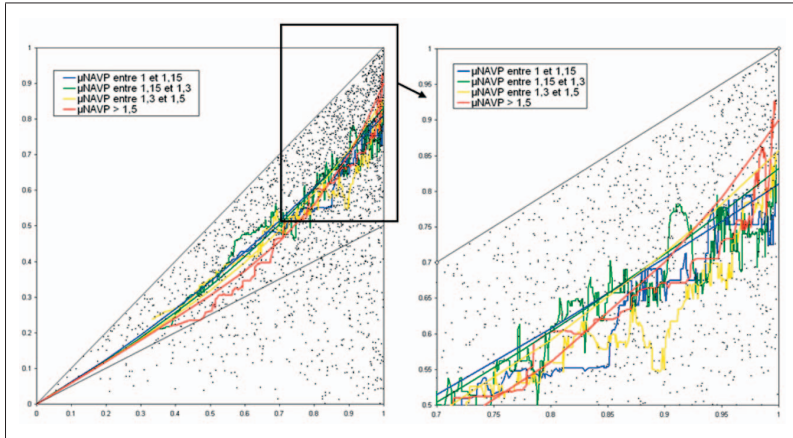


**Figure 20.** Évolution de la fréquence de l'averse principale la plus forte avec le nombre d'averses de l'épisode.

Dans le cas de l'indépendance des averses principales, il est normal d'observer une fréquence de non-dépassement de l'averse principale la plus forte, qui soit croissante avec le nombre d'averses principales. Cependant, on observe des valeurs de fréquences de non-dépassement supérieures à celles que l'on devrait théoriquement obtenir en cas d'indépendance des variables. Lorsque l'on a plusieurs averses principales dans un événement, on a en général plus de chances d'avoir la plus forte des averses principales, plus forte qu'un tirage

aléatoire ne permettrait de le faire. Ce phénomène est constaté sur tous les postes étudiés, et plus ou moins accentué avec le nombre moyen d'averses principales du poste (figure 20).

Lorsqu'il y a plusieurs averses « principales » au cours d'un événement pluvieux, on peut s'intéresser au phénomène de persistance. La figure 21 illustre ce phénomène à travers l'évolution de la fréquence moyenne des averses principales autres que la plus forte de l'événement (notée  $\bar{F}_{\text{secondaires}}$ ), en fonction de la fréquence de la plus forte averse principale (notée  $F_{\text{max}}$ ).



**Figure 21.** Fréquence moyenne des autres averses principales de l'événement en fonction de la fréquence de la plus forte averse principale : valeurs brutes (points) et valeurs médianes (courbes).

Les points représentent les valeurs observées (plus de 2000 valeurs). Chaque point correspond donc à un événement pluvieux composé d'au moins deux averses principales. Les courbes pleines correspondent aux médianes conditionnelles calculées pour des postes appartenant à une même classe de valeurs du paramètre  $\mu\text{NAVP}$ . Enfin, la courbe en pointillés correspond à la courbe théorique que devrait prendre la médiane conditionnelle si les averses principales d'un même événement étaient indépendantes entre elles, en termes de probabilité d'apparition.

Là encore, le graphique montre que les averses principales d'un même événement ne sont pas indépendantes entre elles. Plus l'averse principale la plus forte est rare (fréquence de non-dépassement élevée) plus les autres averses principales de l'événement risquent aussi d'être rares.

Ce phénomène de persistance est marqué quelle que soit la classe de paramètre  $\mu\text{NAVP}$  considérée. Ce phénomène concerne donc tous les postes mais pour un nombre d'événements variables, selon la valeur du nombre moyen d'averses principales du poste concerné (valeur de  $\mu\text{NAVP}$ ). En effet, si le phénomène de persistance des averses principales se caractérise pratiquement de la même façon pour différentes valeurs de  $\mu\text{NAVP}$ , il ne concerne pas le même nombre d'averses. Par exemple, en climat tempéré où la valeur de  $\mu\text{NAVP}$  est proche de 1, la persistance des averses concerne au maximum deux voire trois averses, pour peu d'événements, alors qu'en climat tropical humide où la valeur de  $\mu\text{NAVP}$  peut atteindre 1,8, la persistance des averses concerne plus d'événements, dont le nombre d'averses peut aussi être plus important (jusqu'à une dizaine d'averses).

## MODÉLISATION DE LA PERSISTANCE

La mise en évidence de la dépendance des averses les plus fortes remet en cause le tirage aléatoire de la variable « Volume » caractérisant les averses principales des événements. Il est donc nécessaire de prendre en compte ce phénomène dans la modélisation.

Théoriquement, si les volumes des averses principales d'un événement pluvieux (ayant plus d'une averse principale) étaient indépendants, les fréquences moyennes des averses principales « secondaires » devraient être réparties de façon aléatoire entre 0 et la fréquence du volume maximal. La droite passant par les points (0 ; 0) et (1 ; 0,5) devrait être la fonction représentant la médiane de  $\bar{F}_{\text{sec ondaïres}}$  en fonction de  $F_{\text{max}}$ .

Or on s'aperçoit que la distribution des valeurs de  $\bar{F}_{\text{sec ondaïres}}$  n'est pas conforme avec une distribution résultant de l'indépendance des volumes des averses principales d'un événement. On observe que de façon médiane plus  $F_{\text{max}}$  est forte, plus  $\bar{F}_{\text{sec ondaïres}}$  est supérieure à la valeur théorique liée à une hypothèse d'indépendance. Ce phénomène est observé même pour des valeurs de  $\mu\text{NAVP}$  proches de 1, et il est d'autant plus marqué que la valeur moyenne de  $\mu\text{NAVP}$  augmente, en particulier vers les valeurs fortes de  $F_{\text{max}}$  (graphique de droite de la figure 21).

Cette tendance a été prise en compte dans la modélisation en conditionnant la fréquence des volumes des averses principales secondaires en fonction de la fréquence du volume le plus fort. Pour cela on modélise la courbe représentant au mieux la valeur médiane de  $\bar{F}_{\text{sec ondaïres}}$  en fonction de  $F_{\text{max}}$  avec l'expression suivante. Cette expression a été déterminée de façon empirique de façon à s'ajuster au mieux aux courbes issues des graphiques de la figure 21, en particulier vers les valeurs fortes de  $F_{\text{max}}$  :

$$F_{\text{médiane}} = 0,6 \times F_{\text{max}} + \beta \times F_{\text{max}}^{(1 + \mu\text{NAVP}^3)} \text{ avec } \beta = 0,05 + (\mu\text{NAVP} - 1) / 7$$

Lors de la modélisation d'un événement pluvieux, on procède alors de la façon suivante :

- on génère le nombre d'averses principales de l'événement : NAVP ;
- si on a plus d'une averse principale, on tire au hasard NAVP nombres aléatoires et on retient la valeur maximale :  $F_{\text{max}}$  ;
- on calcule la fréquence médiane ( $F_{\text{médiane}}$ ) de la variable ( $\bar{F}_{\text{sec ondaïres}}$ ) par la relation précédente ;
- on génère ensuite la valeur de  $\bar{F}_{\text{sec ondaïres}}$  en respectant  $F_{\text{médiane}}$ . Pour cela, on tire un nombre au hasard entre 0 et 1. Si ce nombre est inférieur à 0,5 alors  $\bar{F}_{\text{sec ondaïres}}$  sera compris entre 0 et  $F_{\text{médiane}}$ , et s'il est supérieur à 0,5 alors  $\bar{F}_{\text{sec ondaïres}}$  sera compris entre  $F_{\text{médiane}}$  et  $F_{\text{max}}$  ;
- et enfin, on génère les (NAVP – 1) fréquences des averses principales secondaires en respectant la valeur de  $\bar{F}_{\text{sec ondaïres}}$ .

*Remarque 1. La même approche a été réalisée en exprimant la fréquence empirique par sa transformation logarithmique, cela afin de prendre en compte le fait que la variable suit une loi exponentielle. Ce changement de variable permet de « dilater » la représentation graphique de la figure 21 et donner plus de poids aux fréquences rares. Les résultats obtenus n'ont pas apporté une information supplémentaire et n'ont pas été retenus.*

*Remarque 2. La modélisation de la courbe médiane peut introduire un biais dans la génération des variables « volume des averses », en générant trop souvent des fréquences fortes. Ce biais a été supprimé par le tirage de certaines fréquences de façon à respecter une fréquence moyenne de 0,5 pour la variable générée. On vérifie aussi que l'on respecte bien la valeur moyenne des variables « volume des averses ».*

## Résultat de la modélisation de la persistance

La figure 22 montre l'impact de la modélisation de la persistance sur la restitution des distributions de fréquences des pluies maximales en 24 heures, pour trois postes présentant des valeurs différentes du paramètre  $\mu_{NAVP}$ .

On remarque bien sur ces exemples, l'influence de la modélisation de la persistance. Lorsque le nombre moyen d'averses principales est proche de 1, la modélisation de la persistance est presque nulle et n'a donc pas d'effet sur la modélisation. C'est le cas pour les postes du milieu tempéré français. En revanche, plus le nombre moyen d'averses principales augmente, plus l'effet de la modélisation de la persistance se fait sentir. On génère alors des pluies plus fortes en 24 heures, issues de la succession de plusieurs averses principales. Dans le cas extrême des postes les plus pluvieux (île de la Réunion) l'effet de la persistance est très important, dès les fréquences courantes.

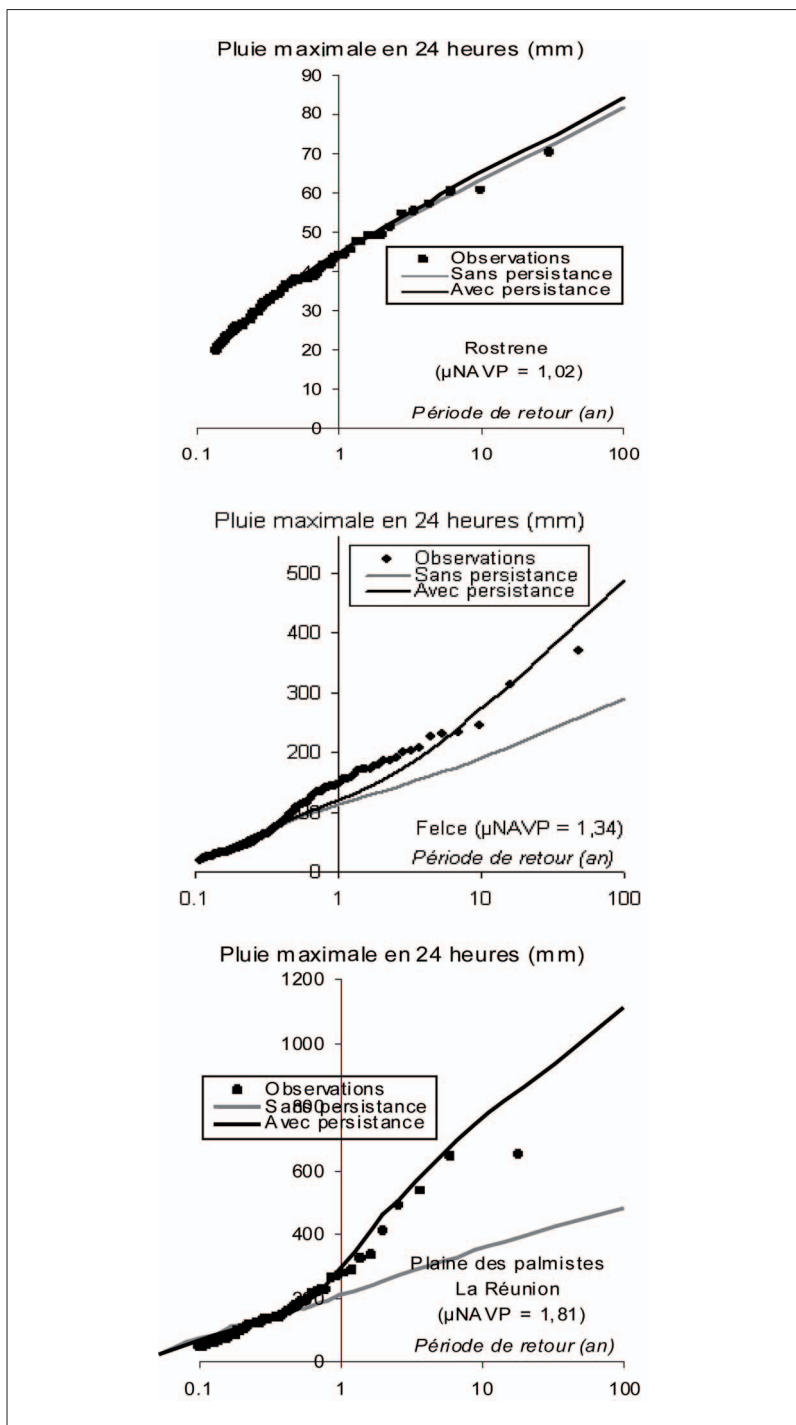
### Persistance des averses principales

Lorsqu'il y a plusieurs averses principales dans un événement pluvieux, ces averses ne sont pas indépendantes les unes des autres. Un phénomène de persistance est observé. On observe que plus le nombre d'averses principales dans un événement est fort, plus on a de chances d'avoir une averse principale très forte. Cette probabilité est supérieure à celle attendue en cas d'indépendance. De plus, quand une averse principale est très forte, c'est généralement l'ensemble des averses principales de l'événement qui sont importantes.

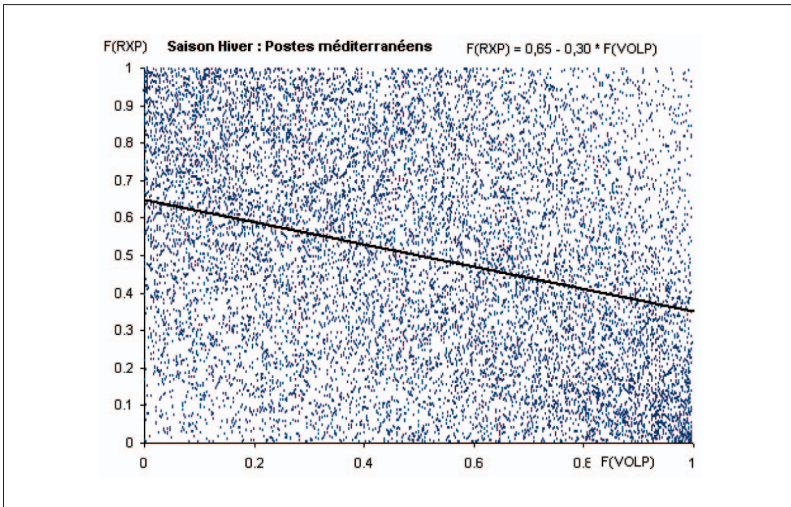
Cette persistance des averses principales, a été observée et paramétrée par le paramètre  $\mu_{NAVP}$ .

## Relation entre les variables RX et VOL

La relation entre les variables RX (rapport entre la pluie horaire maximale et le volume de l'averse) et VOL (volume de l'averse) a été étudiée sur la population des averses principales, qui sont responsables des valeurs maximales des événements. Cette dépendance est illustrée sur la figure 23. Cette figure présente la fréquence de la variable RXP en fonction de la fréquence au non-dépassement de la variable VOLP. Chaque point correspond à une averse principale, observée pour la saison « hiver » sur les données de postes situés sur le pourtour méditerranéen français. On observe une tendance entre les deux variables, puisque le nuage de points n'est pas totalement aléatoire. Les faibles valeurs de RXP sont plus fréquemment associées aux fortes valeurs de VOLP et inversement. La régression linéaire ajustée sur ce nuage de points donne alors l'équation  $F(RXP) = 0,65 - 0,30 * F(VOLP)$ . Si les deux variables étaient totalement indépendantes, on aurait obtenu une équation de régression  $F(RXP) = 0,5$ .



**Figure 22.** Distributions de fréquences des pluies maximales en 24 heures, observées et issues des hétérogrammes simulés selon les deux modélisations : postes présentant différentes valeurs de  $\mu\text{NAVP}$ .



**Figure 23.** Relation entre la fréquence au non-dépassement de la variable RXP et la fréquence au non-dépassement de la variable VOLP pour les postes situés sur le pourtour méditerranéen français.

La variable RX tend à diminuer lorsque le volume de l'averse augmente, mais ce n'est pas vrai dans tous les cas. Les mêmes graphiques réalisés à partir des valeurs de la saison « été » ou pour des postes de climats différents, ne présentent pas toujours le même degré de liaison. Cette liaison étant impossible à caractériser poste par poste, on essaie alors de la déterminer en fonction des deux saisons et en fonction du type de climat.

On regroupe les observations de postes appartenant au même type de climat (en fonction de leur localisation) et d'une même saison, afin d'en tirer une tendance que l'on cherchera à modéliser. L'ajustement d'une droite de régression simple sur le graphique présentant les fréquences des deux variables, donne les résultats présentés dans le tableau 3.

	Hiver	Été
Tous les postes	$F(RXP) = 0,63 - 0,26 * F(VOLP)$	$F(RXP) = 0,52 - 0,04 * F(VOLP)$
Postes tempérés + alpins	$F(RXP) = 0,57 - 0,14 * F(VOLP)$	$F(RXP) = 0,49 + 0,02 * F(VOLP)$
Postes méditerranéens	$F(RXP) = 0,65 - 0,30 * F(VOLP)$	$F(RXP) = 0,53 - 0,06 * F(VOLP)$
Postes tropicaux	$F(RXP) = 0,64 - 0,28 * F(VOLP)$	$F(RXP) = 0,65 - 0,30 * F(VOLP)$

**Tableau 3.** Régression linéaire entre les fréquences des variables RXP et VOLP.

On observe que pour la saison « été », il y a indépendance entre les deux variables. La droite de régression est pratiquement horizontale, de valeur constante proche de 0,5. Pour les postes situés en climat tropical, la saison « été » est cependant marquée par une dépendance des deux variables.

Pour la saison « hiver », une liaison inverse est notée pour les postes méditerranéens et tropicaux. La liaison entre les deux variables semble d'autant plus marquée pour les postes associés à des variables  $\mu NAVP$  fortes.

Pour étudier la tendance existant entre les valeurs de RXP et VOLP, un ajustement linéaire ne semble pas adapté. En effet, si on trace la courbe médiane sur les graphiques présentant les fréquences des valeurs, on observe une accentuation de la liaison vers les valeurs extrêmes (figure 24).

En effet, si on calcule la valeur médiane de F(RXP) pour les valeurs fortes de F(VOLP), on trouve les valeurs présentées dans le tableau 4. Par exemple, pour les postes méditerranéens et pour la saison « hiver », la valeur médiane de F(RXP) associée aux plus fortes valeurs de VOLP (pour F(VOLP) > 0,98) est plus proche de 0,2 que de 0,35 comme le donnerait la droite de régression linéaire.

F (VOLP)	Saison : hiver			Saison : été		
	Tempéré/ alpin	Méditerranéen	Tropical	Tempéré + alpin	Méditerranéen	Tropical
> 0,99	0,38	0,19	0,16	0,64	0,32	0,16
> 0,98	0,47	0,19	0,14	0,57	0,35	0,26
> 0,95	0,5	0,23	0,18	0,56	0,39	0,31
> 0,90	0,47	0,26	0,22	0,55	0,4	0,28

**Tableau 4.** Valeur médiane de F(RXP) pour des valeurs de F(VOLP) fortes.

On choisit donc de modéliser la courbe médiane de la valeur de F(RXP) en fonction des valeurs de F(VOLP) afin de respecter les observations, lors de la modélisation.

La courbe médiane de la valeur de F(RXP) sera alors modélisée par l'expression :

$$\text{médiane de F(RXP)} = 0,5 + a \times (0,5 - \text{Borne}) \times (F(\text{VOLP}) - 0,5)^2$$

Avec  $a = 4$  si  $F(\text{VOLP}) < 0,5$  et  $a = -4$  si  $F(\text{VOLP}) > 0,5$

et où la valeur de Borne correspond à la valeur de F(RXP) lorsque F(VOLP) tend vers 1.

Le tableau 4 donne par exemple une valeur de Borne = 0,19 pour la saison « hiver » des postes méditerranéens. Le graphique de la figure 24 montre l'allure de la courbe médiane observée et théorique (avec Borne = 0,19), que l'on trouve pour la saison « hiver » pour l'ensemble des postes méditerranéens.

On voit que la relation entre les variables RXP et VOLP peut être paramétrée en faisant varier la valeur de « Borne ». Cependant cette valeur ne peut être déterminée localement car elle serait calculée à partir d'un faible nombre de valeurs.

On décide donc de se donner une règle de fluctuation de ce paramètre, qui caractérise la tendance que l'on vient d'observer. La dépendance entre RXP et VOLP semble d'autant plus importante que l'on se trouve en climat pluvieux, caractérisé par des pluies de longue durée fortes. On définit donc le paramètre « Borne » en fonction de la valeur moyenne de NAVP par les relations empiriques suivantes :

$$\begin{aligned} \text{Borne} &= 0,15 + 0,4 \times e^{-5 \times (\mu_{\text{NAVP}} - 1)} && \text{en saison HIVER} \\ \text{et } \text{Borne} &= 0,20 + 0,4 \times e^{-10 \times (\mu_{\text{NAVP}} - 1)} && \text{en saison ÉTÉ} \end{aligned}$$



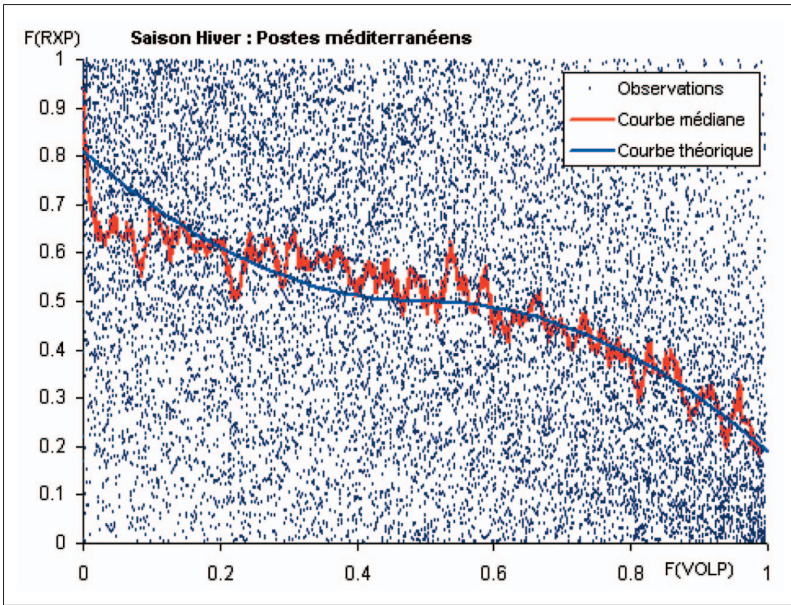


Figure 24. Relation  $F(RXP)$  en fonction de  $F(VOLP)$  pour la saison « hiver » et pour les postes méditerranéens : valeurs observées, courbe médiane observée et courbe médiane modélisée.

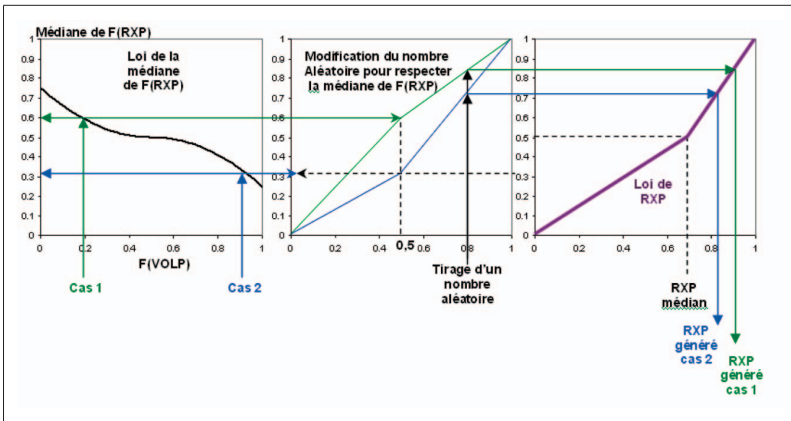


Figure 25. Génération de la variable RXP en tenant compte de sa relation avec la variable « volume de l'averse ».

Pour générer la variable RXP on procède comme le montre le schéma de la figure 25. On calcule d'abord la valeur médiane de sa fréquence en fonction de la fréquence du volume de l'averse principale générée. Puis on génère de façon aléatoire  $F(RXP)$  telle que l'on respecte la valeur médiane calculée. Une fois que la fréquence de la variable RXP est générée, on calcule la valeur de la variable RXP dont la loi théorique est une loi uniforme paramétrée par la valeur médiane de la variable.

On a vérifié que cette procédure permettait de respecter la valeur médiane de la variable RXP observée. Suivant la valeur de  $F(VOLP)$  et l'allure de la courbe

représentant la médiane de  $F(RXP)$ , on peut générer des valeurs de RXP respectant la tendance observée entre les variables RX et VOLUME. Pour simplifier les calculs on a choisi des lois uniformes par morceaux, qui semblent être bien adaptées dans la plupart des cas, lorsque l'on regarde les distributions, poste à poste.

La modélisation de la dépendance entre les variables VOLP et RXP conduit à une nette amélioration de la restitution des quantiles de pluies en une heure. En revanche, elle n'a aucune incidence sur la restitution des quantiles de pluies de plus longue durée. En effet, le volume des averses n'a pas été modifié, c'est juste la répartition de la pluie pendant l'averse qui a été modifiée.

### **Modélisation de la dépendance entre les variables VOLP et RXP**

On observe une dépendance entre les variables RX et VOL, pour les averses principales. Cette dépendance tend à associer préférentiellement des valeurs de RX faibles à des valeurs fortes de VOL. Cette tendance est plus marquée pour la saison « hiver » et pour les postes caractérisés par de fortes pluies.

On choisit alors de contraindre la génération de la variable RXP avec la valeur de la variable VOLP générée. Cette contrainte est alors paramétrée par le biais de la variable  $\mu_{NAVP}$ , représentative des événements longs et forts pour lesquels la contrainte devient plus forte.

La prise en compte de cette relation dans la modélisation, conduit à une nette amélioration dans la restitution des quantiles de pluies maximales en une heure. Le modèle a alors moins tendance à les surestimer sur les postes les plus pluvieux. Cependant les pluies maximales en une heure restent des variables plus facilement soumises aux problèmes d'échantillonnage. Ce point entraîne des écarts qui peuvent être plus importants entre les quantiles issus de la simulation et les quantiles issus d'un ajustement statistique, que pour les pluies maximales de durées plus longues.

## **Application**

### **Les données de pluies horaires**

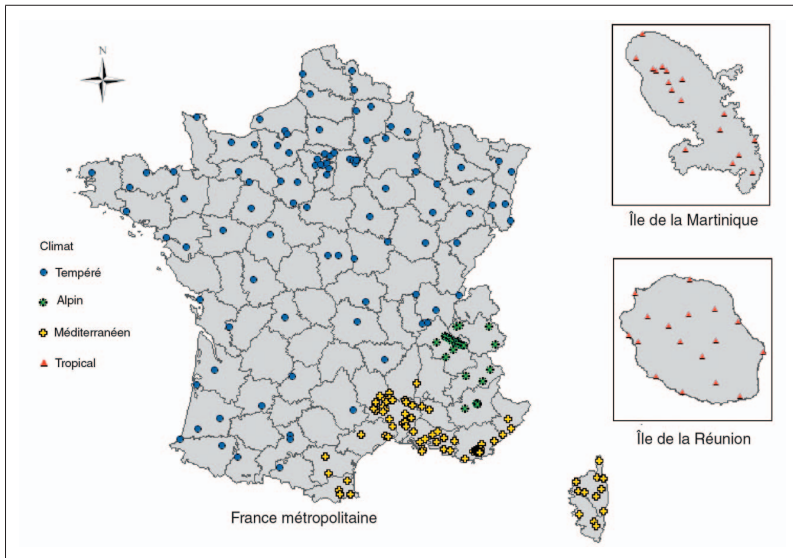
Le développement initial du générateur de pluie a été réalisé sur les données d'une cinquantaine de postes pluviographiques situés sur le pourtour méditerranéen français. Dans le cadre de l'extension de l'approche sur l'ensemble du territoire français, on a appliqué le modèle sur des postes de l'ensemble de la France métropolitaine, ainsi que sur des postes situés sur les îles de la Réunion, de la Martinique et de la Guadeloupe. Au total, on a testé le générateur de pluies horaires sur 251 postes pluviographiques, localisés sur les cartes de la figure 26. De façon indicative, nous avons regroupé ces postes selon leur climatologie. Quatre types de climats liés à la zone géographique sont définis :

- le climat « alpin » : 29 postes sont situés dans le massif alpin, à une altitude moyenne de 950 mètres avec 75 % des postes à plus de 500 mètres ;
- le climat « méditerranéen » : 90 postes sont situés dans les départements des régions du Sud-Est de la France, à proximité de la mer Méditerranée ;

- le climat « tempéré » : 98 postes sont situés sur le reste du territoire français, à une altitude moyenne de 182 m, avec 95 % des postes à moins de 500 mètres ;
- le climat « tropical » : 17 postes sont situés sur l'île de la Réunion et 17 postes sont situés sur l'île de la Martinique.

173 postes étudiés sont des postes pluviographiques de référence de Météo-France (réseau synoptique). Les autres postes sont des postes présentant des chroniques d'observation importantes et des données validées par les services gestionnaires (essentiellement Cemagref, DDE, Diren). Au total, la durée médiane d'observation est de 17,8 ans, avec des périodes d'observation allant de quelques années (durées courtes pour quelques postes alpins) à 78 ans (série pluviographique de Marseille).

Les postes pluviographiques ont été choisis sur des critères de qualité et de durée des observations. Ils ont aussi été choisis pour augmenter la diversité climatique que l'on veut étudier et pour augmenter la couverture spatiale du territoire étudié. On pourra ainsi tester les performances du modèle sur plusieurs types de climat.



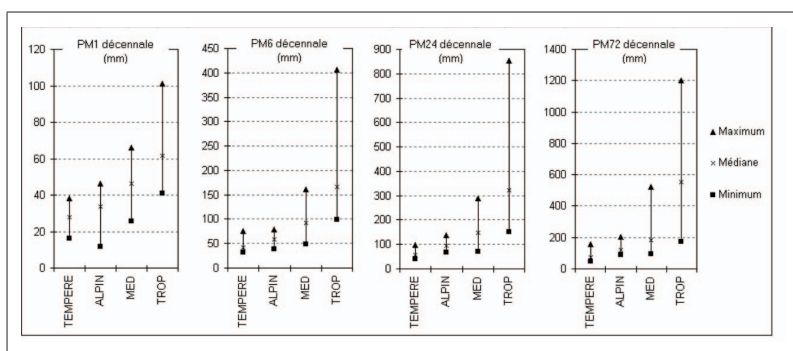
**Figure 26.** Localisation des postes pluviographiques étudiés.

Le tableau 5 et les graphiques de la figure 27 montrent la forte variabilité de la pluviométrie des postes étudiés. Cette variabilité est différente en fonction de la durée de la pluie.

Le rapport moyen entre les pluies de longue durée et les pluies de courte durée est plus faible pour les postes situés en climat « tempéré ». Les postes situés en climat « méditerranéen » et surtout en climat « tropical », sont caractérisés par des pluies de 6 à 72 heures très fortes par rapport aux pluies en une heure. On pressent déjà un phénomène de « persistance » des intensités des pluies fortes sur plusieurs heures, qui devra être pris en compte dans la modélisation.

Rapport moyen des quantiles décennaux	PM6/PM1	PM24/PM1	PM72/PM1
Type tempéré	1,60	2,20	2,90
Type alpin	1,90	3,20	4,50
Type méditerranéen	2,10	3,40	4,60
Type tropical	3,00	6,30	8,50

**Tableau 5.** Quantiles des pluies décennales en 6, 24 et 72 heures normés par la pluie décennale en une heure.



**Figure 27.** Variabilité de la pluviométrie des postes étudiés : valeurs minimales, médianes et maximales des pluies décennales en 1, 6, 24 et 72 heures, en fonction de la typologie des climats.

On remarque aussi la grande variabilité de la pluviométrie qui peut exister sous une même typologie de climat. Par exemple, en climat « tropical », sur uniquement l’île de la Réunion, certains postes sont soumis à une pluviométrie faible (proche des pluies du climat « tempéré ») et d’autres postes sont soumis à une pluviométrie extrême. En effet, c’est sur l’île de la Réunion que l’on a enregistré les records mondiaux de précipitations pour des durées comprises entre 12 heures et 15 jours (Dingman, 1994 ; Soler, 1997).

L’échantillon des données utilisées pour cette étude montre une extrême diversité de la pluviométrie. On pourra ainsi explorer les performances du générateur de pluies horaires dans un contexte pluviométrique extrêmement varié.

## Restitution des quantiles observés

Les lois de probabilité définissant les différentes variables ont été calées sur les différentes chroniques de pluies. La variabilité de certains paramètres peut être importante suivant les climats étudiés. En revanche, d’autres paramètres sont peu variables et traduisent une certaine similitude entre les pluies de différents climats. Par exemple, la durée des averses est relativement constante et dépend plus de la précision des dépouillements que d’une réalité climatique. Toutefois, le nombre de ces averses peut être très variable d’un climat à l’autre. C’est ce paramètre qui conduira à la simulation d’événements plus

ou moins longs, en ajoutant plus ou moins d'averses, indépendamment de leur durée.

La variabilité des paramètres sera étudiée plus loin, dans le paragraphe 3, lors de la régionalisation du modèle. Les résultats présentés ici sont issus du générateur calé localement sur les postes présentant des mesures de pluies à pas de temps fins.

La figure 28 présente les quantiles biennaux et décennaux des pluies maximales en 1, 6, 24 et 72 heures des 251 postes pluviographiques étudiés. On compare les quantiles issus de l'ajustement d'une loi exponentielle sur les échantillons observés, avec les quantiles issus de la simulation des événements pluvieux. Pour l'estimation des quantiles courants (période de retour deux ans) les valeurs obtenues par l'ajustement d'une loi de probabilité restent relativement indépendantes du choix de la loi. Les quantiles obtenus sont alors représentatifs de l'observation. Pour les quantiles plus rares (à partir de la période de retour 10 ans), le choix de la loi théorique ne conduit pas nécessairement à des valeurs représentatives des observations. On peut alors simplement comparer les résultats de SHYPRE avec ce qu'aurait donné l'ajustement d'une loi exponentielle.

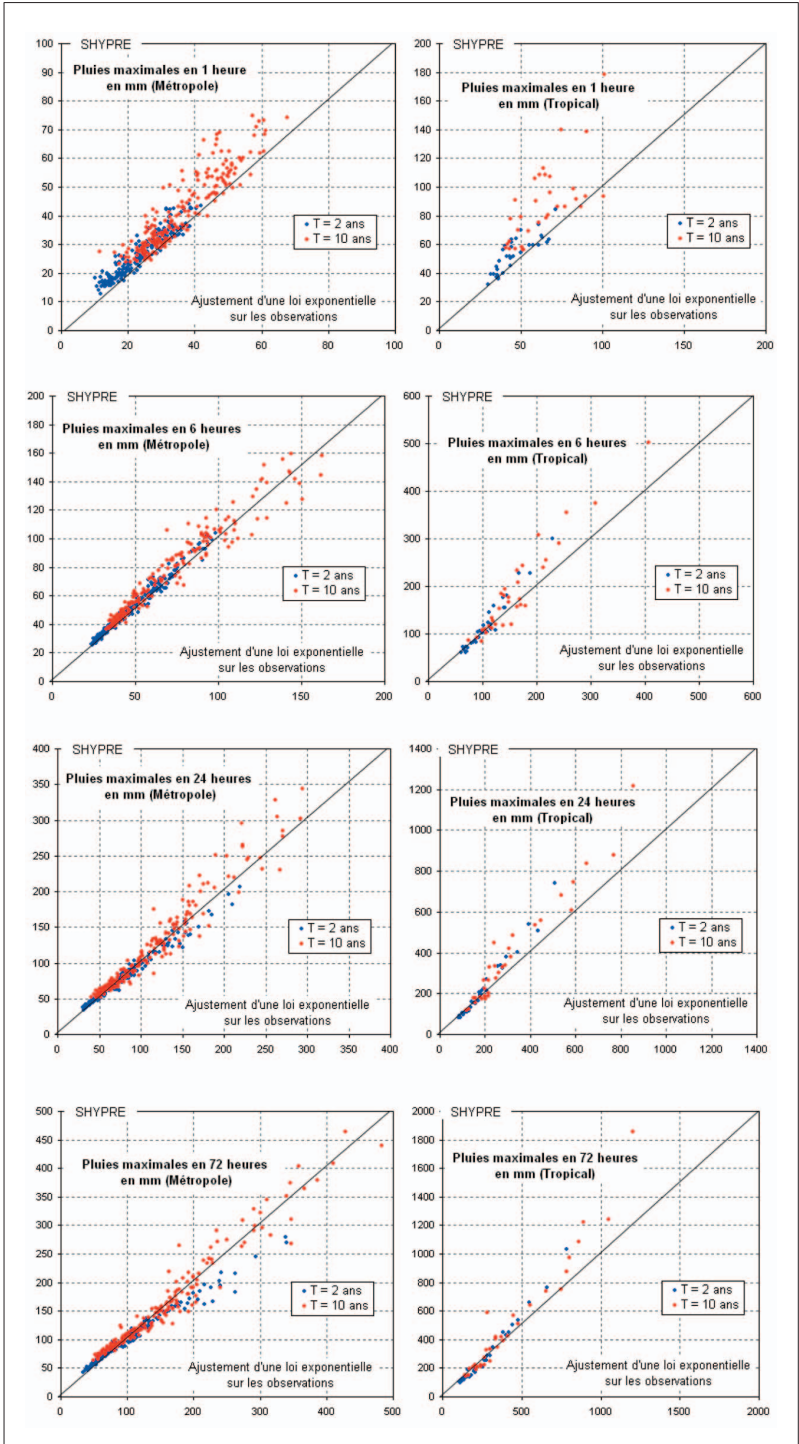
## Conclusion

Les résultats présentés montrent que le générateur de pluie permet d'obtenir des chroniques de pluies horaires statistiquement équivalentes aux pluies observées, et cela sous différents types de climat. Ces performances résultent d'une bonne appréhension des phénomènes observés et de leur modélisation appropriée. On dispose alors d'un générateur performant testé sur une large gamme de pluviométrie.

*Remarque. Pour les pluies maximales en une heure, les quantiles obtenus par SHYPRE semblent en moyenne plus forts que ceux issus de l'ajustement de la loi exponentielle. Ce point peut s'expliquer par la difficulté à modéliser la dépendance entre les variables RX et VOL (voir §2.3.4). On peut aussi noter que près de la moitié des séries pluviographiques dont on dispose sont des chroniques de pluies horaires à pas de temps fixe, qui sous-estiment les pluies maximales en une heure (origine mobile).*

## Robustesse

L'approche par simulation des pluies présente l'intérêt d'être plus robuste face à l'échantillonnage des valeurs extrêmes qu'un simple calage d'une loi de probabilité sur une caractéristique unique. Pour illustrer ce propos, on a étudié les données observées sur la station de Nîmes, sur deux périodes : 1972-1987 et 1972-1990. La seconde période présente une valeur extrême (événement d'octobre 1988). Le modèle de simulation de pluie a été calé sur ces deux périodes. Deux lois de probabilité ont aussi été ajustées sur les deux échantillons : la loi de Gumbel (deux paramètres) à caractère strictement exponentiel et la loi de Jenkinson (trois paramètres). Sur la figure 29, sont reportés les résultats de ces ajustements et ceux obtenus par le générateur de pluie.



**Figure 28.** Comparaison des quantiles décennaux des pluies maximales en 1, 6, 24 et 72 heures sur les 251 postes pluviographiques étudiés : loi exponentielle et simulation SHYPRE.

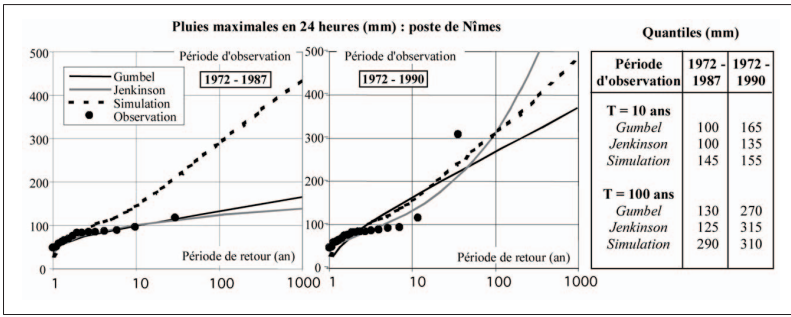


Figure 29. Influence de la présence d'une valeur extrême dans l'échantillon.

On remarque que :

- l'ajustement des lois statistiques de probabilité est extrêmement sensible à la présence d'un événement extrême. Un seul événement modifie considérablement les tracés des distributions de fréquence, et par là même de l'estimation des quantiles. La loi de Jenkinson passe ainsi d'un comportement « supra-exponentiel » à « super-exponentiel » ! Bien que particulier<sup>2</sup>, cet exemple montre bien le danger d'utiliser des lois à trois paramètres sur de courtes durées d'observations, car elles font intervenir des moments d'ordre 3, très sensibles à l'échantillonnage ;
- la distribution de fréquences empiriques des pluies maximales issues de la modélisation des hyétoigrammes horaires est peu sensible à la présence de la valeur extrême d'octobre 1988 : la valeur de la pluie maximale en 24 heures de fréquence centennale varie de 290 mm à 310 mm. Le rapport entre les deux estimations de ce quantile est de 2 avec la loi de Gumbel et de 2,5 avec la loi de Jenkinson.

Cette stabilité des résultats du modèle de génération de hyétoigrammes résulte de deux points :

- d'une part, de la forte taille des échantillons servant à l'ajustement des variables du modèle : on utilise jusqu'à plus de 1 000 observations pour caler les lois des variables du modèle ;
- d'autre part, du choix de lois de probabilité peu sensibles à l'échantillonnage : par exemple la loi des intensités moyennes des averses a été étudiée régionalement par la détermination d'une loi adimensionnelle, par agglomération des valeurs de différents postes réduites par leur moyenne locale, puis recalage local sur les valeurs inférieures à 4 fois la moyenne de la variable, pour ne pas tenir compte des valeurs particulières dues au découpage des averses : Arnaud *et al.*, 1998. Par ailleurs, la plupart des lois utilisées sont paramétrées par la moyenne de la variable (grandeur réputée bien plus stable que l'écart-type).

La surestimation apparente affichée par le modèle calé sur la période 1972-1987 est confortée par l'observation de l'événement d'octobre 1988. Le modèle n'a bien sûr rien de prédictif, mais avait décelé dans la structure interne

2. Cet exemple est particulier parce que l'événement extrême arrive en fin de période, et que l'effectif de l'échantillon reste faible (on rajoute trois années d'observation). On sait qu'il y a une forte incertitude sur l'attribution des fréquences empiriques. Cependant, ce cas de figure est malheureusement assez fréquent.

des épisodes pluvieux la possibilité de génération d'un tel événement. On voit là un des avantages de la méthode de simulation qui permet de recueillir un maximum d'informations à partir d'un minimum d'années d'observation. En effet, la méthode s'attache à décrire au mieux le phénomène observé et à le reproduire. L'analyse des événements extrêmes a permis la caractérisation de ceux-ci par la mise en évidence d'un phénomène de persistance des averses les plus fortes au cours d'un événement. Ce phénomène a été étudié et modélisé dans une approche régionale, permettant une meilleure observation et paramétrisation des phénomènes extrêmes. Ne s'attachant pas à s'ajuster scrupuleusement sur les courbes Intensité-Durée-Fréquence observées mais plutôt à respecter les caractéristiques temporelles du signal de pluie, la méthode n'est donc pas sujette à la fluctuation des distributions empiriques des pluies maximales, avec la présence ou non de valeurs extrêmes dans l'échantillon des valeurs.

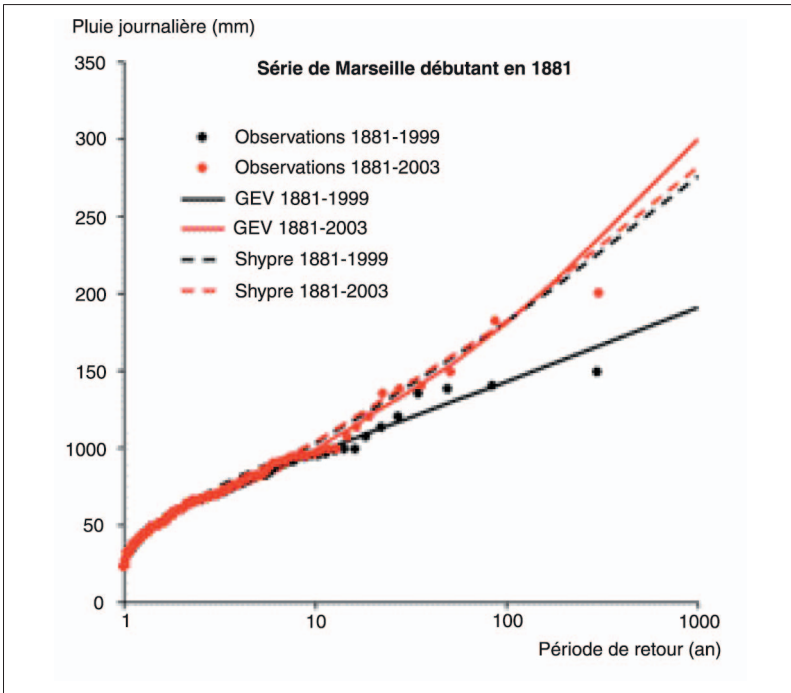
L'exemple présenté sur la figure 30 est une illustration supplémentaire de la robustesse de la méthode et des faiblesses des méthodes classiques d'ajustement. On compare les résultats de l'approche SHYPRE avec l'ajustement d'une loi de probabilité sur l'échantillon des pluies journalières maximales annuelles de la série de Marseille (débutant en 1881). Des tests statistiques ont montré que la série était homogène (Muller, 2006). En se plaçant quelques années en arrière, les résultats obtenus sur la longue série avec 119 années d'observation entre 1881 et 1999 auraient conclu que l'utilisation d'une loi proche de la loi exponentielle semblerait convenir pour extrapoler les distributions de fréquences de pluies journalières vers les valeurs extrêmes. Cependant, comme pour beaucoup d'autres exemples, la présence de quelques années supplémentaires peut conduire à des conclusions très différentes, en termes de lois à utiliser pour l'extrapolation, de paramètres de ces lois et de valeurs des quantiles rares et extrêmes. Concernant les résultats proposés par la méthode par simulation, on constate qu'ils ne sont pas sensibles à la présence de valeurs extrêmes dans l'échantillon des observations. Les résultats obtenus sur la première période d'observation laissent à croire que l'information était déjà disponible pour proposer des valeurs centennales plus « réalistes ». La robustesse de l'extrapolation vient de l'approche régionale faite dans la modélisation des événements, en plus d'une paramétrisation basée sur des valeurs moyennes.

On voit sur ce nouvel exemple que l'observation de longues chroniques de pluies ne fait que décaler les incertitudes liées à l'extrapolation des distributions de fréquence des pluies, et en particulier dans le choix de paramètre de forme que l'on devrait choisir. En effet, si sur la série de Marseille la loi exponentielle pouvait être retenue après plus de 100 ans d'observation, quatre années supplémentaires ont conduit à choisir une autre loi.

### **Robustesse de l'approche**

On voit bien ici l'extrême dépendance des statistiques classiques à la présence de valeurs extrêmes dans l'échantillon des valeurs observées. En particulier lors de l'apparition des valeurs extrêmes après une longue période sans observation de valeurs fortes. À l'inverse, l'approche par simulation, basée sur l'analyse de la structure interne des pluies, et donc sur une information plus abondante, est beaucoup plus robuste.





**Figure 30.** Estimation des quantiles des pluies journalières du poste de Marseille : comparaison des résultats de SHYPRE et de l'ajustement d'une loi GEV sur deux périodes d'observation.

## Formes de hyétoigrammes

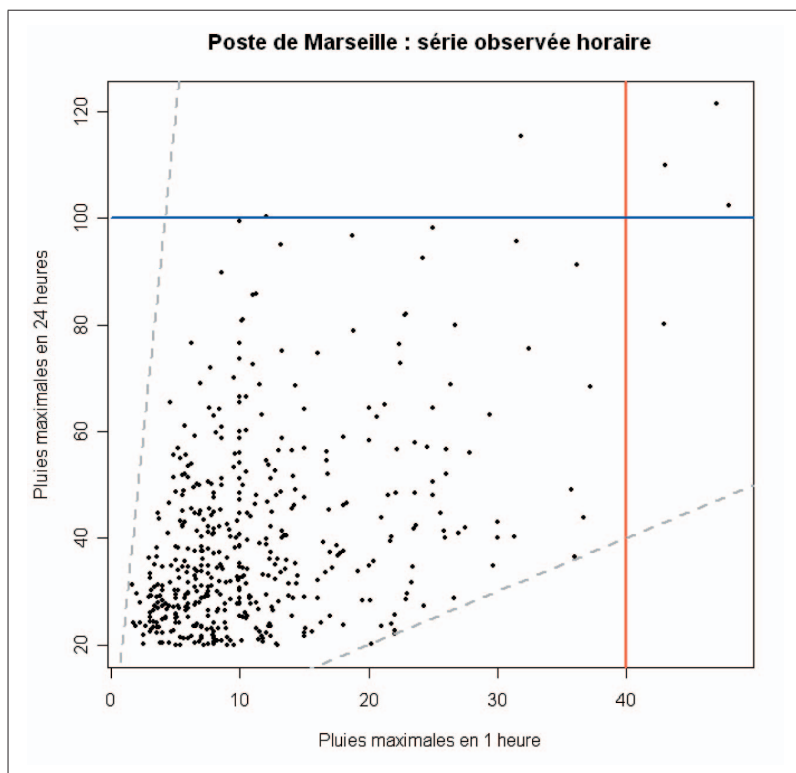
La restitution des statistiques des pluies observées sur différentes durées nous garantit que le générateur reproduit la structure temporelle des hyétoigrammes. Par simulation, on obtient alors des scénarios de pluies (chroniques de pluies horaires) respectant les distributions statistiques des pluies observées. On peut alors proposer différentes formes de hyétoigrammes associés à une même caractéristique de pluie. On s'affranchit ainsi de la notion de pluie de projet en disposant de toute une collection d'événements pluvieux, qui une fois transformés en hydrogrammes, nous permettront de tester le comportement hydraulique des ouvrages soumis au risque de crues.

La notion de pluie de projet est en effet difficile à juger. Par exemple, quelles sont les caractéristiques qui font la pluie de projet décennale ? Est-ce un hyétoigramme caractérisé par une pluie horaire décennale, ou une pluie en 24 heures décennales, ou les deux ?

Pour illustrer la variabilité des caractéristiques associées à un hyétoigramme, on a représenté sur la figure 31, les pluies maximales en 24 heures des 466 événements sélectionnés sur le poste pluviographique de Marseille, en fonction de la pluie maximale horaire du même événement.

Sur ce poste, on estime la pluie maximale en une heure décennale autour de 40 mm et la pluie décennale en 24 heures autour de 100 mm. On voit bien que pour une pluie horaire proche de la décennale (tracé vertical rouge) on

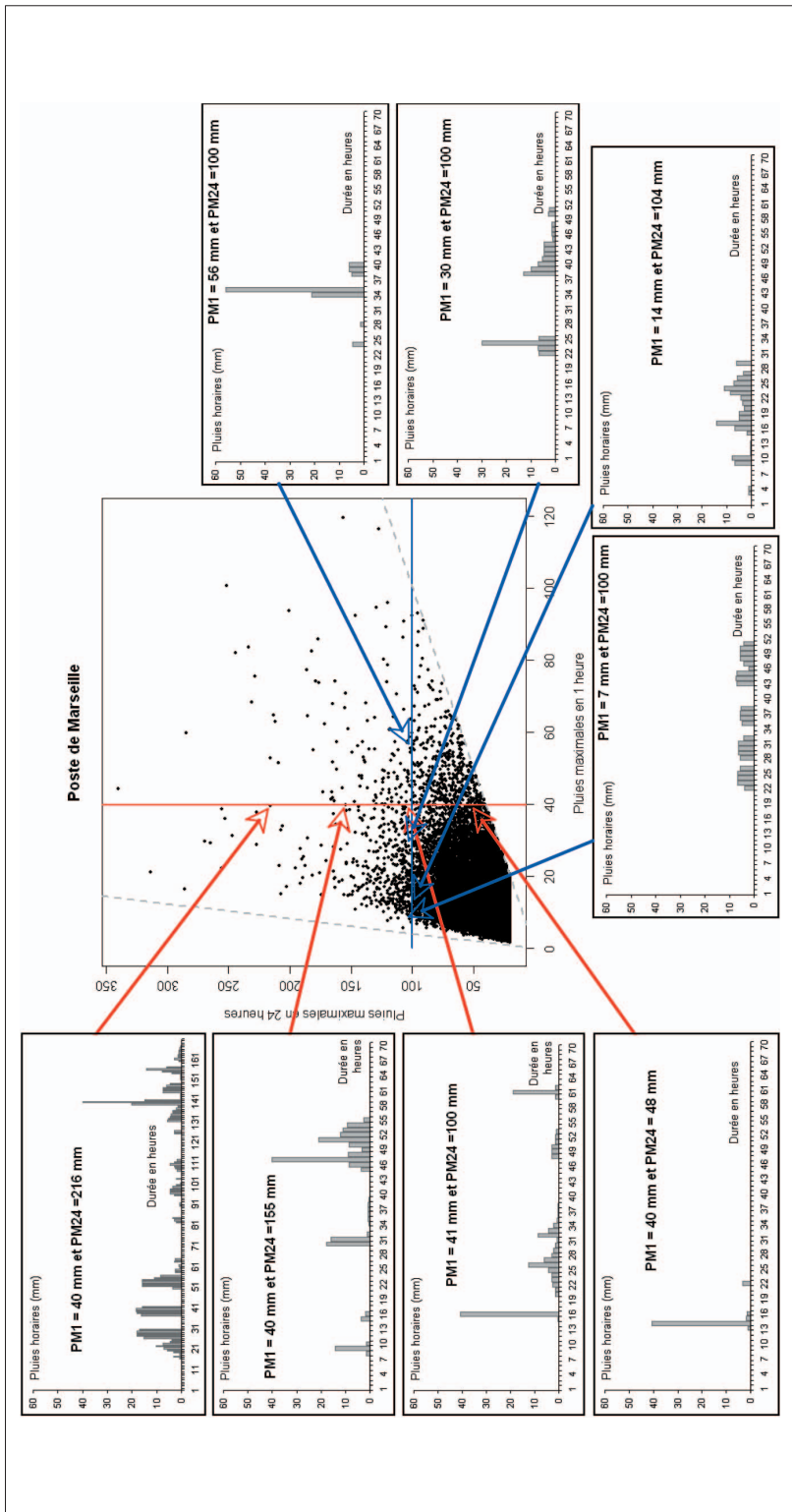
a des pluies en 24 heures qui peuvent fortement varier. De même, pour un événement décennal en 24 heures peut être associé à une pluie maximale horaire très variable. Les valeurs sont bien sûr bornées l'une par rapport à l'autre (traits en pointillés). La pluie maximale en 24 heures est comprise entre la pluie maximale horaire et 24 fois la pluie maximale horaire.



**Figure 31.** Présentation des pluies maximales en 24 heures des (466) événements pluvieux observés sur le poste pluviographique (horaire) de Marseille, en fonction de la pluie maximale en une heure du même événement.

On vérifie bien que les caractéristiques observées sont restituées sur les hyétogrammes simulés. Pour illustrer ce propos, on a simulé 100 échantillons de 50 ans soit l'équivalent de près de 30 000 événements pluvieux (en moyenne 6 événements par an) pour la série horaire de Marseille. On a alors tracé le même graphique que celui de la figure 31, mais pour les valeurs issues de hyétogrammes simulés (figure 32). On présente aussi différents hyétogrammes simulés, choisis au hasard, mais de façon qu'ils aient une pluie horaire proche de la valeur décennale ou une pluie en 24 heures proche de la décennale. Tous ces hyétogrammes ont alors au moins une caractéristique décennale, que ce soit la pluie en une heure ou la pluie en 24 heures.

On voit bien que le choix d'une seule forme de pluie associé à un quantile donné ne peut pas être objectif. Il existe une multitude de hyétogrammes ayant des caractéristiques décennales qui vont conduire à générer des crues pour lesquelles il est impossible d'attribuer directement une période de retour. Ce raisonnement doit aussi prendre en compte les autres durées caractéristiques de pluies.



**Figure 32.** Présentation des pluies maximales en 24 heures de près de 30 000 événements pluvieux simulés sur le poste pluviographique (horaire) de Marseille, en fonction de la pluie maximale en une heure du même événement (graphique du centre) et illustration de quelques hyétogrammes caractérisés par une caractéristique décennale.



La Nartuby en crue, © P. Cantet, 2010.



Vue générale à la sortie de Trans-en-Provence, © P. Cantet, 2010.

# Paramétrisation par l'information journalière

L'extension du modèle de pluies horaires sur l'ensemble du territoire français a d'abord nécessité son adaptation à un contexte climatique plus varié que celui sur lequel il a été développé initialement.

Pour cela une première étape a consisté à introduire des modifications dans la modélisation des pluies horaires afin de conduire à un modèle respectueux des caractéristiques des différents régimes pluviométriques observés sur le territoire français. Les modifications apportées conduisent à une structure unique du modèle, paramétré de façon pertinente afin de permettre la modélisation des pluies allant du climat « tempéré » au climat « tropical ».

Son utilisation en mode « local », c'est-à-dire à partir d'une paramétrisation obtenue de façon locale à partir de l'information des pluies à pas de temps horaire, montre de bons résultats (voir p. 56, « Application »), mais nécessite aussi une information « locale » des pluies horaires.

**Son utilisation sur l'ensemble du territoire français passe par en revanche par une régionalisation des paramètres du modèle (cartographie des paramètres).**

Cette régionalisation du modèle de pluie va se faire en deux phases :

- dans un premier temps on cherche à simplifier la paramétrisation du modèle, afin de faciliter sa régionalisation ;
- dans un deuxième temps on régionalisera les paramètres utilisés dans la paramétrisation simplifiée.

Pour simplifier la paramétrisation, les différents paramètres obtenus localement à partir de l'information des pluies horaires vont être soit fixés à leur valeur médiane, soit expliqués par le biais de variables issues de l'information des pluies journalières. Les paramètres fixés seront ceux qui ne présentent pas une variabilité significative sur la zone étudiée, ou dont la variabilité n'a pas une influence notable sur les résultats de la modélisation. Les paramètres importants, ou très variables, seront expliqués par le biais de régressions avec des variables journalières.

Ce choix a été fait au regard de la disponibilité des données. En effet, **l'information journalière étant nettement plus abondante que l'information horaire, la cartographie de caractéristiques journalières sera plus précise que celle de caractéristiques horaires.** Elle fournira une information régionale plus pertinente, à condition toutefois que les relations entre les variables horaires et les variables journalières soient elles aussi pertinentes et significatives.

Dans cette partie, seulement les données de la France métropolitaine sont utilisées, afin d'aboutir à une cartographie des pluies sur la France métropolitaine. Le même type d'étude a été réalisé sur les autres zones géographiques où l'approche SHYPRE a été régionalisée pour aboutir à une paramétrisation régionale propre à chaque région : Les DOM (La Réunion, La Martinique et la Guadeloupe), le Piémont italien et une région d'Espagne.

## Les paramètres fixés

On rappelle que les neuf variables servant à décrire la structure temporelle des événements pluvieux au pas de temps horaire sont les suivantes : NE (nombre d'événements par saison), NG (nombre de périodes pluvieuses par événement), DIA (durée sèche entre deux périodes pluvieuses), NA (nombre d'averses par période pluvieuse), NAVP (nombre d'averses principales par événement), DA (durée des averses ; on distinguera DAP et DAO suivant le type de l'averse), VOL (volume des averses ; on distinguera VOLP et VOLO suivant le type de l'averse), RX (rapport de la pluie horaire maximale de l'averse sur le volume de l'averse ; on distinguera RXP et RXO suivant le type de l'averse) et RPX (position de la pluie horaire maximale dans l'averse).

Ces neuf variables sont des variables aléatoires dont la distribution de fréquence est décrite par une loi de probabilité théorique. Ces lois de probabilité sont alors définies grâce à vingt paramètres, déterminés pour deux saisons : l'été et l'hiver (voir tableau 2).

Parmi ces vingt paramètres, certains présentent une faible variabilité d'un poste à l'autre, même pour des postes appartenant à des climats différents. De plus, même si cette variabilité existe, elle reste dans une gamme suffisamment faible par rapport à la sensibilité du modèle face à ces variations. Une étude de sensibilité a été menée sur une version antérieure du modèle (Arnaud, 1997). Cette étude a montré que la variabilité constatée de certaines variables n'influait pas les performances du modèle dans sa capacité à restituer les pluies maximales de différentes durées. Les paramètres liés à ces variables sont alors fixés à une valeur médiane quel que soit le poste étudié.

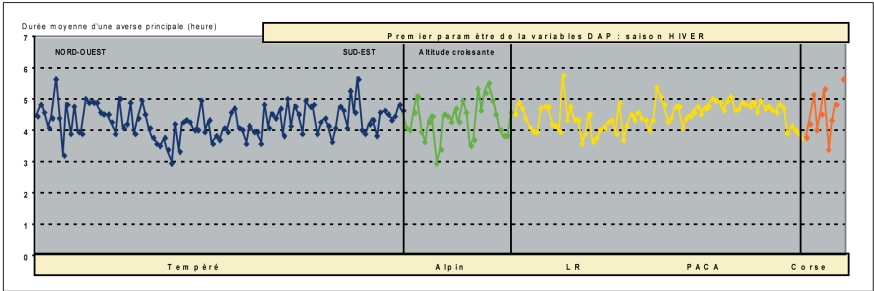
Par exemple, on montre sur la figure 33, la variation du premier paramètre de la variable DAP (durée des averses principales). Ce paramètre correspond à la durée moyenne en heures des averses principales. On voit qu'il existe une certaine variabilité de ce paramètre qui fluctue entre 3 et 5,8 heures.

Cependant, il ne semble pas exister d'explication climatique à cette variabilité puisque l'on retrouve les mêmes gammes de variabilité pour les différents climats présents. La régionalisation de ce paramètre est alors difficilement réalisable. Ce paramètre étant un paramètre moyen calculé sur un grand nombre de valeurs (en moyenne une centaine), cette fluctuation n'est pas seulement due à l'échantillonnage. Elle dépend, en effet, plus de la précision des enregistrements pluviographiques disponibles que d'une réalité climatique.

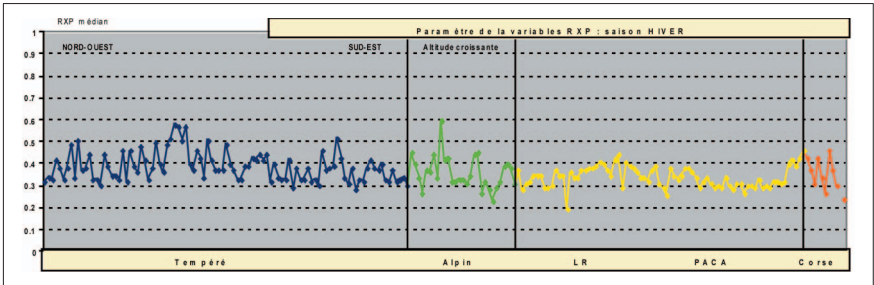
La version du modèle basée sur la description des averses par leur volume, et non plus par leur intensité, réduit considérablement l'influence de la variable « Durée ». C'est pourquoi, la sensibilité du modèle à cette variable étant suffisamment faible, on peut fixer ce paramètre à sa valeur médiane de 4,42 heures.

Si l'on regarde maintenant la variation du paramètre de la loi décrivant la variable RX des averses principales (figure 34), on remarque aussi une variabilité marquée mais difficilement explicable géographiquement. La régionalisation de cette variable est aussi assez difficile.

La variabilité de ce paramètre est suffisamment forte pour influencer les performances du modèle, et en particulier dans sa capacité à reproduire les pluies maximales horaires. Cependant, ce paramètre va être fixé à sa valeur médiane



**Figure 33.** Variation du premier paramètre décrivant la variable DAP sur l'ensemble des 217 postes pluviographiques étudiés.



**Figure 34.** Variation du paramètre décrivant la variable RXP sur l'ensemble des 217 postes étudiés.

de 0,35 pour la saison « hiver » et 0,50 pour la saison « été ». L'impact de cette simplification va cependant être réduit par le fait d'avoir fixé le paramètre lié à la variable DAP. En effet, il existe une liaison marquée entre les paramètres des variables RXP et DAP ( $r^2 = 0,64$  en « hiver » et  $r^2 = 0,53$  en « été »). Les postes présentant une durée moyenne des averses principales forte présentent en même temps une valeur médiane de la variable RX faible, et inversement.

On comprend alors que la variabilité de la variable RX est plus liée à une variabilité de la variable « Durée » qu'à une réalité climatique.

On voit ainsi qu'il est difficile de prendre en compte la variabilité de tous les paramètres, d'autant plus que certains peuvent être liés. Par exemple, on observe aussi une liaison entre les paramètres des variables DAO et NA ( $r^2 = 0,34$  en « hiver » et  $r^2 = 0,40$  en « été »). Dans ce cas, plus les averses « ordinaires » sont courtes, plus il y a d'averses. Cela peut se comprendre, sans y mettre une explication climatique, en fonction d'un problème lié à la précision du dépouillement des enregistrements. En effet, un dépouillement « grossier »<sup>3</sup> peut conduire à regrouper des averses qui auraient été séparées par un minimum relatif avec un dépouillement plus précis.

En partant des résultats d'études antérieures (Arnaud, 1997 ; Fine, 2003), on a donc choisi de fixer les paramètres présentés dans le tableau 6, à leur valeur

3. Les enregistrements « papier » des pluviographes correspondent parfois à des tracés réalisés sur un tambour se déroulant à la vitesse d'un centimètre en quatre heures. L'appréciation des variations d'intensité à l'échelle horaire est alors difficilement appréciable.

médiane obtenue sur l'ensemble des 217 postes pluviographiques étudiés sur la France métropolitaine.

Variable	Médiane saison HIVER	Médiane saison Été
NA : 1/moyenne	0,55	0,62
RXP : médiane	0,35	0,50
RXO : médiane	0,50	0,56
DAP : moyenne en heures	4,42	3,24
DAP : Prob (DAP < 10 heures)	0,97	0,96
DAO : moyenne en heures	1,82	1,56
DAO : Prob (DAO < 6 heures)	0,97	0,98
RPX : moyenne	0,57	0,58
RPX : écart type	0,21	0,21
DIA : 1/moyenne	0,30	0,29
DIA :	0,90	0,89
aVOLP	- 1,98	- 1,84
bVOLP	1,04	0,91
aVOLO	- 1,13	- 1,06
bVOLO	0,15	0,08

**Tableau 6.** Valeur des différents paramètres fixés : valeur médiane obtenue pour les 217 postes pluviographiques étudiés sur la France métropolitaine.

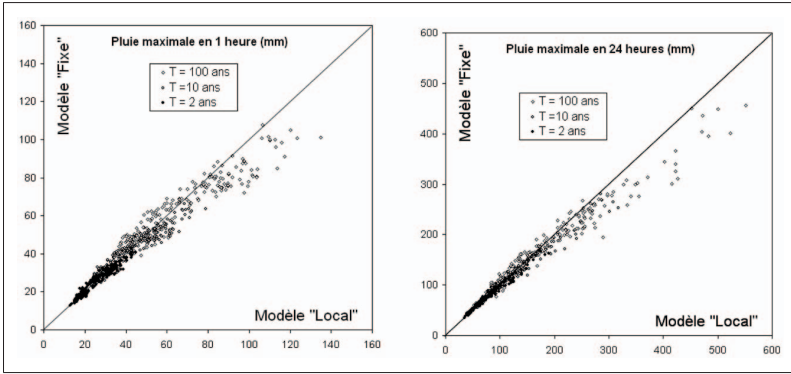
Les graphiques de la figure 35 montrent les performances du modèle de pluies sur les 217 postes métropolitains, en restitution des pluies maximales en 1 et 24 heures de période de retour 2 ans. On compare la restitution de ces quantiles suivant deux types de paramétrisation : la paramétrisation « locale », issue de l'information des pluies horaires de chaque poste, et une paramétrisation dite « fixe », pour laquelle on a fixé 15 paramètres sur les 20. On observe ainsi l'influence de la fixation de 15 des 20 paramètres du modèle.

On rappelle que les cinq autres paramètres sont maintenus à leurs valeurs « locales », c'est-à-dire estimées pour chaque poste à partir de l'information horaire.

On observe des écarts peu différents entre les résultats de la paramétrisation « locale » et ceux obtenus après la fixation de 15 paramètres sur 20.

Les pluies maximales en 1 ou 2 heures sont celles qui sont le plus touchées par le fait de fixer certains paramètres et en particulier le paramètre RX qui caractérise la quantité maximale de pluie d'une averse qui tombe en une heure. Les pluies de longue durée ne sont pas contre pratiquement pas affectées.





**Figure 35.** Influence de la fixation de 15 paramètres du modèle sur la restitution des pluies maximales en 1, 6 et 24 heures de périodes de retour 2 et 10 ans.

### Simplification de la paramétrisation

Les problèmes de dépouillement, d'échantillonnage et de liaison entre paramètres sont autant de points pouvant expliquer la variabilité de certains paramètres. On a donc choisi de fixer 15 paramètres, par saison, dont la variabilité géographique n'est pas nécessairement la conséquence d'une particularité climatique. Ces paramètres sont souvent peu variables, ou dans une gamme où le modèle est peu sensible.

Le fait de fixer certains paramètres n'entraîne pas de dégradations notables des performances du modèle. Ce résultat justifie les hypothèses présentées plus haut qui considéraient les paramètres fixés comme des paramètres peu variables ou peu influents.

## Paramètres « journalisés »

L'objectif de la « journalisation » du modèle est de le paramétrer à partir de l'information pluviométrique au pas de temps journalier.

Certains paramètres ont été fixés et d'autres sont déjà intrinsèquement définis par l'information journalière, comme les paramètres NE et  $\mu_{NAVP}$ .

Il reste donc les paramètres des variables NG (nombre de périodes pluvieuses) et VOL (volume des averses principales et ordinaires) à déterminer à partir de l'information des pluies journalières. Ces paramètres vont alors être « journalisés » par l'intermédiaire d'équations simples les reliant à des caractéristiques journalières des postes pluviométriques.

### DÉFINITION DES CARACTÉRISTIQUES JOURNALIÈRES

Les paramètres journaliers nécessaires à la paramétrisation régionale du modèle de pluie sont les valeurs moyennes des variables définies sur la figure 36.

Pour chaque événement pluvieux on calcule les valeurs des variables suivantes :

- PJMAX (mm) : pluie journalière maximale de l'événement (en millimètre) ;
- DTOT (jour) : durée totale de l'événement (en jour) ;
- NAVP : nombre d'averses principales.

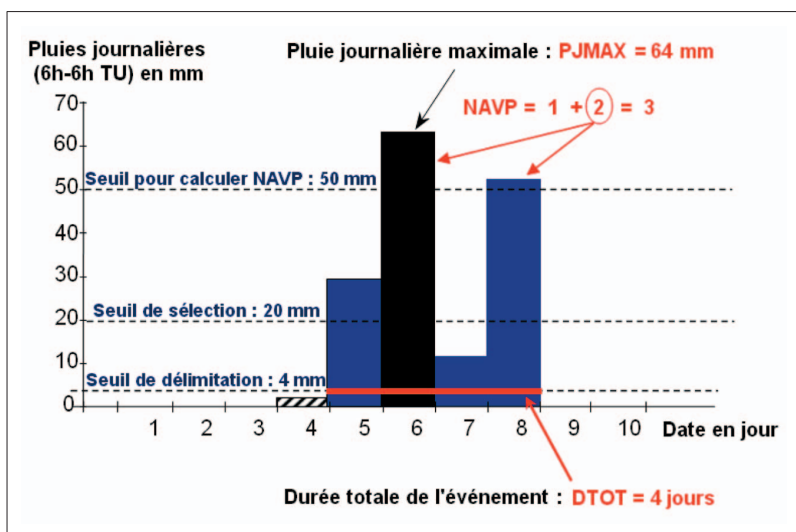


Figure 36. Définition des caractéristiques journalières.

Une fois déterminées pour chaque événement, on calcule la valeur moyenne de ces variables pour les deux saisons :

- $\mu$ PJMAX(été) et  $\mu$ PJMAX(hiver),
- $\mu$ DTOT(été) et  $\mu$ DTOT(hiver),
- $\mu$ NAVP(été) et  $\mu$ NAVP(hiver).

Ces valeurs moyennes correspondent aux paramètres journaliers déterminés pour chaque poste et chaque saison, qui serviront à établir des relations permettant d'estimer les paramètres du modèle de pluie horaire restant à déterminer.

## Le paramètre de la variable NG

Le nombre de périodes pluvieuses par épisode (variable NG) est corrélé avec la durée totale des événements. En effet, plus il y a de périodes pluvieuses dans un événement, plus celui-ci risque d'être long. La recherche d'une relation entre le paramètre NG et la variable journalière DTOT donne les résultats suivants :

Saison	Hiver	Été
Relation	$NG = -0,12 \times \mu DTOT + 0,52$	$y = -0,15 \times \mu DTOT + 0,57$
$r^2$	$r^2 = 0,64$	$r^2 = 0,70$

ou

Relation	$NG = -0,28 \times \ln(\mu DTOT) + 0,49$	$y = -0,32 \times \ln(\mu DTOT) + 0,49$
$r^2$	$r^2 = 0,65$	$r^2 = 0,71$

La corrélation n'est cependant pas optimale. En effet, la durée d'un événement dépend aussi du nombre d'averses par période pluvieuse, ainsi que de la durée de ces averses et de la durée inter-averses.

Afin d'améliorer la détermination du paramètre NG, on le combine à la valeur du paramètre de la variable NA. On approxime alors le nombre moyen

d'averses dans un épisode par le produit des valeurs moyennes des variables NA et NG. Les paramètres NA et NG étant déterminés par l'inverse de la moyenne de la variable, on relie la valeur de  $1/(NA \times NG)$  à la valeur de  $\mu DTOT$ . On trouve alors les relations suivantes :

Saison	Hiver	Été
Relation	$1/(NG \times NA) = 4,18 \times \mu DTOT - 1,79$	$y = 4,83 \times \mu DTOT - 3,46$
$r^2$	$r^2 = 0,83$	$r^2 = 0,85$

Les graphiques de la figure 37 illustrent la liaison entre la variable  $\mu DTOT$  et  $1/(NG \times NA)$ . On peut voir sur ces graphiques les différents postes ayant servi à l'établissement de ces relations : ronds bleus pour les postes classés en climat « tempéré », triangles verts les postes classés en climat « alpin », losanges jaunes les postes classés en climat « méditerranéen » et carrés rouges les postes « corses ».

Le produit  $1/(NA \times NG)$  étant correctement déterminé à partir de la variable  $\mu DTOT$ , on déterminera la valeur de NG par cette relation tout en prenant la valeur fixée pour NA.

### Paramétrisation du nombre d'averses

La procédure présentée ci-dessus permettra en plus de compenser les erreurs commises en fixant le paramètre NA. En effet, si l'on sous-estime NA en prenant une valeur fixe plutôt que sa valeur locale, on surestimera NG de façon à respecter le produit  $1/(NG \times NA)$ , c'est-à-dire le nombre total d'averses dans un événement pluvieux.

## Le paramètre de la variable VOL

La variable « volume » des averses est décrite par une loi de probabilité définie à partir de trois paramètres : la moyenne de la variable (mVOL) et les deux paramètres de la loi exponentielle (aVOL et bVOL) ajustée sur les valeurs réduites (par la moyenne).

Seul le paramètre mVOL est « journalisé ». La variable journalière permettant d'expliquer la variabilité de mVOL est la variable  $\mu PJMAX$ . On trouve, sur les graphiques de la figure 39 et la figure 38, les points ayant servi à établir la relation entre la variable journalière  $\mu PJMAX$  et le paramètre de la loi des volumes des averses.

Pour les averses ordinaires, on trouve les relations suivantes :

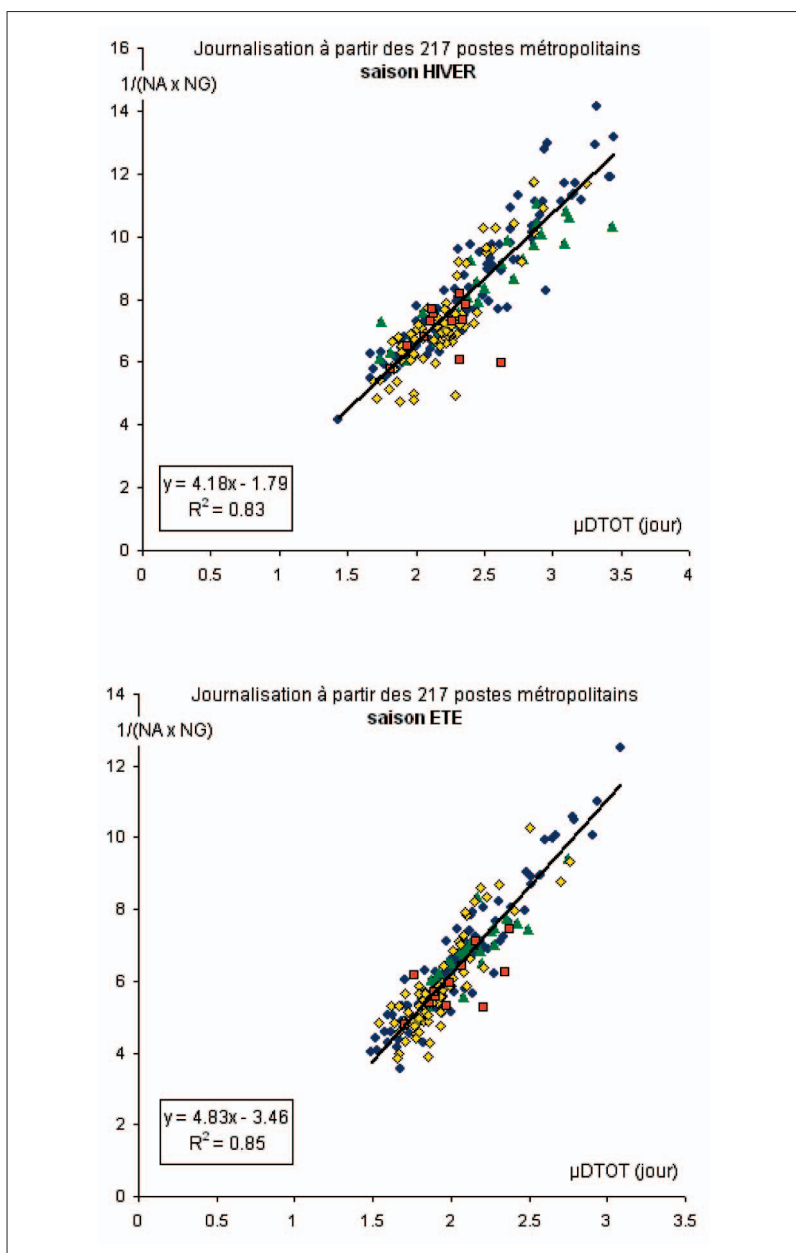
Saison	Hiver	Été
Relation	$mVOLO = 0,12 \times \mu PJMAX + 0,31$	$mVOLO = 0,10 \times \mu PJMAX + 1,09$
$r^2$	$r^2 = 0,79$	$r^2 = 0,76$

ou

Relation	$mVOLO = 4,33 \times \ln(\mu PJMAX) - 10,8$	$mVOLO = 3,93 \times \ln(\mu PJMAX) - 9,37$
$r^2$	$r^2 = 0,81$	$r^2 = 0,79$

Pour les averses principales, qui sont les plus influentes sur les résultats de la modélisation, on trouve les relations suivantes :

Saison	Hiver	Été
Relation	$mVOLP = 0,46 \times \mu PJMAX + 2,21$	$mVOLP = 0,54 \times \mu PJMAX + 2,27$
$r^2$	$r^2 = 0,84$	$r^2 = 0,89$

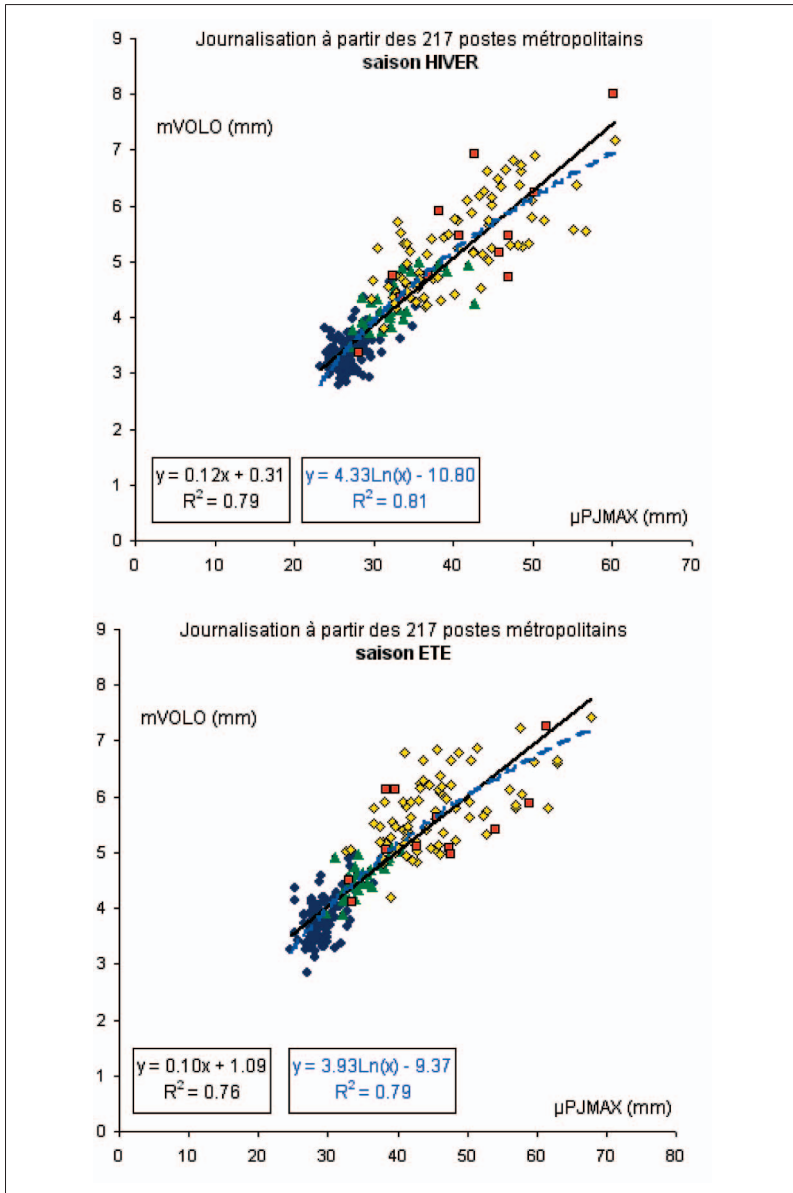


**Figure 37.** Relation entre la variable journalière  $\mu DTOT$  et le produit  $1/(NG \times NA)$ .

ou

$$\begin{array}{ll} \text{Relation} & mVOLP = 16,65 \times \ln(\mu PJMAX) - 40,41 & mVOLP = 21,5 \times \ln(\mu PJMAX) - 54,7 \\ r^2 & r^2 = 0,85 & r^2 = 0,90 \end{array}$$

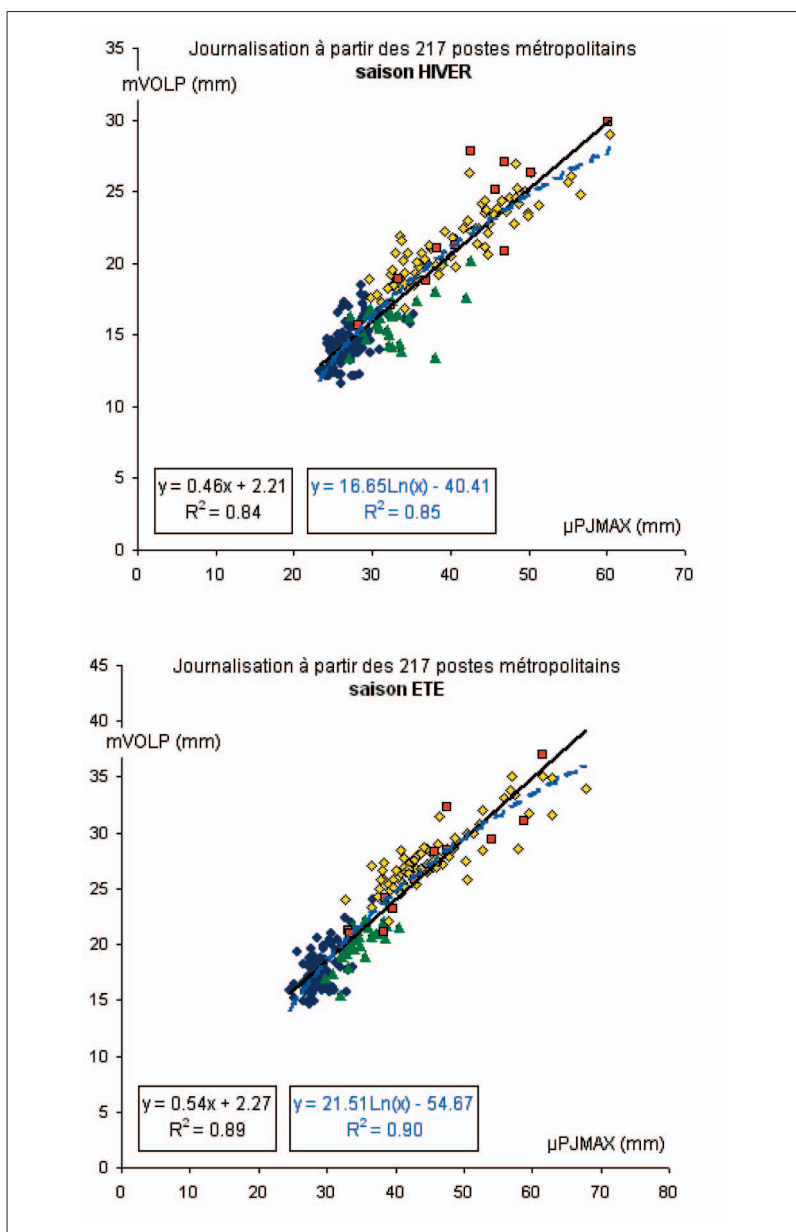
Les régressions sont largement significatives avec des coefficients de détermination compris entre 0,79 et 0,90 (relations de type logarithmique entre les variables).



**Figure 38.** Relation entre la variable  $\mu$ PJMAX et mVOLO pour les deux saisons.

Ces relations sont bonnes car les variables présentent une forte variation sur l'ensemble du territoire métropolitain. Si l'on avait établi une relation par grandes zones géographiques, elles auraient été moins marquées sur certaines zones où la pluviométrie est moins variable (zone du climat tempéré).

Cependant pour des raisons de continuité dans la régionalisation du modèle, il a été choisi de rester sur une relation unique entre les variables horaires et journalières pour l'ensemble du territoire métropolitain. Dans ce cas, on peut déjà remarquer que pour les postes situés en « climat alpin » (triangles verts



**Figure 39.** Relation entre la variable  $\mu\text{PJMAX}$  et mVOLP pour les deux saisons.

sur les graphiques de la figure 37 et de la figure 38) la relation établie avec la variable  $\mu\text{PJMAX}$  tend à légèrement surestimer la variable mVOLP.

Au final, pour expliquer la variable mVOL, on a choisi une relation du type logarithmique qui semble être la plus appropriée, bien que dans la gamme des paramètres observés cette relation soit proche d'une régression linéaire.

## Relations avec les caractéristiques journalières

Avec des coefficients de détermination ( $r^2$ ) compris entre 0,8 et 0,9, on arrive à expliquer correctement la variabilité des paramètres horaires à partir de l'information pluviométrique journalière.

Pour des raisons de continuité dans la régionalisation, il a été choisi d'établir des relations uniques pour l'ensemble du territoire métropolitain, entre les variables journalières et horaires.

## Corrélations supplémentaires

Les relations présentées précédemment suffisent pour définir une paramétrisation du modèle basée sur l'unique connaissance de l'information pluviométrique journalière. Cependant, on observe des corrélations supplémentaires pour certaines variables que l'on va prendre en compte afin d'améliorer les performances du modèle dans sa version « journalisée ».

- Corrélacion entre les paramètres aVOLP et  $\mu$ PJMAX :

Il existe une faible corrélation, pour la saison « été » entre le paramètre aVOLP et le paramètre journalier  $\mu$ PJMAX ( $r^2 = 0,41$ ). La relation qui existe entre ces deux paramètres sera alors prise en compte afin d'améliorer légèrement les performances du modèle « journalisé ».

- Corrélacion entre les paramètres aVOLP et bVOLP :

Il existe une forte corrélation entre les paramètres aVOLP et bVOLP ( $r^2 = 0,98$  en hiver et  $r^2 = 0,99$  en été), de même qu'entre aVOLO et bVOLO ( $r^2 = 0,97$  en hiver et  $r^2 = 0,95$  en été). Cette corrélation s'explique par le fait que les lois adimensionnelles de la variable « volume » passent par un même point pivot, conduisant à une augmentation de l'ordonnée à l'origine quand la pente de la droite représentant la loi diminue. Le fait de faire varier la variable aVOLP avec  $\mu$ PJMAX, nous oblige alors de ne plus considérer bVOLP comme constant, mais de le faire varier avec aVOLP suivant la relation déterminée.

- Corrélacion entre les paramètres  $\mu$ NAVp et  $\mu$ PJMAX :

On trouve une liaison forte entre les paramètres  $\mu$ NAVp et  $\mu$ PJMAX. Cette liaison est relativement logique. En effet, on s'attend à ce que plus la pluviométrie d'un poste soit forte, plus on ait de chances d'avoir plusieurs jours dépassant le seuil de 50 mm. Cependant, de fortes valeurs de  $\mu$ NAVp peuvent aussi être observées sur des postes présentant des événements longs (forts  $\mu$ DTOT) mais ne présentant pas forcément une pluviométrie très intense. C'est le cas des régions ayant des événements longs et relativement pluvieux, mais rarement extrêmes, comme les régions des Vosges, du Jura, des Alpes du Nord ou des Pyrénéens-Atlantiques.

Afin de laisser au paramètre  $\mu$ NAVp un caractère lié à la génération d'événements « persistants », on préférera le déterminer par les relations suivantes, prenant en compte  $\mu$ PJMAX.

Hiver

$$\mu\text{NAVp} = 0,015 \times \mu\text{PJMAX} + 0,61$$

Été

$$\mu\text{NAVp} = 0,011 \times \mu\text{PJMAX} + 0,71$$

## Paramétrisation journalière

Les paramètres qui restent variables d'un poste à l'autre sont déterminés à partir de trois variables journalières qu'il faudra régionaliser à partir de l'information de la pluie journalière. Chacune de ces variables journalières correspond en fait à des caractéristiques climatiques bien distinctes, permettant de définir la pluviométrie locale :

- NE : traduit l'occurrence des événements de chaque saison ;
- $\mu$ PJMAX : traduit l'intensité de la pluviométrie de chaque saison ;
- $\mu$ DTOT : traduit la durée des événements pluvieux de chaque saison.

La cartographie de ces trois paramètres journaliers permettra alors de paramétrer le modèle de génération de pluies horaires en tout point de la zone étudiée, à partir des relations présentées.

## Résultats en simulation

Afin de tester la pertinence des relations établies précédemment entre les paramètres horaires et les caractéristiques journalières, on a comparé les résultats des simulations effectuées sur les 217 postes pluviographiques disponibles, à partir des deux types de paramétrisation : paramétrisation « locale » et paramétrisation « journalisée » (figure 40).

Les résultats présentés proviennent de simulations effectuées à partir des relations permettant d'expliquer au mieux la variation des paramètres horaires. Les graphiques de la figure 40 montrent que la paramétrisation « journalisée » fournit des résultats très proches de ceux obtenus par la paramétrisation « locale ». Des écarts sont cependant plus marqués pour les pluies maximales horaires. Ce résultat déjà observé lorsque l'on avait seulement fixé certains paramètres, vient de la difficulté rencontrée pour expliquer la variabilité du paramètre RX.

On observe aussi une tendance à sous-estimer les plus fortes pluies en 24 heures et des plus fortes pluies totales.

Enfin, on remarque que les écarts produits par les deux paramétrisations sont équivalents quelle que soit la période de retour entre 2 et 100 ans.

**De façon globale, on peut déjà constater les bonnes performances de la « journalisation » des variables du modèle.**

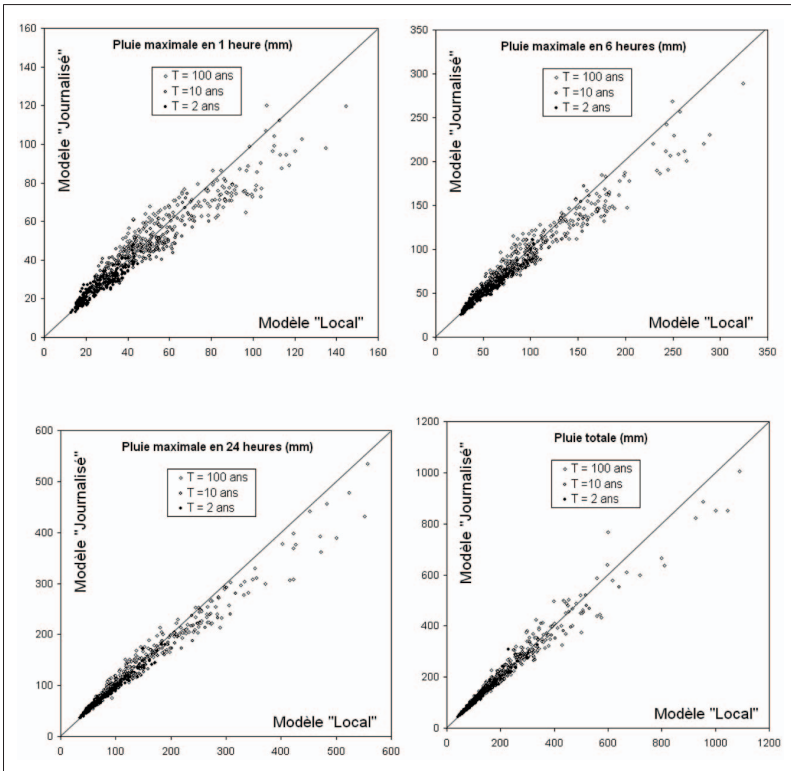
## Corrections systématiques

Bien que l'on ait vu que la version « journalisée » du générateur conduise à de bonnes performances, on peut se poser la question suivante.

**Faut-il caler la paramétrisation du générateur pour retrouver les quantiles de pluies obtenus par un ajustement statistique des pluies maximales observées ?**

Dans les études précédentes sur la journalisation du modèle SHYPRE, les écarts entre les quantiles observés (issus de l'ajustement d'une loi de probabilité) et les quantiles simulés (obtenus à partir des hyétogrammes simulés) étaient réduits par l'intermédiaire du calcul du paramètre PLIMS (Paramètre





**Figure 40.** Influence de la paramétrisation « journalisée » par rapport à la paramétrisation « locale » sur la restitution des quantiles de pluies maximales en 1, 6 et 24 heures, et des pluies totales sur l'événement, pour les périodes de retour 2, 10 et 100 ans.

local indicateur du modèle de simulation). Ce paramètre correspond en fait à la valeur du paramètre  $\mu$ PJMAX qui permet de restituer au mieux les quantiles locaux de pluies maximales journalières de périodes de retour 2, 5 et 10 ans obtenus par ajustement d'une loi exponentielle (Fine, 2003 ; Fouchier *et al.*, 2002).

Le paramètre PLIMS était alors optimisé sur chaque poste de façon à ce que la distribution des pluies journalières obtenues par simulation se rapproche au mieux de la loi exponentielle calée sur la distribution des valeurs observées. Dans ce cas, on émet l'hypothèse que les quantiles issus de l'ajustement d'une loi exponentielle sur la distribution de fréquences des valeurs observées servent de référence.

Cependant, il faut avoir conscience que les écarts entre les valeurs dites « observées » et les valeurs dites « simulées » sont dus à la fois aux problèmes de modélisation et de régionalisation, mais aussi au fait que ces quantiles sont issus de deux approches d'estimation différentes.

Si le calcul du paramètre PLIMS permet de corriger les erreurs moyennes liées aux problèmes du modèle et de sa régionalisation, il ne conduit pas forcément à une valeur de référence. En effet, il sous-entend que la distribution des

pluies doit être exponentielle et surtout il reste soumis aux problèmes d'ajustement face à l'échantillonnage des valeurs et plus particulièrement face à la présence de valeurs extrêmes dans la distribution des pluies observées.

**C'est pourquoi, afin de ne pas attribuer à l'ajustement de la loi exponentielle le rôle de référence pour la détermination des quantiles de pluies, il a été choisi de ne pas se caler localement sur les quantiles issus de cet ajustement pour recalibrer la paramétrisation du modèle « journalisé ».**

**En revanche, afin de prendre en compte les problèmes liés à la « journalisation », on regarde s'il existe un biais systématique du modèle « journalisé », que l'on corrigera.**

En effet, on observe une tendance systématique du modèle « journalisé » à surestimer les pluies des postes présentant des valeurs faibles de  $\mu\text{PJMAX}$  et à sous estimer les pluies des postes présentant des fortes valeurs de  $\mu\text{PJMAX}$ .

Ce biais peut être dû à la recherche d'une relation unique entre les paramètres horaires et journaliers sur l'ensemble de la France, alors qu'il peut y avoir des relations ou des paramètres fixés différents suivant les zones climatiques.

#### Correction du biais

On appliquera donc à la valeur du paramètre  $\mu\text{PJMAX}$  calculé localement, une correction systématique, indépendante de la localisation du poste étudié.

### Correction du $\mu\text{PJMAX}$ : $\mu\text{PJMAX}_c$

Afin de déterminer la correction « optimale », on a cherché les valeurs que devrait prendre le paramètre  $\mu\text{PJMAX}$  pour que les simulations conduisent à une distribution de fréquence des pluies journalières la plus proche de la distribution empirique. Pour cela on a calculé pour chaque poste la valeur du paramètre PLIMS qui optimise la restitution des quantiles 2 et 5 ans.

On étudie donc la relation entre la valeur du PLIMS obtenue et la valeur du paramètre  $\mu\text{PJMAX}$  (figure 41).

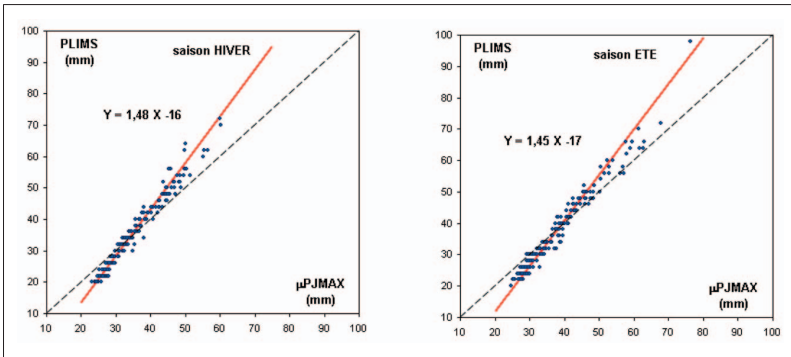
On s'aperçoit de la bonne corrélation entre les valeurs du paramètre PLIMS et celles du paramètre  $\mu\text{PJMAX}$ , ce qui nous conforte dans le choix d'effectuer une correction systématique, mais pas locale. Cette correction systématique va permettre de corriger les biais de la modélisation et de la « journalisation », par l'intermédiaire des relations suivantes :

pour la saison Hiver  $\mu\text{PJMAX}_c = 1,48 \times \mu\text{PJMAX} - 16$

pour la saison Été  $\mu\text{PJMAX}_c = 1,45 \times \mu\text{PJMAX} - 17$

Ces relations conduisent à augmenter les valeurs les plus fortes du paramètre  $\mu\text{PJMAX}$  et à diminuer les valeurs les plus faibles. Cette tendance est appliquée à l'ensemble de la zone étudiée.

Les relations présentées permettent ainsi d'obtenir une valeur corrigée de  $\mu\text{PJMAX}$  notée  $\mu\text{PJMAX}_c$ . Cependant, si on regarde la fluctuation des valeurs du paramètre PLIMS autour de la droite de régression, on observe une légère relation entre ces écarts et le paramètre  $\mu\text{DTOT}$ . Ces relations, conduisent à augmenter les valeurs de  $\mu\text{PJMAX}$  pour ces postes présentant des événements



**Figure 41.** Relation entre les paramètres PLIMS et  $\mu$ PJMAX pour les 217 postes pluviographiques.

en moyenne plutôt courts, et à réduire les valeurs de  $\mu$ PJMAX pour ces postes présentant des événements en moyenne plutôt longs. On apporte alors une correction supplémentaire pondérée par la durée moyenne des événements, correction qui va dans le même sens que celle qui nous a conduits à estimer le paramètre  $\mu$ NAVp par le paramètre  $\mu$ PJMAX.

Ce sont les zones du pourtour atlantique et des reliefs en dehors de la zone méditerranéenne qui sont le plus touchées par cette correction tendant à diminuer la valeur du paramètre  $\mu$ PJMAX. En effet, ces zones présentant des événements pluvieux longs, sont souvent caractérisées par des valeurs de  $\mu$ PJMAX assez élevées par rapport à la valeur du paramètre PLIMS. Ce sont des postes qui peuvent présenter de forts cumuls de pluie, mais pas nécessairement extrêmes.

Au final, on choisit d'effectuer une correction systématique, du paramètre  $\mu$ PJMAX grâce aux relations empiriques suivantes.

Pour la saison Hiver :  $\mu$ PJMAX corrigé =  $1,48 \times \mu$ PJMAX -  $3 \times \mu$ DTOT - 8,5.

Pour la saison Été :  $\mu$ PJMAX corrigé =  $1,45 \times \mu$ PJMAX -  $4 \times \mu$ DTOT - 7,5.

Afin d'apprécier les corrections effectuées, on représentera par la suite les résultats du modèle en termes de cartographie des écarts relatifs entre les quantiles de pluies simulées (issus de SHYPRE) et les quantiles dits « observés » (c'est-à-dire issus de l'ajustement d'une loi exponentielle sur les distributions observées des N plus fortes valeurs, où N est le nombre d'années d'observation).

On voit par exemple sur les cartes de la figure 42, la répartition des écarts relatifs de pluies maximales en 1, 6 et 24 heures pour la période de retour 2 ans, issue du modèle journalisé « brut ». On retrouve en fait les valeurs des graphiques de la figure 40.

Sur ces cartes, on retrouve la surestimation des pluies en une heure qui est présente sur tout le territoire, sauf en milieu méditerranéen. Cette surestimation est due à la difficulté d'appréhender la valeur de la variable RXP, qui est alors fixée pour l'ensemble du territoire.

On retrouve aussi la tendance à la surestimation en milieu tempéré, où les pluies sont les plus faibles et une sous-estimation en milieu méditerranéen où les pluies sont les pluies fortes.

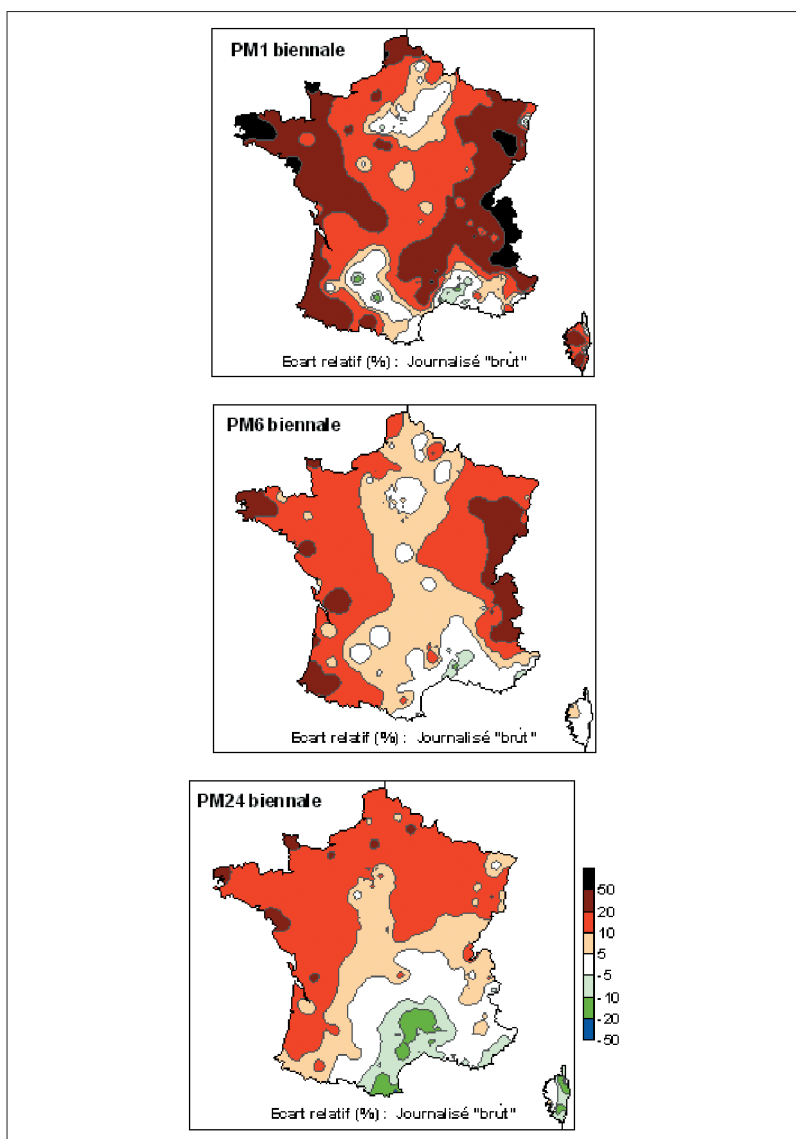


Figure 42. Écarts relatifs sur les quantiles : modèle journalisé « brut ».

On essaie alors de dé-biaisé la version journalisée du modèle, en corrigeant les valeurs de la variable  $\mu\text{PJMAX}$  par les relations présentées précédemment. Les cartes de la figure 43 montrent les nouveaux écarts relatifs obtenus avec le paramètre  $\mu\text{PJMAX}$  corrigé.

On observe une très nette amélioration de la restitution pluies maximales de 6 et 24 heures, mais aussi sur toutes les durées de 3 à 72 heures, pour la période de retour biennale. Ce quantile plus faiblement soumis aux problèmes d'échantillonnage que des quantiles de périodes de retour plus fortes, nous sert à valider les performances du modèle.

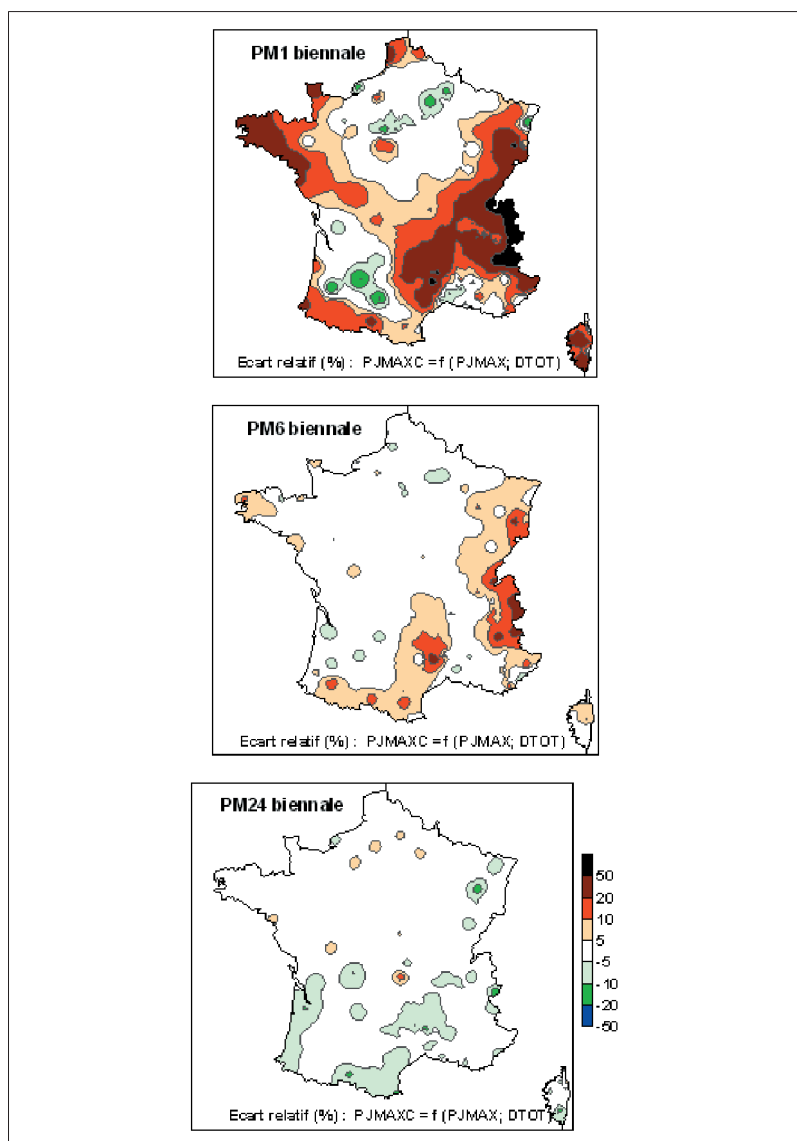


Figure 43. Écarts relatifs sur les quantiles : modèle journalisé « corrigé ».

### Performances du modèle

Avec la correction effectuée sur la variable  $\mu PJMAX$ , le modèle présente de très bonnes performances dans la restitution des quantiles biennaux des pluies de différentes durées avec des écarts relatifs généralement compris entre  $\pm 5\%$  et rarement en dehors de l'intervalle  $\pm 10\%$ .

Seule la restitution des pluies maximales en une heure présente des difficultés liées à la variabilité inexplicable par l'information journalière de la variable RXP.

## Correction des pluies en 6 heures

L'analyse de la carte des écarts relatifs sur les pluies maximales en 6 heures montre une surestimation dans les zones où les événements ont tendance à durer (fortes valeurs du paramètre  $\mu\text{DTOT}$ ). Ce problème est lié au fait qu'il existe une liaison entre la variable « volume des averses » et leur durée, qui n'est pas prise en compte et qui est accentuée avec la paramétrisation « journalisé » qui fixe les paramètres de la variable « durée ».

On corrige alors le modèle « journalisé » en fonction du paramètre journalier «  $\mu\text{DTOT}$  », de façon à prendre en compte ce problème. On arrive aux cartes présentées sur la figure 44.

Les modifications apportées dans ce cas conduisent à une amélioration de la restitution des quantiles de pluies maximales en six heures sur les zones où la pluviométrie est persistante, en particulier sur la région cévenole et les Alpes du Sud. La restitution des autres quantiles reste inchangée.

## Correction des pluies en une heure

Le problème de la restitution des pluies en une heure a été réduit suite aux améliorations apportées dans la restitution des pluies maximales sur des durées plus longues. Cette variable reste cependant celle qui présente les plus grands écarts relatifs.

Une modulation de la variable RXP a alors été introduite en fonction des variables journalières  $\mu\text{PJMAX}$  et  $\mu\text{DTOT}$ , pour réduire sa valeur sur les zones présentant les plus gros écarts. Cette correction a été effectuée de façon empirique au regard des cartes des paramètres  $\mu\text{PJMAX}$  et  $\mu\text{DTOT}$ , et de la carte des écarts relatifs. Elle se traduit par les relations suivantes :

En saison « hiver » :  $\text{RXP} = 0,35 + \Delta$   
avec  $\Delta = 0,15 \times (1 - e^{5 \times (\mu\text{DTOT} - 2,9)})$  si  $\mu\text{DTOT} < 2,9$  jours  
et  $\Delta = 0$  si  $\mu\text{DTOT} > 2,9$  jours

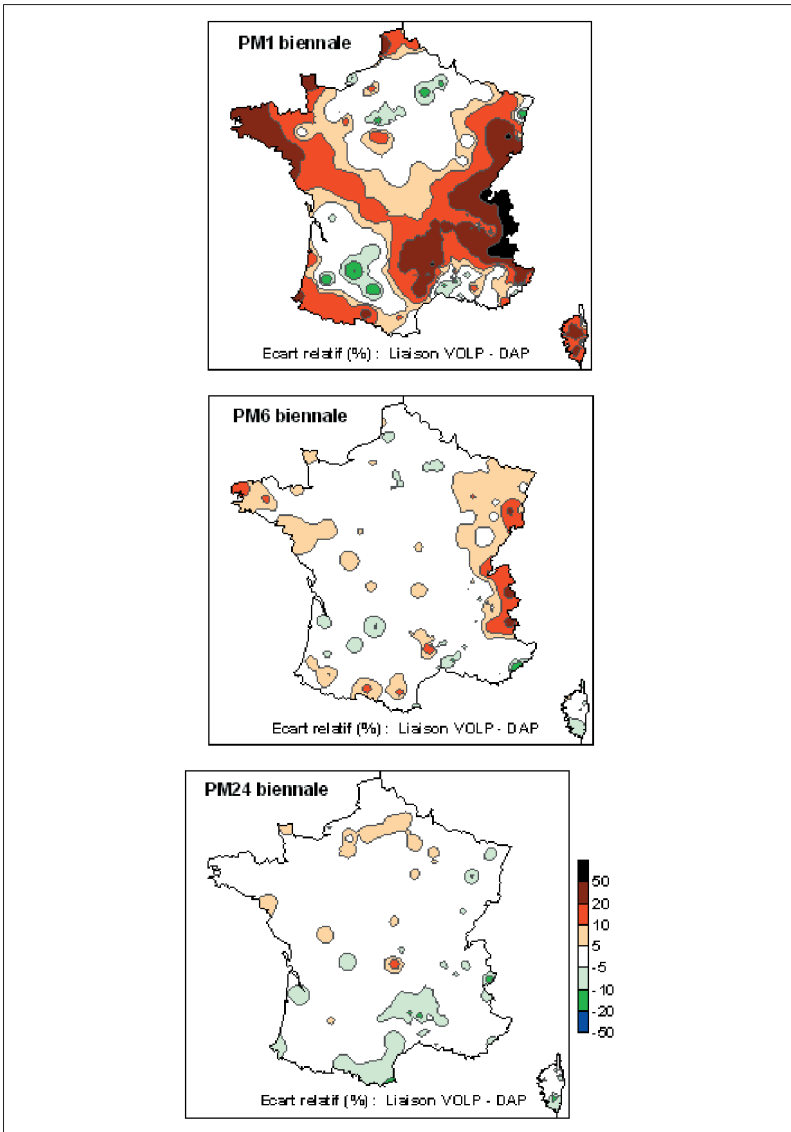
En saison « été » :  $\text{RXP} = 0,35 + \Delta_1 + \Delta_2$   
avec  $\Delta_1 = 0,3 \times (1 - e^{8 \times (\mu\text{DTOT} - 2)})$  si  $\mu\text{DTOT} < 2$  jours,  
 $\Delta_1 = 0,4 \times (e^{-8 \times (\mu\text{DTOT} - 2)} - 1)$  si  $\mu\text{DTOT} > 2$  jours,  
 $\Delta_2 = 0,25 \times (1 - e^{0,2 \times (\mu\text{PJMAX} - 45)})$  si  $\mu\text{PJMAX} < 45$  mm,  
 $\Delta_2 = 0,35 \times (e^{-0,2 \times (\mu\text{PJMAX} - 45)} - 1)$  si  $\mu\text{PJMAX} > 45$  mm

Une amélioration notable est visible sur la région cévenole, mais aussi sur les Alpes et sur la Bretagne (figure 45).

## Application sur 2 812 postes journaliers

Afin de tester les performances du modèle « journalisé » ou plus précisément de la paramétrisation journalière du modèle de génération de pluies horaires, on va le caler et contrôler ses performances sur les postes pluviométriques du réseau de Météo-France sur l'ensemble de la France.

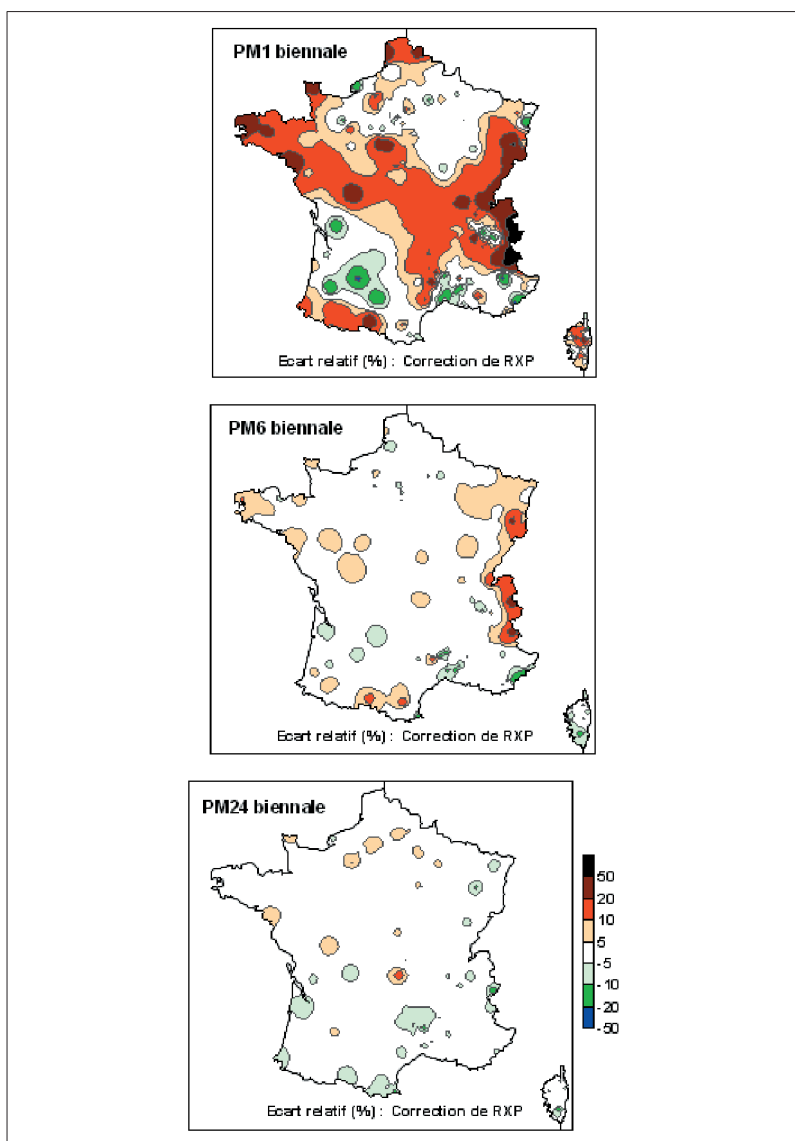
Le choix des postes pluviométriques a été effectué par Météo-France en sélectionnant les postes présentant au moins 120 mois de mesures par saison sur la période du 1/12/1977 au 30/11/2002 (soit l'équivalent de 20 ans d'observation sur une période commune de 25 ans). Un mois est considéré comme



**Figure 44.** Écarts relatifs sur les quantiles : amélioration des pluies maximales en 6 heures.

valide s'il a moins de 10 valeurs manquantes. En procédant ainsi, on optimise le nombre de postes utilisés (nécessaire pour faire la régionalisation des variables) et les chances d'avoir une période d'observation homogène.

On dispose alors de 2 812 postes pluviométriques contenant l'équivalent d'au moins 20 années de mesures sur la période 1975-2002. Ces postes sont localisés sur la figure 46.

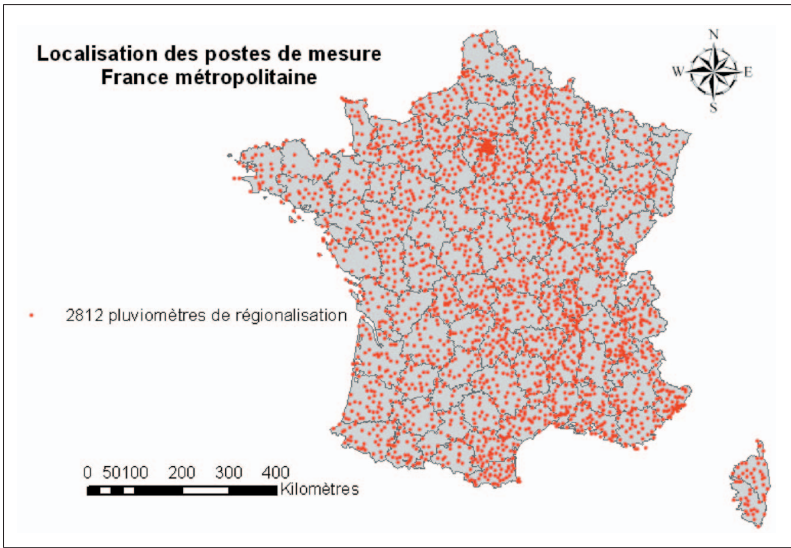


**Figure 45.** Écarts relatifs sur les quantiles : amélioration des pluies maximales en une heure.

## Cartographie sommaire de paramètres

À partir des chroniques journalières des 2 812 postes pluviométriques, Météo-France nous a fourni les valeurs des paramètres journaliers obtenus sur ces 2 812 postes. Une première cartographie de ces variables, obtenue par un krigeage simple des valeurs, nous a permis de tracer les cartes représentées sur la figure 47, p. 88 et sur la figure 48, p. 89.





**Figure 46.** Localisation des 2 812 postes pluviométriques.

*Remarque. Cette cartographie ne constitue pas la régionalisation des variables journalières, mais présente la répartition spatiale des paramètres journaliers à partir desquels seront faites les simulations présentées plus loin.*

À partir de ces données, une première série de simulations a été effectuée pour chaque poste, sans corriger la valeur du paramètre  $\mu\text{PJMAX}$ . Suite à ces simulations, on a calculé les valeurs du paramètre PLIMS pour effectuer la même analyse que celle faite au paragraphe p. 80 « Correction... ».

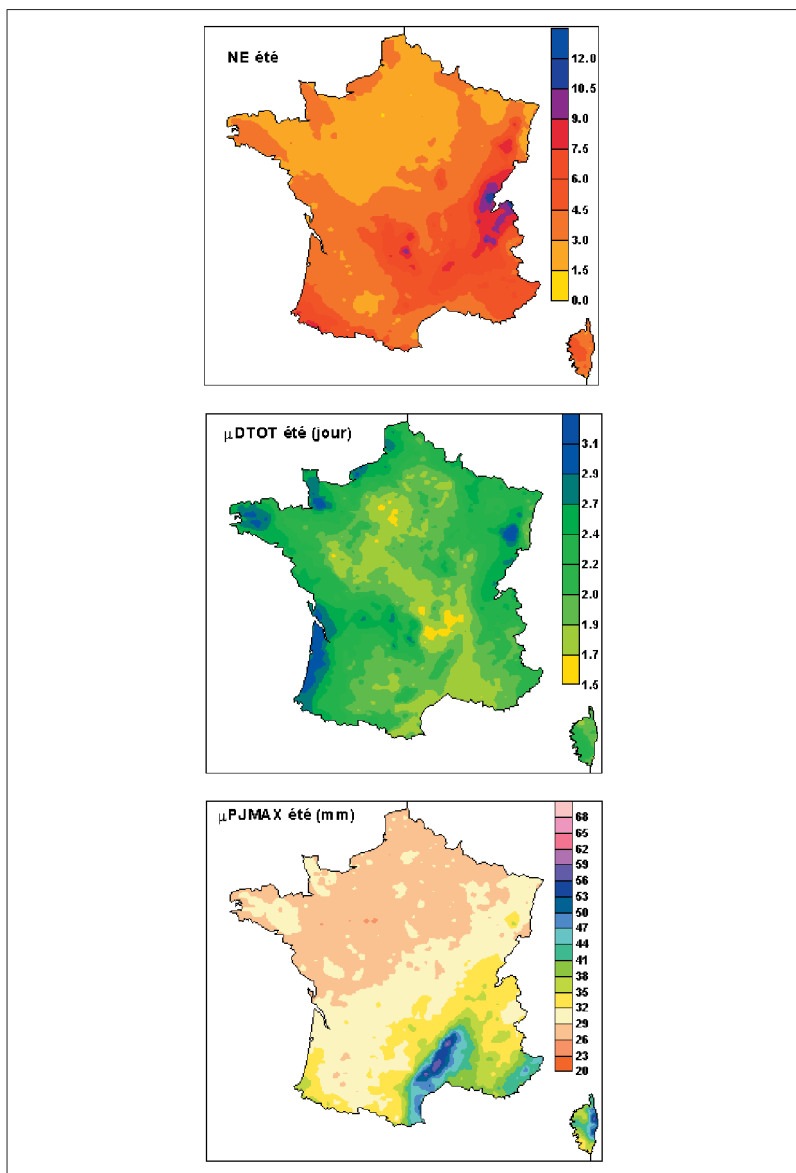
La comparaison des valeurs du PLIMS et du paramètre  $\mu\text{PJMAX}$  conduit aux graphiques de la figure 49 et de la figure 50. En reportant les relations déterminées dans le paragraphe 3.3.1, on voit qu'elles sont parfaitement adaptées pour réaliser la correction du paramètre  $\mu\text{PJMAX}$  à partir de sa valeur mesurée et du paramètre  $\mu\text{DTOT}$ .

On utilisera alors les relations déterminées dans le paragraphe 3.3.1 pour effectuer la correction des valeurs du paramètre  $\mu\text{PJMAX}$  :

- pour la saison hiver  $\mu\text{PJMAX corrigé} = 1,48 \times \mu\text{PJMAX} - 3 \times \mu\text{DTOT} - 8,5$  ;
- pour la saison été  $\mu\text{PJMAX corrigé} = 1,45 \times \mu\text{PJMAX} - 4 \times \mu\text{DTOT} - 7,5$ .

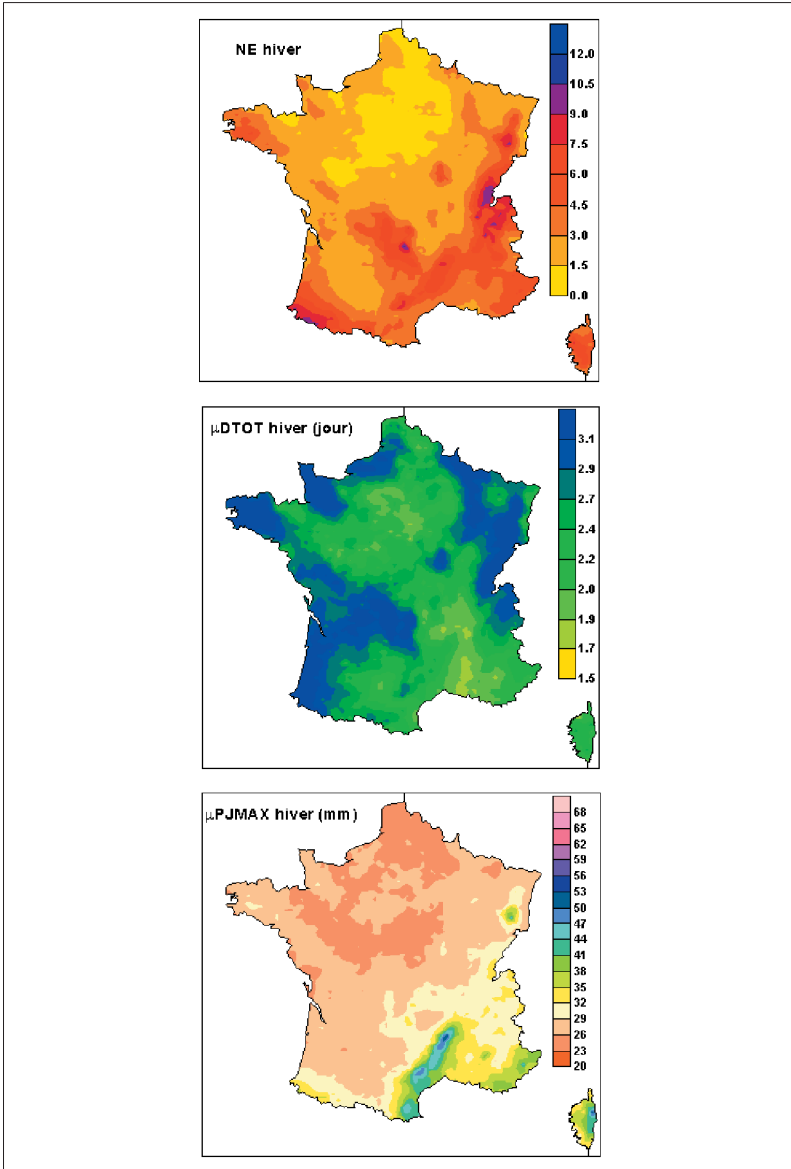
Sur les cartes de la figure 51, on représente la répartition spatiale du paramètre  $\mu\text{PJMAX}$ , du paramètre  $\mu\text{PJMAX corrigé}$  et du paramètre PLIMS pour la saison « été » et la saison « hiver ».

On s'aperçoit de l'effet de la correction du paramètre  $\mu\text{PJMAX}$  qui conduit à une cartographie de ce paramètre proche de celle du paramètre PLIMS. À ce stade de l'analyse, il est difficile de juger de la pertinence du paramètre  $\mu\text{PJMAX corrigé}$  par rapport au paramètre PLIMS tant leur cartographie est ressemblante. On peut noter cependant que la détermination du paramètre  $\mu\text{PJMAX corrigé}$  est basée sur le calcul de valeurs moyennes qui devraient être plus « robustes », face à l'échantillonnage, que les valeurs du PLIMS basées sur un calcul prenant en compte l'écart-type des pluies journalières maximales.



**Figure 47.** Variabilité spatiale des paramètres journaliers de la saison « été ».

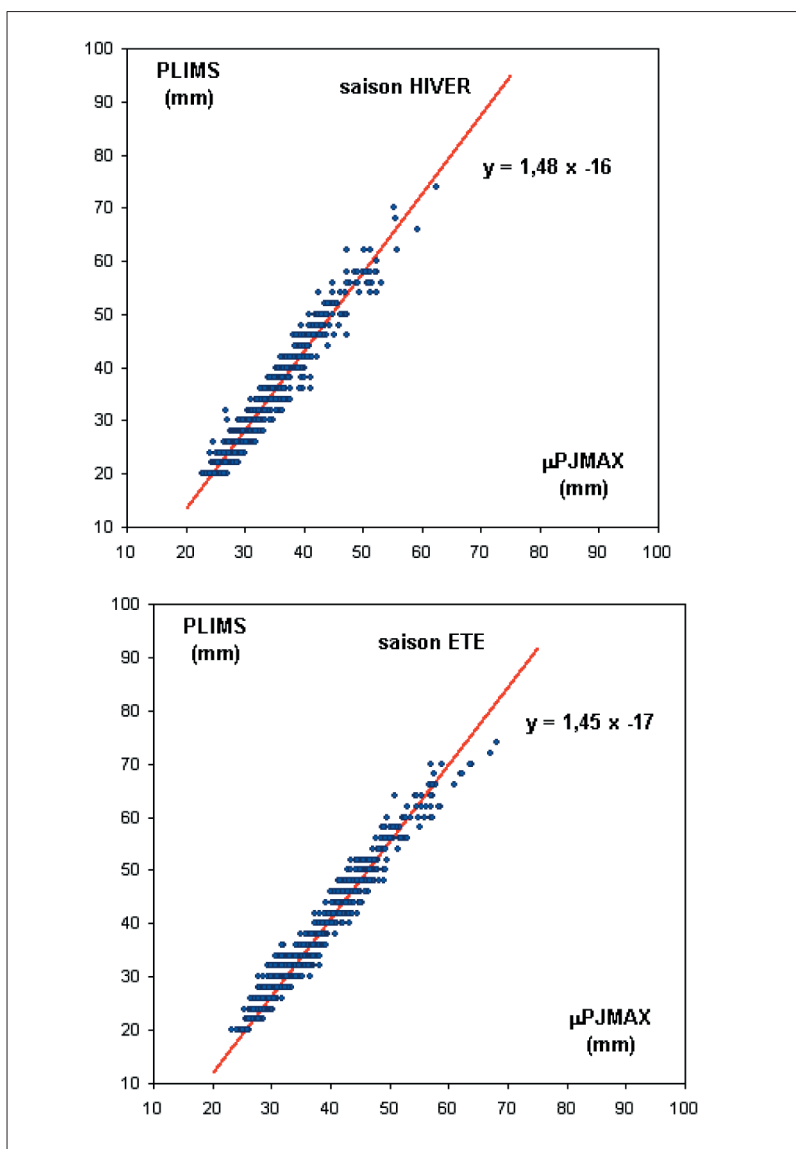
L'analyse des paramètres journaliers nous a conduits à faire des classes de valeurs pour chacun d'entre eux, afin de réaliser des simulations correspondant à chaque classe de valeur. La détermination des courbes Intensité-Durée-Fréquence (IDF) en chaque poste pluviométrique sera effectuée en repérant les simulations associées aux classes auxquelles appartiennent les paramètres du poste. C'est la même procédure qui sera effectuée pour la restitution des courbes IDF en chaque pixel de 1 km<sup>2</sup> sur lesquels la régionalisation des paramètres a été effectuée.



**Figure 48.** Variabilité spatiale des paramètres journaliers de la saison « hiver ».

Pour le paramètre NE, nous avons choisi une valeur de 15, qui correspond à une borne maximaliste sur la zone d'étude. Les distributions obtenues avec cette valeur permettent de retrouver facilement toutes les distributions associées à des valeurs de NE plus faibles. Il suffit pour cela de faire une transformation de la période de retour d'un facteur  $15/NE$ . Par exemple, le quantile de période de retour 10 ans obtenu avec un NE de 15 aura une période de retour de 30 ans avec un NE de 5.

Pour les autres paramètres, on a pris une plage de variation qui englobe aussi les valeurs pouvant être prise en milieu tropical.



**Figure 49.** Relation entre le paramètre PLIMS et  $\mu$ PJMAX : 2 812 postes pluviométriques.

Les plages de variation des paramètres observés et les classes de paramètres choisies sont représentées dans le tableau 7, p. 93.

## Résultats de simulation

À partir des paramètres journaliers déterminés localement pour chacun des 2 812 postes pluviométriques on a recherché les quantiles de pluies journalières associés aux simulations appropriées. Les graphiques de la figure 52, p. 93, présentent les quantiles obtenus par SHYPRE « journalisé » en fonction

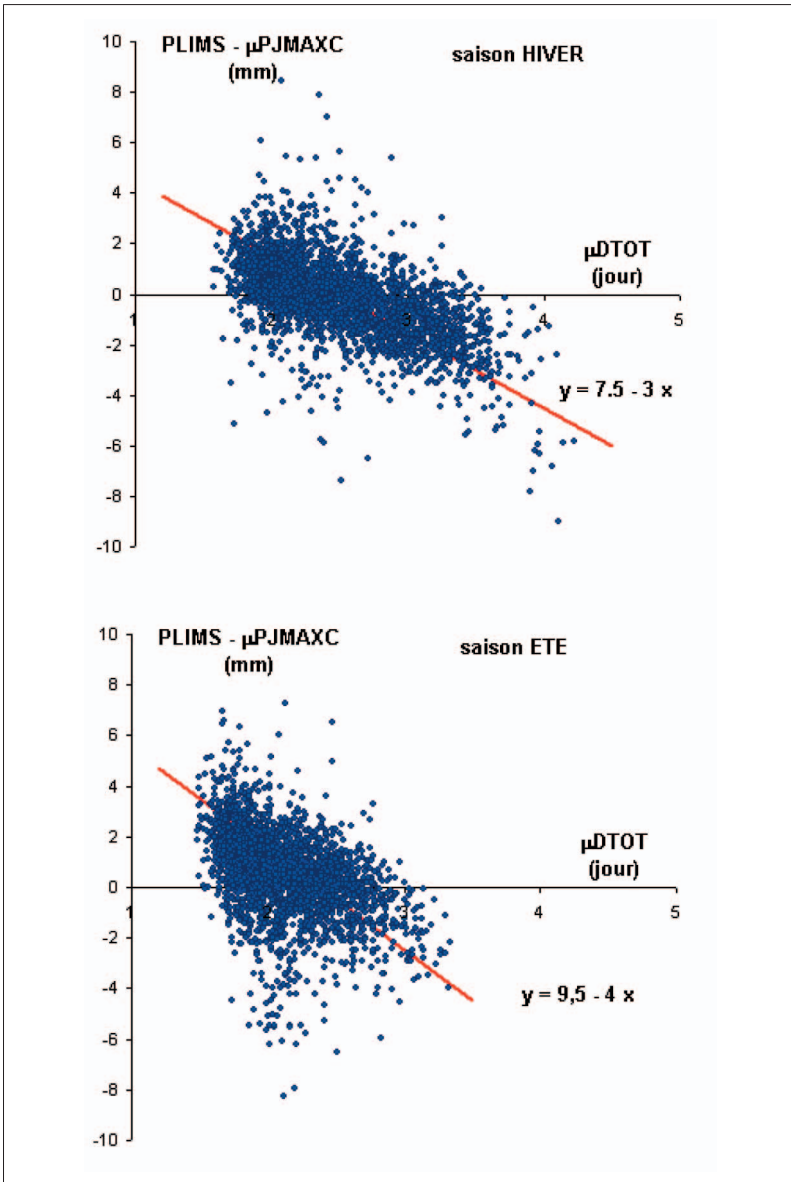
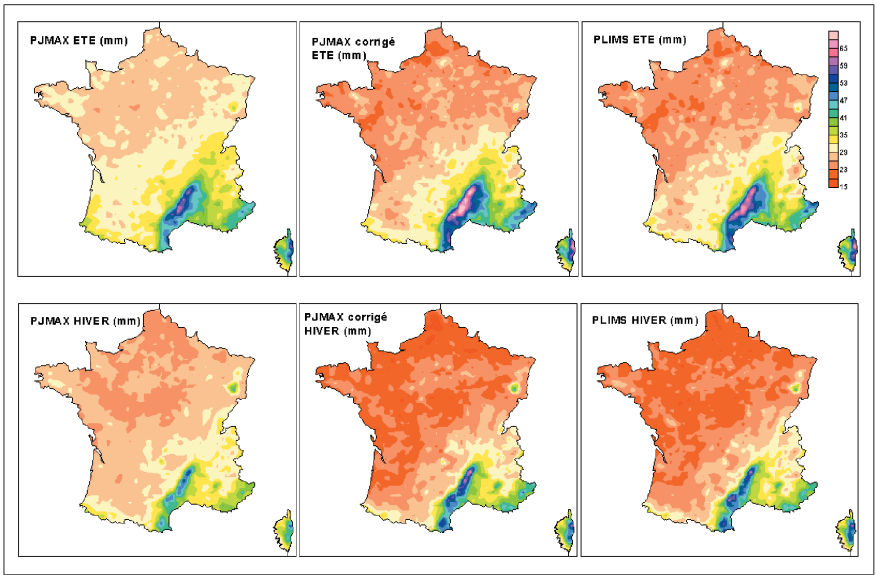


Figure 50. Entre l'écart (PLIMS-μPJMAXC) et μDTOT : 2 812 postes pluviométriques.

des quantiles issus de l'ajustement d'une loi exponentielle sur la distribution des valeurs observées<sup>4</sup>, issues des événements sélectionnés suivant les critères définis dans le paragraphe « Paramètres journalisés » p. 71.

4. L'ajustement de la loi exponentielle a été effectué sur un échantillon constitué des N plus fortes valeurs observées, N étant le nombre d'années d'observation. Cet échantillonnage par valeurs supérieures à un seuil (seuil variable d'un poste à l'autre) permet de prendre en compte plusieurs valeurs fortes une même année sans donner trop de poids aux valeurs faibles.



**Figure 51.** Variabilité spatiale des paramètres  $\mu$ PJMAX,  $\mu$ PJMAX corrigé et PLIMS : saison « été » et « hiver ».

## Conclusion

On observe sur ces graphiques la très bonne restitution des quantiles courants (période de retour 2 ans) par le modèle SHYPRE journalisé. On rappelle à cette occasion que ces résultats proviennent d'une version unique du modèle, et de relations entre les paramètres horaires et journaliers, uniques pour l'ensemble du territoire français. Seule la connaissance locale des paramètres moyens NE,  $\mu$ PJMAX et  $\mu$ DTOT conduit à ces résultats. Les quantiles plus rares (valeurs décennales) sont aussi proches des valeurs obtenues par l'ajustement d'une loi exponentielle sur les distributions observées. On remarque seulement une tendance du modèle à fournir des quantiles décennaux légèrement supérieurs à ceux issus d'une extrapolation strictement exponentielle. Ce résultat est une caractéristique du modèle qui sera encore plus marquée en extrapolation vers les valeurs extrêmes.

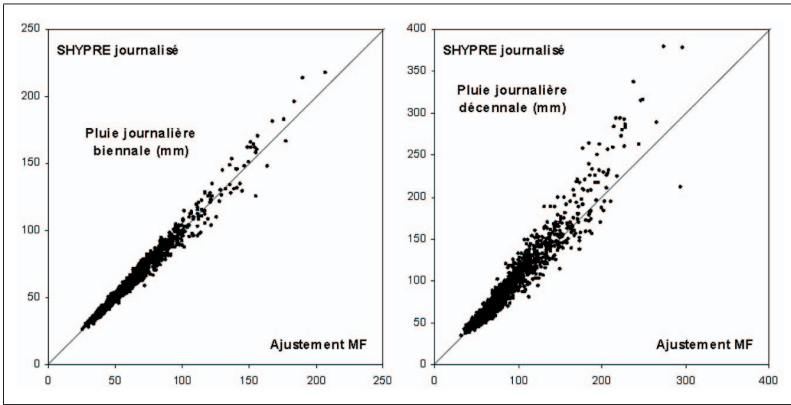
La répartition spatiale de ces quantiles est présentée sur les cartes de la figure 53 et de la figure 54. On observe la forte similitude des cartes de quantiles obtenues par les deux approches : modélisation des pluies horaires et ajustement de valeurs maximales.

Cependant, afin de mieux analyser les écarts qu'il peut y avoir entre les deux méthodes, nous avons cartographié les écarts relatifs entre les quantiles obtenus par les deux approches. À partir des valeurs de quantiles présentées sur les cartes précédentes, on calcule l'écart relatif suivant :

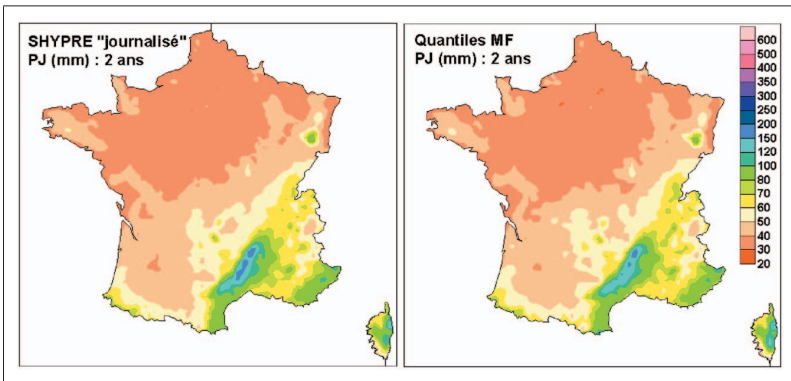
$$\text{Écart (\%)} = \frac{\text{Quantile SHYPRE} - \text{Quantile MF}}{\text{Quantile MF}} \times 100$$

	217 postes pluviographiques	2 812 postes pluviométriques	Classes choisies
NE hiver	de 0,4 à 10,2	de 0,3 à 12,4	NE = 15
NE été	de 1,2 à 12,9	de 0,9 à 12,9	NE = 15
PJMAX hiver	de 23,2 à 60,3	de 22,9 à 62,3	40 classes : 16-18-20-...-94
PJMAX été	de 24,6 à 76,4	de 23 à 68	40 classes : 16-18-20-...-94
DTOT hiver	de 1,43 à 3,45	de 1,58 à 4,23	10 classes : 1, 4/1, 6/1, 8/2/2, 5/3/3, 5/4/4, 5/5
DTOT été	de 1,49 à 3,09	de 1,48 à 3,34	10 classes : 1, 4/1, 6/1, 8/2/2, 5/3/3, 5/4/4, 5/5

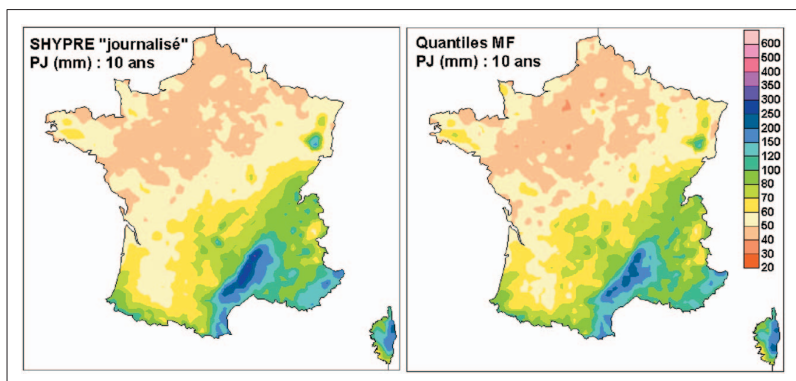
**Tableau 7.** Plages de variation des paramètres observés et les classes de paramètres.



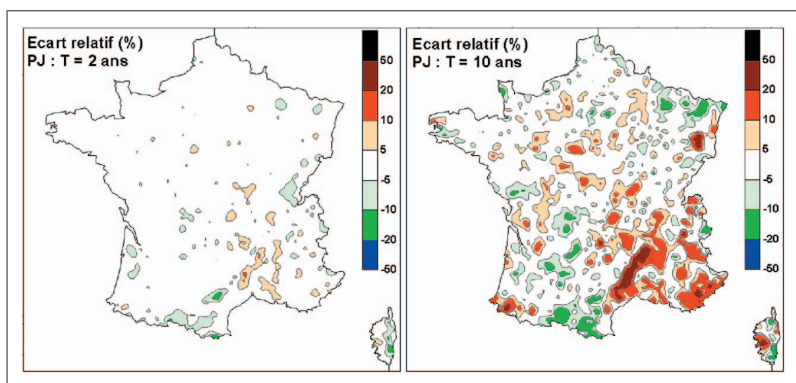
**Figure 52.** Comparaison des quantiles obtenus par les simulations de SHYPRE « journalisé » et ceux issus de l’ajustement statistique des distributions de fréquences observées fait par Météo-France.



**Figure 53.** Cartographie des pluies journalières biennales, issue des 2 812 postes journaliers : valeurs obtenues par SHYPRE « journalisé » et valeurs obtenues par Météo-France par ajustement d’une loi exponentielle.



**Figure 54.** Cartographie des pluies journalières décennales, issue des 2 812 postes journaliers : valeurs obtenues par SHYPRE « journalisé » et valeurs obtenues par Météo-France par ajustement d'une loi exponentielle.



**Figure 55.** Cartes issues des 2 812 postes pluviographiques, des écarts relatifs entre les quantiles de pluies journalières obtenus par SHYPRE « journalisé » et les quantiles issus de l'ajustement d'une loi exponentielle, pour les périodes de retour 2 et 10 ans.

Les cartes de la figure 55 présentent la répartition spatiale des écarts relatifs calculés sur les valeurs biennales et décennales, pour les pluies journalières maximales.

L'examen de ces cartes permet d'observer les zones où les quantiles obtenus par SHYPRE sont supérieurs à ceux issus de l'ajustement statistique (zones à dominante rouge) et les zones où les quantiles obtenus par SHYPRE sont inférieurs à ceux issus de l'ajustement statistique (zones à dominante verte).

La carte des écarts obtenus pour les valeurs biennales montre de très faibles valeurs, les écarts relatifs étant très largement dans la gamme  $\pm 5\%$ , et très rarement en dehors de la gamme  $\pm 10\%$ . Ces résultats montrent que le modèle « journalisé » conduit au respect de valeurs courantes de pluies journalières. Pour les valeurs décennales, les écarts sont évidemment un peu plus importants mais restent toujours très largement dans la gamme de  $\pm 10\%$ .



L'analyse des écarts conduits à distinguer deux cas :

- certains écarts semblent aléatoires : on observe la présence de petits « îlots » d'écarts positifs au milieu d'écarts négatifs et inversement. Ces écarts difficilement explicables peuvent être liés à des problèmes d'échantillonnage avec par exemple la présence de valeurs extrêmes dans l'échantillon des valeurs observées ;
- d'autres semblent plus « organisés » dans l'espace : on observe en particulier, sur la zone du pourtour méditerranéen et sur les Cévennes, des écarts positifs plus importants qui peuvent être liés à la divergence des deux approches vers les fréquences rares.

### **Vers la régionalisation**

La journalisation du générateur de pluies horaires consiste à le paramétrer à partir d'une information journalière. Trois paramètres journaliers, climatiquement pertinents, permettent de paramétrer le modèle de pluie horaire. L'application de la version « journalisée » du générateur de pluie sur près de 3 000 postes journaliers en France, montre sa capacité à reproduire très correctement les quantiles de pluies journalières, tout en fournissant une information aux pas de temps plus fins.



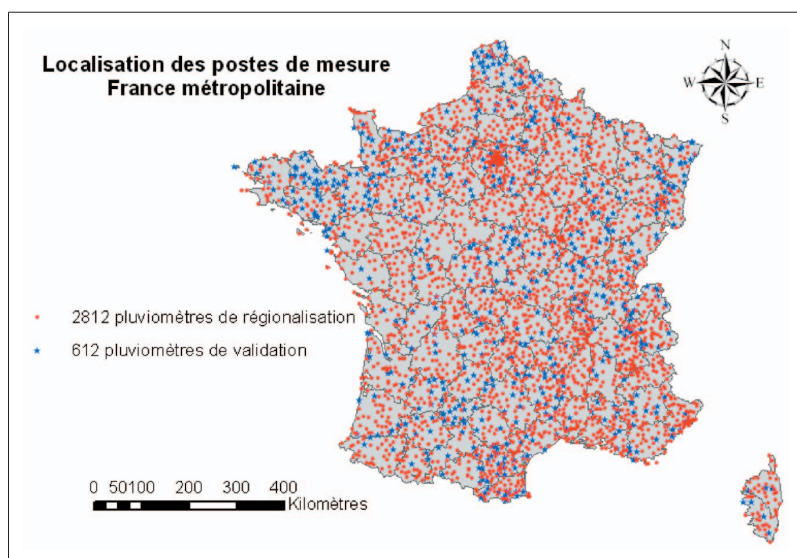
Chemin détruit par les ruissellements dus aux pluies intenses du 16 juin 2010,  
© P. Cantet, 2010.

# Régionalisation des paramètres journaliers

Afin d'être applicable en tout point du territoire, les paramètres du générateur de pluies horaires doivent être régionalisés. Pour cela, on régionalisera les paramètres journaliers dont l'information est plus largement disponible.

Les paramètres journaliers ont été régionalisés par Météo-France à partir de l'information de 2 812 postes journaliers, présentés précédemment.

À ces postes s'ajoutent 612 postes supplémentaires utilisés pour le contrôle de l'approche, et caractérisés par la présence de 15 à 20 ans d'observation sur la même période commune du 1/12/1977 au 30/11/2002 (figure 56).



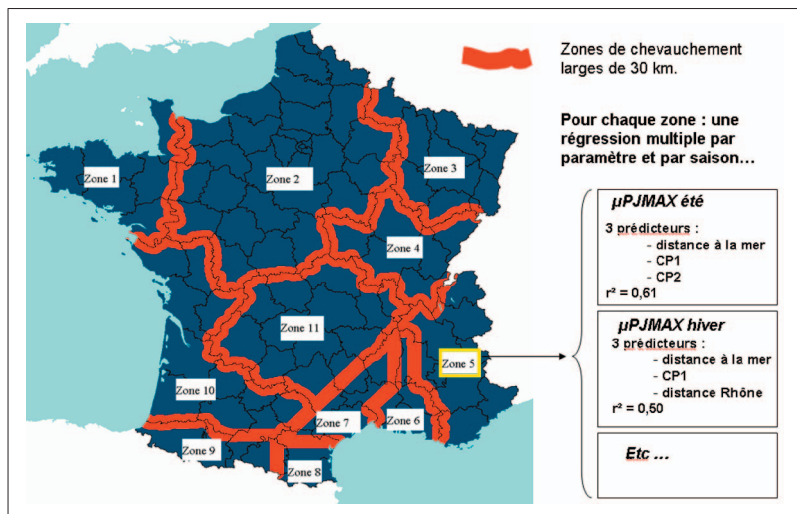
**Figure 56.** Localisation des postes pluviométriques ayant servi à la régionalisation des paramètres journaliers et à la validation de l'approche.

À partir des chroniques de pluies journalières, les paramètres NE,  $\mu\text{PJMAX}$  et  $\mu\text{DTOT}$  ont été calculés sur les 2 812 postes pluviométriques servant à la régionalisation du modèle. La régionalisation de ces trois paramètres a été réalisée par Météo-France.

## Descriptif de la méthode utilisée

Pour plus de détails concernant la méthode de spatialisation des variables journalières on pourra se référer à l'étude réalisée par Météo-France (Sol, Desouches, 2005).

La cartographie des paramètres journaliers a été réalisée sur la maille du kilomètre carré. Pour cela, la France métropolitaine a été découpée en 11 zones homogènes par rapport aux paramètres à régionaliser. Sur chaque zone, une régression linéaire multiple (appelée fonction géographique) a été étudiée entre chaque paramètre et des variables caractérisant « l'environnement géographique ».



**Figure 57.** Localisation des 11 zones « homogènes » déterminées par Météo-France pour la cartographie des paramètres journaliers du générateur de pluies horaires.

Des bandes de chevauchement sont définies de part et d'autre des limites des 11 zones, sur une largeur totale de 30 kilomètres, pour éviter les problèmes de discontinuité entre les zones. Sur ces bandes, les paramètres seront calculés par une moyenne pondérée des valeurs associées aux zones voisines à la bande.

Les variables géographiques pouvant intervenir dans l'élaboration de la fonction géographique sont :

- **l'altitude lissée calculée sur une maille de 5 par 5 kilomètres.** En moyennant le relief sur les 25 km<sup>2</sup> l'avoisinant, on réduit une variabilité du relief sans signification réelle vis-à-vis du comportement des précipitations (fonds de petites vallées peu encaissées par exemple) ;
- **des variables de proximité** (distances) : distance à la mer ou aux océans, distance aux grands cours d'eau (Rhône, Loire...), distance aux grandes crêtes (Cévennes, Ardennes, Alpes...) ;
- **des composantes principales du relief** : un « paysage » relatif<sup>5</sup> est déterminé pour chaque pixel de la zone. L'analyse en composantes principales de ce paysage permet de condenser cette information (121 valeurs d'altitudes moyennes par point) en une somme de paysages de base (les vecteurs propres) affectés de coefficients (les composantes principales). Cette analyse

5. Pour un pixel donné, un « paysage » est déterminé à partir de l'altitude lissée (moyenne sur 9 km<sup>2</sup>) des 121 points situés dans le carré de 30 km de côté, centré sur le pixel en prenant un point tous les trois kilomètres en latitude et longitude. À chaque altitude des 121 points on ôte ensuite l'altitude du pixel central, pour obtenir le paysage « relatif » du pixel.

a conduit à retenir 12 composantes principales et donc 12 vecteurs propres (« paysage de base ») servant à reconstituer les paysages relatifs. Le premier vecteur propre, représente une cuvette ou un monticule (selon le signe du coefficient que l'on y affecte), le deuxième une pente NW/SE, le troisième une pente SW/NE, le quatrième une colline au NE et au SW, une vallée au NW et au SE, le cinquième des collines à l'ouest et à l'est, des vallées au nord et au sud ; le sixième, au SE une vallée profonde et, plus loin, une montagne, des collines au NNE et à l'WSW, des vallées au NW... etc. Les paysages de base se compliquent de plus en plus pour les dernières composantes.

Ces variables géographiques permettent d'estimer les valeurs des paramètres uniquement à partir de l'environnement géographique. On choisit généralement de limiter le nombre de variables explicatives en fonction du gain que l'on peut avoir en termes de variance expliquée du paramètre. Le nombre de prédicteurs retenus varie alors d'une zone à l'autre, pour les différents paramètres. Lorsque les corrélations trouvées sont trop faibles, le nombre de prédicteurs choisis peut être nul.

On retrouve dans le tableau 8, les coefficients de détermination ( $r^2$ ) des fonctions géographiques trouvées, ainsi que le nombre de prédicteurs retenus (nombre entre parenthèses) pour établir ces fonctions. Quand le coefficient de détermination est faible, on réalise un krigeage direct des valeurs.

Zones	$\mu$ DTOT (été)	$\mu$ DTOT (hiver)	NE (été)	NE (hiver)	$\mu$ PJMAX (été)	$\mu$ PJMAX (hiver)
1	0,50 (5)	0,45 (3)	0,46 (4)	0,49 (4)	0,20 (2)	0,25 (3)
2	0,64 (7)	0,44 (4)	0,45 (3)	0,58 (5)	0	0
3	0,39 (1)	0,52 (2)	0,57 (2)	0,50 (1)	0,58 (5)	0,68 (5)
4	0,64 (3)	0,64 (3)	0,78 (2)	0,73 (2)	0,77 (5)	0,70 (5)
5	0,71 (6)	0,74 (4)	0,74 (8)	0,54 (6)	0,61 (3)	0,50 (3)
6	0,42 (3)	0,39 (3)	0,72 (3)	0,64 (3)	0,63 (4)	0,43 (4)
7	0,40 (3)	0,67 (6)	0,74 (3)	0,67 (6)	0,76 (4)	0,69 (5)
8	0,42 (3)	0,38 (2)	0,44 (3)	0,35 (3)	0,81 (4)	0,80 (4)
9	0,62 (2)	0,69 (1)	0,77 (3)	0,81 (5)	0,34 (1)	0,63 (3)
10	0,71 (5)	0,60 (5)	0,54 (5)	0,61 (5)	0,44 (2)	0,43 (4)
11	0,47 (2)	0,60 (2)	0,41 (1)	0,36 (2)	0,50 (4)	0,44 (5)

**Tableau 8.** Coefficient de détermination ( $r^2$ ) et nombre de prédicteurs (entre parenthèses) des fonctions géographiques retenues (Sol, Desouches, 2005).

On retient dans la plupart des cas 3 à 4 prédicteurs. Les coefficients de détermination ( $r^2$ ) sont aussi très variables et fluctuent entre 0,20 et 0,81, avec une valeur médiane de 0,58. On peut donc considérer de façon très globale, que la variation des paramètres journaliers peut s'expliquer à 60 % par les variables caractéristiques de l'environnement géographique.

Les résidus aux différentes régressions sont alors calculés aux 2 812 points de mesure. Ils sont ensuite interpolés par krigeage simple et ajoutés à la fonction géographique pour obtenir les champs finaux spatialisés à la maille de 1 km<sup>2</sup>.

Notons que la méthode de krigeage choisie pour interpoler les résidus, est une méthode d'interpolation exacte, c'est-à-dire qui conduit à une carte passant par les points de mesure. Dans ce cas, on verra que les valeurs des paramètres aux pixels contenant un poste pluviométrique sont proches des valeurs observées localement sur ce poste.

## Cartes résultantes

La régionalisation étant effectuée par Météo-France, nous avons récupéré les valeurs des trois paramètres journaliers, pour les deux saisons considérées, pour l'ensemble de la France métropolitaine, à la résolution d'un pixel d'un kilomètre carré.

Ces valeurs conduisent aux cartes présentées ci-après, des paramètres journaliers.

## Cartographie du paramètre $\mu$ DTOT

Les cartes de la figure 58 présentent l'interpolation spatiale du paramètre  $\mu$ DTOT à une maille de 1 km<sup>2</sup>.

Ce paramètre correspond à la valeur moyenne de la durée des événements pluvieux observés sur la saison considérée. Ce paramètre sert à déterminer le nombre de périodes pluvieuses d'un événement. Le nombre d'averses par période pluvieuse étant fixé dans la version régionale du modèle,  $\mu$ DTOT conduit à déterminer le nombre moyen d'averses à modéliser pour chaque saison.

On observe déjà des valeurs plus importantes en « hiver » qu'en « été ». Les événements d'hiver sont généralement plus longs que les événements d'été, et seront donc modélisés par une succession d'averses plus nombreuses.

On observe aussi une répartition spatiale, représentative des différents climats. Les zones influencées par le climat océanique et continental présentent des valeurs plus importantes que les zones influencées par le climat méditerranéen. On trouve les événements pluvieux les plus longs, le long des côtes atlantiques et le long des côtes de la Manche, de même que sur les reliefs de la zone continentale (partie ouest du massif central, Vosges, Ardennes, Jura et Alpes du Nord).

Les zones méditerranéennes sont par contre caractérisées par des événements pluvieux plutôt courts.

## Cartographie du paramètre NE

Les cartes de la figure 59 présentent l'interpolation spatiale du paramètre NE à une maille de 1 km<sup>2</sup> : nombre d'événements pluvieux par saison présentant au moins une pluie journalière supérieure à 20 mm. Il caractérise donc l'occurrence des événements pluvieux les plus forts et non pas l'occurrence des jours de pluies qui présente, elle, une répartition spatiale et temporelle différente.

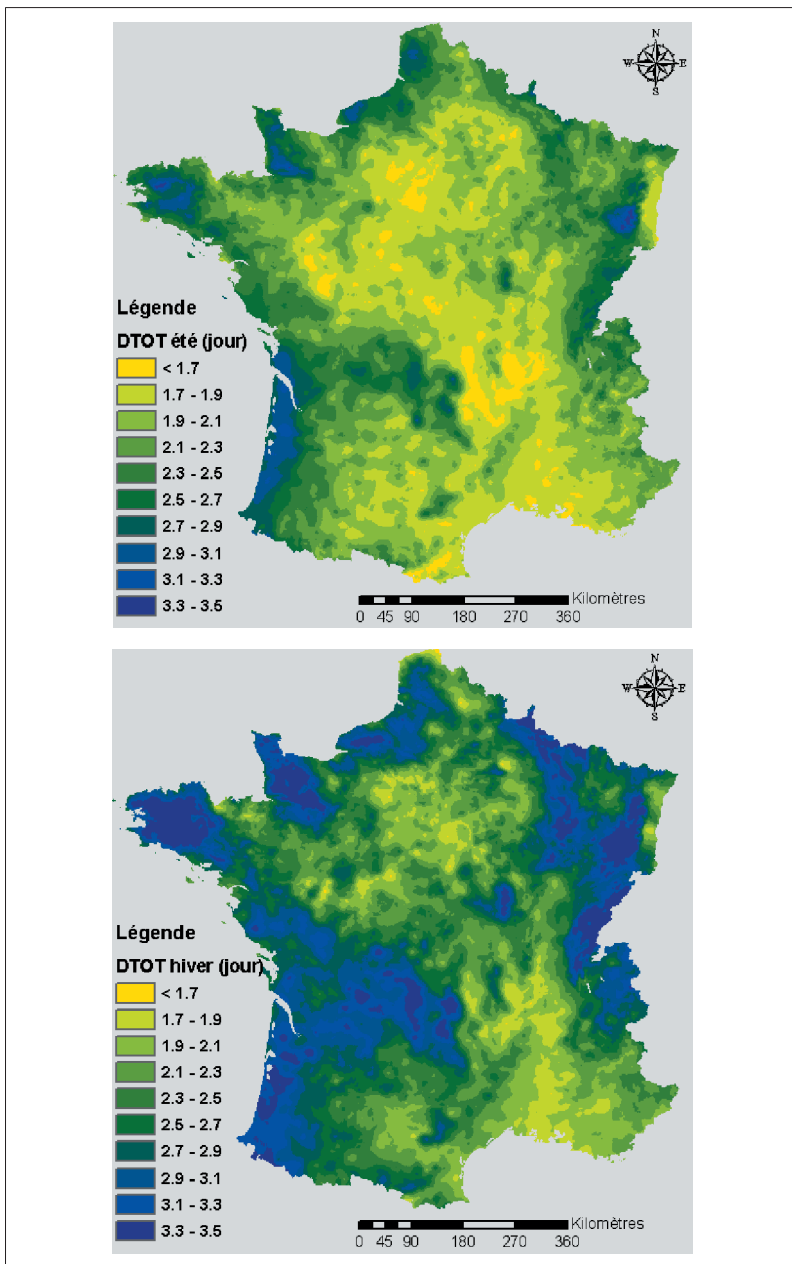
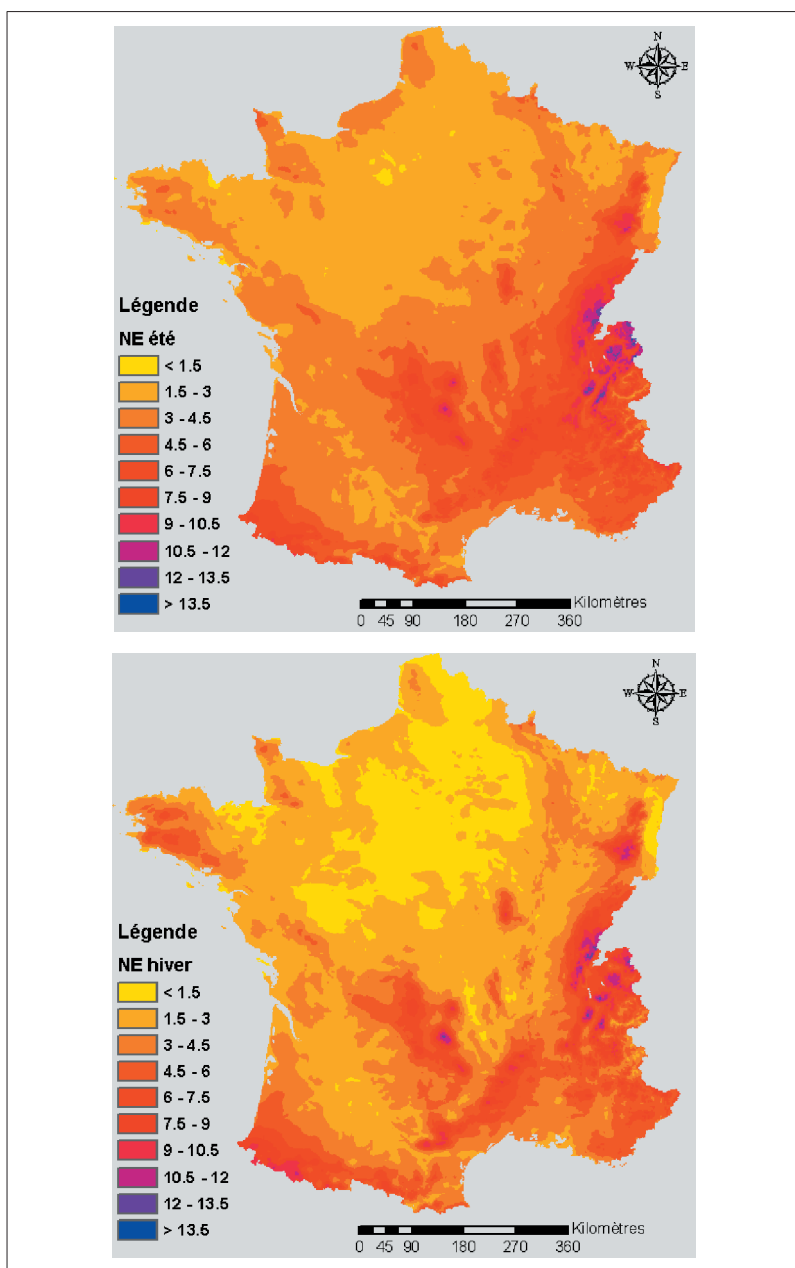


Figure 58. Cartographie au kilomètre carré du paramètre  $\mu$ DTOT (jour).

Bien que la différence soit moins marquée que pour la variable  $\mu$ DTOT, on observe une différence des valeurs prises par le paramètre NE suivant les deux saisons. Généralement, la saison « été » présente plus d'événements que la saison « hiver ». Rappelons que la saison « été » est définie par les mois de juin à novembre, et correspond en général à la période des événements pluvieux les plus forts (orages d'été en milieu tempéré et épisodes du type « cévenol » en milieu méditerranéen).



**Figure 59.** Cartographie au kilomètre carré du paramètre NE.

En termes de répartition spatiale du paramètre NE, on trouve les plus fortes valeurs sur les reliefs, avec un maximum dans les Alpes du Nord, le Jura, les Vosges et les monts du Cantal...

Les valeurs les plus faibles se retrouvent dans le centre de la France, le Bassin parisien et le Nord.



Une forte corrélation avec l'altitude est visible. Cependant, à altitude égale, les régions côtières présentent des valeurs de NE plus fortes que les régions plus continentales.

## Cartographie du paramètre $\mu\text{PJMAX}$

Les cartes de la figure 60 présentent l'interpolation spatiale du paramètre  $\mu\text{PJMAX}$  à une maille de 1 km<sup>2</sup>. Ce paramètre, défini comme la moyenne des pluies journalières maximales des événements pluvieux, traduit l'intensité de la pluviométrie. Il est borné à 20 mm au minimum par définition et augmente avec l'importance des pluies les plus fortes.

En termes de répartition temporelle, on observe des valeurs plus importantes en « été », considérée comme la saison la plus pluvieuse (sauf peut être sur les sommets des Vosges). Il s'agit généralement des orages d'été en milieu tempéré et alpin, et des pluies d'automne (septembre à octobre) en milieu méditerranéen.

La répartition spatiale de ce paramètre est beaucoup plus tranchée que celle des autres paramètres. On remarque un détachement net de la région cévenole et d'un grand quart Sud-Est de la France, avec des valeurs très élevées. Ces zones correspondent bien à des zones à pluviométrie très intense.

On observe aussi une très forte homogénéité sur le reste de la France, avec des valeurs relativement peu variables. L'influence du relief est cependant notable, en particulier sur les Alpes et sur les Vosges.

### Pertinence climatique des paramètres journaliers

La cartographie des paramètres journaliers conduit à une cartographie de caractéristiques pluviométriques pertinentes concernant la modélisation des pluies horaires fortes en zones climatiques hétérogènes. On retrouve différentes zones qui se différencient par la durée moyenne des événements pluvieux ( $\mu\text{DTOT}$ ), leur occurrence (NE) et leur intensité ( $\mu\text{PJMAX}$ ).

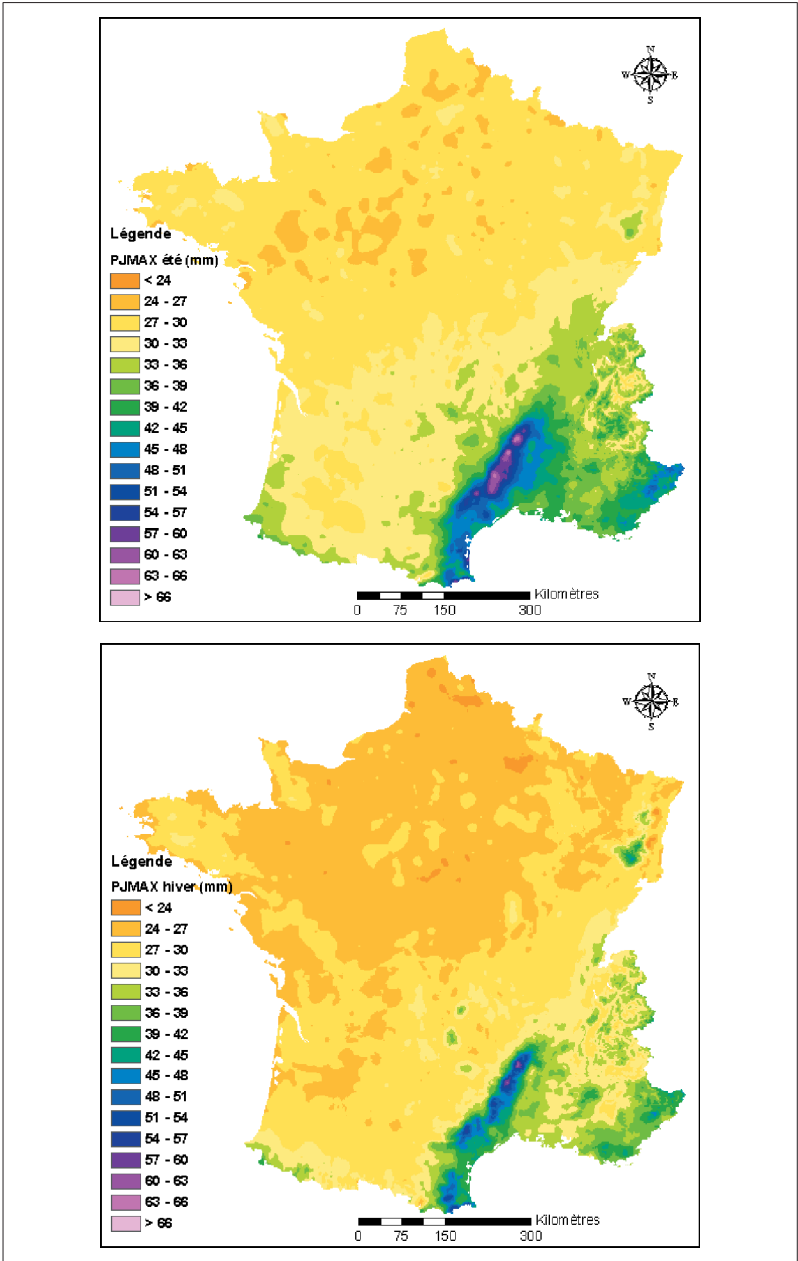
Les régions méditerranéennes sont caractérisées par une forte pluviométrie, intense, conduisant à de forts cumuls journaliers, d'autant plus marqués sur les reliefs cévenols. Ces régions ne présentent pas pour autant les plus fortes occurrences de pluie, ni les plus fortes durées.

En dehors des régions méditerranéennes, les régions côtières (côte atlantique et côte de la Manche) ou les régions d'altitude présentent une pluviométrie marquée par des événements plutôt longs mais pas nécessairement intenses.

L'occurrence des événements est, elle, d'autant plus forte sur les régions alpines.

Ce sont les régions du Bassin parisien et du Nord de la France qui présentent les plus faibles valeurs pour l'ensemble des trois paramètres.

La paramétrisation journalière du modèle de pluie horaire est réalisée à partir des trois paramètres journaliers présentés (NE,  $\mu\text{DTOT}$  et  $\mu\text{PJMAX}$ ). Ces paramètres permettent de déterminer en tout point la valeur des paramètres variables du générateur de pluie horaire, à partir de relations linéaires simples. Leur variabilité spatiale est donc directement liée à la variabilité spatiale du paramètre qui les détermine.



**Figure 60.** Cartographie au kilomètre carré du paramètre  $\mu$ PJMAX.

D'autres variables du générateur de pluies horaires sont déterminées par la combinaison des deux paramètres journaliers  $\mu$ DTOT et  $\mu$ PJMAX. C'est le cas des variables RXP et  $\mu$ NAVP dont nous présentons ici la spatialisation.

## Carte du paramètre RXP

Ce paramètre est un paramètre horaire qui est défini comme étant le rapport moyen entre la pluie maximale horaire d'une averse pluvieuse « principale » et son volume. Ce paramètre traduit approximativement le rapport moyen entre les pluies horaires et les pluies maximales en 4 heures (durée moyenne des averses principales).

Ce paramètre, difficilement explicable par des paramètres journaliers, devait être fixé. Cependant, il présentait une certaine variabilité qu'il était intéressant de prendre en compte. Bien que les relations permettant d'expliquer la variabilité de ce paramètre restent faiblement significatives, leur détermination à partir des paramètres journaliers est préférable, plutôt que de les fixer sur l'ensemble du territoire français.

À partir des résultats des simulations effectuées sur les 217 postes pluviographiques, une règle de calcul du paramètre RXP en fonction des paramètres journaliers a été déterminée pour minimiser les erreurs dans la restitution des pluies maximales horaires (voir p. 84). Cette règle a été appliquée sur l'ensemble du territoire français et a conduit à la cartographie du paramètre RXP pour la saison « été » présentée sur la première carte de la figure 61.

Bien que la cartographie du paramètre RXP ne soit pas fondée sur une liaison forte avec les paramètres journaliers, on retrouve globalement une correction du paramètre RXP qui va dans le sens des valeurs observées. Plutôt que de fixer la valeur du paramètre RXP à 0,50 (par exemple pour la saison « été »), on le module à la hausse sur les zones de relief et en bordure océanique, et on le module à la baisse en zone continentale de faible altitude.

Le paramètre RXP pour la saison « été », est ainsi modulé entre les valeurs 0,25 et 0,70 ; le paramètre RXP pour la saison « hiver », est modulé entre les valeurs 0,2 et 0,4.

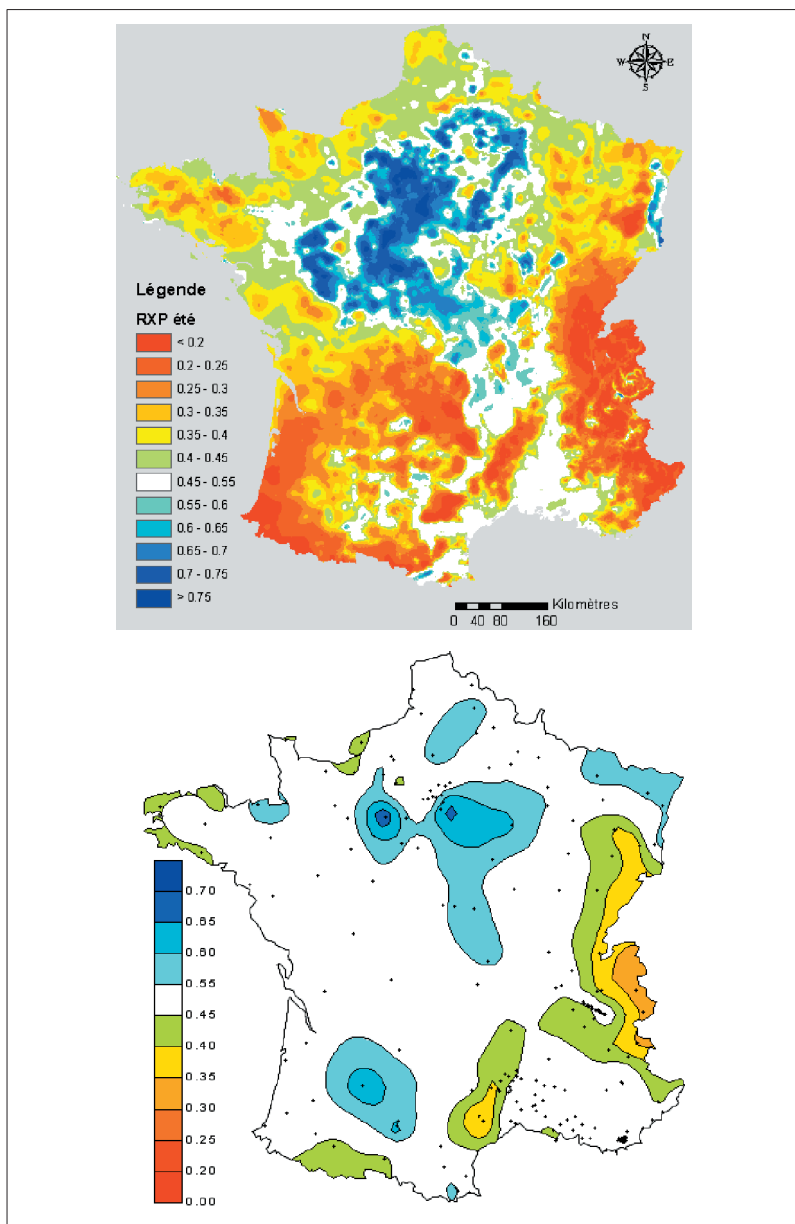
*Remarque importante. Le paramètre RXP contribue certes au calcul des pluies en une heure, mais son influence est surtout marquée pour les périodes de retour courantes. En effet, la liaison mise en avant entre la valeur du paramètre et la fréquence de l'averse (p. 51 « Relation entre les variables RX et VOL ») est plus importante dans la modélisation des pluies horaires de périodes de retour élevées.*

## Carte de $\mu$ NAVVP

Le paramètre  $\mu$ NAVVP est un paramètre journalier déduit du paramètre  $\mu$ PJMAX. Il correspond au nombre moyen d'averses principales dans un événement pluvieux. Il caractérise alors le phénomène de persistance des averses au sein d'un événement pluvieux.

La relation entre  $\mu$ NAVVP et  $\mu$ PJMAX étant linéaire, leur répartition spatiale est strictement identique.

En se référant à la cartographie du paramètre  $\mu$ PJMAX, on voit que le phénomène de persistance est surtout modélisé en climat méditerranéen, et en particulier sur la région cévenole.



**Figure 61.** Cartographie du paramètre RXP, saison « été », déduit des paramètres journaliers et directement interpolé à partir des 217 valeurs issues des observations horaires.

# Contrôle de la régionalisation des paramètres

Les cartes des paramètres journaliers sont à la base de la paramétrisation du générateur de pluies horaires pour son fonctionnement en tout point du territoire français. Le contrôle de la spatialisée proposée est nécessaire avant de procéder aux simulations. Les performances en termes de restitution des quantiles de pluies dépendront de la pertinence de la régionalisation des paramètres du modèle.

## Restitution des paramètres journaliers

On dispose de deux jeux de données sur lesquelles vont être effectués les tests :

- les 2 812 postes journaliers ayant servi à la régionalisation des paramètres journaliers (20 ans d'observations sans lacune sur une période commune de 25 ans du 1/12/1977 au 30/11/2002 ;
- les 612 postes journaliers de contrôle, n'ayant pas servi à la cartographie des paramètres (15 ans d'observations sans lacune sur une période commune de 25 ans du 1/12/1977 au 30/11/2002).

À partir des cartes des différents paramètres journaliers, on a extrait les valeurs des paramètres aux pixels contenant la localisation des postes pluviométriques pour lesquels on a calculé localement les valeurs des paramètres journaliers.

On peut ainsi comparer les valeurs observées et les valeurs issues de la régionalisation, des différents paramètres, aux postes de régionalisation et de validation (figure 62, figure 63 et figure 64, p. 108 à 110).

La méthode d'interpolation choisie étant une méthode exacte, la carte obtenue passe par les points ayant servi à son élaboration. Les très faibles écarts observés aux postes de régionalisation proviennent en fait de l'approximation concernant la localisation des postes pluviométriques sur une grille à la résolution du kilomètre carré.

On observe des écarts un peu plus importants aux postes de validation. Cependant, ces écarts restent largement acceptables, en particulier en ce qui concerne les paramètres les plus importants comme  $\mu_{PJMAX}$  et NE pour la saison « été ». Les écarts concernant le paramètre  $\mu_{DTOT}$  auront beaucoup moins d'impact sur la restitution des quantiles de pluies, et ne sont donc pas préoccupants.

## Justification de la méthode retenue

Afin de juger de la pertinence de la méthode d'interpolation choisie, nous avons comparé les résultats qu'auraient fournis d'autres méthodes d'interpolation plus « rustiques ». Pour cela, à partir des valeurs aux 2 812 postes ayant servi à la régionalisation, nous avons réalisé d'autres cartographies à partir des méthodes suivantes :

- krigeage direct des valeurs : notée « Krigeage » ;
- ajustement d'une fonction spline : notée « Spline » ;
- interpolation par pondération par l'inverse de la distance : notée «  $1/d^2$  ».

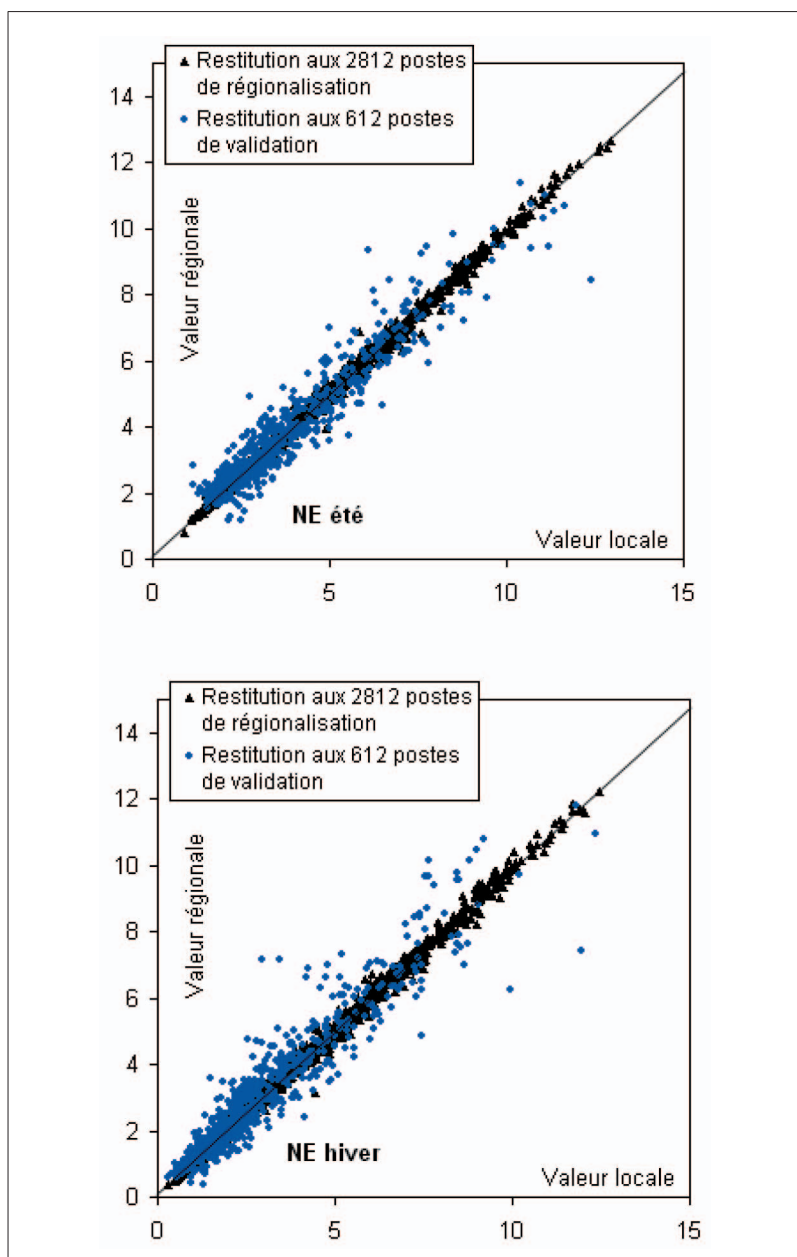


Figure 62. restitution du paramètre NE.

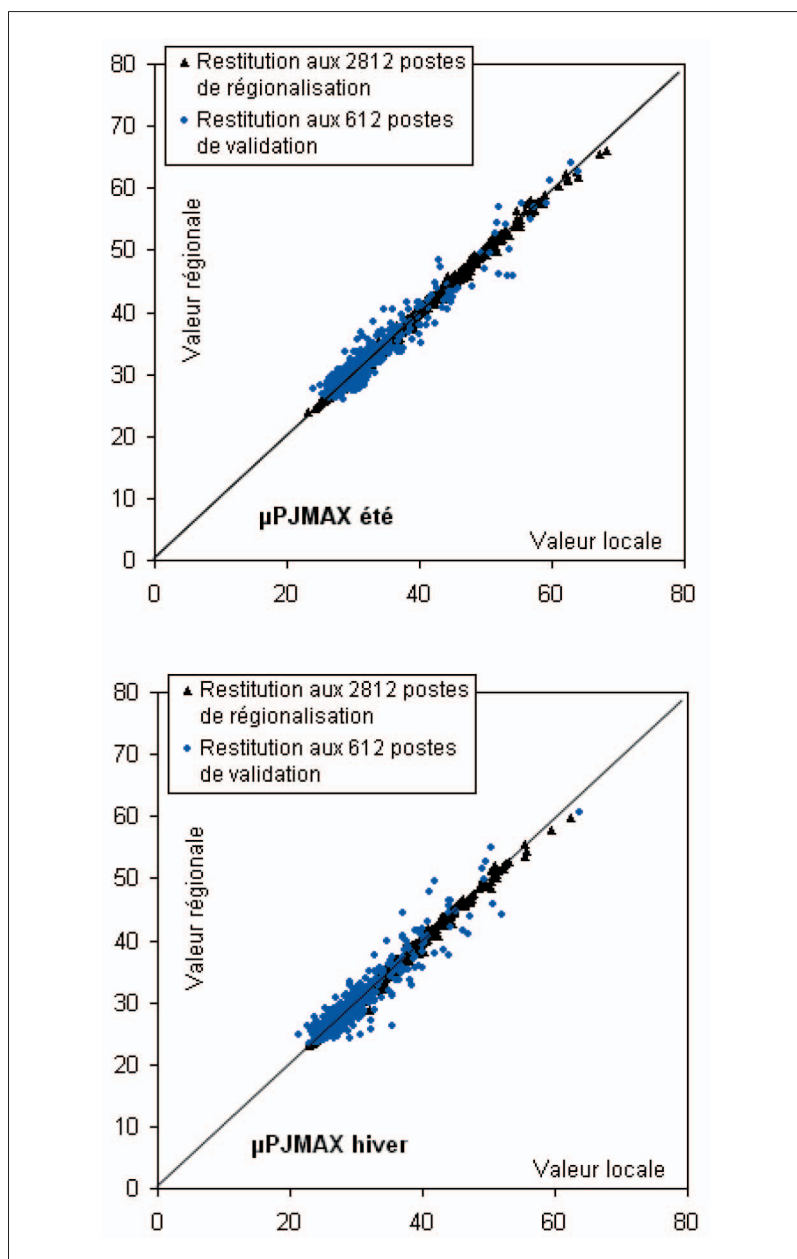
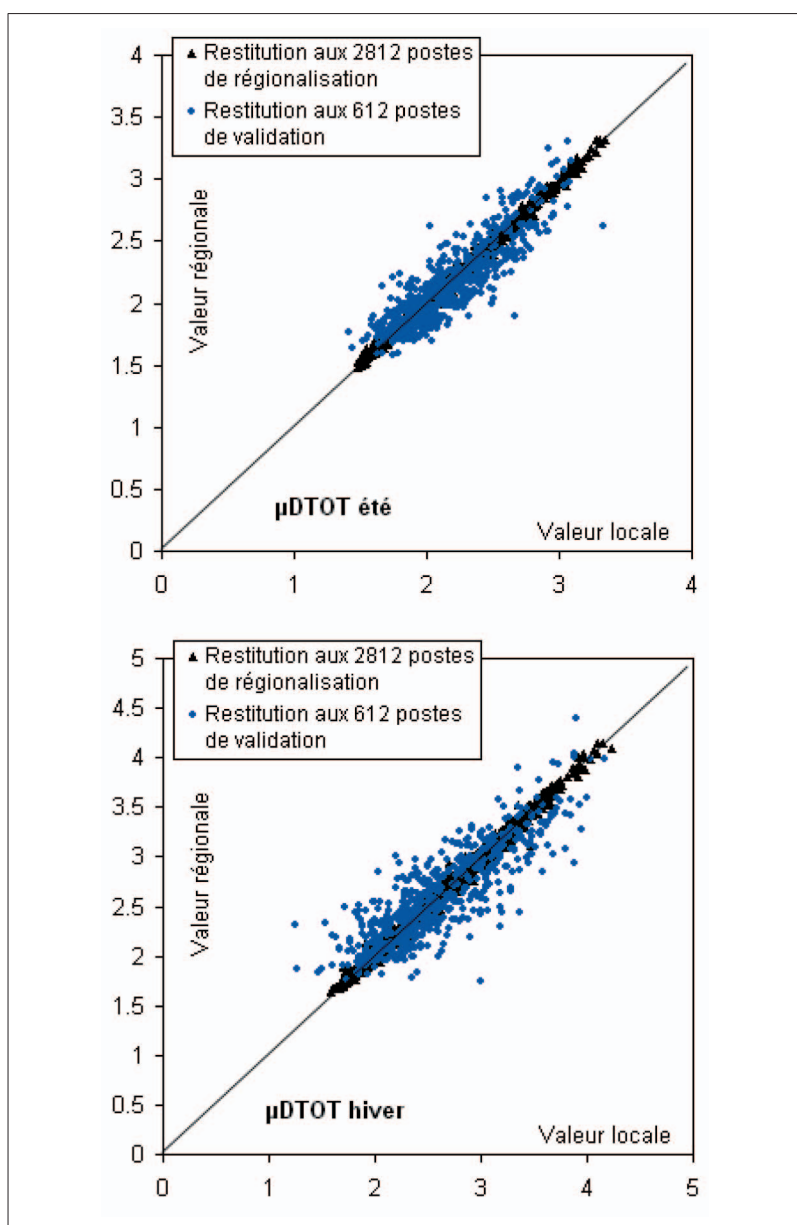


Figure 63. Restitution du paramètre  $\mu\text{PJMAX}$ .



**Figure 64.** Restitution du paramètre  $\mu$ DTOT.

Puis, nous avons comparé leurs performances en restitution des paramètres journaliers sur les 612 postes de contrôle.

La méthode d'interpolation proposée par Météo-France conduit aux meilleurs résultats concernant la restitution des paramètres NE et  $\mu$ PJMAX (paramètres les plus importants). Les autres méthodes, plus simples, comme la pondération par l'inverse de la distance, conduisent aussi à de bons résultats. Ces résultats satisfaisants sont dus à la densité relativement forte de postes pluviométriques utilisés pour la cartographie de paramètres moyens.



	NE hiver	$\mu$ NE été	$\mu$ PJMAX hiver	$\mu$ PJMAX été	$\mu$ DTOT hiver	$\mu$ DTOT été
AURELHY	87,5	90,1	87,7	91,9	72,8	76,9
Krigeage	83,5	88,8	87,7	91,7	72,3	72
Spline	85,7	88,4	86,6	91,2	68,6	75
1/d <sup>2</sup>	86	89,2	87,4	91,5	73,2	77,3

**Tableau 9.** Coefficient d'efficacité (critère de NASH en %) des différentes méthodes d'interpolation testées, pour spatialiser les différents paramètres journaliers.

Notons cependant que pour certaines zones et pour certaines variables, aucune « fonction géographique » n'a pu être trouvée. Dans ce cas, une interpolation par krigeage simple a été réalisée. La méthode de la présente étude ne présente d'intérêt que pour les zones à relief prononcé. Les résultats présentés dans le tableau 9 sont donc *a priori* encore meilleurs si l'on enlève les zones de plaine.

## Restitution des quantiles journaliers

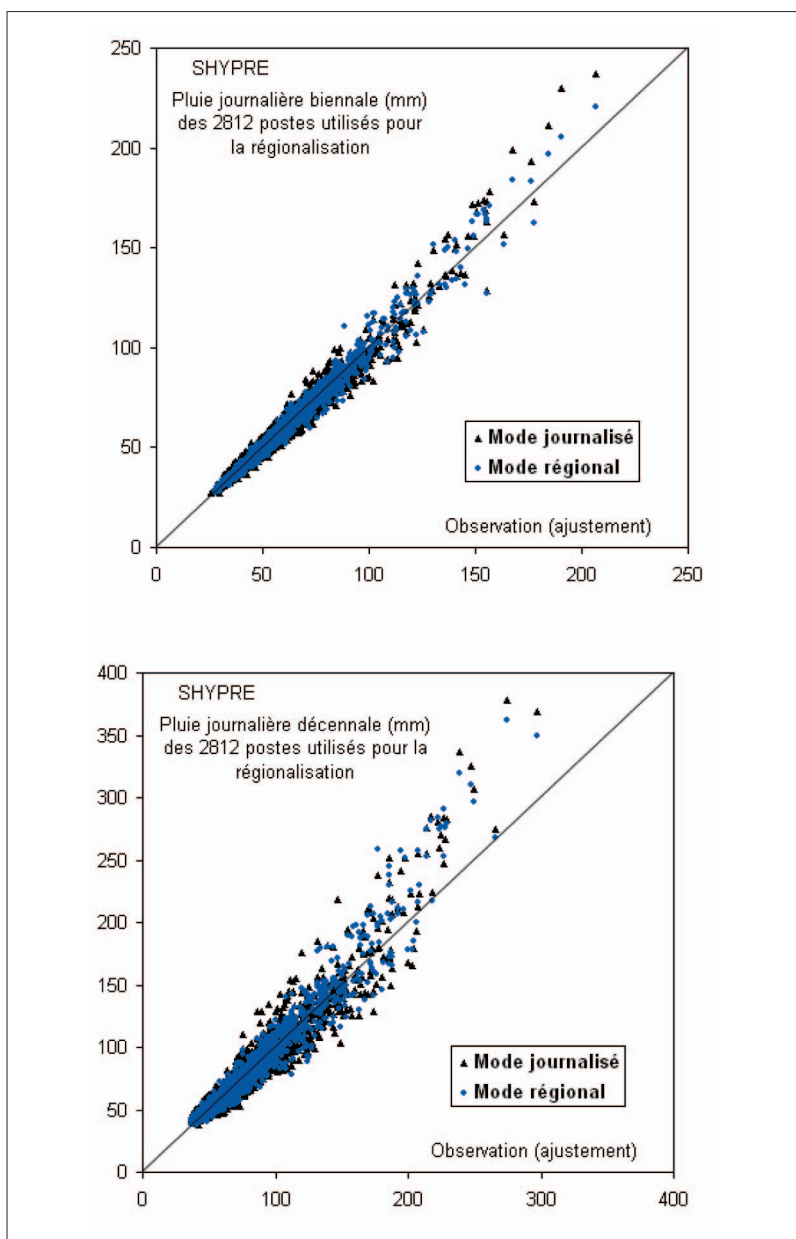
La cartographie des paramètres journaliers permet de faire fonctionner le générateur de pluies horaires en tout point du territoire étudié. On a donc généré des chroniques de pluies horaires en chaque pixel d'un kilomètre carré de la France métropolitaine. À partir de ces simulations, on extrait les variables de validation du modèle de pluies horaires, c'est-à-dire les pluies maximales de 1 à 72 heures. On détermine alors les quantiles de pluies à la maille du kilomètre carré.

De la même façon que pour les paramètres journaliers, nous allons alors comparer les quantiles de pluies obtenus par la méthode SHYPRE régionalisée, aux valeurs issues d'un ajustement statistique local d'une loi exponentielle aux pluies observées aux postes pluviométriques.

## Restitution sur les 2 812 postes de régionalisation

Les graphiques de la figure 65 présentent les quantiles de pluies journalières obtenus par le générateur de pluie en fonction des quantiles de pluies journalières obtenus par un ajustement local des distributions de fréquences des valeurs observées. Les quantiles obtenus par le générateur sont ceux issus de l'étape de « journalisation » et ceux issus de l'étape de « régionalisation ». Les quantiles obtenus en mode « régional » correspondent aux quantiles issus des cartes de pluies simulées, en prenant les valeurs des quantiles aux pixels contenant le poste pluviométrique.

On constate que les quantiles issus de la régionalisation sont équivalents à ceux obtenus par une paramétrisation journalière locale (mode « journalisé »). Ce résultat est directement lié à la méthode de régionalisation des paramètres qui passe par les points de mesure. Les écarts entre les quantiles



**Figure 65.** Restitution des pluies journalières biennales et décennales sur les 2 812 postes utilisés pour la régionalisation, en fonction du mode de paramétrisation : paramétrisation « journalisée » et paramétrisation « régionale ».

issus du mode « journalisé » et ceux issus du mode « régional » proviennent de l'approximation déjà observée entre la position exacte des pluviomètres et la position du pixel (centre du pixel) auxquels ils sont associés.

L'écart entre les quantiles de pluies journalières issus des simulations et issus de l'ajustement des distributions observées reste donc le même aux postes

de régionalisation. Il dépend uniquement des performances du modèle « journalisé ».

## Restitution sur les 612 postes de contrôle

Les graphiques de la figure 66 présentent les quantiles de pluies journalières obtenus suite à la régionalisation du générateur de pluies horaires, aux 612 postes de contrôle. On compare les quantiles issus de SHYPRE au pixel contenant un poste pluviométrique de contrôle, avec les quantiles obtenus par un ajustement des pluies observées sur ce poste.

On observe une bonne restitution des quantiles de pluies journalières, proches de ceux issus d'un ajustement statistique sur les distributions de valeurs observées. Les écarts observés sont bien sûr plus importants pour les valeurs décennales, car ils prennent aussi en compte les divergences de la méthode de SHYPRE par rapport à l'ajustement d'une loi de probabilité purement exponentielle.

Notons que les résultats présentés ici intègrent l'ensemble des incertitudes liées à la régionalisation de la méthode :

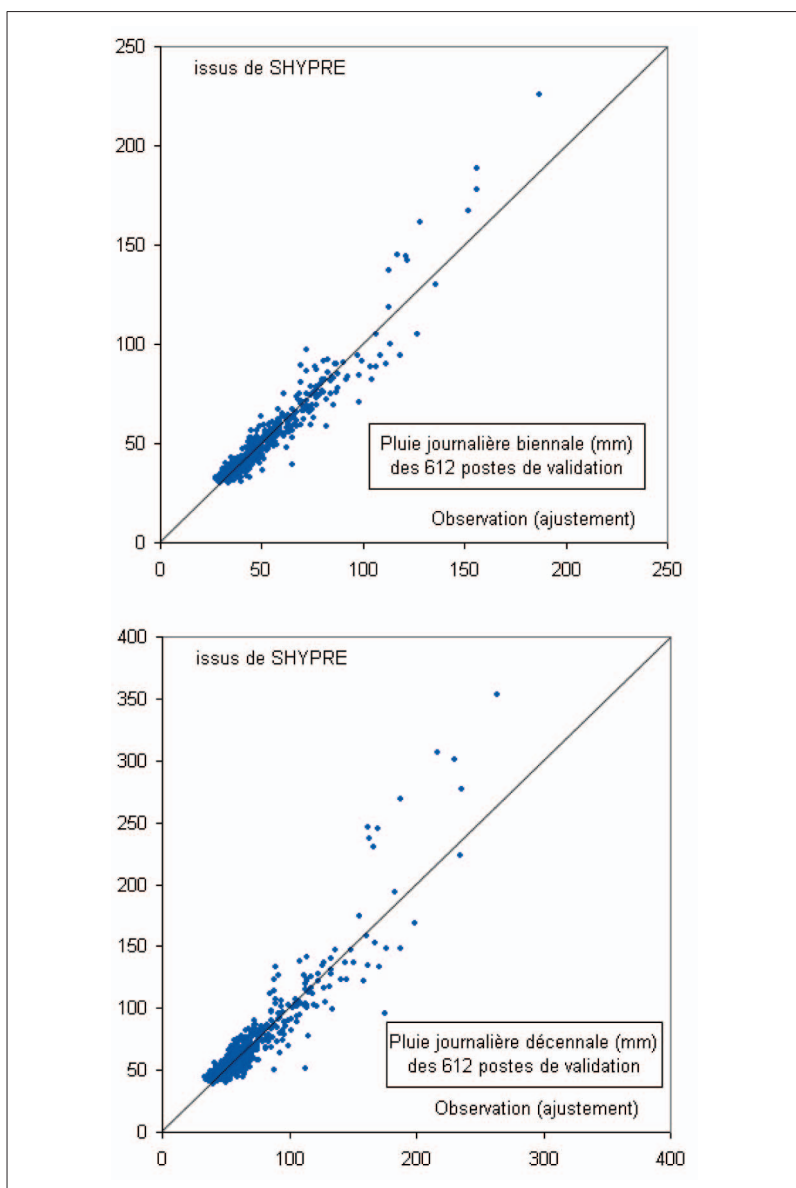
- on compare des quantiles déduits de la simulation de chroniques de pluies horaires, à des quantiles ajustés (« observation ») : écarts relatifs aux deux approches ;
- le modèle est paramétré par une information uniquement journalière : écarts liés à la « journalisation » du modèle ;
- les paramètres journaliers du modèle, aux postes de validation, sont issus d'une interpolation spatiale : écarts liés à la régionalisation des variables.

Sur les graphiques de la figure 67 on peut comparer les distributions des écarts relatifs concernant la restitution des quantiles de pluies journalières sur les postes ayant servi à la régionalisation et sur les postes de validation.

Les écarts relatifs calculés sur les 2 812 postes ayant servi à la régionalisation ne sont pas dus à la régionalisation de la méthode, mais aux écarts liés à la « journalisation » du modèle SHYPRE et à la divergence de méthode par rapport à l'ajustement d'une loi exponentielle sur la distribution des valeurs observées. Ces résultats servent donc de référence et sont comparés avec ceux obtenus sur les 612 postes de validation pour juger de la régionalisation de la méthode SHYPRE.

Dans ce cas, on s'aperçoit que la régionalisation de la méthode conduit à des écarts largement acceptables. Pour 95 % des postes de validation on obtient un écart relatif sur les pluies journalières biennales compris entre +/- 20 % par rapport aux quantiles obtenus localement par l'ajustement d'une loi de probabilité. De même, 80 % des postes présentent un écart inférieur à +/- 10 %.

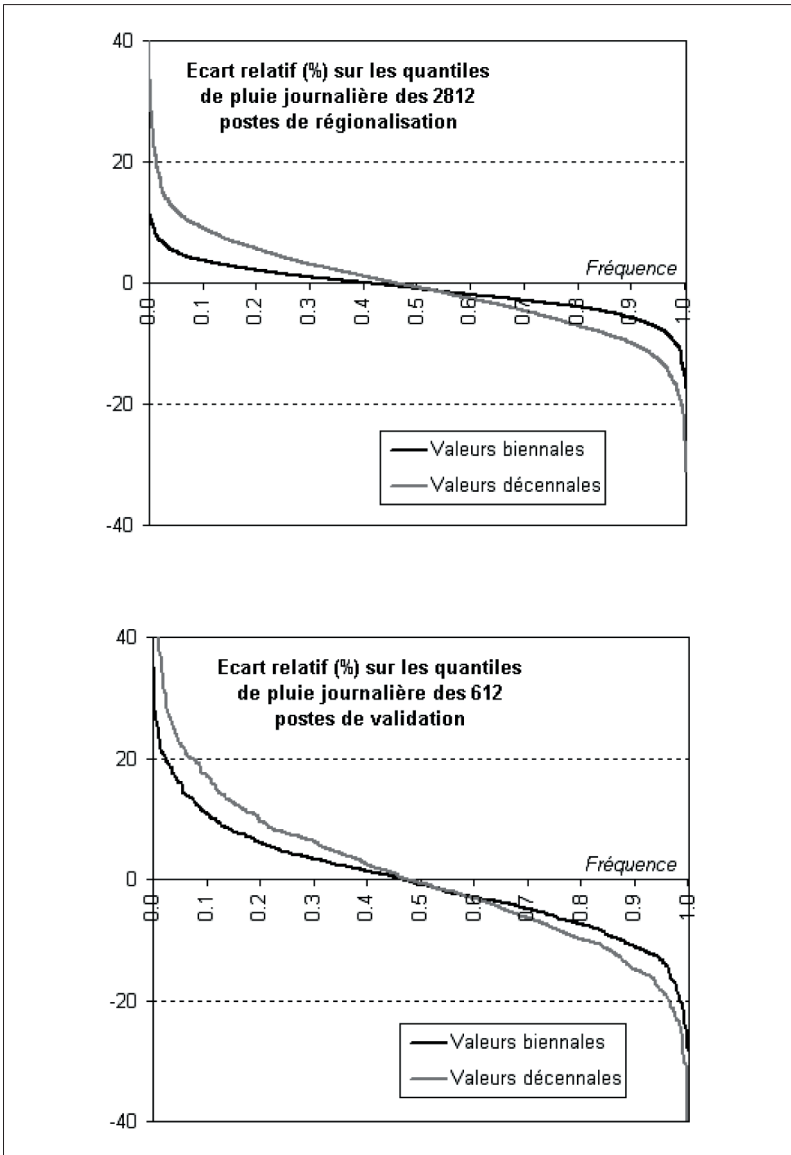
Pour les pluies journalières décennales l'écart est un peu plus grand puisque 90 % des postes de validation présentent un écart relatif compris entre +/- 20 %. Cependant, pour les quantiles décennaux cet écart plus important traduit surtout la divergence des deux approches (SHYPRE et ajustement) pour les périodes de retour élevées.



**Figure 66.** restitution des pluies journalières biennales et décennales aux 612 postes pluviométriques de contrôle.

## Comparaison avec une interpolation directe des quantiles

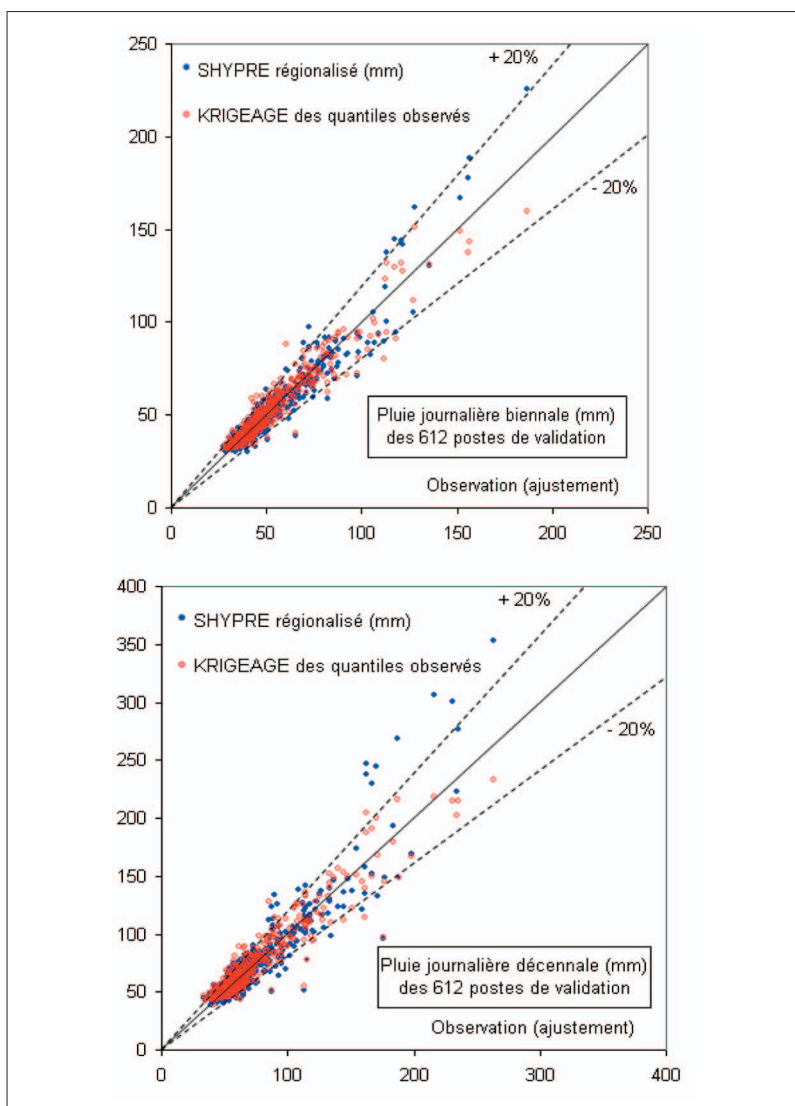
Afin de discuter de la méthode proposée, nous avons voulu regarder ce que nous aurait donné une cartographie directe des quantiles de pluies journalières. Nous avons donc interpolé, par krigeage ordinaire, les quantiles de pluies issus des ajustements de la loi exponentielle sur les 2 812 postes de



**Figure 67.** Distribution des écarts relatifs concernant la restitution des quantiles de pluies journalières, sur les 2 812 postes ayant servi à la régionalisation et sur les 612 postes de validation.

régionalisation. Les cartes ainsi obtenues nous permettent de trouver les quantiles de pluies aux 612 postes de validation grâce au krigeage direct des quantiles « observés ». Ces quantiles sont alors comparés aux quantiles observés localement sur ces 612 postes de validation.

Les graphiques de la figure 68 comparent les performances des deux approches régionales possibles : régionalisation de l'approche SHYPRE et cartographie directe des quantiles. En termes de coefficient d'efficacité (critère de NASH) on trouve les valeurs suivantes :



**Figure 68.** Quantiles de pluies journalières, obtenus par SHYPRE régionalisé et directement par une cartographie des quantiles ajustés par une loi exponentielle : comparaison aux quantiles obtenus localement par l’ajustement d’une loi exponentielle.

- pour l’approche SHYPRE régionalisé : Nash = 91,2 % pour les pluies journalières biennales et 81,8 % pour les pluies journalières décennales ;
- pour l’approche par Krigeage direct des quantiles : Nash = 92,2 % pour les pluies journalières biennales et 89,4 % pour les pluies journalières décennales.

Ces résultats auraient tendance à attribuer de meilleures performances au krigeage direct des quantiles de pluies. Cependant, ces performances sont établies dans la capacité des deux approches à restituer les quantiles

obtenus localement par l'ajustement d'une loi exponentielle. Les écarts alors observés pour les quantiles obtenus par la méthode SHYPRE sont alors aussi dus à la divergence des deux approches, en particulier pour le quantile décennal.

On voit bien par exemple que pour les postes les plus pluvieux, les pluies journalières décennales obtenues par la méthode SHYPRE régionalisée sont supérieures aux quantiles obtenus localement. Cependant cet écart positif est plus lié à la tendance sur-exponentielle obtenue par SHYPRE qu'à un problème de régionalisation de l'approche.

Pour être plus objectif, nous avons alors comparé les quantiles obtenus pour les deux approches en version régionalisée, par rapport aux quantiles obtenus par les deux mêmes approches en version « locale ». On obtient alors des graphiques présentés sur la figure 69.

Dans ce cas on trouve :

- pour l'approche **SHYPRE régionalisée** : Nash = 93,3 % pour les pluies journalières biennales et 93,9 % pour les pluies journalières décennales ;
- pour l'approche par Krigeage direct des quantiles, toujours les mêmes valeurs : Nash = 92,2 % pour les pluies journalières biennales et 89,4 % pour les pluies journalières décennales.

#### Avantages de la régionalisation de la méthode

On s'aperçoit, qu'en termes de régionalisation, la méthode SHYPRE est plus stable qu'une cartographie directe des quantiles de pluies issus d'un ajustement statistique. Ce résultat s'explique par deux points :

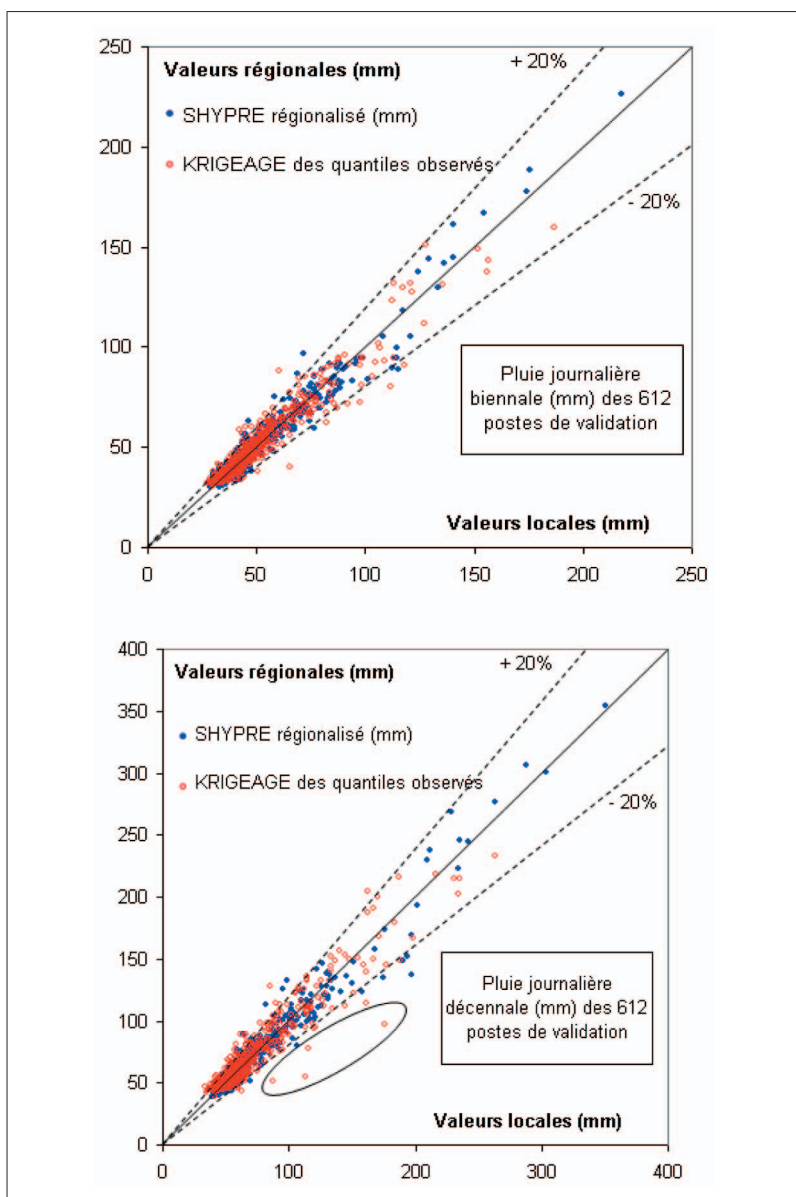
- les paramètres de SHYPRE sont des paramètres moyens plus stables que des quantiles, et dont la variabilité spatiale est plus facilement appréhendable ;
- la cartographie des paramètres de SHYPRE est issue d'une méthode d'interpolation spatiale plus performante qu'un krigeage direct des valeurs observées.

De plus...

Il faut aussi remarquer que bien que les résultats présentés puissent paraître relativement proches, ils ne correspondent pas réellement aux mêmes performances.

En effet, la régionalisation de la méthode SHYPRE ne vise pas à restituer uniquement les quantiles de pluies journalières mais l'ensemble des quantiles de pluies de différentes durées. On obtient une seule cartographie des paramètres pour refaire l'ensemble de pluies de toutes durées et de toutes périodes de retour. C'est donc la même cartographie qui a conduit par exemple au contrôle des quantiles biennaux et décennaux.

En revanche, la régionalisation directe des quantiles de pluies nécessite une cartographie pour chaque durée de pluies et pour chaque période de retour, à moins de travailler avec des lois multi-durées (par exemple, les lois de Montana).



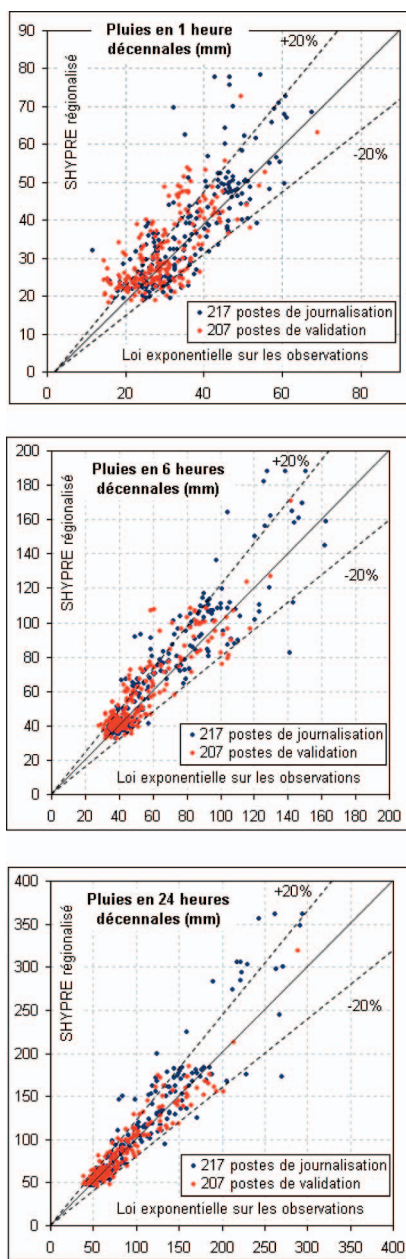
**Figure 69.** Quantiles de pluies journalières, obtenus par SHYPRE régionalisé et directement par une cartographie des quantiles ajustés par une loi exponentielle : comparaison aux quantiles obtenus localement par chaque des méthodes.

## Analyse des pluies horaires

### Restitution des quantiles horaires

Une validation sur les pas de temps fins a été réalisée grâce à l'information de 207 postes horaires supplémentaires de Météo-France, ayant une dizaine d'années d'observations en moyenne.





**Figure 70.** Comparaison des pluies maximales décennales en 1, 6 et 24 heures obtenues par la méthode SHYPRE régionalisée à ceux issus d'un ajustement d'une loi exponentielle sur les distributions des valeurs observées.

Les graphiques de la figure 70, comparent les quantiles décennaux des pluies maximales en 1, 6 et 24 heures, issus de la régionalisation du simulateur à ceux obtenus par ajustement d'une loi exponentielle sur les distributions des valeurs observées. Les points correspondant aux 217 postes ayant servi à la

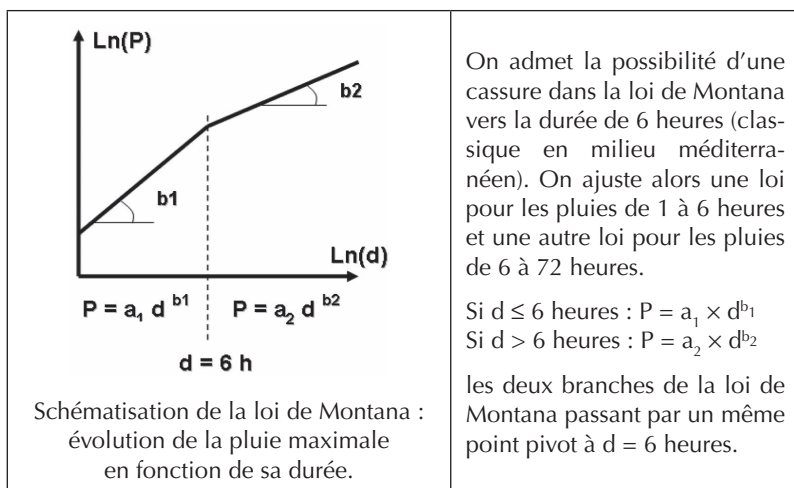
« journalisation », correspondent à une validation de la régionalisation, ces postes n'ayant pas participé à la régionalisation des paramètres journaliers, mais uniquement à l'élaboration des régressions entre les paramètres horaires et journaliers. Les points correspondant aux 207 postes de contrôle permettent de juger à la fois de la « journalisation » et de la régionalisation.

On observe les bonnes performances de l'approche, avec des écarts relatifs généralement inférieurs à 20 %. Les écarts plus importants pour les pluies en une heure, restent les mêmes qu'au cours de la journalisation et proviennent d'une part de la difficulté à régionaliser la variable RX et d'autre part de la sensibilité des pluies en une heure à l'échantillonnage. La similitude des écarts pour les deux échantillons de postes horaires montre cependant la pertinence des relations établies lors de la journalisation, n'entraînant pas d'erreurs supplémentaires sur des postes qui n'y ont pas participé. Notons que les résultats présentés ici intègrent toujours l'ensemble des incertitudes liées à la régionalisation de la méthode (divergence des deux approches vers les fréquences rares, écarts liés à la « journalisation » et à la régionalisation des variables).

## Diversité des régimes pluviographiques

Le générateur de pluie développé permet de générer des chroniques de pluies dans un contexte climatique très varié, sur l'ensemble de la France. La régionalisation du générateur étant basée sur la cartographie de variables journalières moyenne, on peut s'interroger d'une part sur la pertinence de l'utilisation de valeurs moyennes pour « retrouver » les pentes des distributions de pluies, et d'autre part sur la pertinence de l'utilisation de valeurs journalières pour « retrouver » la diversité des pluies à pas de temps fins.

Afin de vérifier la cohérence des résultats obtenus, ainsi que la diversité des régimes pluviographiques modélisés, on a cartographié les valeurs des coefficients de Montana calculés sur les courbes IDF résultant de la modélisation des hyétogrammes horaires.



Les paramètres de la loi de Montana sont calculés pour les différentes périodes de retour étudiées. On présente sur les cartes de la figure 71, la variation spatiale des paramètres  $a_1$ ,  $b_1$  de la loi de Montana et du rapport  $b_1/b_2$ , pour la période de retour 10 ans.

Notons que le paramètre  $a_1$  est un paramètre « absolu » qui permet de situer les lois de Montana en valeurs absolues, alors que les paramètres de pente étudiés sont des paramètres plus « relatifs ».

- La variabilité de  $a_1$  est par définition identique à celle de la pluie maximale horaire ( $d = 1$ ).
- Le paramètre  $b_1$  correspond à la première pente de la loi de Montana et représente l'évolution de la pluie avec la durée, pour les faibles pas de temps ( $d < 6h$ ). Plus  $b_1$  est grand, plus la pluie augmente vite avec la durée.
- Le rapport  $b_1/b_2$  permet d'apprécier la cassure de la loi de Montana pour les valeurs supérieures à six heures. Un rapport proche de 1 signifie que l'évolution de la pluie avec la durée est identique pour les pluies courtes et les pluies longues. Un rapport supérieur à 1 signifie par contre que l'évolution des pluies longues est proportionnellement moins forte avec la durée que pour les pluies courtes.

L'analyse des coefficients de la loi de Montana, permet de distinguer quatre types de régimes pluviographiques modélisés et schématisés sur les graphiques de la figure 72.

On retrouve les deux grands types de climat, à partir de la cartographie du paramètre  $a_1$ .

– Le climat de type « tempéré » concerne l'ensemble de la France sauf le quart Sud-Est, avec de faibles valeurs de  $a_1$ . On peut alors distinguer dans cette catégorie, deux régimes pluviométriques :

- « tempéré sec » : avec de faibles valeurs  $b_1$  et  $b_2$  (rapport  $b_1/b_2$  proche de 1), on a une faible évolution de la pluviométrie avec la durée des pluies. Les régions concernées sont les régions du centre de la France, pour lesquelles les pluies de toutes durées sont les plus faibles ;
- « tempéré humide » : avec des valeurs de  $b_1$  plus fortes, ce régime conduit à des pluies de longue durée proportionnellement plus fortes que les pluies courtes. Les régions concernées sont les zones à relief et proches des côtes, où d'ailleurs la variable  $\mu\text{DTOT}$  était forte.

– Le climat de type « méditerranéen » concerne le quart Sud-Est et la Corse, avec des valeurs de  $a_1$  plus fortes. On peut aussi distinguer deux régimes pluviométriques :

- « méditerranéen sec » : les régions concernées sont les régions méditerranéennes de plaine et la vallée du Rhône. Elles présentent des valeurs de  $b_1$  modérées (par rapport à la partie cévenole) et surtout une rupture très marquée dans la loi de Montana vers la durée 6 heures. Ces régions sont caractérisées par des pluies très intenses sur quelques heures. La rupture de pente traduit le fait que l'intensité de la pluviométrie n'est pas persistante et dure généralement moins de 12 heures ;
- « méditerranéen humide » : avec des valeurs maximales pour  $a_1$  et  $b_1$ , les intensités pluvieuses sur quelques heures sont très fortes. Les relativement faibles valeurs de  $b_1/b_2$  montrent que les pluies de longue durée continuent à être intenses, marquant ainsi le phénomène de persistance des intensités au-delà de 6 heures, voire de 24 heures. On arrive alors à des pluies sur un ou deux jours, les plus fortes du territoire. Les régions concernées sont les zones méditerranéennes à relief comme les Cévennes, mais aussi les sommets de la Corse et des Pyrénées-Orientales.

On retrouve aussi ces résultats lorsque l'on cartographie le rapport entre les quantiles des pluies courtes et les quantiles des pluies longues. Les cartes de la

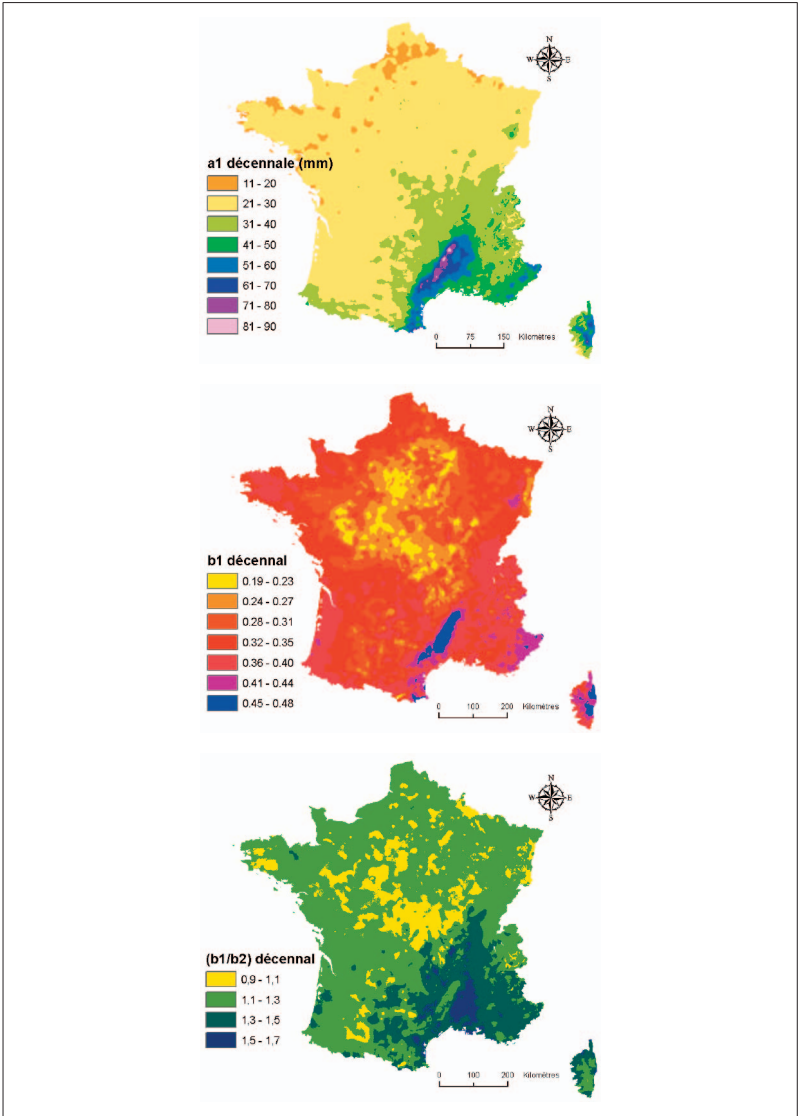


Figure 71. Cartographie des paramètres  $a_1$ ,  $b_1$  et du rapport  $b_1/b_2$  pour la période de retour décennale.

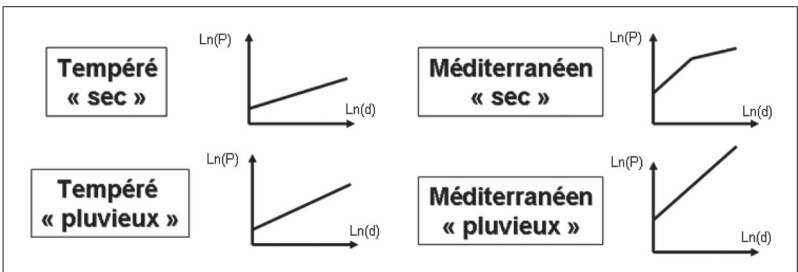


Figure 72. Schématisation des quatre types de régimes pluviographiques modélisés.

figure 73 montrent par exemple la répartition spatiale des rapports des pluies maximales en une heure et en six heures biennales avec les pluies maximales en 24 heures biennales, issue de SHYPRE et issue d'ajustement sur les postes d'observation.

On retrouve l'écart entre les zones du climat « tempéré sec » qui présente les plus fortes valeurs du rapport PM1/PM24, et la zone du climat « méditerranéen humide » présentant les valeurs les plus faibles.

Même si les pluies maximales horaires sont corrélées avec les pluies maximales journalières, la relation qui les lie n'est pas linéaire. Cette relation non linéaire, rapportée par la modélisation, est due à différents facteurs modélisés :

- la variable RXP n'est pas constante (liaison moyenne entre la pluie en une heure et la pluie en quelques heures) ;
- la loi conditionnelle de la variable RXP est accentuée avec le paramètre  $\mu$ NAVP (modélisation des pluies en une heure vers les fréquences rares) ;
- la persistance des averses est plus marquée en milieu pluvieux, conduisant à des pluies de longue durée proportionnellement plus fortes que les pluies de courte durée.

Sur les observations on a la même tendance, sauf qu'il existe une variabilité spatiale beaucoup plus forte. Deux postes voisins peuvent avoir des ratios entre quantiles de différentes durées très variables sans qu'une explication climatique puisse être mise en avant. On peut voir ici un effet de l'échantillonnage sur les observations, qui est moins marqué sur les quantiles issus de la modélisation des pluies.

### **Respect des régimes pluviographiques**

Les résultats présentés ici montrent bien que la paramétrisation du générateur de pluies horaires, bien que basée sur une information strictement journalière, permet de rendre compte de différents régimes pluviographiques. La diversité de ces régimes s'exprime à travers la diversité des lois de Montana déduites des simulations et caractérisant la pluviométrie à différentes échelles de temps, d'une heure à plusieurs jours.

La cohérence des simulations avec les observations est mise en évidence, bien que la variabilité spatiale des observations, liée à l'échantillonnage, rende l'évaluation difficile.

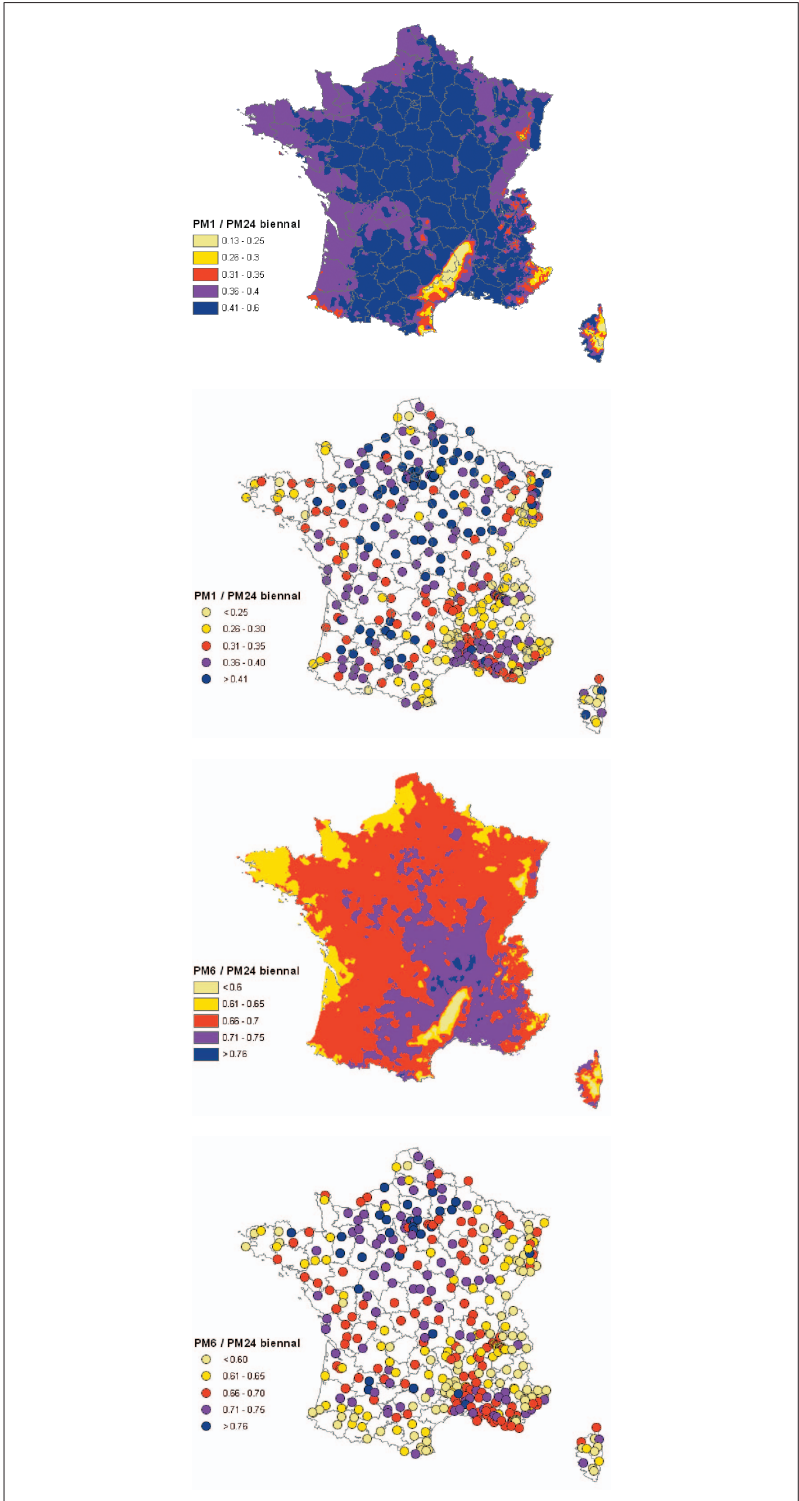


Figure 73. Cartographie des rapports PM1/PM24 et PM6/PM24 biennaux.

# Analyse des résultats.

## Discussion

Nous allons présenter quelques résultats permettant de valider et justifier les choix faits dans l'approche SHYPRE régionalisée.

### Les cartes induites

Le premier intérêt de l'approche est de fournir une information homogène et complète de l'aléa pluviographique sur l'ensemble du territoire français. Cette information est complète puisque l'on dispose de cartes de quantiles de pluies de toutes durées (de 1 à 72 heures) et pour différentes périodes de retour. Rappelons que ces cartes sont déduites de la simulation de chroniques de pluies. C'est donc la cartographie unique des trois paramètres du générateur qui conduit aux cartes de tous les quantiles de pluies. Il n'y a donc pas besoin de refaire une régionalisation pour chaque durée et chaque période de retour, contrairement aux approches basées sur la cartographie de quantiles ponctuels.

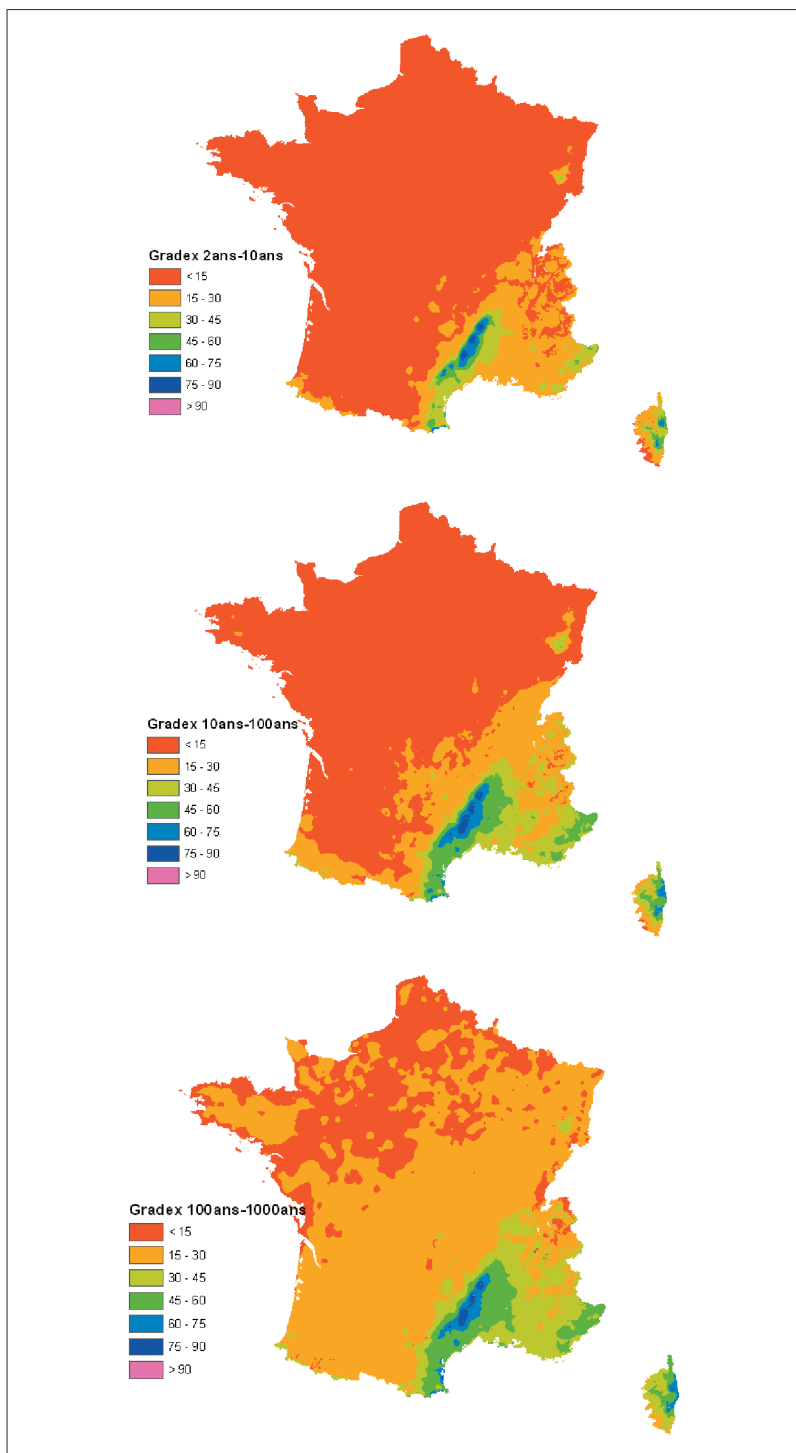
De même, on peut construire d'autres cartes telles que celles des Gradex, des coefficients de Montana, des rapports de quantiles, sans simulation ni interpolation supplémentaire.

Les distributions de fréquences des pluies générées par simulation sont des distributions « empiriques ». Elles ne sont donc pas associées à une loi de probabilité théorique. On peut cependant calculer un « Gradex », correspondant à une approximation de la pente de la distribution empirique entre deux périodes de retour.

On a donc calculé un Gradex sur les pluies journalières à partir des différents quantiles pour voir leur évolution dans l'espace et avec la période de retour. Le Gradex entre deux ans et dix ans est par exemple calculé par  $(\text{Quantile } 10 \text{ ans} - \text{Quantile } 2 \text{ ans}) / \ln(5)$ .

Les cartes de la figure 74 présentent le Gradex de la pluie journalière obtenu par SHYPRE, entre les périodes de retour 2 et 10 ans, 10 et 100 ans, et 100 et 1 000 ans.

On observe la forte variabilité spatiale du Gradex des pluies journalières qui reste inférieur à 15 mm sur les régions tempérées et qui monte à 80 mm dans les régions cévenoles. On observe aussi que le Gradex évolue avec la période de retour. Il tend à augmenter avec la période de retour, en accord avec le comportement plus qu'exponentiel du générateur. Au-delà de la période de retour 100 ans, le Gradex des régions tempérées tend plus fréquemment à dépasser les 15 mm.



**Figure 74.** Cartographie du Gradex des pluies journalières calculé entre différentes périodes de retour : (a) entre 2 et 10 ans, (b) entre 10 et 100 ans et (c) entre 100 et 1 000 ans.



Pour mieux évaluer l'évolution du Gradex avec la période de retour, on a calculé le ratio entre les différents Gradex. Un premier ratio montre l'évolution du Gradex autour de la période de retour 10 ans (ratio [Gradex entre 10 et 100 ans] / [Gradex entre 2 et 10 ans]) et un second montre l'évolution du Gradex autour de la période de retour 100 ans (ratio [Gradex entre 100 et 1 000 ans] / [Gradex entre 10 et 100 ans]).

Ces ratios sont présentés sur les cartes de la figure 75. En couleur dans les tons rouges on trouve une augmentation du Gradex avec la période de retour, et dans les tons verts on trouve une diminution du Gradex avec la période de retour.

On observe les points suivants :

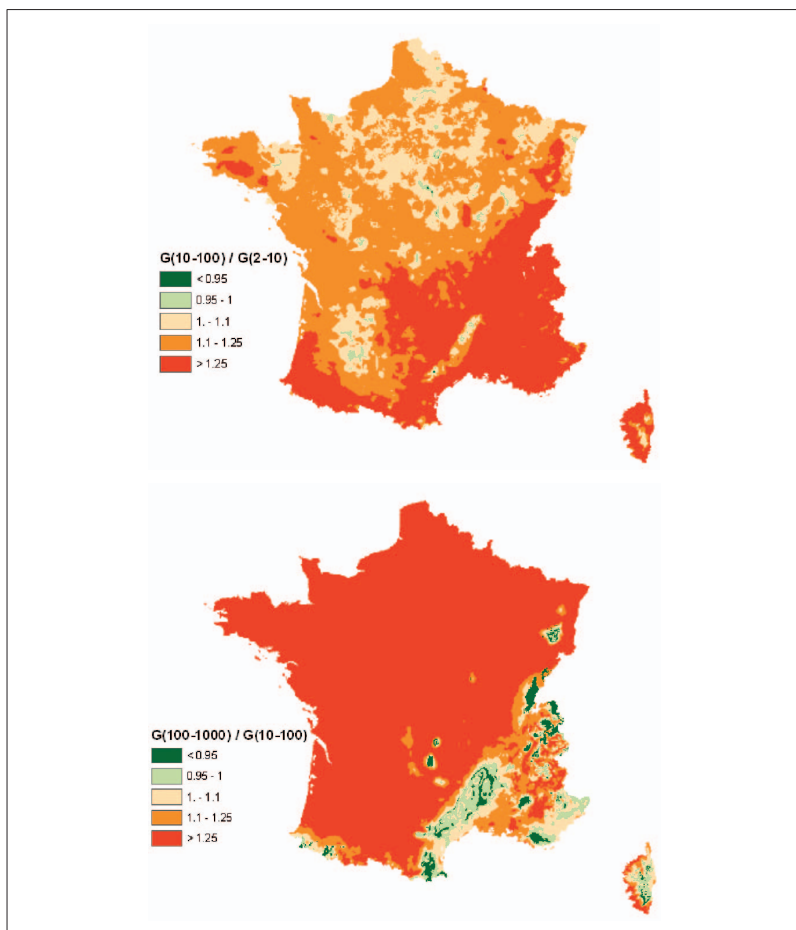
- le Gradex des pluies journalières augmente plus vite dans les régions du Sud de la France autour de la période de retour 10 ans que dans les régions du Nord ;
- à l'inverse, autour de la période de retour 100 ans, le Gradex augmente moins vite dans les régions du Sud, bien qu'en valeur absolue les valeurs soient nettement plus élevées ;
- on observe même une légère décroissance du Gradex des pluies journalières dans les régions à forte pluviométrie.

Ces résultats sont en adéquation avec les résultats présentés p. 46, sur la modélisation de la persistance.

En effet, c'est la modélisation de la persistance des averses qui conduit à un comportement « hyper-exponentiel » du générateur. L'effet de la persistance étant lié au nombre d'averses principales et indirectement à la pluviométrie, il apparaît pour des périodes de retour courantes en milieu très pluvieux (climat méditerranéen et zones d'altitude) conduisant à une forte augmentation du Gradex autour des périodes de retour de 2 à 10 ans. L'effet de la persistance n'apparaît que pour les périodes de retour plus rares en milieu moins pluvieux (climat tempéré) conduisant à une augmentation du Gradex après la période de retour 10 ans. Puis, une fois que la persistance a atteint son maximum, l'évolution des pluies suit une asymptote plutôt exponentielle, due au choix de la loi exponentielle pour modéliser la variable « volume » des averses.

## Robustesse

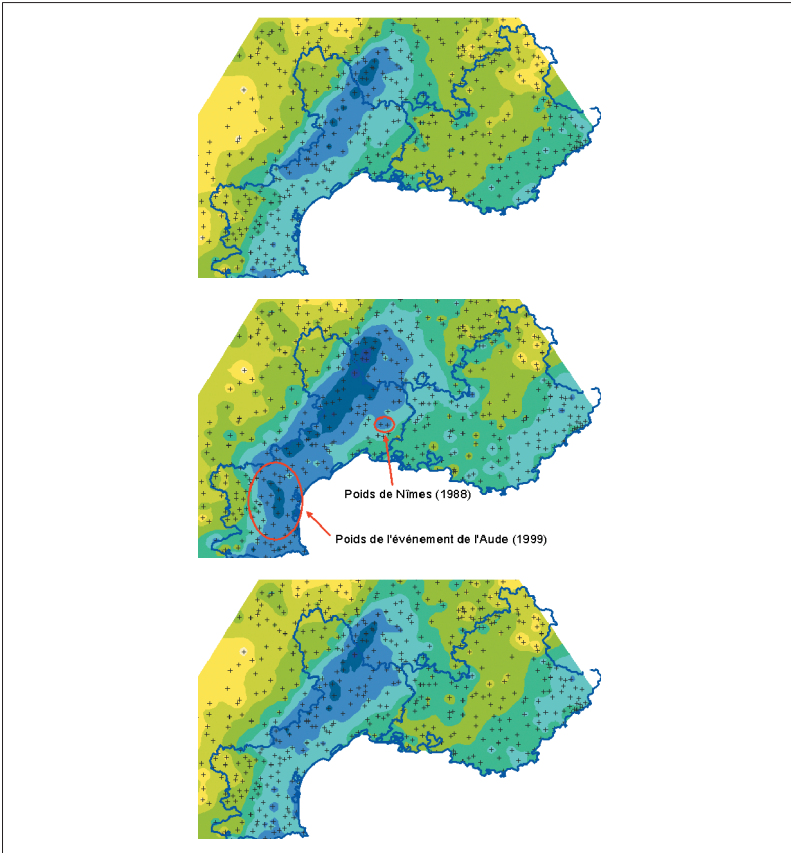
Les performances du modèle « journalisé » sont jugées sur sa capacité à reproduire les quantiles de pluies journalières dits « ajustés localement » par une loi exponentielle. Comme cela a été dit précédemment, il ne faut pas oublier que ces quantiles sont issus de l'ajustement d'une loi théorique sur la distribution de fréquence des valeurs observées. Bien que ces quantiles servent de « base » pour valider les performances de la modélisation, ils n'en restent pas moins, eux-mêmes, soumis à des choix de « modélisation ». En effet, comme il a été montré en introduction, les valeurs que peuvent prendre les quantiles issus d'un ajustement statistique dépendent à la fois du mode d'échantillonnage des valeurs maximales, de la loi de probabilité théorique choisie et de la méthode utilisée pour la détermination de leurs paramètres.



**Figure 75.** Évolution des Gradex avec la période de retour : (a) rapport du Gradex calculé entre 10 et 100 ans par le Gradex calculé entre 2 et 10 ans, (b) rapport du Gradex calculé entre 100 et 1 000 ans par le Gradex calculé entre 10 et 100 ans.

Pour illustrer ce problème, on a comparé les quantiles de pluies journalières des postes des régions Languedoc-Roussillon et Provence-Alpes-Côte d'Azur, issus de différentes études :

- étude « 3 000 postes » (Météo-France, 1999) : les quantiles « ajustés localement » de pluies journalières ont été estimés par la méthode du renouvellement utilisant la loi exponentielle, à partir d'observations couvrant la période 1961-1998 ;
- étude « LR-PACA » (Fouchier *et al.*, 2002) : les quantiles de pluies journalières « ajustés localement » ont été estimés par la loi de Gumbel par la méthode des moments, à partir d'observations couvrant la période 1980-1999, échantillonnées suivant une valeur maximale annuelle ;
- étude « Guide SHYPRE »(2008) : les quantiles « ajustés localement » de pluies journalières ont été estimés par la loi exponentielle, par la méthode des moments, à partir d'observations couvrant la période 1978-2002, échantillonnées en prenant les N plus fortes valeurs observées (N étant le nombre d'années d'observations).



**Figure 76.** Cartographie des pluies journalières décennales « observées », sur les régions LR et PACA, à partir de différentes études.

Les cartes de la figure 76, présentent les quantiles « ajustés localement » de pluies journalières décennales (c'est-à-dire obtenues par ajustement statistique d'une loi de probabilité sur la distribution des valeurs observées), sur les régions LR et PACA, issus des trois études précitées. Ces cartes ont été établies à partir des 535 postes pluviométriques communs aux trois études.

### Étude « 3 000 postes (1999) », étude « LR-PACA » (2002), étude « Guide SHYPRE (2008) »

Ces cartes illustrent l'instabilité qui peut exister dans l'estimation des quantiles de pluies par l'ajustement statistique d'une loi de probabilité théorique. La fluctuation des valeurs obtenues dépend de la méthode d'ajustement et de l'échantillonnage des valeurs.

Les quantiles obtenus dans l'étude « guide SHYPRE » et dans l'étude de « 3 000 postes » restent cependant assez proches. Cela est dû à la méthode d'ajustement choisie. Dans les deux cas, on a utilisé des méthodes « robustes » face aux problèmes d'échantillonnage. En revanche, l'estimation des quantiles de pluies par la loi de Gumbel par la méthode des moments, conduit à une

estimation des quantiles décennaux plus fortement soumise aux problèmes d'échantillonnage. On voit par exemple apparaître sur la carte, l'influence des événements extrêmes conduisant à une augmentation importante de l'estimation des quantiles : comme la présence de l'événement de Nîmes (1988) ou de l'Aude (1999) dans la chronique des pluies observées. Notons que ces événements sont aussi pris en compte dans l'étude « guide SHYPRE », mais que leur présence influence moins la méthode d'ajustement choisie.

Les graphiques de la figure 77 présentent les mêmes résultats mais, sur l'ensemble des postes communs, soit 2 536 postes communs entre l'étude « guide SHYPRE » et l'étude « 3 000 postes » et 535 postes communs entre l'étude « actuelle » et l'étude « LR-PACA ».

On s'aperçoit que les quantiles déterminés dans l'étude « guide SHYPRE » sont un compromis avec les quantiles issus des deux autres études : l'ajustement d'une loi de Gumbel par la méthode des moments conduisant à une surestimation de certains quantiles lors de la présence de valeurs extrêmes dans l'échantillonnage, et la méthode du renouvellement et la loi exponentielle conduisant parfois à une estimation faible des quantiles rares, à cause du poids trop important donné aux valeurs courantes.

### Quelle est la vraie référence ?

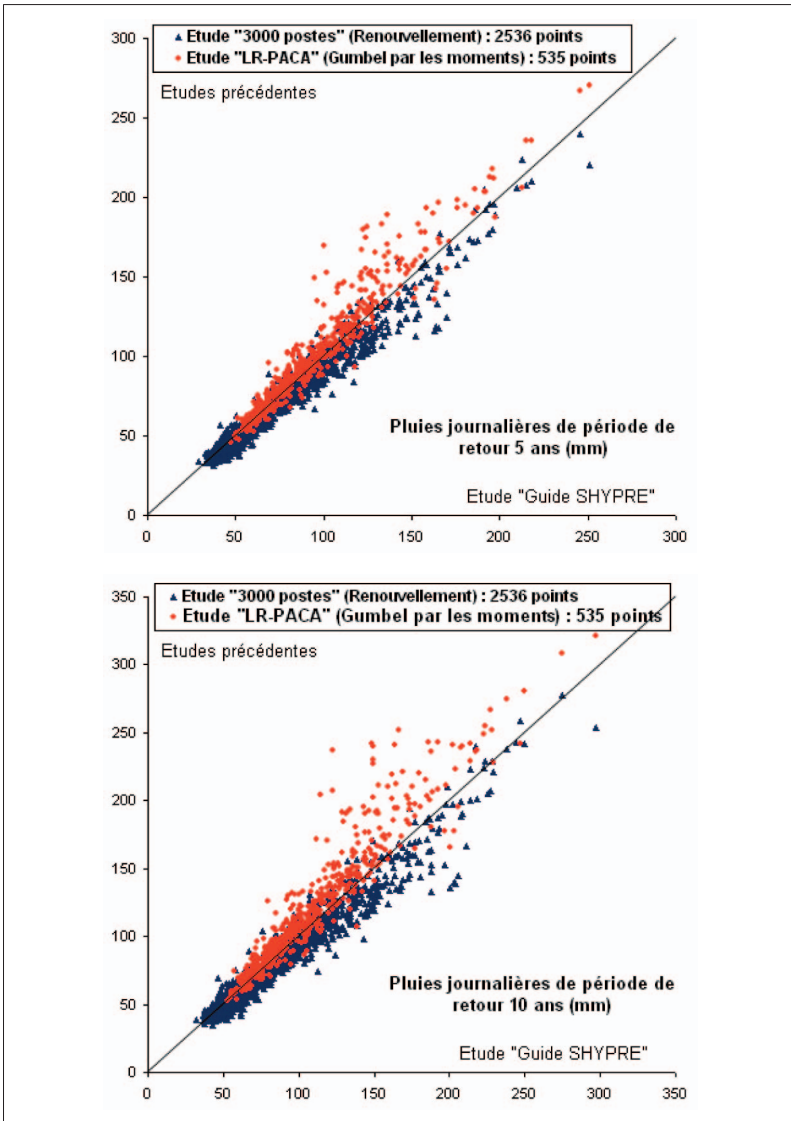
Ce paragraphe permet surtout de relativiser les résultats de la modélisation présentés jusque là. En effet, les quantiles dits « ajustés localement » ne sont que le résultat d'un ajustement d'une loi théorique, avec tous les problèmes d'échantillonnage et de choix des méthodes d'ajustement sous-jacents. Il faut donc garder à l'esprit que la comparaison des quantiles « observés » avec ceux issus de la simulation SHYPRE correspond à des écarts provenant de la différence des deux approches, et ne traduit pas nécessairement un écart entre des valeurs dites « simulées » à d'autres valeurs dites « observées ». Cette différence est peu sensible pour la période de retour deux ans, où la plupart des ajustements sont équivalents. Elle devient plus sensible pour la période de retour 10 ans.

L'instabilité des quantiles obtenus par ajustement d'une loi de probabilité théorique est surtout importante vers les fréquences rares. Elle est aussi généralement liée à la présence (ou l'absence) d'une valeur extrême dans l'échantillon des valeurs observées.

L'une des caractéristiques intéressantes de l'approche par simulation, et particulièrement la version « journalisée », est d'être peu sensible à l'échantillonnage des valeurs extrêmes. Cette « robustesse » est liée à sa paramétrisation essentiellement basée sur le calcul de valeurs moyennes.

Afin d'illustrer ce propos, on peut regarder par exemple l'évolution des paramètres journaliers du modèle, calculés sur les 535 postes communs aux deux études : période 1978-2002 (guide SHYPRE) et période 1980-1999 (étude LR-PACA) (figure 78). On s'aperçoit que ces paramètres restent stables d'une étude à l'autre, car ils correspondent au calcul d'une valeur moyenne sur généralement plus de 60 valeurs. La variation observée vient de la différence des périodes d'observation entre les deux études.

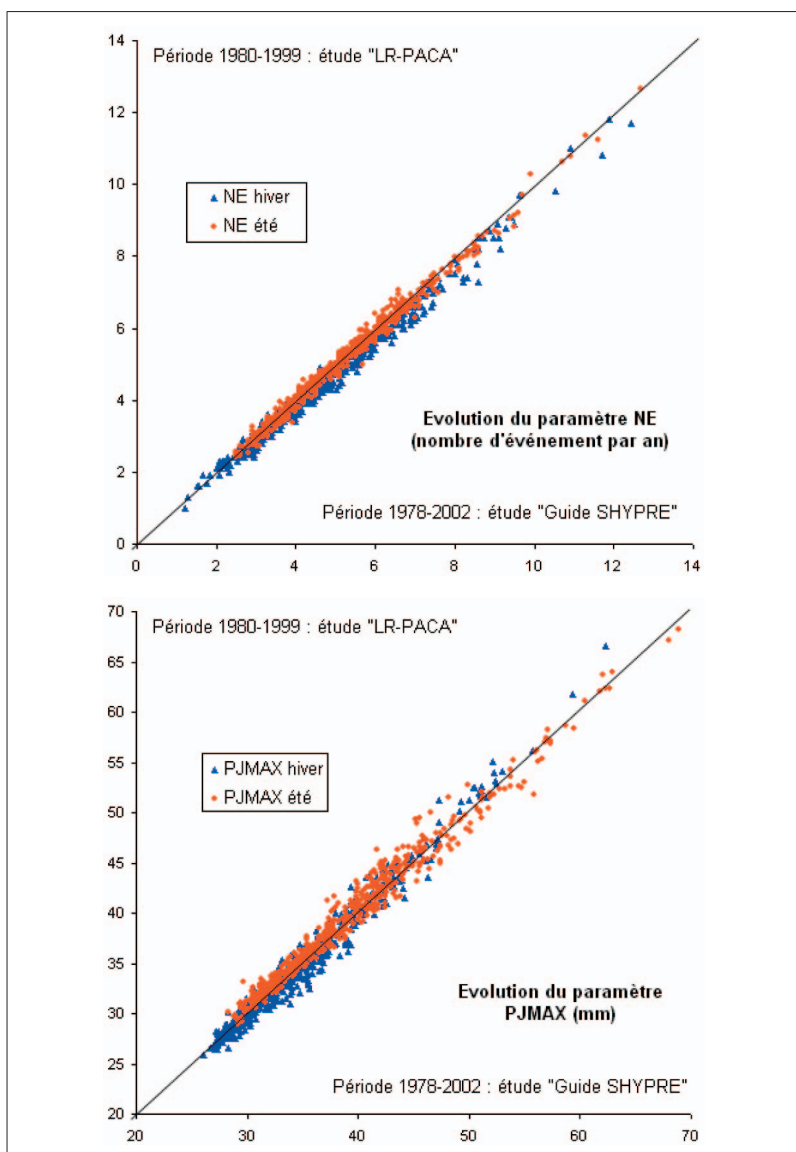
Les paramètres journaliers locaux sont donc beaucoup plus stables que les quantiles de pluies que l'on peut obtenir pour différentes périodes



**Figure 77.** Comparaison des quantiles de pluies journalières quinquennaux et décennaux, issus d'un ajustement statistique, sur l'ensemble des postes communs aux différentes études.

d'observation, avec le problème d'estimation que l'on connaît. **La stabilité des paramètres du modèle de simulation est le garant de la robustesse de l'approche face à l'apparition de valeurs extrêmes dans l'échantillon des valeurs observées.**

À titre d'exemple, si l'on prend les valeurs du paramètre  $\mu_{PJMAX}$  déterminées pour les deux saisons et pour les 2 812 postes journaliers de l'étude « actuelle », et qu'on le recalcule sans la valeur de la variable PJMAX la plus forte observée, on trouve que dans 99 % des cas le paramètre  $\mu_{PJMAX}$  a



**Figure 78.** Évolution des paramètres journaliers NE et  $\mu$ PJMAX calculés sur les 535 postes communs des deux études : périodes 1978-2002 (guide SHYPRE) et 1980-1999 (étude LR-PACA).

diminué de moins de 6 %. La diminution maximale du paramètre  $\mu$ PJMAX est de l'ordre de 12 %, pour trois postes du département de l'Aude et un poste de la Corse où la valeur de la pluie journalière la plus forte a été supérieure à 400 mm.

La robustesse du paramètre journalier  $\mu$ PJMAX est un argument supplémentaire pour préférer baser la régionalisation de la méthode sur celui-ci plutôt que sur le calcul du PLIMS.

## Stabilité spatiale

Si l'on compare les cartes des quantiles de pluies journalières « ajustés localement » avec celles des quantiles de pluies journalières « simulées », on observe certaines différences traduites sur la figure 55. Si les écarts sont faibles pour les quantiles biennaux, traduisant ainsi une bonne modélisation des pluies, les écarts peuvent être plus forts pour les quantiles décennaux, traduisant alors la divergence des approches vers les fréquences plus rares.

Indépendamment des écarts entre les quantiles provenant des deux approches, on regarde dans ce paragraphe la variabilité spatiale des quantiles obtenus par chaque approche prise isolément. On cherche alors à étudier la « stabilité spatiale » des deux approches afin de juger de leur robustesse pour la régionalisation.

Pour cela, on regarde l'écart entre les valeurs des quantiles estimés localement et les valeurs obtenues pour les postes voisins.

À partir de la grille des valeurs de quantiles interpolés sur la France, on crée une grille « lissée » qui calcule la moyenne des valeurs des pixels les plus proches, créant ainsi une carte lissée des quantiles de pluie. On calcule ensuite l'écart relatif entre la carte des valeurs ponctuelles interpolées et la carte lissée. Ces cartes d'écarts relatifs, présentées sur la figure 79 ont été calculés sur les quantiles des pluies journalières décennales : issus de l'ajustement de la loi exponentielle et issus de la simulation par SHYPRE.

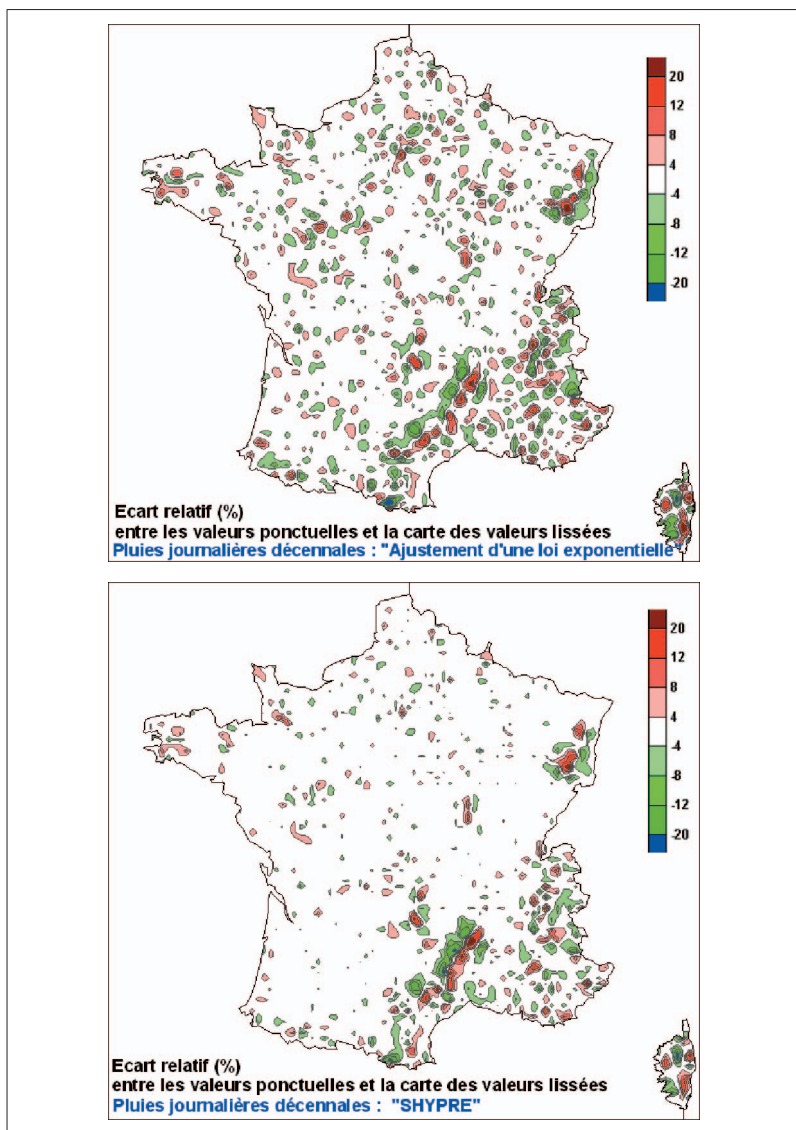
On observe des zones à très forte variabilité à l'approche des reliefs : Cévennes, Corse, Vosges, Alpes et Pyrénées. Ces zones, marquées pour les deux approches, s'expliquent climatiquement avec la présence d'un fort gradient pluviométrique à l'approche des reliefs.

Toutefois, les zones ayant un faible gradient pluviométrique présentent une variabilité pour le quantile décennal, difficilement explicable autrement que par le fait de l'échantillonnage. On observe alors une plus forte variabilité spatiale des quantiles obtenus par l'ajustement statistique, montrant le problème de sensibilité à l'échantillonnage, malgré le choix d'une méthode d'ajustement relativement peu sensible.

Si on prend par exemple le cas de trois postes voisins du département du Loir-et-Cher, département où la variabilité de la pluie journalière décennale est faible, on trouve les valeurs présentées dans le tableau 10. Ces trois postes sont alignés et le poste central d'Herbault est situé à 13 km du poste de Blois et à 26 km du poste de Montoire.

Nom	Altitude (m)	Pluie moyenne annuelle (mm)	$\mu$ PJ/MAX annuel (mm)	Max (PJ) mm	PJ10 (mm) ajustement	PJ100 (mm) ajustement	PJ10 (mm) SHYPRE	PJ100 (mm) SHYPRE
Blois	110	613	28	103	54	81	51	76
Herbault	145	651	25	50	39	51	41	58
Montoire	73	653	28	128	62	91	53	77

**Tableau 10.** Exemple de variabilité des quantiles des pluies sur une région relativement « homogène ».



**Figure 79.** Écarts relatifs entre les valeurs ponctuelles et les valeurs « lissées » des quantiles de pluies journalières décennales : issus de l'ajustement de la loi exponentielle et issus de la simulation par SHYPRE.

Dans cet exemple, les caractéristiques des trois postes sont très proches : altitudes comparables, pluies moyennes annuelles proches, valeurs du paramètre  $\mu_{PJMAX}$  proches. Seule la valeur  $Max(PJ)$  de la plus forte pluie journalière observée est très différente d'un poste à l'autre, et peut expliquer les différences dans l'estimation des quantiles de pluies journalières décennales et centennales. On voit par contre que les quantiles estimés par la méthode SHYPRE présentent moins de variabilité, celle-ci étant fortement liée à la stabilité du paramètre  $\mu_{PJMAX}$ .



## Robustesse spatiale

Liés à la robustesse d'estimation du paramètre  $\mu_{PJMAX}$ , les quantiles estimés par la méthode SHYPRE présentent une variabilité spatiale plus faible sur les zones de pluviométrie homogène. En revanche, les quantiles issus d'un ajustement statistique, même en utilisant une méthode d'ajustement relativement peu sensible à la présence de valeurs extrêmes, présentent une variabilité spatiale plus forte, pouvant conduire à des quantiles très différents entre deux postes voisins et situés en zone de pluviométrie relativement « homogène ».

## Comportement à l'infini

L'examen des quantiles biennaux et décennaux, effectué jusqu'à présent, nous a permis de comparer les résultats de la « journalisation » de SHYPRE par rapport aux résultats d'un ajustement des distributions de pluies observées.

Les quantiles biennaux et décennaux, bien que soumis à des incertitudes pouvant être très fortes (pour le décennal), restent cependant dans la gamme des fréquences observées.

Si l'examen des quantiles biennaux conduit à de très faibles écarts entre les deux approches, on a vu que les quantiles décennaux présentent de plus forts écarts. Ces écarts semblent traduire à la fois la différence qui existe entre les deux approches, mais aussi semblent liés aux problèmes de l'échantillonnage des fortes valeurs.

Nous allons poursuivre la comparaison des quantiles obtenus par SHYPRE en regardant les valeurs centennales (figure 80).

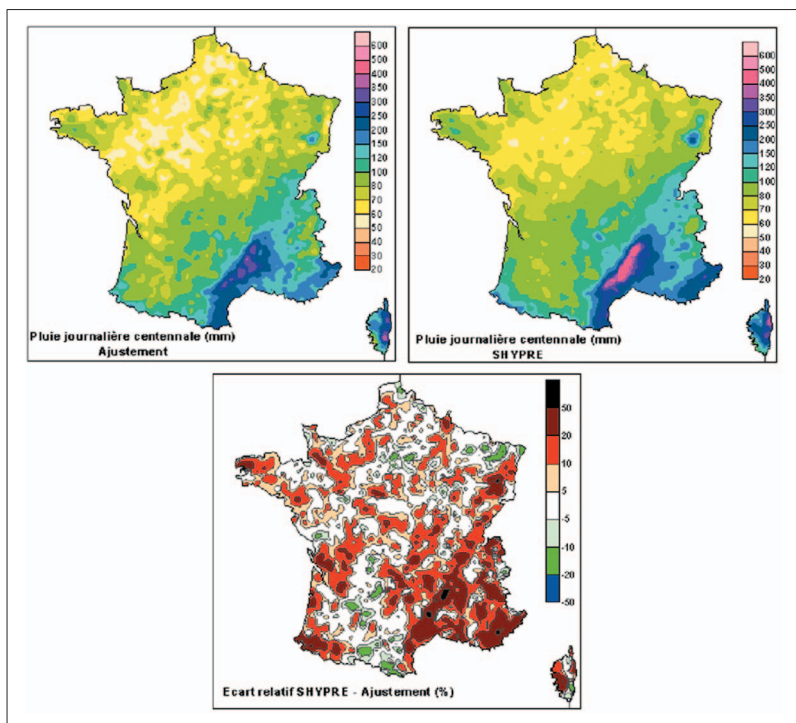
La carte des écarts sur les quantiles de pluies journalières centennales montre des écarts assez fortement positifs surtout sur les régions du Sud de la France. La majeure partie de la zone étudiée conduit à des valeurs centennales issues de SHYPRE, modérément supérieures à celles issues d'un ajustement statistique par une loi exponentielle. Seules quelques zones localisées présentent des quantiles centennaux obtenus par SHYPRE inférieurs à ceux issus de l'ajustement d'une loi exponentielle (écart relatif le plus faible étant de l'ordre de 20 %).

## Comportement en extrapolation vers les fréquences rares

Ces résultats montrent que les simulations faites par SHYPRE conduisent à une distribution des pluies plus qu'exponentielle. La différence entre les quantiles centennaux issus de SHYPRE et ceux obtenus par ajustement d'une loi exponentielle est d'autant plus marquée sur les régions du Sud de la France, où les valeurs de la variable  $\mu_{NAVP}$  sont élevées.

Pour les autres régions, le comportement plus qu'exponentiel n'est pas toujours visible pour les quantiles centennaux.

De façon synthétique, le tableau 11 compare les quantiles issus de l'ajustement par la loi exponentielle et ceux issus de SHYPRE, en comptabilisant le nombre de fois où les premiers sont supérieurs aux seconds et inversement.



**Figure 80.** Cartes des pluies centennales obtenues par un ajustement d'une loi exponentielle et par la méthode SHYPRE, et cartes de leur écart relatif.

Période de retour	Nombre de fois où Quantile SHYPRE > loi exponentielle (sur 2 812 valeurs)	Pourcentage de valeurs où Quantile SHYPRE > loi exponentielle
T = 2 ans	1 479	53 %
T = 10 ans	1 421	51 %
T = 100 ans	2 067	74 %
T = 1 000 ans	2 631	94 %

**Tableau 11.** Effectif des postes où les quantiles issus de l'ajustement de la loi exponentielle sont supérieurs à ceux issus de la simulation des pluies par SHYPRE.

On voit bien que pour les périodes de retour inférieures à 10 ans les deux approches ont le même comportement moyen. En revanche, la divergence apparaît fortement pour les périodes de retour supérieures à 100 ans, les résultats obtenus par SHYPRE étant issus d'un comportement plus qu'exponentiel.

Afin d'évaluer la pertinence du comportement à l'infini obtenu par la modélisation des pluies, nous avons comparé les quantiles obtenus, aux records de pluies journalières observés, sachant que l'on ne dispose que d'une valeur par station. On calcule alors le nombre de fois où ce maximum observé est plus que centennal, et le nombre de fois où ce maximum observé est plus que millennial.

Si l'on émet l'hypothèse d'indépendance temporelle et spatiale de toutes les valeurs de records observés, on peut estimer le nombre de valeurs centennales et le nombre de valeurs millennales que l'on a théoriquement dû observer. Les valeurs observées nous ont été fournies par Météo-France : les 2 812 valeurs de pluies journalières maximales (« records ») des postes pluviométriques utilisés pour la journalisation du modèle, et les 612 valeurs pour les postes de validation. Suivant l'échantillon de poste considéré, ces valeurs sont issues d'un certain nombre d'années-stations. Par exemple, pour l'échantillon des 2 812 postes ayant servi à la régionalisation, cela correspond à l'observation de 69 005 années-stations. On peut alors estimer de façon simpliste qu'il doit y avoir dans cet échantillon la présence de 690 valeurs centennales (fréquence de dépassement de 1 %) et la présence de 69 valeurs millennales (fréquence de dépassement de 0,1 %).

Le tableau 12 présente alors le nombre de « records » supérieurs aux quantiles centennaux et millennaux, pour différents échantillonnages des stations.

Échantillon	T = 100 ans			T = 1 000 ans		
	« Théorique »	Exponentielle	SHYPRE	« Théorique »	Exponentielle	SHYPRE
2 812 postes	690	909	593	69	117	43
612 postes	106	144	102	11	25	9
Postes distants de 25 km (519 postes)	126	151	104	13	20	7
Postes « Climat méditerranéen » (557 postes)	130	210	94	13	48	20
Postes « Climat alpin » (199 postes)	46	49	18	5	7	1
Postes « Climat tempéré » (2 706 postes)	620	794	583	62	87	31

**Tableau 12.** Décompte du nombre de « records » supérieurs aux quantiles centennaux et millennaux, suivant différents échantillonnages.

On retrouve le fait que SHYPRE conduit à des quantiles plus forts que la loi exponentielle. Les résultats restent cependant difficiles à exploiter, même si l'on remarque que SHYPRE aura tendance à minimiser la période de retour des événements extrêmes, là où un ajustement par la loi exponentielle tend à maximiser la période de retour des événements extrêmes.

Les différents échantillonnages réalisés permettent de nuancer les résultats :

- l'échantillonnage de postes distants au minimum de 25 km, minimise le fait que deux valeurs observées puissent être dépendantes. Si l'on avait pu avoir la date de chaque record, on aurait pu plus facilement justifier de l'indépendance des valeurs. Cependant cet échantillonnage ne semble pas influencer les résultats ;

- l'échantillonnage par grandes zones climatiques conduit aussi à des résultats peu variés. Il semble toutefois que la zone « alpine » soit la plus « surestimée » par SHYPRE et que la zone « méditerranéenne » soit la plus « sous-estimée » par l'ajustement exponentiel ;
- l'échantillonnage sur les 612 n'ayant pas servi à la mise en œuvre de la méthode, montre aussi qu'il n'y a pas de biais du à l'échantillon des postes utilisés pour caler et régionaliser l'approche.

Ces résultats, bien qu'intéressants, ne sont que qualitatifs et mériteraient d'être repris dans le cadre d'une validation plus poussée (avec comparaison à d'autres approches) du comportement à l'infini de la méthode. Pour cela une information plus complète sur les observations est nécessaire :

- il faut disposer des dates des valeurs maximales ;
- il faut disposer d'estimations des valeurs « records » non répertoriées dans les chroniques continues de pluies journalières ;
- il faut aussi disposer au moins des deux plus fortes valeurs de la série, pour ne pas sous-estimer les valeurs dépassant, par exemple, le quantile centennal (même si ce cas est statistiquement rare...).

## Intervalle de confiance (Muller, 2006)

L'analyse des incertitudes présente un intérêt évident, bien qu'en pratique elle soit rarement réalisée. Ce paragraphe traitant des incertitudes de la méthode SHYPRE résulte du travail de thèse d'Aurélien Muller (2006).

De façon générale, les incertitudes peuvent être dues aux données ou au modèle : incertitudes sur les données d'entrée du modèle (échantillonnage ou erreurs de mesure), incertitudes sur les paramètres du modèle lors de son calage, incertitudes sur la structure du modèle statistique ou stochastique utilisé. Cependant, peu d'études (voire aucune) traitent de l'analyse des incertitudes en considérant simultanément les différentes sources d'erreurs. Notons au préalable que selon plusieurs études, les mesures de pluies causent de plus grandes incertitudes que les paramètres (Storm *et al.*, 1988 ; Thorsen *et al.*, 2001) ou la structure du modèle (Krzysztofowicz, 1999).

On présente ici, l'analyse de la sensibilité du paramétrage à l'échantillonnage, traduisant la robustesse des estimations par rapport à la taille et au contenu en valeurs extrêmement fortes de la série d'observations.

## Méthode utilisée

Afin d'évaluer les incertitudes fournies par la méthode SHYPRE, on a comparé les résultats de cette approche, à ceux fournis par une approche probabiliste s'appuyant sur la théorie des valeurs extrêmes (ajustement d'une loi de probabilité Pareto généralisée GPD, par la méthode des moments pondérés PWM).

L'analyse des incertitudes est menée dans un cadre Bayésien, en exploitant les données de pluies journalières de la longue série de Marseille (122 ans, 1882-2003), dans le Sud de la France. On considère donc que l'estimation des paramètres d'un modèle est tributaire des données d'entrée, et en particulier de l'incertitude liée à l'échantillonnage de ces données. La distribution

*a posteriori* des paramètres représente la variabilité aléatoire des paramètres étant donné les observations avec lesquelles ils ont été estimés.

– Dans le cas de SHYPRE paramétré par l’information journalière, les lois *a posteriori* marginales des paramètres sont des lois normales, puisque les paramètres journaliers sont estimés par des moyennes d’un grand nombre de valeurs. Sur la série de Marseille, la corrélation estimée entre les paramètres est de 0,22 entre  $\mu$ PJMAX et  $\mu$ DTOT, et inférieure à 5 % entre NE et les deux autres paramètres. On modélise alors la distribution *a posteriori* des paramètres par une loi normale en trois dimensions, avec des lois marginales indépendantes, centrées sur les estimateurs des paramètres et de variances égales aux variances d’échantillonnage des paramètres. On simule ensuite des pluies par SHYPRE, avec différents jeux de paramètres (NE,  $\mu$ PJMAX et  $\mu$ DTOT) tirés dans les lois *a posteriori*. On retient enfin la médiane et les intervalles de confiance des quantiles déduits des chroniques de pluies horaires générées par SHYPRE, et plus particulièrement pour les pluies journalières supérieures à un seuil.

– Dans le cas de la loi GPD, la loi *a posteriori* des paramètres, notée *post*, est calculée dans le cadre Bayésien par la relation :

$$\text{post}(\alpha, k | x_1, \dots, x_n) = \text{prior}(\alpha, k) \text{ vrais}(x_1, \dots, x_n | \alpha, k) / C$$

où :

- $\text{prior}(\alpha, k)$  est la distribution *a priori* des paramètres, elle traduit les connaissances relatives aux paramètres, sans tenir compte des observations ;
- C est une constante ;
- $\text{vrais}(x_1, \dots, x_n | \alpha, k)$  désigne la vraisemblance des observations ( $x_1, \dots, x_n$ ) supérieures au seuil  $u$  sous l’hypothèse de la distribution GPD de paramètres  $\alpha, k$  :

$$\text{vrais}(x_1, \dots, x_n | \alpha, k) = \prod_{i=1}^n g(x_i | \alpha, k)$$

avec

$$g(x | \alpha, k) = [1 - k(x - u) / \alpha]^{1/k - 1} / \alpha \quad \text{si} \quad k \neq 0$$

$$g(x | \alpha, k) = \exp[-(x - u) / \alpha] / \alpha \quad \text{si} \quad k = 0$$

La loi *a priori* utilisée ici est la loi lognormale de paramètres 0 et 100 pour  $\alpha$ , uniforme entre  $-1$  et  $1$  pour  $k$ . Ces distributions *a priori* sont semblables à celles de Coles *et al.*, (2003) pour les paramètres de la GEV. L’algorithme utilisé pour simuler la loi *a posteriori* des paramètres est un algorithme MCMC combinant 1 000 itérations de Gibbs-Metropolis puis 80 000 itérations de Metropolis (Muller *et al.*, 2008b ; Renard *et al.*, 2006).

## Application à la série de 122 ans de Marseille

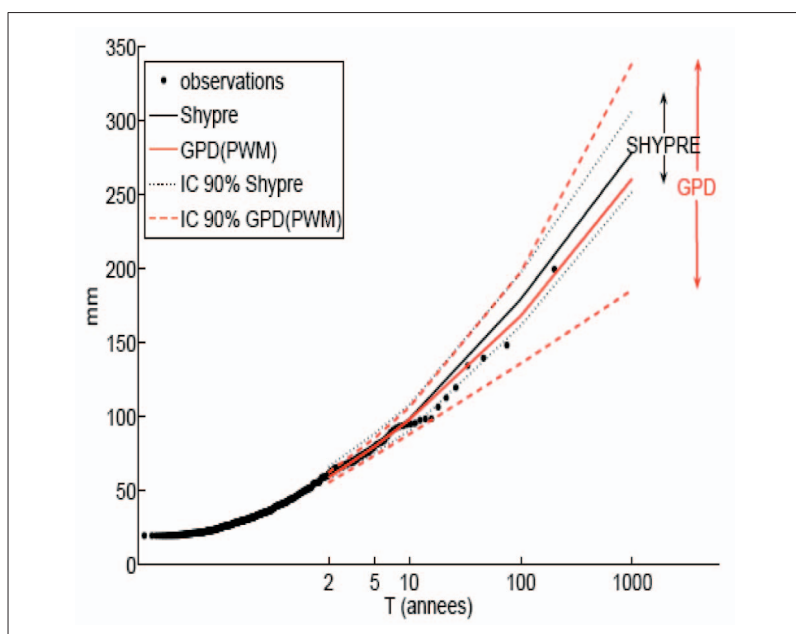
Les intervalles de confiance associés aux estimations des quantiles de pluies journalières ont été calculés pour les deux approches : simulation de chroniques (SHYPRE), ajustement statistique classique (loi GPD).

– Pour SHYPRE, les paramètres sont calculés sur 122 années de mesures, ce qui correspond à 444 événements pluvieux sur la saison considérée (de juin à novembre). Les variances marginales des paramètres sont donc faibles, étant donné le grand nombre d’années et d’événements.

– Pour la loi GPD, les paramètres ont été calculés sur les 444 dépassements du seuil 20 mm, supposés indépendants car la pluie journalière entre deux événements passe sous le seuil des 4 mm.

Les résultats sont présentés figure 81. On constate que les intervalles de confiance des quantiles simulés par SHYPRE sont plus étroits que ceux de la distribution estimée par la loi GPD.

Ceci est dû au fait que SHYPRE n'utilise que des moments d'ordre 1 pour l'estimation de ses paramètres ( $NE$ ,  $\mu PJMAX$  et  $\mu DTOT$ ), tandis que la loi GPD utilise les moments d'ordres 2 et 3 dans l'estimation par maximum de vraisemblance des paramètres  $\alpha$  et  $k$ . On peut aussi remarquer que l'intervalle de confiance est plus faible pour SHYPRE bien que l'on ait trois paramètres ( $NE$ ,  $\mu PJMAX$  et  $\mu DTOT$ ) et donc trois degrés de liberté au lieu de deux pour la loi GPD ( $\alpha$  et  $k$ ). Les écarts entre les deux approches seraient un peu plus marqués si le même nombre de paramètres avait été utilisé.



**Figure 81.** Intervalles de confiance à 90 % calculés pour l'estimation des quantiles de pluies journalières, par la méthode SHYPRE et par l'ajustement d'une loi GPD.

## Étude de sous-séries de 20 ans

En général, l'étude du risque pluviométrique ou hydrologique repose sur l'analyse de séries chronologiques d'une taille d'une vingtaine à une trentaine d'années. Les séries peuvent d'ailleurs être encore plus courtes pour les pluies enregistrées à des pas de temps inférieurs à la journée. Par le biais de l'échantillonnage, une chronique de 20 ans peut contenir aucune, une ou plusieurs valeurs extrêmement fortes. On cherche à étudier ici cet aspect de l'échantillonnage, en découpant la série de 122 ans en six sous-séries successives de 20 ans. On compare alors les résultats fournis par les deux modèles. La méthode utilisée est encore la méthode bayésienne présentée dans le paragraphe précédent, mais appliquée cette fois aux sous-séries de 20 ans.

Le tableau 13 présente les statistiques associées à chaque sous-série, ainsi que les valeurs fortes (pluies journalières supérieures à 100 mm).

Période	1882-1901	1902-1921	1922-1941	1942-1961	1962-1981	1982-2001
Série	1	2	3	4	5	6
Ne	3,79 (1,67)	3,74 (1,77)	3,63 (2,03)	3,88 (1,71)	3,58 (1,84)	3,53 (1,43)
Pjx (mm)	36,7 (16,3)	44,2 (22,9)	42,4 (24,2)	39,0 (20,0)	37,5 (20,7)	39,5 (25,9)
Dtot (j)	1,65 (0,85)	1,69 (0,92)	1,78 (0,96)	1,69 (0,94)	1,84 (1,26)	1,74 (1,05)
Valeurs > 100 (mm)	Aucune	148	113 120 135	107	140	200

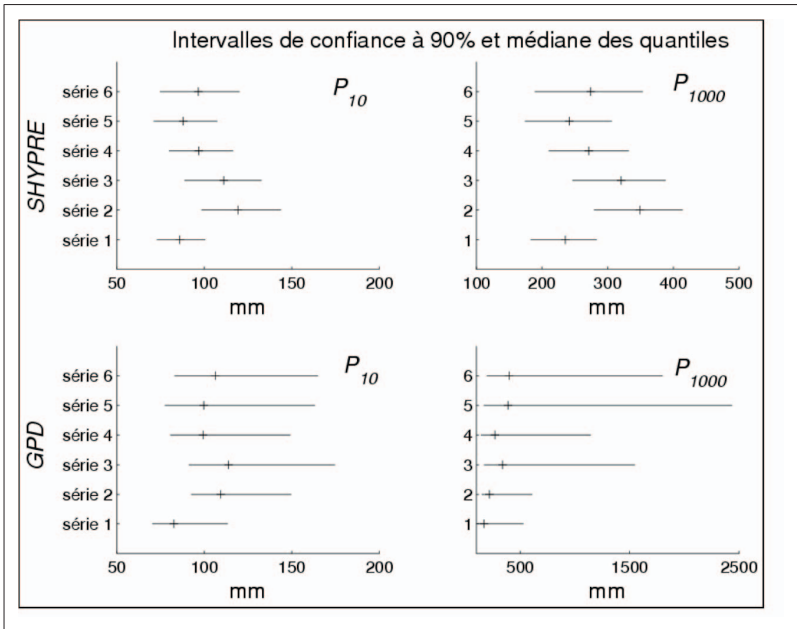
**Tableau 13.** Valeurs de paramètres journaliers caractéristiques de six sous-séries de 20 ans du poste pluviométrique de Marseille.

Si chaque sous-série de 20 ans est équivalente en termes d'occurrence des événements (paramètre Ne), elles peuvent être sensiblement différentes en termes de hauteur d'eau précipitée et en termes d'occurrence de valeurs extrêmes. Les séries 1 et 2 sont à ce titre les deux séries présentant les plus forts écarts de paramètres pour la méthode SHYPRE, alors que ce sont les séries 1 et 3 ou 5 pour la loi GPD.

Sur la figure 82, les distributions *a posteriori* des quantiles simulés par SHYPRE sont d'allure symétrique, tandis qu'on retrouve des distributions à forte dissymétrie pour la loi GPD. La forme de ces distributions marginales dépend d'une part de la distribution des paramètres et d'autre part de la sensibilité des modèles à leurs paramètres. Si les distributions des paramètres sont relativement symétriques pour les deux modèles, la sensibilité aux paramètres n'est pas la même. Les formes des distributions des paramètres et des quantiles simulés par SHYPRE sont semblables : la relation entre paramètres et quantiles simulés semble être linéaire. Les résultats de la loi GPD sont eux par contre, très sensibles aux valeurs du paramètre de forme k. Bien que ce paramètre ait une distribution à faible dissymétrie, la sensibilité de la loi GPD aux valeurs négatives de k conduit à une dissymétrie positive forte pour les quantiles rares.

Les incertitudes des quantiles se révèlent plus fortes avec la GPD qu'avec SHYPRE, au vu de la longueur des intervalles de confiance. La forte incertitude de la GPD sur les séries 5 et 6 conduit à de forts coefficients d'asymétrie (2.4 ; 3.9) et d'aplatissement (10.7 ; 23) sur ces séries (pour une loi exponentielle, l'asymétrie vaut 2 et l'aplatissement vaut 6). On peut noter que les intervalles de confiance donnés par la loi GPD peuvent conduire à des valeurs physiquement irréalistes. En effet, la borne de 2 500 mm d'eau précipité en un jour est inconcevable sur la région d'étude, même pour une valeur millénaire. Les valeurs maximales observées sur la région atteignant difficilement les 500 mm. En France, le record de pluie journalière observé est de l'ordre de 1 000 mm et de 1 800 mm dans le monde (sur l'île de la Réunion).

Les écarts relatifs entre les estimations médianes des quantiles décennaux des six sous-séries sont relativement faibles, et similaires pour les deux approches. Suivant les sous-séries étudiées, on estime les pluies journalières décennales



**Figure 82.** Intervalle de confiance à 90 % des quantiles de pluies journalières décennales et millénales des six sous-séries de 20 ans sur les pluies de Marseille.

entre 80 et 120 mm, avec les deux modèles. Ils donnent donc des estimations équivalentes dans la gamme des valeurs observées.

En revanche, en extrapolation, les résultats des deux modèles divergent. L'estimation des valeurs millénales médianes varient de 100 à 350 mm (rapport de 1 à 3,5) pour la loi GPD et de 220 à 350 mm pour SHYPRE (rapport de 1 à 1,6).

Si l'on regarde les intervalles de confiance, ils deviennent très larges et variables pour la loi GPD. Si l'extrême largeur des intervalles de confiance de la loi GPD vers les fréquences rares garantit leur recouvrement, dans le cas où différentes sous-séries sont utilisées en entrée, ils n'en sont pas plus réalistes et informatifs. En effet, suivant la valeur du paramètre de forme obtenu, on peut avoir des intervalles de confiance conduisant à des valeurs négatives (inconcevables pour la pluie) ou à des valeurs si grandes qu'elles deviennent irréalistes par rapport à la physique du phénomène étudié. Ce cas illustre le problème de la sensibilité de la loi GPD, en particulier dans son utilisation en extrapolation vers les fréquences rares.

### Robustesse face à l'échantillonnage

Ces résultats traduisent la robustesse de l'approche SHYPRE face à l'échantillonnage. Ce point est directement lié à la plus forte stabilité des paramètres de SHYPRE, estimés par des valeurs moyennes, par rapport aux paramètres de la loi GPD estimés par des moments d'ordre 2 et 3. SHYPRE est donc moins biaisé et moins sensible à l'échantillonnage que la loi GPD calée localement.



## Comparaison à une loi GPD dont le paramètre de forme est fixé

L'approche SHYPRE « journalisée » peut être comparée à une approche régionale, puisqu'elle utilise l'information de plusieurs postes pour modéliser certaines de ses composantes, comme la persistance des averses. Certaines caractéristiques sont donc calculées de façon locale et d'autres de façon régionale. C'est une raison pour laquelle SHYPRE présente des intervalles de confiance plus étroits qu'une loi de probabilité ajustée localement.

Pour comparer les intervalles de confiance de SHYPRE avec une loi GPD régionale, Muller (2008a) a refait les calculs présentés dans les paragraphes précédents, en prenant une valeur constante pour le paramètre de forme de la loi GPD.

En prenant un paramètre de forme constant, dit « régional », égal à  $-0,119$  pour la série de Marseille, on trouve les intervalles de confiance présentés sur les graphiques de la figure 83.

Dans ce cas on voit que les intervalles de confiance de la loi GPD deviennent nettement plus réduits que précédemment. Ils deviennent alors proches de ceux de SHYPRE, bien que légèrement plus grands pour les périodes de retour courantes.

Le fait d'avoir fixé le paramètre de forme de la loi GPD, permet de se rapprocher d'une approche régionale. Cependant, même dans une approche régionale, ce paramètre garde un certain intervalle de confiance. Certes, il est bien plus faible que dans une approche locale, mais il n'est pas nul comme le suggère l'application avec une valeur fixée.

Cette application permet d'envisager l'intérêt des approches dites « régionales », qui sont des méthodes moins sensibles à l'échantillonnage des valeurs extrêmes, car elles prennent en compte l'information disponible sur des postes proches et comparables, pour réduire l'effet de l'échantillonnage.

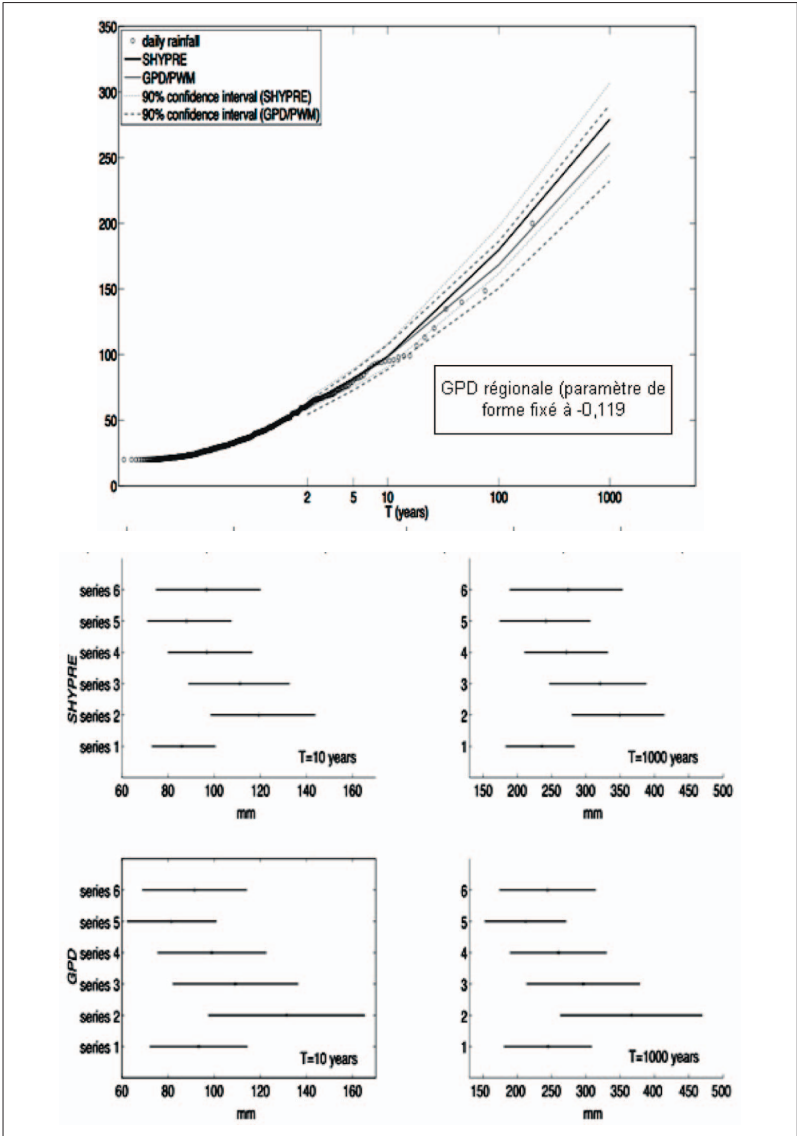
Ce point va être plus largement discuté dans le paragraphe suivant.

## Comparaisons à une approche statistique régionale d'après Neppel *et al.*, 2007

L'objectif de cette partie est de confronter les distributions de pluies journalières générées par SHYPRE (que l'on notera par la suite méthode M1) aux distributions issues d'une approche régionale (notée M2), développée en région méditerranéenne, afin de confirmer ou d'infirmer le comportement hyperexponentiel des distributions, et afin d'apprécier les écarts auxquels peuvent conduire l'application de ces deux approches. Ces résultats sont repris de l'article de Neppel *et al.* (2007).

### Loi régionale des pluies journalières (méthode M2)

L'approche régionale est basée sur la constitution d'un échantillon unique issu d'observations réalisées à différentes stations de mesure. Elle repose sur



**Figure 83.** Intervalle de confiance à 90 % des quantiles de pluies journalières, obtenus par SHYPRE et par l'ajustement d'une loi « GPD régionale » (paramètre de forme fixé).

les hypothèses suivantes : i) l'homogénéité spatiale de la variable étudiée, ii) l'indépendance spatiale des observations et iii) les distributions des variables dans chaque station ne diffèrent que par un facteur d'échelle (Cunnane, 1988 ; Onibon *et al.*, 2004 ; Saint-Hilaire *et al.*, 2003).

Cette approche a été appliquée en région méditerranéenne, sur la zone des Cévennes (figure 84). La période d'observation est 1958-2002 et seuls les postes d'au moins 15 années de mesures sont retenus. La pluie journalière maximale annuelle a ainsi été échantillonnée sur les 165 postes pluviométriques du Gard, de l'Hérault et de la Lozère répondant à ce critère.

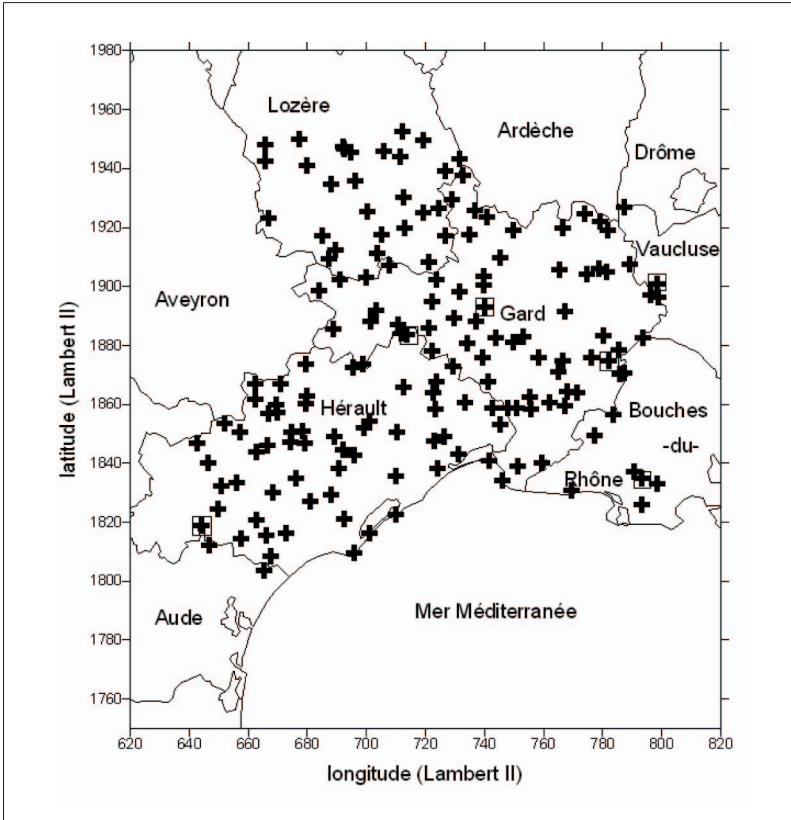


Figure 84. Localisation des 165 postes journaliers utilisés pour la loi régionale.

L'homogénéité spatiale est vérifiée par la méthode présentée par (Hosking, Wallis, 1993). Elle repose sur une mesure de discordance établie à partir des L-moments à partir des coefficients de variation, d'aplatissement et d'asymétrie. Elle conduit à rejeter six postes jugés non-homogènes. L'indépendance spatiale des stations est obtenue en ne retenant que les stations distantes de plus de 80 km lorsque le maximum annuel correspond au même événement sur différents postes. Les maxima annuels à chaque station sont ensuite réduits par la moyenne des pluies journalières maximales annuelles de la station, paramètre noté  $\mu_{PJMoy}$ . Un échantillon régional des 5 120 valeurs réduites est ensuite constitué. Une loi GEV, dont la fonction de répartition est donnée par l'équation suivante,  $y$  est ajustée par la méthode des L-moments (Dominguez *et al.*, 2005).

$$F(x) = \exp \left\{ - \left[ 1 - k \left( \frac{x - u}{\alpha} \right) \right]^{1/k} \right\} \quad \text{avec} \quad \begin{array}{l} \hat{u} = 0,79 \\ \hat{\alpha} = 0,29 \\ \hat{k} = -0,13 \end{array}$$

De cette loi régionale sont donc déductibles des quantiles réduits pour des périodes de retour données, puis les quantiles correspondants pour tout site en multipliant la valeur réduite par le paramètre  $\mu_{PJMoy}$ . Pour une application dite locale, notée « M2 loc », le paramètre  $\mu_{PJMoy}$  est calculé à un poste de mesures. Pour une application dite régionale, notée « M2 rég », le

paramètre  $\mu_{PJMOY}$  résulte d'une interpolation spatiale des valeurs calculées aux différents postes de mesures.

### Convergence vers un comportement hyperexponentiel

On notera le coefficient de forme ( $\hat{k}$ ) négatif qui traduit un comportement hyperexponentiel de la loi régionale (figure 85), concordant ainsi avec les résultats de SHYPRE mais aussi de nombreuses études récentes (Chaouche *et al.*, 2002 ; Coles *et al.*, 2003 ; Koutsoyiannis, 2004).

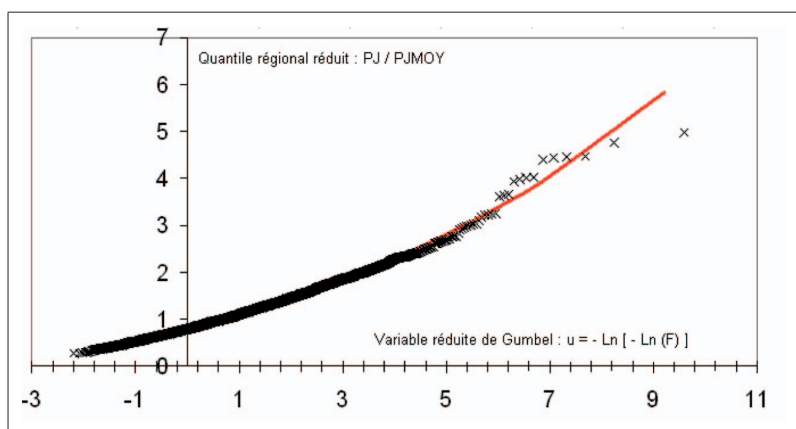


Figure 85. Loi régionale des quantiles réduits de pluie journalière maximale annuelle.

L'interpolation du paramètre  $\mu_{PJMOY}$  est réalisée suivant le principe de la méthode Aurelhy, comme pour les paramètres de SHYPRE. On utilise 109 postes disposant d'au moins 20 ans de mesures. Les caractéristiques géographiques du relief sont les mêmes que celles utilisées pour la régionalisation des paramètres de SHYPRE (cf. § 4.1p.)

Les caractéristiques du relief expliquant le plus la variabilité spatiale de  $\mu_{PJMOY}$  sont estimées par une régression *stepwise*. Les trois premières composantes principales de l'ACP, les distances à l'Aude, aux Cévennes et à la mer Méditerranée sont les variables retenues. Elles expliquent 90 % de la variance spatiale de  $\mu_{PJMOY}$ . Les résidus de cette régression sont krigés. Ainsi en un point quelconque de l'espace, le paramètre  $\mu_{PJMOY}$  est obtenu en sommant l'estimation de  $\mu_{PJMOY}$  par la régression multiple et la valeur du résidu interpolé en ce point.

La comparaison des approches est réalisée avec les versions régionalisée (« rég ») et locale (« loc ») des deux modèles, sur le domaine d'étude commun composé des départements de l'Hérault et du Gard, zone particulièrement soumise à une forte variabilité du risque pluviométrique.

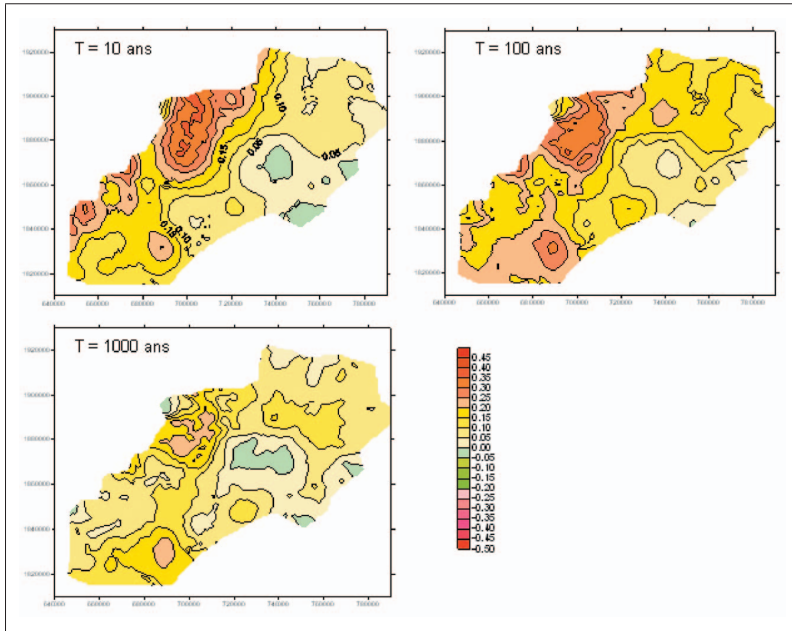
## Comparaison des versions régionalisées

Sur le domaine de comparaison, on calcule sur chaque maille de 1 km<sup>2</sup>, les quantiles décennaux, centennaux et millénaux avec le modèle SHYPRE régionalisé (M1 rég) et la loi régionale régionalisée (M2 rég). Les écarts relatifs notés ER et exprimés en % sont définis par :

$$ER(T) = \frac{(Quantile_{M1r\acute{e}g(T)} - Quantile_{M2r\acute{e}g(T)})}{Quantile_{M2r\acute{e}g(T)}} \times 100$$

où T est la période de retour associée aux quantiles de pluies journalières.

On constate que les écarts relatifs sont, en valeur absolue, toujours inférieurs à 30 % pour 95 % des 8 523 pixels de la zone d'étude (figure 86).



**Figure 86.** Cartographie des écarts relatifs (Quantiles SHYPRE – Quantiles loi régionale) / Quantiles loi régionale, pour T = 10, 100 et 1 000 ans.

C'est sur la zone de l'Aigoual que les écarts relatifs sont les plus importants. Les écarts relatifs sont variables suivant la période de retour, mais tendent à se réduire pour les fortes périodes de retour : pour 80 % de la zone d'étude les quantiles décennaux estimés par les deux méthodes diffèrent de moins de 20 %, 70 % pour les quantiles centennaux et pour seulement 4 % de la zone d'étude les quantiles de période de retour 1 000 ans diffèrent de plus ou moins 20 %. Les écarts les plus importants concernent toujours des zones de relief, avec des quantiles du modèle « M1 rég » supérieurs à ceux du modèle « M2 rég ». On peut noter ici que les écarts relatifs les plus forts sont obtenus pour la période de retour 100 ans. Ce point s'explique par le phénomène de persistance plus ou moins marqué sur la région étudiée, qui influence le comportement hyper-exponentiel de SHYPRE. Suivant l'importance de ce phénomène, le comportement hyper-exponentiel commence à se produire dans la gamme des périodes de retour comprises entre 10 et 100 ans, conduisant à des ratios quantile 100 ans sur quantile 10 ans variables, alors que la loi régionale implique un ratio constant.

## Convergence des approches

Un premier résultat intéressant est que les deux approches conduisent à des estimations très proches dans leur domaine d'extrapolation, alors que c'est dans le domaine des quantiles extrêmes que l'on observe généralement des divergences marquées entre les méthodes.

## Comparaison des versions locales

La comparaison des deux modèles a été effectuée sur 12 postes journaliers « tests » choisis aléatoirement sur la zone d'étude, sur lesquels le calage des méthodes a été réalisé avec l'information locale. Dans ce cas, la loi régionale est appliquée avec une valeur du paramètre  $\mu_{PJMOY}$  calculée à partir des mesures disponibles des pluies journalières. Pour l'approche SHYPRE dont l'objectif premier est l'estimation de quantiles de pluies de courte durée, le calage sur une station donnée nécessite normalement une information horaire. Dans le cas où l'on ne dispose que de l'information journalière, on utilise les relations déterminées pour la régionalisation, qui permettent d'estimer les paramètres horaires avec des variables journalières. Cependant, les résultats vont être altérés suivant la qualité des régressions entre les paramètres horaires et les variables journalières. C'est pourquoi, dans le cadre de cette étude, l'application « locale » de SHYPRE sur des postes journaliers a été réalisée par un calage des paramètres journaliers de façon à respecter le quantile de période de retour 5 ans de la distribution des pluies journalières observées. Dans ce cas, l'interpolation spatiale des paramètres n'intervient plus, et les paramètres des modèles sont calés pour reproduire au mieux les distributions des pluies journalières dans la gamme des fréquences courantes.

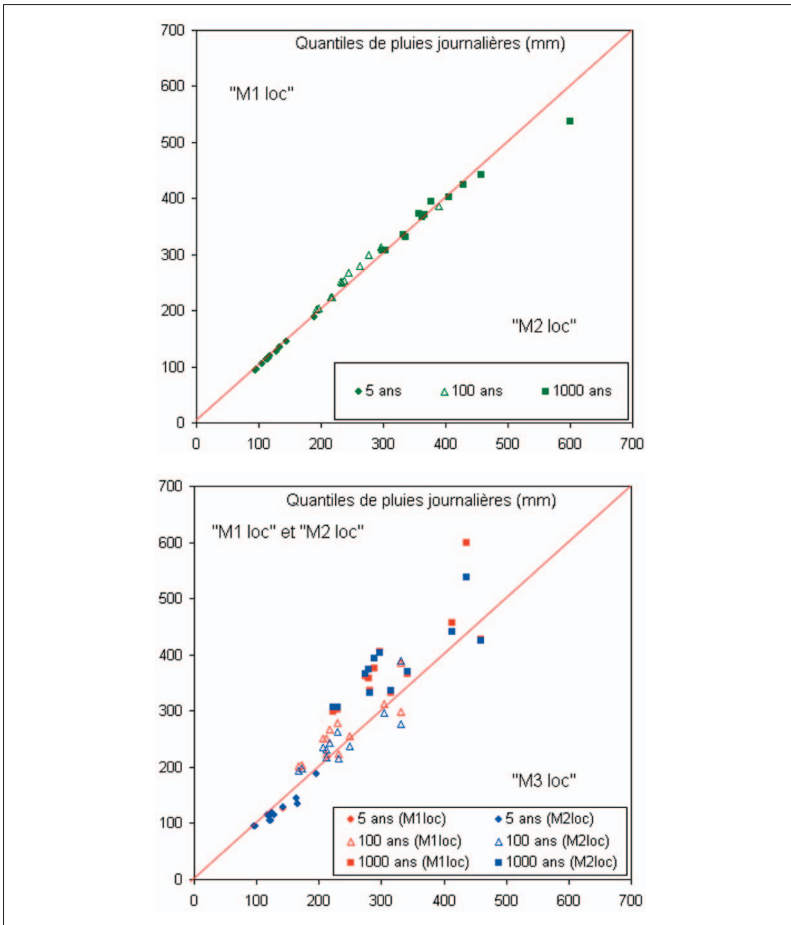
Les résultats des deux approches ont aussi été comparés à un ajustement statistique classique d'une loi exponentielle calée localement, notée modèle « M3 loc ».

Le graphique de gauche de la figure 87 montre que les modèles M1 loc et M2 loc ont un comportement très proche. Calés sur les mêmes quantiles pour la période de retour 5 ans, ils conduisent pratiquement aux mêmes estimations des quantiles de 100 et 1 000 ans. Le graphique de droite de la figure 87 compare les deux modèles M1 loc et M2 loc au modèle plus classique M3 loc. On observe alors un écart croissant des estimations des quantiles extrêmes obtenus par les méthodes M1 et M2 par rapport à ceux de la méthode M3.

*Remarque. Bien que le comportement des modèles M1 et M2 soit qualifié d'hyper-exponentiel, l'ajustement local d'une loi d'exponentielle peut parfois conduire à des quantiles plus forts en présence d'une valeur extrême dans l'échantillon.*

## Discussion

Une première analyse des résultats montre que dans leur version régionalisée les deux méthodes sont très proches en termes d'extrapolation vers les fréquences rares (moins de 20 % d'écart). Ce point est particulièrement intéressant quand on sait que c'est généralement en extrapolation que des méthodes peuvent diverger. Les écarts obtenus pour les quantiles millénaux sont d'un



**Figure 87.** Comparaison des quantiles 5 ans, 100 ans et 1 000 ans obtenus sur 12 postes pluviométriques « tests » de la zone d'étude, par les trois approches dans leur version locale.

ordre de grandeur nettement inférieur aux intervalles de confiance que l'on peut calculer pour cette période de retour. On peut alors raisonnablement penser que les extrapolations auxquelles conduisent les deux méthodes convergent vers les mêmes estimations.

Les écarts observés entre les deux approches sont essentiellement dus à la régionalisation de leurs paramètres respectifs. Les paramètres étant déterminés par des valeurs moyennes, les écarts entre les approches sont directement proportionnels aux écarts entre ces paramètres. Lorsque ces écarts sont faibles, les écarts entre les deux approches restent faibles.

Les résultats obtenus avec les versions locales des deux approches sont particulièrement intéressants. En partant d'approches totalement différentes dans leur principe, on obtient pratiquement les mêmes résultats. Si ce point paraît évident pour les fréquences courantes (on retrouve les quantiles « observés »), il n'était pas nécessairement attendu en extrapolation vers les valeurs extrêmes.

Le point commun de ces deux approches est la prise en compte de l'information régionale des pluies, réduisant ainsi leur sensibilité à l'échantillonnage. Elles conduisent à une estimation des quantiles rares plus robuste et plus cohérente spatialement.

Pour juger de la cohérence spatiale des approches M1 et M2, nous avons contrôlé les résultats sur 16 postes journaliers supplémentaires non utilisés lors de leur développement. Pour cela les méthodes M1, M2 et M3 sont appliquées dans leur version « locale » et dans leur version « régionale ». Par ailleurs, la régionalisation des approches a été réalisée suivant deux méthodes d'interpolation : par l'inverse de la distance (notée *idw*) et par la méthode Aurelhy prenant en compte des paramètres du relief (notée *relief*).

Dans le cas des méthodes M1 et M2 ce sont les paramètres estimés localement sur 115 postes qui sont interpolés sur la zone d'étude. Cela permet d'estimer les quantiles de pluies en tout point de la zone par les deux approches (M1 et M2), notamment sur les 16 postes de comparaison. Dans le cas de la méthode M3 ce sont les quantiles 10, 100 et 1 000 ans obtenus localement par la méthode M3 sur ces 115 postes qui sont interpolés sur la zone d'étude.

Sur les 16 postes non utilisés dans la phase de régionalisation, on calcule un critère de Nash pour chaque méthode M et pour les trois quantiles de pluies journalières (10, 100 et 1 000 ans), P<sub>JXAT</sub>, de façon à juger de la robustesse des méthodes utilisées dans leur version régionalisée :

$$C_M = 1 - \frac{\sum_{i=1}^{16} (P_{JXAT_{i,M \text{ loc}}} - P_{JXAT_{i,M \text{ reg}}})^2}{\sum_{i=1}^{16} (P_{JXAT_{i,M \text{ loc}}} - \overline{P_{JXAT_{i,M \text{ reg}}}})^2} \times 100$$

où  $P_{JXAT_{i,M,loc}}$  est l'estimation locale au poste *i* avec la méthode M, du quantile de pluie maximale annuelle de période de retour T et  $P_{JXAT_{i,M,reg}}$  est son estimation au poste *i* avec la méthode M régionalisée.

Le tableau 14 synthétise les résultats obtenus.

Interpolation des paramètres	Inverse de la distance : « <i>idw</i> »			Régression multiple avec variables descriptives du relief : « <i>relief</i> »		
	10	100	1 000	10	100	1 000
Période de retour (an)	10	100	1 000	10	100	1 000
Shypre (M1)	74 %	69 %	68 %	82 %	82 %	82 %
Loi régionale (M2)	75 %	75 %	75 %	86 %	86 %	86 %
Ajustement loi expo (M3)	57 %	39 %	33 %	49 %	46 %	39 %

**Tableau 14.** Critère de Nash pour chaque méthode comparant une estimation des quantiles P<sub>JXAT</sub> avec la version régionale du modèle M à la version locale de ce même modèle, calculé sur les 16 postes de validation.



On constate que les méthodes SHYPRE (M1) et loi régionale (M2) régionalisées restent robustes quelle que soit la méthode d'interpolation des paramètres qui est utilisée : les critères de Nash varient peu avec la période de retour et sont au moins de l'ordre de 70 % avec une méthode « idw » et de plus de 80 % avec la méthode prenant en compte le relief. En revanche, une régionalisation directe des quantiles ne permet pas une estimation aussi bonne : le critère de Nash est au maximum de 57 %, et décroît nettement avec la période de retour (33 % pour les quantiles millénaux). Ceci traduit la sensibilité de la méthode M3 à l'échantillonnage, contrairement aux méthodes régionales M1 et M2 dont la régionalisation repose sur des paramètres qui sont des moyennes, moins sensibles à l'échantillonnage.

### Avantages des méthodes régionales

On illustre ainsi clairement la supériorité des méthodes basées sur des approches régionales selon deux critères :

- une bien moins grande dispersion des estimations des quantiles extrêmes ;
- une meilleure cohérence spatiale des quantiles estimés sur une région.

## Conclusion

Le générateur de pluies horaires développé dans la méthode SHYPRE permet d'estimer l'aléa pluviographique par une approche originale. Les avantages de cette approche ont été présentés :

- **approche robuste face à l'échantillonnage des valeurs extrêmes**, grâce à l'analyse d'un grand nombre d'événements et une paramétrisation basée sur des valeurs moyennes ;
- **approche prenant en compte différentes durées de pluies**, de 1 heure à plusieurs jours, de façon implicite et cohérente ;
- **approche régionalisée sur l'ensemble du territoire français**, permettant d'estimer le risque pluvial à l'échelle du km<sup>2</sup>, grâce à la prise en compte de la quasi-totalité de l'information pluviométrique disponible ;
- **approche robuste spatialement**, grâce à la prise en compte de l'information régionale de pluies, avec une paramétrisation basée sur des moyennes et à une méthode d'interpolation prenant en compte de caractéristiques géographiques.

Outre ses avantages indéniables, l'originalité de l'approche réside dans la modélisation de chroniques de pluies. L'analyse des chroniques de pluies observées, structurées en succession d'averses et de périodes sèches, conduit à leur modélisation, permettant ainsi de créer des scénarios pluvieux statistiquement équivalents aux observations. La représentativité climatique des variables étudiées (durée des averses, nombre d'averse, volume des averses, durées sèches, etc.) et l'information temporelle générée présente un intérêt supplémentaire à cette approche.

L'utilisation de l'approche pour étudier d'autres problèmes que la caractérisation de l'aléa pluviographique peut alors être explorée.

- **Passage aux débits** : le couplage du générateur de pluie à une modélisation hydrologique permet l'étude de l'aléa hydrologique, avec l'estimation classique de quantiles de crues, mais aussi en gardant toute l'information

temporelle des crues (hydrogrammes complets). L'utilisation des différentes formes de crues permet de s'affranchir de la définition arbitraire d'une crue de projet basée sur une seule caractéristique temporelle, ou sur plusieurs caractéristiques de même fréquence.

– De plus, la régionalisation du modèle de transformation de la pluie permet l'estimation de l'aléa hydrologique en tout point du réseau hydrographique.

– **Passage aux pas de temps fins** : la désagrégation des pluies horaires permettrait de passer à des pas de temps plus fins que l'heure. Ce pas de temps nécessaire pour l'étude des petits bassins versants ou en hydrologie urbaine, permettrait aussi d'appréhender des problèmes liés par exemple à l'érosion.

– **Application au changement climatique** : il est actuellement difficile d'évaluer l'impact du changement climatique sur les phénomènes hydrométéorologiques extrêmes par des méthodes statistiques classiques. Le problème vient du manque de recul dans les chroniques de pluies observées pour évaluer une tendance sur les phénomènes extrêmes et de la difficulté rencontrée par les modèles climatiques globaux pour modéliser les processus météorologiques extrêmes. Une approche originale du changement climatique peut être abordée par la modélisation des processus générateurs des crues, à travers l'utilisation du générateur stochastique de pluies horaires. La détection de l'évolution climatique est faite à travers les paramètres du générateur de pluie, qui sont directement liés à des caractéristiques climatiques moyennes. Les tests de scénarios futurs sont alors possibles pour évaluer l'évolution du risque pluviographique mais aussi hydrologique. La régionalisation du modèle de pluies, avec l'hypothèse du climat actuel, sur le territoire français, permettra d'apprécier l'évolution de la cartographie du risque sous l'hypothèse de changement climatique.

Au-delà de ses qualités intrinsèques, le générateur de pluie permet des applications multiples grâce à l'information temporelle complète qu'il produit, grâce à la mise en fréquence de ses résultats et grâce à sa régionalisation permettant son utilisation en tout point du territoire.

Le développement du générateur de pluie résulte de nombreux travaux de recherche publiés. Son application dans différents contextes climatiques a montré des résultats pertinents et sa régionalisation conduit à l'élaboration d'une base de données complète permettant la détermination opérationnelle de l'aléa pluvial en tout point du territoire français. Ce travail ne conclut pas pour autant les recherches concernant le générateur de pluie. De nouveaux travaux de recherche peuvent être envisagés tels que :

– l'amélioration du formalisme de la persistance ou de certaines dépendances entre variables ;

– l'étude de la sensibilité du générateur à ces hypothèses intrinsèques (définition des variables, notion d'averses, choix des saisons...) ;

– son couplage avec des générateurs de champs spatiaux de précipitations ;

– poursuite de comparaisons avec d'autres approches ;

– etc.

Ce travail ne marque donc pas un point d'arrêt à nos recherches sur le générateur de pluie, mais fait un état des lieux de la méthode ayant conduit à la base de données des quantiles de pluies à pas de temps fins, mise en ligne dans la « climathèque » de Météo-France.

# Références bibliographiques

Acreman, M.C., 1990. A Simple Stochastic Model of Hourly Rainfall for Farnborough, England. *Hydrological Sciences Journal*, 35(2): 119-148.

Arnaud, P., 1997. Modèle de prédétermination de crues basé sur la simulation – Extension de sa zone de validité, paramétrisation horaire par l'information journalière et couplage des deux pas de temps. Thèse de doctorat de l'université Montpellier II, 286 pages.

Arnaud, P., Lang, M., Lavabre, J., 1998. Comparaison des méthodes SHYPRE, AGREGEE et QDF. Rapport d'étude Cemagref : 55 pages + annexes.

Arnaud, P., Lavabre, J., 1999. Using a Stochastic Model for Generating Hourly Hyetographs to Study Extreme Rainfalls. *Hydrological Sciences Journal*, 44(3): 443-445.

Beven, K., 1987. Towards the use of catchment geomorphology in flood frequency predictions. *Earth Surf. Processes Landforms*, 12: 69-82.

Beven, K., Kirkby, M.J., 1979. A physically based variable contributing area model of basin hydrology. *Hydrological Sciences Bulletin*, 24(1): 43-69.

Blazkova, S., Beven, K., 1997. Flood frequency prediction for data limited catchments in the Czech Republic using a stochastic rainfall model and TOPMODEL. *Journal of Hydrology*, 195: 256-278.

Buishand, T.A., 1978. Some remarks on the use of daily rainfall models. *Journal of Hydrology*, 36: 295-308.

Burlando, P., Rosso, R., 1996. Scaling and multiscaling models of depth-duration-frequency curves for storm precipitation. *Journal of Hydrology*, 187(1-2): 45-64.

C.I.G.B., 1992. Choix de la crue de projet. Méthodes actuelles. Bulletin 82 de la Commission Internationale des Grands Barrages: 233.

Cadavid, L., Obeyseker, J.T.B., Shen, H.W., 1991. Flood frequency derivation from kinematic wave. *Journal of Hydraulic Engineering*, 117(4): 489-510.

Cernesson, F., 1993. Modèle simple de prédétermination des crues de fréquences courantes à rares sur petits bassins versants méditerranéens. Thèse de doctorat de l'université Montpellier II, 240 pages.

Cernesson, F., Lavabre, J., Masson, J.-M., 1996. Stochastic model for generating hourly hyetographs. *Atmospheric Research*, 42(1-4): 149-161.

Chauouche, A., Hubert, P., Lang, G., 2002. Graphical characterisation of probability distribution tails. *Stochastic Environmental Research and Risk Assessment*, 16(5): 342-357.

Coles, S., Pericchi, L.R., Sisson, S., 2003. A fully probabilistic approach to extreme rainfall modeling. *Journal of Hydrology*, 273(1-4): 35-50.

Collier, C.G., Hardaker, P.J., 1996. Estimating probable maximum precipitation using a storm model approach. *Journal of Hydrology*, 183(3-4): 277-306.

Cowpertime, P.S.P., 1991. Further developments of the Neyman-Scott clustered point process for modelling rainfall. *Water Resources Research*, 27(7): 1431-1438.

Cunnane, C., 1988. Methods and merits of regional flood frequency analysis. *Journal of Hydrology*, 100(1-4): 269-290.

Dingman, S.L., 1994. *Physical Hydrology*. Prentice Hall. Englewood Cliffs, NJ, 575 pages.

Dominguez, R., Bouvier, C., Neppel, L., Niel, H., 2005. Approche régionale pour l'estimation des distributions ponctuelles des pluies journalières dans le Languedoc-Roussillon (France). *Hydrological Sciences Journal*, 50(1): 17-29.

Eagleson, P.S., 1972. Dynamics of flood frequency. *Water Resources Research*, 8(4): 878-898.

Econopouly, T.W., David, D.R., Woolhiser, D.A., 1990. Parameter transferability for a daily rainfall disaggregation model. *Journal of Hydrology*, 118: 209-228.

Ferrer, J.-P., 1992. Analyse statistique de pluies maximales journalières – Comparaison de différentes méthodes et application au bassin Guadalhorce (Espagne). *Hydrologie continentale*, 7(1): 23-31.

Fine, J.-A., 2003. Synthèse des débits de crue sur l'île de la Réunion. Phase 1 : La pluviométrie. Rapport d'étude Hydris.

Fine, J.-A., Lavabre, J., 2002. Synthèse des débits de crue sur l'île de la Réunion. Phase I : la pluviométrie. Éléments de régionalisation du générateur de pluies. Rapport d'étude Cemagref, 24 pages.

Fisher, R.A., Tippett, L.H.C., 1928. Limiting forms of the frequency distribution of the largest and smallest member of a sample. *Proc. Cambridge Phil. Soc.*, 24: 180-190.

Fouchier, C., Lavabre, J., Sol, B., Desouches, C., 2002. Synthèse des débits de crue sur les régions Provence-Alpes-Côte d'Azur et Languedoc-Roussillon. Phase I : la pluviométrie. Rapport d'étude Cemagref, Convention n° 57/2000 MATE/Météo- France/CETE Méditerranée/Cemagref, 84 pages.

Georgakakos, K.P., Bras, R.L., 1984. A hydrologically useful station precipitation model : 1 – Formulation. *Water Resources Research*, 20(11): 1585-1596.

Glasbey, C.A., Cooper, G., McEchan, M.B., 1995. Disaggregation of daily rainfall by conditional simulation from a point-process model. *Journal of Hydrology*, 165: 1-9.

Gnedenko, B.V., 1943. Sur la distribution limite du terme maximum d'une série aléatoire. *Annals of Mathematics*, 44: 423-453.

Guillot, P., Duband, D., 1967. La méthode du Gradex pour le calcul de la probabilité des crues à partir des pluies. I.A.S.H., Publication, 84.

Hebson, C., Wood, E.F., 1982. A derived flood frequency distribution using Horton order ratios. *Water Resources Research*, 18(5): 1509-1518.

Hershenson, J., Woolhiser, D.A., 1987. Disaggregation of daily rainfall. *Journal of Hydrology*, 95: 299-322.

Hershfield, D.M., 1961. Estimating the Probable Maximum Precipitation. Proc. American Society of Civil Engineers, *Journal Hydraulics Division*, 87: 99-106.

Hosking, J.R.M., wallis, J.R., 1993. Some statistics useful in regional frequency analysis. *Water Resour. Res.*, 29(2): 271-281.

Koutsoyiannis, D., 1999. A probabilistic view of hershfield's method for estimating probable maximum precipitation. *Water Resources Research*, 35(4): 1313-1322.

Koutsoyiannis, D., 2004. Statistics of extremes and estimation of extreme rainfall: II. Empirical investigation of long rainfall records. *Hydrological Sciences Journal*, 49(4): 591-610.

Krzysztofowicz, R., 1999. Bayesian theory of probabilistic forecasting via deterministic hydrologic model. *Water Resources Research*, 35(9): 2739-2750.

Lang, M., Lavabre, J., 2007. Estimation de la crue centennale pour la prévention des risques d'inondations, éditions Quae.

Lebel, T., 1984. Moyenne spatiale de la pluie sur un bassin versant : estimation optimale, génération stochastique et gradex des valeurs extrêmes. Thèse INPG, Grenoble, 350 pages.

LeCam, L., 1961. A stochastic description of precipitation. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, edited by J. Neyman: 165-186.

Météo-France, 1999. Météo-France. Estimation des hauteurs de précipitations d'occurrence rare pour des durées de cumul de 1 à 10 jours sur 3 000 postes français – Méthode du renouvellement. Rapport d'étude Météo-France., Service Central d'Exploitation de la Météorologie Division Climatologie et Bases de Données.

Miquel, J., 1983. Crue : un modèle d'estimation des probabilités des débits de crue. *La houille blanche*, 2: 95-103.

Muller, A., 2006. Analyse du comportement asymptotique de la distribution des pluies extrêmes en France. Thèse de doctorat de l'université Montpellier II, 182 pages.

Muller, A., Arnaud, P., Lang, M., Lavabre, J., 2008a. Uncertainties in extreme rainfall distribution using a stochastic rainfall model. Soumis à *Hydrological Sciences Journal*.

Muller, A., Bacro, J.-N., Lang, M., 2008b. Bayesian comparison of different rainfall depth-duration-frequency relationships. *Stochastic Environmental Research and Risk Assessment*, 22(1): 33-46.

Naulet, R., 2002. Utilisation de l'information des crues historiques pour une meilleure prédétermination du risque d'inondation. Application au bassin de l'Ardèche à Vallon-Pont-d'Arc et St-Martind'Ardèche. Thèse, université Joseph Fourier Grenoble, université du Québec, Cemagref Lyon, Inrs-Ete Québec, (2002), 322p. Thèse, université Joseph Fourier, Grenoble – université du Québec, 322 pages.

Naulet, R. et al., 2005. Flood frequency analysis on the Ardèche river using French documentary sources from the two last centuries. *Journal of Hydrology*, Special Issue "Applications of palaeoflood hydrology and historical data in flood risk analysis" (312): 58-78.

Neppel, L., Arnaud, P., Lavabre, J., 2007. Connaissance régionale des pluies extrêmes : comparaison de deux approches appliquées en milieu méditerranéen. *Comptes-rendus de l'Académie des sciences Paris Sciences de la terre et des planètes*, Soumis.

Neyman, J., Scott, E., 1958. Statistical approach to problems of cosmology. *Journal of Royal Statistical Society, series b (methodological)*, 20(1): 1-43.

Onibon, H. et al., 2004. Analyse fréquentielle régionale des précipitations journalières maximales annuelles au Québec, Canada. *Hydrological Sciences Journal*, 49(4): 717-735.

Onof, C., Chandler, R., kakou, A., Northop, P., 1995. Rainfall modelling using Poisson-cluster process. *Recueil de publications à l'International Conférence in honour of Jacques Bernier, Chap. II, Paris, Unesco.*

Ouarda, T.B.M.J., Rasmussen, P.F., Bobée, B., Bernier, J., 1998. Utilisation de l'information historique en analyse hydrologique fréquentielle. *Revue de Sciences de l'Eau*, 11 (n° spécial) : 41-49.

Pickands, J., 1975. Statistical inference using extreme order statistics. *Ann. Stat.*, 3: 119-131.

Renard, B., Garreta, V., Lang, M., 2006. An application of Bayesian analysis and Markov chain Monte Carlo methods to the estimation of a regional trend in annual maxima. *Water Resources Research*, 42(W12422, doi:10.1029/2005WR004591): 17 pages.

Rodriguez-Iturbe, I., Cox, D.R., Isham, F.R.S.e.V., 1987. Some models for rainfall based on stochastic point processes. *Proc. Royal Society London*, A410: 269-288.

Rodriguez-Iturbe, I., Cox, D.R., Isham, V., 1988. A point process for rainfall : further developments. *Proc. Roy. Soc. London*, A417: 283-298.

Rodriguez-Iturbe, I., Valdes, J.B., 1979. The Geomorphologic structure of Hydrologic Response. *Water Resources Research*, 15(6): 1409-1420.

Rossi, F., Fiorentino, M., Versace, P., 1984. Two component extreme value distribution for flood frequency analysis. *Water Resources Research*, 20(7): 847-856.

Saint-Hilaire, A. et al., 2003. La régionalisation des précipitations: une revue bibliographique des développements récents. *Revue de Sciences de l'Eau*, 16(1): 27-54.

Shen, H.W., Koch, G.J., Obeysekera, J.T.B., 1990. Physically-based flood features and frequencies. *Journal of Hydraulic Engineering*, 116(4): 494-514.

Sighomnou, D., Desbordes, M., 1988. Recherche d'un modèle de pluie de projet adapté aux précipitations de la zone tropicale africaine. *Hydrologie continentale*, 3(2): 131-139.

Sol, B., Desouches, C., 2005. Spatialisation à résolution kilométrique sur la France de paramètres liés aux précipitations. Rapport d'étude Météo-France., Convention Météo-France DPPR n° 03/1735, 41 pages.

Soler, O., 1997. Atlas climatique de la Réunion. Rapport d'étude Météo-France : 77 pages.

Storm, B., Jensen, K., Refsgaard, J.C., 1988. Estimation of catchment rainfall uncertainty and its influence on runoff predictions. *Nordic Hydrology*, 19: 77-88.

Thorsen, M. *et al.*, 2001. Assessment of uncertainty in simulation of nitrate leaching to aquifers at catchment scale. *Journal of Hydrology*, 242(3-4): 210-227.

Tourasse, P., 1981. Analyses spatiales et temporelles des précipitations et utilisation opérationnelle dans un système de prévision de crues. Application aux régions cévenoles. Thèse INPG, Grenoble : 190 pages.

Waymire, E., Gupta, V.K., 1981. The mathematical structure of rainfall representations: a review of the stochastic rainfall models. *Water Resources Research*, 17(5): 1261-1272.

Édition, maquette et couverture : éditions Quæ  
Mise en pages : Desk  
Infographie : J. Delbrayère



Comment estimer les pluies des plus courantes aux plus extrêmes, en tout point d'un territoire ? Comment bien prendre en compte l'aléa pluvial dans l'aménagement du territoire, vis-à-vis des risques d'inondation ?

La méthode SHYPRE (simulation d'hydrogrammes pour la prédétermination des crues) développée par le Cemagref a conduit à mettre au point une base de données de quantiles de pluies de durées allant d'une heure à 72 heures. Ce guide méthodologique vient compléter et expliquer la méthode dont les résultats fournis sur l'ensemble du territoire français sont mis en ligne via la climathèque de Météo-France. Actuellement, c'est l'unique méthode de cartographie du risque pluvial appliquée à l'ensemble du territoire français (DOM compris).

Après une analyse des différentes approches et des problèmes associés à la prédétermination de l'aléa pluviométrique, l'ouvrage détaille les étapes de conception, de calage, de régionalisation du générateur de pluie. Il propose aussi une synthèse des différentes publications scientifiques associées au développement de cette méthode.

Ce guide servira de référence à tous les acteurs opérationnels (services de l'État et bureaux d'études) utilisateurs de la base de données associée à la méthode et distribuée par Météo-France. Il intéressera également les chercheurs concernés par les simulations de pluies extrêmes.

**Patrick Arnaud** est ingénieur-chercheur en hydrologie au Cemagref à Aix-en-Provence. Responsable du groupe de recherche en hydrologie, il a travaillé entre autres sur la méthode de prédétermination des débits de crue, basée sur des approches par simulation.

**Jacques Lavabre** est ingénieur-chercheur en hydrologie au Cemagref à Aix-en-Provence. Expert en hydrologie de surface sur les questions de prévention et prévision des crues et des étiages, il est à l'initiative de l'ensemble des travaux de recherche sur la méthode SHYPRE.

En couverture : débordement du Réal Martin dans le Var © Aubert Yoann, juin 2010.



20 €

e-ISBN : 978-2-7592-0862-3

éditions  
**Quæ**



Éditions Cemagref, Cirad, Ifremer, Inra  
[www.quae.com](http://www.quae.com)

ISSN : 1952-2770  
Réf. : 02202 num