

Estimation de la crue centennale pour les plans de prévention des risques d'inondations

Michel Lang,  Jacques Lavabre,
coordinateurs

Estimation de la crue centennale pour les plans de prévention des risques d'inondations

Michel Lang et Jacques Lavabre, coordinateurs
avec les contributions d'Éric Sauquet et Benjamin Renard

Éditions Quæ
c/o Inra, RD 10, 78026 Versailles Cedex

Collection *Update Sciences & Technologies*

Conceptual Approach to the Study of Snow Avalanches.
Maurice Meunier, Christophe Ancey, Didier Richard,
2005, 262 p.

Qualité de l'eau en milieu rural.
Savoirs et pratiques dans les bassins versants,
Philippe Mérot, coordinateur
2006, 352 p.

Biodiversity and Domestication of Yams in West Africa.
Traditional Practices Leading to *Dioscorea rotundata* Poir,
Alexandre Dansi, Roland Dumont, Philippe Vernier, Jeanne Zoundjihèkpon,
2006, 104 p.

Génétiquement indéterminé.
Le vivant auto-organisé
Sylvie Pouteau, coordinatrice,
2007, 172 p.

L'éthique en friche.
Dominique Vermersch,
2007, 116 p.

Agriculture de précision.
Martine Guérif, Dominique King, coordinateurs,
2007, 292 p.

Territoires et enjeux du développement régional.
Amédée Mollard, Emmanuelle Sauboua, Maud Hirczak, coordinateurs,
2007, 240 p.

© Éditions Quæ, 2007

ISBN : 978-2-7592-0364-2

ISSN :1767-7923

Toute reproduction ou représentation intégrale ou partielle, par quelque procédé que ce soit, notamment sa rediffusion sous forme numérique ou imprimée, ou la création de liens hypertextes pointant vers ladite œuvre, faite sans l'autorisation de l'Éditeur ou de l'auteur, est illicite et constitue une contrefaçon, soumettant son auteur et toutes les personnes responsables aux sanctions pénales et civiles prévues par la loi. Seules sont de plein droit autorisées les reproductions strictement réservées à l'usage du copiste et non destinées à une utilisation collective, ainsi que les courtes citations justifiées par le caractère scientifique ou d'information de l'œuvre dans laquelle elles sont incorporées. Tous droits réservés

Sommaire

Préambule	7
Remerciements	9
Chapitre 1. Prédétermination d'événements de référence de crue ..	11
Cadre réglementaire	11
Maximisation de l'aléa	12
Crue historique de référence.....	13
Crue centennale	14
Cadre probabiliste et difficultés d'interprétation	14
Différentes étapes d'une analyse probabiliste.....	16
Plan du guide	20
Chapitre 2. Ajustement d'une loi de probabilité à un échantillon de débits de pointe de crue	23
Le débit : la variable à étudier	23
La métrologie des débits	23
Les bases de données de débit	27
La critique d'une chronique de débits.....	27
Représentation d'un échantillon	30
L'échantillonnage	30
Les horsains ou valeurs singulières.....	32
Construction d'un graphique de fréquence cumulée	33
Exemples d'ajustement sur la distribution du débit de pointe	35
Le Rhône à Perrache : un cas d'école	36
Non-stationnarité ou fluctuations d'échantillonnage : l'Aude à Belviane.....	40
La Berre à Villeneuve-les-Corbières : un événement exceptionnel !	43
Recommandations pratiques pour l'ajustement d'une loi de probabilité.....	45
Chapitre 3. Méthodes probabilistes spécifiques d'estimation des débits de crue de référence sur un site jaugé	51
Établissement de courbes débit-durée-fréquence à partir d'observations hydrométriques	51
Intérêt de l'étude de la durée des crues et des courbes QdF	51
Le modèle QdF convergent.....	54
Méthodes d'extrapolation de la distribution des crues	58

Nécessité d'informations complémentaires pour l'extrapolation des crues...	58
Extrapolation de la distribution des crues par la méthode du Gradex.....	59
Extrapolation de la distribution des crues par le modèle Agregee.....	65
La méthode Speed.....	66
Extrapolation de la distribution des crues par le modèle Shypre.....	68
Analyse des crues historiques.....	72
Construction d'hydrogrammes de référence.....	76
Formulations analytiques.....	77
Hydrogramme synthétique monofréquence (HSMF).....	77
Hydrogramme moyen monofréquence.....	78
Hydrogramme d'apport par différence.....	79
Recommandations pratiques pour l'étude des événements extrêmes et la constitution d'hydrogrammes de crue de référence.....	80
Information pluviométrique.....	80
Information sur les crues historiques.....	81
Hydrogrammes de crue de référence.....	81
Chapitre 4. Méthodes probabilistes spécifiques d'estimation des débits de crue de référence sur un site non jaugé.....	83
Régressions multiples.....	84
La formulation de Bressand-Golossof (1995).....	85
Les courbes-enveloppes de Francou-Rodier (1967).....	85
Synthèse des débits de période de retour 1 000 ans calculés par la méthode du Gradex (Duband <i>et al.</i> , 1988).....	85
Formule Crupedix (Ctgreg <i>et al.</i> , 1980-1982).....	86
Formule Socose (Ctgreg <i>et al.</i> , 1980-1982).....	87
Modèles préparamétrés.....	88
La méthode rationnelle (Stifel W.J. et Mc Pherson M.B., 1964).....	89
La méthode SCS (Soil Conservation Service, 2003).....	91
Modèles QdF de référence.....	92
Estimation de trois paramètres locaux représentatifs du bassin versant.....	92
Choix d'un modèle de référence QdF adimensionnel.....	93
Reconstitution du régime des crues par dénormalisation du modèle de référence.....	94
Modèle QdF régional convergent.....	96
Modèle Shyreg : une méthode régionale pour l'estimation des débits de crue ..	98
Le générateur de hyétogrammes.....	98
La modélisation de la pluie en débit.....	101
La régionalisation du modèle pluie-débit.....	101
Agglomération de l'information à l'échelle du bassin versant.....	101
Contrôle des performances du modèle régional.....	102
Recommandations pratiques pour l'estimation de débits de crue de référence en site non jaugé.....	104

Chapitre 5. Exemples d'application	105
Construction d'un hydrogramme de projet relatif à la crue centennale.	
Cas du bassin versant de la Moselle à Rupt-sur-Moselle	105
Examen de la dynamique de crue	105
Les volumes générés sur différentes durées.....	106
Courbes QdF dans le domaine des observations.....	109
Extrapolation des courbes QdF aux fréquences rares	112
Hydrogramme de projet centennal (HSMF)	116
Comparaison avec un hydrogramme de projet centennal issu d'une méthode sommaire.....	116
Construction d'un hydrogramme de projet relatif à la crue centennale.	
Cas d'un bassin versant à crues lentes du Nord de la France	118
Distribution des crues fréquentes à moyennes.....	122
Extrapolation de la distribution des crues pour les événements rares	128
Construction d'hydrogrammes de référence.....	135
Conclusion	138
Applications de méthodes de régionalisation en zone méditerranéenne française	138
Présentation des approches testées.....	138
Régressions multiples	141
Modèles préparamétrés	143
Modèle QdF de référence.....	146
Modèle QdF régional convergent	148
Application de la méthode Shyreg à la zone méditerranéenne française.....	149
Comparaison des différents résultats	152

ANNEXES

Annexe 1. Principaux logiciels d'ajustements de lois de probabilité en hydrologie	157
Logiciel Extrêmes.....	157
Logiciel Hydro	158
Logiciel Hydrolab.....	159
Logiciel Hyfran	159
Logiciel PRD' crue.....	160
Logiciel Prethy	161
Logiciel Safarhy	161
Logiciel Winfap-Feh.....	162
Annexe 2. Éléments de probabilités et statistiques appliquées à l'hydrologie	163
Introduction	163
Statistiques descriptives.....	164
Quelques définitions	164
Comment résumer un échantillon ?	165

Étude de deux variables	169
Le modèle probabiliste.....	174
Quelques rappels de théorie des ensembles	174
Premières définitions : probabilité et variables aléatoires.....	174
Loi de probabilité d'une variable aléatoire	176
Statistique inférentielle	185
Retour sur la notion d'échantillon	185
L'estimation	190
Les tests statistiques	194
Applications à l'étude des crues	202
Les données.....	202
L'échantillonnage.....	203
Estimation des paramètres des distributions	207
Quantiles, périodes de retour et intervalles de confiance.....	214
Tables usuelles	218
Fonction de répartition de la loi normale réduite	219
Fractiles de la loi normale centrée réduite	220
Fractiles pour la loi du Chi-2 à <i>ndl</i> degrés de liberté	221
Centiles <i>tp</i> pour la distribution <i>t</i> de Student à <i>ndl</i> degrés de liberté.....	222
Distribution <i>F</i> (percentile 95 %)	223
Tables du test de Kolmogorov-Smirnov	224
Annexe 3. Références bibliographiques.....	227
Références citées dans le texte.....	227
Quelques ouvrages francophones de référence en hydrologie et en statistiques.....	232
Sites Web	232

Préambule

Le risque d'inondation touche près de cinq millions de personnes en France. Les événements récents de l'Aude en 1999, du Gard en 2002 et du Rhône en 2003 nous ont rappelé que les enjeux humains et économiques confrontés à ce risque étaient toujours plus importants. Les inondations constituent le premier risque traité par les Plans de prévention des risques (PPR), outil réglementaire de l'État pour la prise en compte des risques naturels dans l'aménagement du territoire : en 2007, 87 % des communes couvertes par un PPR prescrit ou approuvé le sont pour un PPR intégrant le risque inondation, et 67 % de l'ensemble des PPR traitent du seul risque inondation.

Outre l'état des risques, devenu obligatoire avec la mise en place de l'information « acquéreur locataire » au 1^{er} juin 2006¹, et l'action d'information communale au moins une fois tous les deux ans, le PPR impose des servitudes sur l'occupation du sol. Les contraintes qui en résultent rendent d'autant plus indispensable la garantie que tous les citoyens seront traités de façon équitable. Il est donc nécessaire de pouvoir assurer la plus grande rigueur et la plus grande cohérence, aussi bien en ce qui concerne les études d'aléas que l'analyse des enjeux lors de l'élaboration du PPR.

Concernant le risque d'inondation par débordement de cours d'eau, en l'absence de longues séries d'observations, c'est-à-dire dans l'immense majorité des cas, les études hydrologiques laissent une large place à l'expérience et à la subjectivité de l'expert. Il en résulte que l'aléa hydrologique dans le cadre de PPR dépend étroitement de celui qui les détermine et de l'époque de détermination. Ainsi, il n'est pas rare de devoir réévaluer les débits de référence, généralement à la hausse, après quelques années d'observation, et en particulier après un événement conséquent. Ceci n'est pas nécessairement de nature à remettre en cause la qualité des études précédentes, mais illustre bien la complexité de la

1. Loi n° 2003-699 relative à la prévention des risques naturels et technologiques et à la réparation des dommages. Décret n° 2005-134 du 15 février 2005 relatif à l'information des acquéreurs et des locataires de biens immobiliers sur les risques naturels et technologiques majeurs.

réalisation et de la diffusion des études hydrologiques et de la mise en cohérence des pratiques au sein des services instructeurs de PPR.

Ce guide a comme destinataires principaux les services instructeurs de PPR qui sont confrontés aux difficultés liées à la réalisation de ces documents. Il a pour objectif d'aider ces services dans leurs missions, en leur proposant un inventaire le plus exhaustif possible des techniques en hydrologie, ainsi que des recommandations sur leur utilisation.

L'hydrologie n'est pas une science exacte. Un panel très vaste de méthodes existe pour calculer des débits de période de retour donnée, ou inversement pour affecter une période de retour à un débit mesuré. Chaque méthode présente ses spécificités, ses champs de compétences, ses limites et ses résultats. Cet ouvrage n'a pas vocation à privilégier une méthode plutôt qu'une autre, mais à présenter aux lecteurs un état de l'art des méthodes disponibles, ainsi que des recommandations sur leur utilisation, le donneur d'ordre devant être capable de piloter une étude, même s'il n'est pas capable de la réaliser. Ces méthodes sont en constante évolution. Cet ouvrage ne présente qu'une photographie au moment où a été rédigé cet état de l'art. Le lecteur est incité à se renseigner directement auprès des concepteurs des différents outils pour connaître les récentes modifications apportées.

Remerciements

Le guide a été réalisé à l'initiative du ministère de l'Écologie, du Développement et de l'Aménagement durables (direction de la Prévention des Pollutions et des Risques, bureau des Risques Naturels), avec un suivi de l'avancement du travail par Marie Renne. La rédaction, confiée au Cemagref, a été coordonnée par Michel Lang et Jacques Lavabre, avec les contributions d'Éric Sauquet (modèles QdF et exemples d'application) et de Benjamin Renard (annexes statistiques).

Un second guide, pour le pilotage des études hydrauliques, a été rédigé par le CETMEF sous maîtrise d'ouvrage du ministère de l'Équipement (direction générale de l'Urbanisme, de l'Habitat et de la Construction). Ces deux ouvrages ont été validés par un comité, présidé par le conseil général des Ponts et Chaussées, avec l'appui du conseil général de l'Agriculture, de l'Alimentation et des Espaces ruraux et de l'inspection générale de l'Environnement. Ce comité de validation, placé sous la présidence de Daniel Burette, et composé d'experts en hydraulique ou en hydrologie, ainsi que d'utilisateurs potentiels du guide, a œuvré à une simplification, une amélioration de la compréhension et de la précision de leur contenu, et a permis d'assurer leur cohérence et leur complémentarité.

La composition de ce comité de validation est la suivante :

Daniel Burette	CGPC, président du comité de validation
Yannick Cadet	Diren Franche-Comté
Patrick Chasse	CETMEF
Dominique Chatillon	DDAF du Vaucluse
Philippe Dayet	DDE de la Drôme
Isabelle Descatoires	CETMEF
Michel Lang	Cemagref
Alain Limandat	Ingerop
René Marcaud	Silene
Xavier Martin	IGE
Xavier Pin	CGAAER

Stéphane Piney	LRPC Blois
Michel Ravier	MTETM/DGUHC
Marie Renne	MEDD/DPPR
Patrick Simon	MEDD/DPPR
Hichem Tachrift	LRPC Clermont-Ferrand

Le guide hydrologique a été plus particulièrement analysé, amendé et validé par :

Jacky Astier	BRL Ingénierie Nîmes
Philippe Bois	INPG/LTHE Grenoble
Patrick Fourmigue	CETE Méditerranée Aix-en-Provence
Rémy Garcon	EDF/DTG Grenoble
François Nau	CGPC

Qu'ils en soient tous ici remerciés.

Chapitre 1

Prédétermination d'événements de référence de crue

Cadre réglementaire

Les Plans de prévention des risques naturels (PPRN), institués par la loi n° 95-101 du 2 février 1995, relative au renforcement de la protection de l'environnement, et le décret n° 95-1089 du 5 octobre 1995 remplacent les anciens dispositifs existants, à savoir :

- les Plans de surface submersible (PSS) créés en 1935 (décret d'application du 20 octobre 1937) ;
- les périmètres de risque prévus par l'article R. 111-3 du Code de l'urbanisme (décret n° 77-755 du 7 juillet 1977) ;
- les Plans d'exposition aux risques (PER), introduits par la loi n° 82-600 du 13 juillet 1982 relative à l'indemnisation des victimes des catastrophes naturelles, et définis par le décret n° 84-328 du 3 mai 1984 relatif à l'élaboration des plans d'exposition aux risques naturels prévisibles.

Le PPR a pour objectif de réglementer les usages du sol dans les zones à risques. Contrairement au Plan local d'urbanisme (PLU) ou au schéma de cohérence territoriale (SCOT), le PPR n'est pas un document d'aménagement du territoire ; il prend en compte l'aléa et les enjeux tels qu'ils sont connus au moment de son élaboration. En revanche, le PPR est une servitude d'utilité publique, et doit être annexé aux PLU des communes sur lesquelles le PPR est approuvé. Le PPR comprend une note de présentation (explication et motivation de la démarche), un zonage réglementaire et un règlement associé à ce zonage. Le zonage réglementaire est élaboré à partir d'une analyse de l'aléa et des enjeux sur le territoire sur lequel le PPR est prescrit. À partir de ce zonage, le règlement prévoit différents types de mesures :

- (i) des règles sur les projets (nouveaux bâtiments ou extensions-modifications d'un bâtiment existant) ;
- (ii) des mesures sur le bâti existant ;
- (iii) des mesures de prévention, de protection et de sauvegarde.

Dans le cadre des PPR inondations, l'aléa de référence est celui qui correspond à la plus forte crue connue et bien documentée, ou l'aléa centennal si celui-ci correspond à une crue plus forte. Pour le scénario de référence retenu, le niveau d'aléa sur une zone donnée est défini en croisant la hauteur d'eau et la vitesse attendues. Une grille de définition des niveaux d'aléa a été proposée dans le guide méthodologique pour l'élaboration des PPR inondation.

La qualification de l'aléa dans le cadre de l'élaboration d'un PPRI suppose donc que l'on connaisse la période de retour de la plus forte crue historique bien documentée, et si cette période de retour est inférieure à 100 ans d'évaluer la crue centennale. La doctrine pour la qualification de l'aléa disponible dans le guide pour l'élaboration des PPR inondation ne précise pas quel paramètre (hauteur, débit moyen sur n jours ou n heures, débit maximum, volume...) doit être retenu pour la qualification de la crue historique et/ou pour l'élaboration du scénario centennal. Si l'on utilise classiquement hauteur d'eau et débit maximum, l'attention du lecteur doit être attirée sur l'existence d'autres critères qui peuvent être très significatifs pour l'appréciation de l'aléa en fonction du contexte local.

Ce cahier de recommandations ne traitera donc que des études hydrologiques visant à associer une période de retour à une crue donnée, et à construire une crue centennale. Les méthodes présentées peuvent être utilisées dans un autre cadre que celui traité dans ce document. Dans ce cas, les critiques et recommandations formulées dans ce guide ne sauraient être utilisées en l'état.

Maximisation de l'aléa

Une approche intuitive consiste à raisonner sur la dernière grande crue dommageable de la région et à appliquer un coefficient de sécurité sur le débit de référence. Pour réduire l'arbitraire dans le choix de l'événement historique et de la marge de sécurité à retenir, certains auteurs ont reporté sur un graphique les records de crue observés, avec en abscisse la superficie du bassin versant (km^2) et en ordonnée le débit de pointe ($\text{m}^3/\text{s} \cdot \text{km}^2$). Ils tracent ensuite une courbe enveloppe supérieure qui donne un premier ordre de grandeur d'une crue maximale en fonction de la taille du bassin versant.

Ce type d'approche reste très approximatif : le choix d'une large zone géographique — par exemple, la France — permet de disposer d'un échantillon étoffé de valeurs de crues, mais il présente l'inconvénient de mélanger des valeurs issues de régimes climatiques différents ; et le choix d'une zone hydrologiquement homogène, mais plus restreinte, conduit à retenir une courbe enveloppe qui doit être régulièrement revue à la hausse en fonction de nouvelles catastrophes. Le découpage hydrologique en zones homogènes est par ailleurs loin d'être aisé (cf. *figure 1.1*), avec un choix discutable de l'auteur qui dissocie le Languedoc du littoral méditerranéen. Francou et Rodier (1967) ont proposé et formalisé une approche de même type à partir des plus fortes crues observées dans le monde aux XIX^e et XX^e siècles.

L'approche hydrogéomorphologique (Garry *et al.*, 1996) permet, à partir d'une analyse des cartes topographiques au 1/25 000^e, de stéréographies aériennes et de reconnaissances de terrain par un géomorphologue, de distinguer les différentes unités du lit d'écoulement de la rivière : le lit mineur correspond à la section usuelle d'écoulement, le lit moyen assure la transition entre le lit mineur et le lit majeur, et le lit majeur, qui fonctionne pour

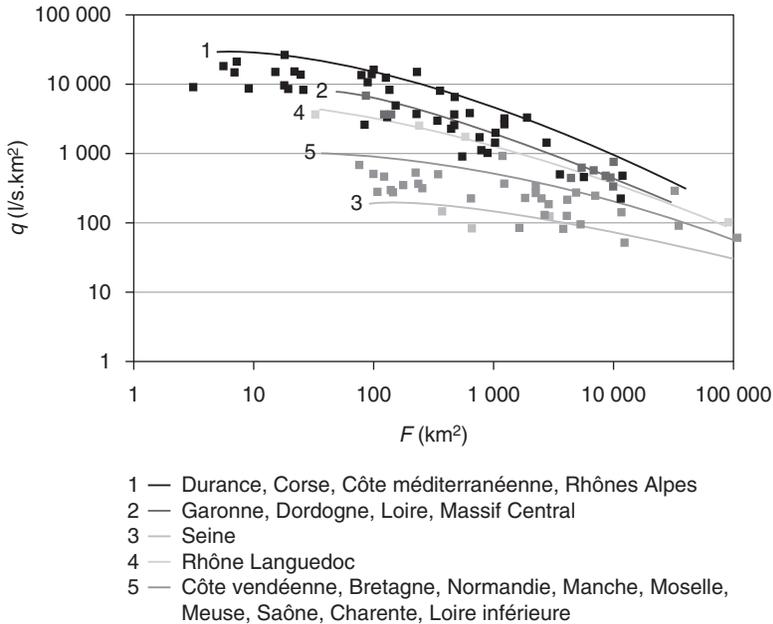


Figure 1.1. Courbes enveloppes de records de crue sur la France (Stanescu, 1997).

les crues importantes, correspond à la plaine alluviale. Les limites de ce dernier lit donnent une idée de l'extension maximale du champ d'inondation. Une approche complémentaire, appelée paléo-hydrologie, permet par ailleurs la reconstitution du niveau atteint par les plus fortes crues depuis plusieurs millénaires, en s'intéressant aux dépôts laissés par les crues, soit directement dans le champ d'inondation, soit dans des grottes surplombant le lit mineur. Des techniques d'analyse stratigraphique et de datation des dépôts (C^{14} , thermoluminescence) sont alors utilisées pour positionner en altitude les plus forts événements, et proposer une chronologie de crues (House *et al.*, 2002 ; Sheffer *et al.*, 2003 ; Benito *et al.*, 2004).

Ces approches naturalistes sont intéressantes dans la mesure où elles permettent d'élargir de façon considérable l'échelle chronologique d'investigation, et de donner des éléments objectifs sur les crues majeures survenues sur le bassin versant. Elles soulèvent toutefois des difficultés d'interprétation pour déterminer si les niveaux de crues atteints dans les conditions climatiques et géomorphologiques du passé restent représentatifs du risque d'inondation futur (*cf.* variations climatiques à l'échelle pluriséculaire, divagation latérale du cours d'eau dans la plaine d'inondation, changement de l'occupation du sol sur le bassin).

Crue historique de référence

Dans le cas d'un grand bassin versant, on dispose généralement d'une information suffisante sur les inondations pour établir un classement des crues ayant occasionné les plus fortes cotes de submersion. Une première difficulté consiste alors à choisir dans cet inven-

taire la crue historique la plus forte connue. Faut-il remonter le plus loin dans le passé, quitte à sélectionner un événement qui ne serait plus représentatif des conditions d'écoulement actuelles ? C'est l'analyse de l'historique des aménagements sur le bassin versant et la rivière qui doit permettre de déterminer le classement des crues en termes de débit, et de retenir la crue la plus pénalisante dans le contexte actuel. Par ailleurs, on observe généralement sur un grand bassin que les contributions de chacun des affluents sont rarement équivalentes, et que le jeu des concomitances peut induire des conséquences très différentes en progressant vers l'aval. Il peut être nécessaire de raisonner sur plusieurs scénarios hydrométéorologiques en fonction de la genèse des crues, comme c'est le cas par exemple sur le Rhône, où l'on distingue les crues de types océanique (secteur amont), cévenol (secteur intermédiaire) et méditerranéen (secteur aval), avec le cas extrême d'une crue généralisée (Safege, 2000).

Dans le cas d'un petit bassin versant, l'information documentaire sur les inondations est en général très limitée, et seules des approches naturalistes (géomorphologie, paléohydrologie) pourront donner des indications sur les crues majeures survenues sur le bassin.

Crue centennale

Cadre probabiliste et difficultés d'interprétation

L'analyse probabiliste repose sur un cadre théorique complet et sur un ensemble d'observations très détaillées. Elle suppose que les phénomènes peuvent être traités comme une variable aléatoire dont les réalisations peuvent revenir indéfiniment, ce qui n'est pas toujours le cas, comme par exemple pour les aléas gravitaires — chutes de bloc, glissements de terrain, laves torrentielles. Dans le cas des inondations, cette hypothèse est acceptable, avec un nombre significatif d'événements, car les rivières sortent plusieurs fois par an de leur lit mineur. Il est donc essentiel de valoriser le patrimoine d'informations sur les inondations pour une meilleure prédétermination de la crue centennale.

Toutefois, la manipulation des concepts probabilistes n'est pas immédiate en l'absence de formation technique appropriée. La période de retour d'un événement souffre de mauvaises interprétations. Elle est souvent considérée à tort comme indicative d'une récurrence régulière du phénomène : « la crue centennale arrive une fois par siècle, je suis tranquille pour quelques années... ». On peut citer le contre-exemple de la Loire-Moyenne, où trois crues supérieures à la centennale ont été observées au XIX^e siècle, en vingt ans (1846, 1856 et 1866), qui n'ont plus jamais été dépassées depuis. Si l'on définit un événement comme le dépassement d'un débit Q_p de référence par le maximum annuel, on peut lui associer deux grandeurs :

- la probabilité annuelle p de non-dépassement du débit Q_p : $p = \text{Prob}[Q < Q_p]$;
- le nombre moyen d'années T séparant deux occurrences d'événement.

La relation entre ces deux variables,

$$p = 1 - \frac{1}{T}, \quad (1.1)$$

montre que le risque d'être confronté à une crue supérieure à l'événement de référence ne dépend ni de la date considérée, ni de la date de la crue précédente. Le *tableau 1.1* montre comment ce risque évolue lorsqu'on augmente la durée de référence :

$$\text{Prob}[Q < Q_p \text{ au moins une fois sur } N \text{ années}] = 1 - \left(1 - \frac{1}{T}\right)^N \quad (1.2)$$

Tableau 1.1. Risque d'observer un événement supérieur à la crue de référence sur une durée donnée.

Quantile Q_p		Probabilité de dépasser au moins une fois le quantile Q_p sur N années				
Probabilité p	Période de retour T (ans)	$N = 1$ an	$N = 10$ ans	$N = 20$ ans	$N = 50$ ans	$N = 100$ ans
0,9	10	10^{-1}	2/3	0,88	0,995	$1 - (3 \times 10^{-5})$
0,99	100	10^{-2}	10^{-1}	0,18	0,40	2/3
0,999	1 000	10^{-3}	10^{-2}	0,02	0,05	10^{-1}

On voit par exemple que la crue centennale, qui revient en moyenne une fois par siècle, a une chance sur cent d'être dépassée chaque année, environ une chance sur cinq en 20 ans (18 %) et un peu moins d'une chance sur deux en 50 ans (39 %), ce qui est loin d'être négligeable à l'échelle de la vie humaine ou de la durée de vie des biens exposés aux inondations. Ce dernier résultat peut paraître surprenant pour un riverain qui pense à tort qu'il a peu de chance de voir une crue centennale en cinquante ans.

Une seconde difficulté réside dans l'échelle géographique de référence sur laquelle on considère les phénomènes. Si un événement centennal défini localement se produit bien en moyenne une fois tous les cent ans au point considéré, il en va autrement à l'échelle du département — ou de la région — auquel il appartient. Si par exemple on s'intéresse au risque annuel d'observer une crue au moins centennale sur une zone constituée de 50 bassins versants dont les crues sont supposées indépendantes, on obtient une probabilité bien supérieure au risque local de 10^{-2} : $1 - 0,99^{50} = 0,40 = 1/2,5$. En moyenne, on devrait observer sur cette zone une crue au moins centennale tous les deux ans et demi. Ainsi, la répétition de crues centennales à une large échelle spatiale n'est ni forcément anormale, ni imputable au changement climatique, à une modification des conditions d'écoulement, ou à une erreur dans l'étude hydrologique. C'est seulement à l'issue d'une analyse détaillée des phénomènes à l'échelle régionale que l'on peut émettre un diagnostic sur le risque d'occurrence sur différentes échelles spatiales.

Une dernière source d'erreur réside dans l'appréciation des probabilités de plusieurs variables. Le débit de crue étant la conjonction d'un état initial de saturation du bassin versant et d'un épisode pluvieux, *il n'y a aucune raison a priori pour que les périodes de retour de la pluie et du débit d'un même événement soient identiques*. On trouve pourtant un certain nombre d'approches de prédétermination basées sur ce principe. Le phénomène de concomitance des crues conduit par ailleurs à des résultats non immédiats. Il s'agit du risque que la pointe de crue d'un affluent soit concomitante avec celle du cours principal ; lorsqu'il y a homogénéité climatique et que les superficies des deux sous-bassins sont voisines, cela se produit assez souvent. En revanche, pour de grands bassins versants, il est peu vraisemblable qu'une crue de même ampleur survienne sur l'ensemble du bassin. À titre d'exemple, l'étude hydrologique des crues du Rhône (Safege, 2000) a montré que

pour obtenir une crue centennale sur le Rhône à Beaucaire (95 600 km²) il suffisait que les crues des principaux affluents soient seulement de fréquence moyenne ($2 < T < 10$ ans).

Différentes étapes d'une analyse probabiliste

À l'issue du travail de recensement des crues anciennes et du choix d'une crue historique de référence, il reste à la comparer à la crue centennale dont le débit a une chance sur cent d'être dépassé chaque année. Le principe de l'analyse probabiliste consiste (cf. *tableau 1.2*) :

- (A) à extraire d'une chronique de valeurs journalières ou instantanées un échantillon de valeurs maximales indépendantes ;
- (B) à valider le choix d'une loi de probabilité adaptée à l'échantillon étudié ;
- (C) à extrapoler la distribution pour des événements rarement ou pas observés.

Chacun de ces points comporte une part d'arbitraire liée à la formulation d'hypothèses dont la validité n'est que partiellement vérifiée, et une part d'incertitudes d'échantillonnage, d'autant plus importante que le niveau d'extrapolation est élevé.

Tableau 1.2. Les points clés d'une analyse locale des probabilités.

Étapes	Points clés
A. Échantillonnage	<ul style="list-style-type: none">• Extraction de valeurs maximales respectant plusieurs hypothèses : caractère aléatoire, valeurs indépendantes, échantillon homogène, processus stationnaire.
B. Validation d'une distribution théorique	<ul style="list-style-type: none">• Choix d'une distribution et d'une méthode d'estimation des paramètres.• Test d'adéquation (statistique, visuel).
C. Extrapolation aux événements non observés	<ul style="list-style-type: none">• Limite de validité à fixer, en fonction de la qualité des données, de la taille de l'échantillon et de la bonne adéquation de la distribution théorique aux observations.

A. Échantillonnage

L'échantillonnage peut être mené par extraction de la valeur maximale annuelle ou de toutes les valeurs supérieures à un seuil. Il est supposé permettre l'obtention d'une collection de valeurs issues d'une variable aléatoire, qui vérifient les hypothèses d'indépendance, d'homogénéité (toutes les valeurs suivent la même distribution) et de stationnarité (les paramètres de la distribution ne varient pas avec le temps). Ceci conduit à corriger éventuellement les valeurs influencées, à éliminer les valeurs manifestement liées de l'échantillon (même épisode de crue), à tester le caractère saisonnier du régime des crues (si nécessaire, scinder en sous-échantillons de valeurs homogènes), et à vérifier la stationnarité du processus (le cas échéant, scinder en deux périodes). Le cas des valeurs singulières (*outlier* en anglais) est plus délicat ; il nécessite un traitement particulier qui peut être abordé par une analyse historique, régionale ou statistique.

B. Validation d'une distribution théorique

Le choix d'une loi de probabilité est guidé par des considérations théoriques et pratiques. Gumbel (1958) et Pickands (1975), à partir de développements théoriques sur les propriétés asymptotiques des valeurs extrêmes, recommandent respectivement l'usage de la loi des valeurs extrêmes généralisée (GEV en anglais) pour les valeurs maximales annuelles, et de la loi de Pareto généralisée (GP en anglais) pour les valeurs supérieures à un seuil. De nombreuses autres distributions ont par ailleurs été proposées (lois de Gauss, Gamma, Weibull, Halphen...), ainsi que des changements de variables sur les valeurs de l'échantillon (puissance, logarithme...). On retiendra l'existence de trois familles de distributions (cf. lois GEV ou Pareto généralisée), suivant la valeur du paramètre de forme ($k > 0$; $k = 0$; $k < 0$), dont le cas des lois de probabilité avec $k = 0$ qui, s'il présente l'avantage de la simplicité et de la robustesse, n'est cependant pas forcément adapté à l'ensemble des phénomènes étudiés.

À titre d'exemple, nous avons reporté le tracé de différentes lois de probabilité (cf. figures 1.2 à 1.4) calées sur un même échantillon de la valeur maximale annuelle du débit journalier de la Seine à Bar-sur-Seine (bassin versant de 2 340 km²). La période d'observation s'étend de 1950 à 1979, soit 30 valeurs. La construction de tels graphiques sera détaillée dans le chapitre 2 (§ Construction d'un graphique de fréquence cumulée). Les débits figurent en ordonnée et les périodes de retour en abscisse. L'échelle de l'abscisse est une échelle logarithmique. Les points représentent les observations, et plusieurs lois à deux paramètres ont été reportées : loi normale, loi de Weibull, loi des fuites, loi Gamma incomplète et loi de Gumbel (citées par ordre croissant d'estimation de la crue centennale). Sur la figure 1.2, le débit de la crue décennale (période de retour 10 ans) vaut environ 200 m³/s. Ce débit est celui qui a une chance sur dix d'être dépassé chaque année, et qui est dépassé en moyenne une fois tous les 10 ans. On constate sur la figure 1.2 que l'ensemble

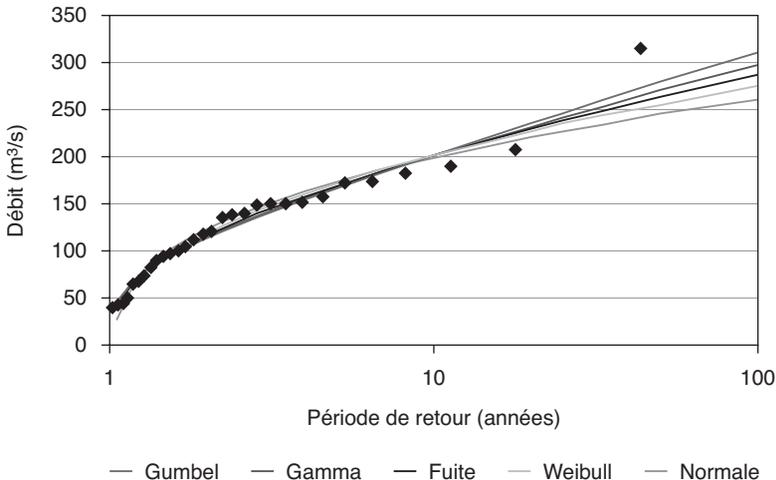


Figure 1.2. La Seine à Bar-sur-Seine. Ajustement de différentes lois de probabilité à l'échantillon des maxima annuels des débits journaliers.

des lois suivent bien la distribution de fréquence empirique des débits, sans accorder trop de poids à la valeur maximale qui s'éloigne quelque peu de l'alignement.

En zoomant sur les périodes de retour élevées (entre 10 et 100 ans) (cf. figure 1.3), on constate toutefois que les différentes lois proposent des extrapolations sensiblement différentes : l'estimation centennale du débit varie entre 260 m³/s selon la loi normale, et 310 m³/s selon la loi de Gumbel, soit un écart relatif de l'ordre de 20 %. À l'expérience, ces différentes estimations de la crue centennale restent comparables compte tenu des erreurs sur la reconstitution des débits (cf. chapitre 2, § Le débit : la variable à étudier) et de l'incertitude d'échantillonnage (cf. annexe 2, § *Quantiles, périodes de retour et intervalles de confiance*).

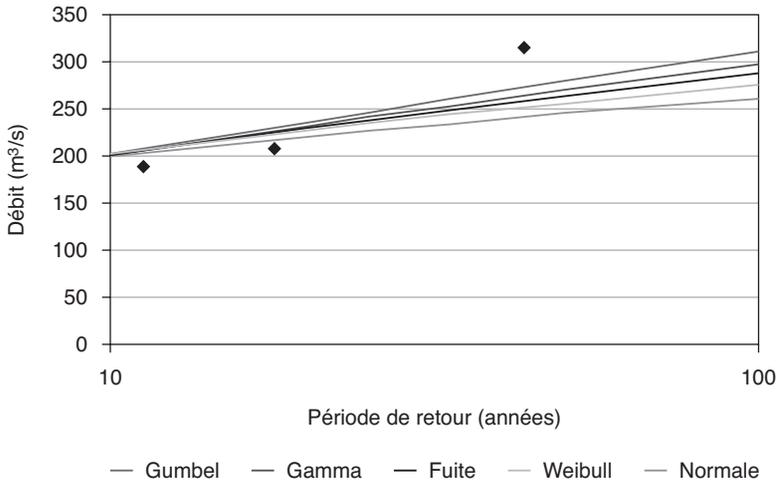


Figure 1.3. Distribution des crues de la Seine à Bar-sur-Seine (zoom sur les périodes de retour élevées).

Afin de coller au mieux à la distribution empirique, les hydrologues effectuent parfois des changements de variables — en général, une transformation logarithmique ou racine carrée des débits. La figure 1.4 présente les courbes de la loi normale et de la loi de Gumbel obtenues après transformation des variables. La loi de Gumbel, avec une transformation racine carrée, et la loi normale, avec une transformation logarithmique (loi de Galton), se rapprochent du plus fort débit. La courbe de la loi normale-racine carrée passe en dessous du point le plus fort, et la loi de Gumbel avec transformation logarithmique (loi de Fréchet) conduit à une extrapolation très supérieure aux autres. L'estimation centennale du débit serait, selon cette loi, de 890 m³/s. On mesure sur cet exemple le poids des hypothèses qui peuvent être faites.

Outre le choix de la distribution, *il existe également plusieurs méthodes d'estimation des paramètres de la distribution*. Les plus connues sont la méthode des moments ou du maximum de vraisemblance (cf. annexe 2, § *L'estimation*), avec également la méthode des moments pondérés (Greenwood *et al.*, 1979), censée être robuste en présence dans l'échantillon de valeurs singulières. On retiendra ici que l'incertitude liée au choix d'une

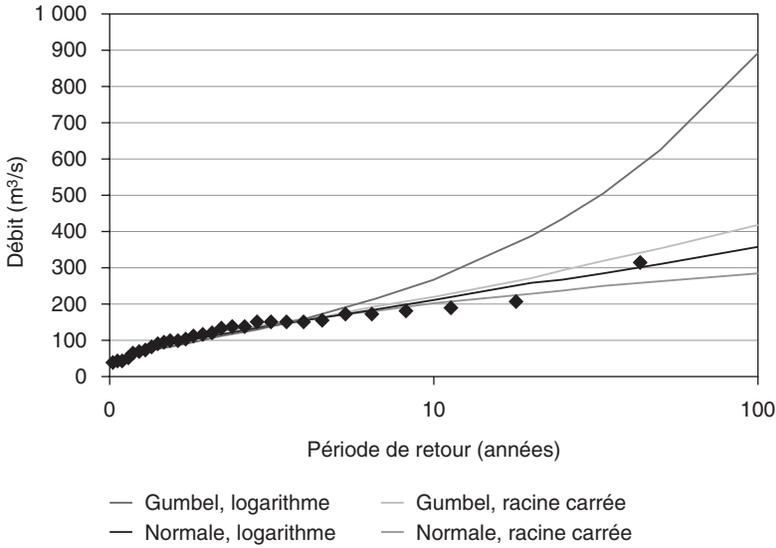


Figure 1.4. Distribution des crues de la Seine à Bar-sur-Seine (lois de Gauss et Gumbel avec transformation racine carrée et logarithmique des débits).

méthode d'estimation est généralement inférieure à celle liée au choix d'une distribution, et qu'elle dépend principalement de la présence de valeurs singulières dans l'échantillon.

Au final, on vérifie que le binôme (loi théorique, méthode d'ajustement) représente correctement le phénomène étudié, et ce, en effectuant un test d'adéquation entre la loi théorique et la distribution empirique. Un premier contrôle visuel sur le graphique de la distribution (*cf. figures 1.2 à 1.4*) permet généralement d'éliminer des lois non adaptées, mais il reste trop subjectif pour départager des lois candidates voisines. Il est ensuite conseillé d'introduire une règle de décision objective basée sur le test de deux hypothèses alternatives (la loi théorique est adaptée ou pas à l'échantillon, avec un risque d'erreur maximal admissible). On peut vérifier l'appartenance des valeurs expérimentales de l'échantillon à l'intervalle de confiance des quantiles, ou utiliser des tests non paramétriques basés sur les valeurs classées d'un échantillon. Ces différents tests statistiques sont par nature négatifs (rejet ou pas d'une hypothèse), et ne permettent pas d'éviter le risque d'erreur dit de seconde espèce qui consiste dans le cas présent à valider le choix de la loi de probabilité alors que le phénomène est distribué différemment. Ils sont surtout utiles pour éliminer des lois théoriques non adaptées au régime des crues.

C. Extrapolation aux événements non observés

L'étude statistique des événements extrêmes est par nature délicate. Ces événements sont rarement observés localement, et l'hydrologue a peu de chance de disposer d'une collection suffisante de catastrophes. Or, le principe de l'analyse probabiliste consiste à analyser une série de réalisations événementielles, à en déduire un comportement probabiliste et à l'extrapoler ensuite à l'ensemble de la population des crues. La mise en place généralisée d'un réseau hydrométrique en France a débuté après la fin de la Seconde Guerre

mondiale (dimensionnement des grands barrages hydroélectriques) ; elle a pris un deuxième essor dans les années soixante, avec la création des réseaux des Services régionaux d'aménagement des eaux (SRAE) du ministère de l'Agriculture et des Services hydrologiques centralisateurs (SHC) du ministère de l'Équipement. L'essentiel de ce réseau est actuellement géré par les directions régionales du ministère de l'Écologie, du Développement et de l'Aménagement durables. Les séries à la disposition de l'hydrologue couvrent en général une période de quelques dizaines d'années, ce qui est insuffisant pour extrapoler de façon fiable l'analyse probabiliste des observations aux événements extrêmes.

La première raison de cette difficulté vient du comportement différencié du régime des crues, avec des événements courants fortement influencés par le filtre du bassin versant, et des événements extrêmes conditionnés principalement par les précipitations. Des séries d'observation de quelques dizaines d'années ne permettent pas d'observer avec précision ce changement, et il est difficile de départager les lois de probabilités candidates, car elles donnent des résultats voisins sur le corps de la distribution expérimentale, et divergent seulement pour les événements forts à extrêmes. La seconde difficulté vient des données elles-mêmes qui ne présentent pas le même degré de précision suivant la gamme de débit considérée. La courbe de tarage reliant la hauteur observée et le débit estimé est établie à partir des jaugeages disponibles. Elle est extrapolée au-delà du plus fort débit mesuré, en général inférieur à la crue décennale, avec une marge d'erreur plus forte que sur les débits courants et des discontinuités pouvant apparaître en présence d'écoulements dans le lit majeur.

Ces difficultés d'extrapolation ont conduit les hydrologues à rechercher des informations complémentaires, en élargissant l'échelle spatiale d'investigation (approche régionale), le cadre chronologique (approche historique) et les paramètres explicatifs (approche par les processus d'écoulement, la pluie étant alors un paramètre explicatif fort des débits). Ces méthodes d'extrapolation feront l'objet d'une présentation spécifique dans le chapitre 3.

Plan du guide

Le ministère de l'Écologie, du Développement et de l'Aménagement durables a souhaité la mise au point d'une série de recommandations pour le calcul des aléas hydrologiques dans le cadre des Plans de prévention des risques d'inondations (PPRI). Ce dernier document a pour objectif d'améliorer la qualité des études hydrologiques, d'assurer une certaine homogénéité dans le mode de détermination des crues de référence à l'échelle du territoire national, et d'exploiter au mieux les informations hydrologiques disponibles sur les événements passés.

La *figure 1.5* récapitule la démarche générale à suivre pour estimer la crue de référence des PPRI. Après une phase de recherche des données disponibles, on distingue le cas où les données hydrométriques sont disponibles près du site à étudier de celui où les données sont insuffisantes ou inexistantes.

Dans le premier cas de figure, il sera possible de réaliser un ajustement probabiliste sur l'échantillon des débits de crue. Le chapitre 2, après un rappel sur les incertitudes liées à la métrologie des débits, donne les éléments de base pour comprendre le mode de construction d'un graphique de fréquence cumulée, avec une série d'exemples d'ajustements. Les éléments plus techniques figurent en annexe. Le chapitre 3 présente une série de

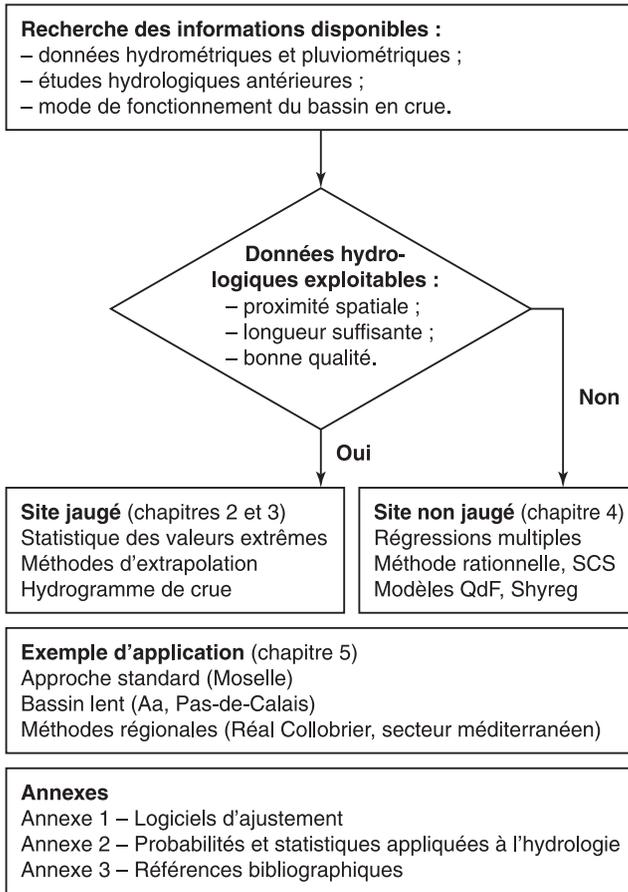


Figure 1.5. Plan général du guide d'estimation des probabilités de crues.

méthodes spécifiques à l'hydrologie qui permettent d'établir les courbes débit-durée-fréquence à partir d'observations hydrométriques, d'extrapoler la distribution des crues pour les événements extrêmes et le mode de construction d'hydrogrammes de référence.

Dans le second cas de figure, il est nécessaire d'utiliser des approches régionales qui permettent d'estimer l'aléa hydrologique de référence sur un site non jaugé, ou lorsque l'on dispose d'une information réduite. Le chapitre 4 présente une synthèse de méthodes développées en France pour transférer spatialement des calculs de crue de référence d'un site jaugé à un site quelconque du territoire.

Une série d'exemples d'applications est ensuite présentée au chapitre 5, avec :

- (a) le cas standard où un hydrogramme de projet centennal est construit à partir d'observations hydrométriques et pluviométriques ;
- (b) le cas particulier d'un bassin à crues lentes du Nord de la France, où le non-respect des hypothèses statistiques de base impose un traitement particulier ;

(c) la comparaison de méthodes régionales d'estimation des quantiles de crue sur un bassin méditerranéen.

Les annexes comportent la présentation des principaux logiciels d'ajustements de lois de probabilité dédiés à l'hydrologie, puis des éléments techniques nécessaires à l'étude probabiliste des crues (statistiques descriptives, modèle probabiliste, statistique inférentielle et application à l'étude des crues), et enfin une partie bibliographique avec les références citées dans le texte, ainsi que la mention de quelques ouvrages francophones de référence en hydrologie et en statistiques.

Notons enfin que ne sont pas abordés dans ce guide toute une série de cas particuliers (bassin urbain, perméable, karstique, nival, côtier, de montagne ou avec un régime des eaux influencé...) qui nécessitent des approches spécifiques.

Chapitre 2

Ajustement d'une loi de probabilité à un échantillon de débits de pointe de crue

L'objectif de ce chapitre est de présenter les éléments de base qui permettent de décrire le régime des pointes de crue sous forme probabiliste. Nous aborderons successivement la chaîne de traitement des débits (§ *Le débit : la variable à étudier*), en rappelant les incertitudes liées à la métrologie des crues, le mode de constitution et de représentation graphique d'un échantillon de valeurs de pointes de crue (§ *Représentation d'un échantillon*), quelques exemples d'ajustements fréquentiels (§ *Exemples d'ajustement sur la distribution du débit de pointe*), et une série de recommandations pratiques (§ *Recommandations pratiques pour l'ajustement d'une loi de probabilité*). Les éléments plus techniques sur les ajustements figurent en annexe 2, avec notamment des rappels sur les statistiques, les probabilités et leur application à l'étude des crues.

Le débit : la variable à étudier

La métrologie des débits

Avant d'aborder l'analyse statistique d'une variable, il n'est pas inutile de rappeler comment les observations de cette variable sont réalisées, ainsi que les incertitudes qui entourent ces observations. Les écoulements d'un bassin versant sont appréhendés à travers le suivi des débits qui transitent à l'exutoire. La courbe qui montre l'évolution des débits dans le temps est appelée hydrogramme. Les hydrogrammes au cours d'une année présentent des variations plus ou moins marquées en liaison avec la climatologie et la géologie du bassin versant, l'existence d'ouvrages de régulation ou de prélèvement... Globalement, les hydrogrammes se composent d'une succession de périodes de hautes eaux (réponse plus ou moins rapide à la pluviométrie, à la fonte des neiges...) et de basses eaux, aussi appelées étiages. Nous nous intéressons ici aux forts débits qui sont observés lors des crues. Nous y reviendrons plus loin, mais l'origine des débits de crue n'est pas *a priori* à négliger si l'on doit conduire une étude statistique. Il est en effet vivement conseillé d'effectuer des statistiques sur une variable homogène, ce qui n'est pas forcément

le cas si l'analyse mélange des crues d'origines différentes, par exemple des crues nivales et des crues pluviales.

La métrologie des débits est appelée hydrométrie. Elle consiste à passer d'une mesure en continu de la hauteur d'écoulement à la production de séries chronologiques de débit. On trouvera dans Bertrand-Krajewski *et al.* (2000), et Benoist et Birgand (2002) un récapitulatif des dispositifs de mesure de la hauteur d'écoulement avec différents types de capteurs. On retiendra qu'une bonne gestion des équipements de terrain — contrôle du bon fonctionnement des appareils, entretien des capteurs et de la centrale d'acquisition, collecte des données, recalage des dérives des capteurs, comblement des lacunes, stockage sur base de données... — est nécessaire si l'on veut disposer d'une chronique en continu de la hauteur d'écoulement.

Le passage de cette chronique de hauteur à celle des débits repose généralement sur une relation hauteur-débit, appelée courbe de tarage. Elle est établie à partir de mesures de débit (jaugeages) réalisées par exploration du champ des vitesses (moulinet, courantomètre, ADCP) ou par dilution d'un traceur chimique, et, dans la mesure du possible, calée à l'aide d'un modèle hydraulique à surface libre. On se reportera au guide de pilotage des études hydrauliques (ministère de l'Écologie, du Développement et de l'Aménagement durables, 2007) pour obtenir des informations détaillées sur l'utilisation de tels modèles. Nous allons ici très rapidement passer en revue les principaux éléments à prendre en compte pour la transformation des hauteurs en débits.

Relation hauteur-débit

Les débits s'expriment généralement en m^3/s . L'équation aux dimensions est de fait un volume que divise un temps. Bien que dans le cas particulier de très faibles débits la mesure puisse effectivement se ramener à la mesure d'un volume pendant un certain temps (jaugeage par empotement, par exemple), les débits d'un cours d'eau résultent généralement d'une exploration du champ des vitesses dans une section connue. Le débit s'exprime alors par la relation :

$$Q = V_{\text{moy}} \times S, \quad (2.1)$$

où :

Q représente le débit (en m^3/s) ;

V_{moy} la vitesse moyenne de l'eau dans la section mouillée (en m/s) ;

S la section mouillée du cours d'eau (en m^2).

Toute modification du débit d'un cours d'eau se traduit, au droit de la station de jaugeage, par une modification de la vitesse de l'eau et de la section mouillée. La connaissance en continu du débit impliquerait, d'après l'équation 2.1, que l'on suive à tout instant les évolutions de la vitesse de l'eau et de la section mouillée.

La section mouillée peut se déduire de la mesure de la hauteur de l'eau au droit de la section, à partir de la connaissance de la topographie de la section d'écoulement. Cependant, des incertitudes significatives peuvent entacher cette détermination, notamment lors des périodes de hautes eaux qui envahissent les lits majeurs du cours d'eau. D'autres phénomènes, tels que la non-stabilité du fond du cours d'eau (creusement du lit lors de la crue, et colmatage lors de la décrue), peuvent aussi entacher d'incertitudes cette détermination.

En hydrométrie, la mesure directe et permanente de la vitesse moyenne de l'eau s'effectue par une mesure du temps de transit d'ondes ultrasonores, ou par une estimation déduite du décalage de fréquence dû à l'effet Doppler. L'information n'est le plus souvent disponible qu'en surface. Aussi, a-t-on généralement recours, plus ou moins directement, aux lois de l'hydraulique à surface libre et aux relations qui lient la vitesse de l'eau à la charge hydraulique et/ou à la pente de la ligne d'eau, en tenant compte des pertes de charges par un terme dit de rugosité.

Lorsque la section d'écoulement reste de dimension modeste, il est possible d'installer des dispositifs de mesure hydraulique : venturi, Parshall, déversoirs... Pour ce genre d'équipement, la vitesse moyenne est liée, par des relations connues, à la charge hydraulique en amont du seuil. Il convient toutefois de vérifier — et/ou modifier si nécessaire — les valeurs empiriques des coefficients de débit de ces relations. Ce peut être aussi le cas de grands cours d'eau pour lesquels on peut estimer la vitesse moyenne au droit de la section de contrôle, comme un seuil de moulin, par exemple. L'enregistrement en continu de la hauteur d'eau permet alors de connaître la section en travers et la vitesse moyenne, donc le débit du cours d'eau.

Si l'on ne bénéficie pas d'installations hydrauliques, l'estimation des débits est un peu plus délicate. Quand les caractéristiques de l'écoulement sont stables — régime permanent avec un débit constant et de faibles variations longitudinales de la hauteur d'écoulement —, la vitesse moyenne est fonction de la rugosité, du rayon hydraulique de la section et de la pente de la ligne d'eau (formule de Strickler). Cela impliquerait *a minima* la mesure de la hauteur d'eau en amont et en aval de la section pour apprécier la pente hydraulique. Dans l'immense majorité des cas, une seule hauteur d'eau H est enregistrée, et l'écoulement est en régime fluvial ($Fr = \frac{V}{\sqrt{gH}} < 1$, Fr étant le nombre de Froude) : le niveau H est alors

fonction à la fois du débit et d'un niveau en aval H_{aval} . En l'absence de section de contrôle aval stable (cf. ouvrage hydraulique avec passage en écoulement critique, $Fr = 1$), la courbe de tarage $Q(H, H_{\text{aval}})$ ne peut alors être correctement appréhendée à partir de la seule mesure de hauteur H . Ce problème se rencontre sur des stations hydrométriques situées à proximité de la mer (influence de la marée sur les écoulements) ou dans la zone d'influence d'un ouvrage hydraulique régulé ou d'une autre rivière plus en aval.

Par ailleurs, le phénomène de propagation des ondes de crue peut induire un phénomène analogue à une courbe d'hystérésis, avec une courbe de tarage en forme de boucle dans la phase crue/décru. Ce phénomène, lié à la variation de la pente d'énergie au cours de la crue, est d'autant plus accentué que la pente du cours d'eau est faible (figure 2.1a) ou que la crue est rapide. Il a été peu observé en métropole à ce jour, mais cela peut s'expliquer par la difficulté à réaliser en crue une série de jaugeages rapprochés. Les nouvelles techniques de jaugeage par ADCP, beaucoup plus rapides, devraient permettre de lever cette incertitude.

Outre les problèmes possibles de détarage de la station en période courante, lorsque la relation hauteur-débit change périodiquement (gel progressif de la section, évolution saisonnière due à la végétation...), progressivement (incision/rehaussement du profil en long) ou brusquement (modification brusque et/ou durable des conditions hydrauliques à la suite de travaux), il faut mentionner ici des problèmes spécifiques liés à l'estimation des débits en crue.

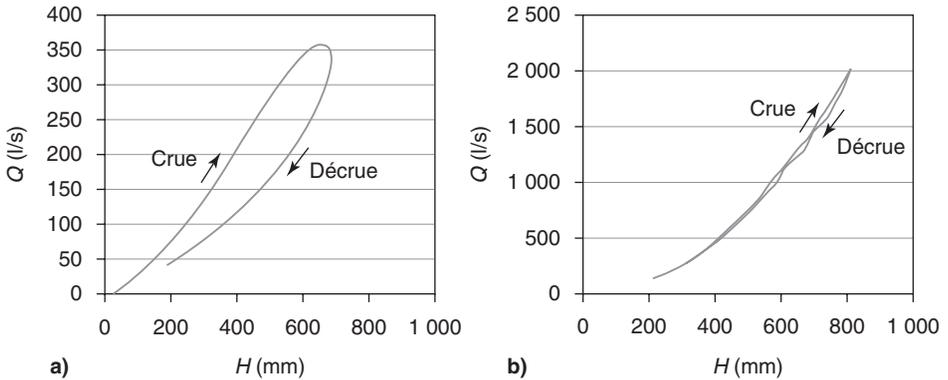


Figure 2.1. Exemple de courbe de tarage : (a) faible pente ; (b) forte pente (source : Benoist et Birgand, 2002).

Les contraintes opérationnelles liées à la réalisation de jaugeages (connaissance de la survenance d'une crue, délai de mobilisation et de déplacement des opérateurs, possibilité physique d'effectuer la mesure et sécurité des équipes) font que le plus souvent la hauteur associée au plus fort débit mesuré est bien inférieure à celle des crues moyennes ($T < 10$ ans), notamment sur les rivières à crues rapides du Sud-Est de la France (Lang *et al.*, 2006). Il est alors nécessaire d'extrapoler la courbe de tarage établie à partir des débits couramment observés. Une erreur courante consiste à extrapoler linéairement la courbe de tarage, alors que les formules hydrauliques sont du type $Q = a(H - H_0)^b$, où l'exposant b est supérieur à 1 ($3/2$ pour un seuil épais rectangulaire dénoyé, $5/3$ pour la formule de Strickler). Une seconde source d'erreur réside dans la discontinuité éventuelle de la section d'écoulement (cas des débordements dans le lit majeur notamment), qui induit une discontinuité de la relation $Q(H)$. Dans ce cas de figure, il est conseillé d'extrapoler séparément les courbes $S(H)$ et $V_{\text{moy}}(H)$. D'autres erreurs peuvent provenir d'un fonctionnement hydraulique spécifique au droit de la station en période de crue : submersion de la section de contrôle ou des berges, impliquant une influence aval des écoulements ; mise en charge de l'écoulement au droit d'un pont suite à des embâcles ou à la montée des eaux ; contournement de la station via le lit majeur ou des bras remis en activité ; modification de la section d'écoulement suite au transport solide... On trouvera dans Jaccon (1987) et dans Forray *et al.* (1998) une série de recommandations pour l'établissement des courbes de tarage.

Nous ne voulons pas ici exagérément noircir le tableau, mais rappeler au lecteur et *a fortiori* au statisticien qui manipule des chroniques de débit que la mesure des débits est en soi une opération délicate, et qui comporte des incertitudes non négligeables (Lang *et al.*, 2006). Il convient dès lors d'être vigilant et de s'assurer que la qualité des données est compatible avec les enjeux de l'étude, avant de se lancer dans les statistiques proprement dites. En plus de ces incertitudes propres à la métrologie, diverses sources d'incertitudes peuvent se superposer :

- Erreurs aléatoires : enregistrements erronés des hauteurs d'eau, erreurs dans les fichiers...

– Absence chronique des enregistrements des plus fortes crues survenues à la station, consécutivement à son mauvais fonctionnement pendant les plus fortes hauteurs d'eau, ou parfois à la destruction de la station.

Signalons enfin qu'une mesure même parfaite d'un débit de crue n'est pas obligatoirement représentative du comportement hydrologique du bassin versant. Il est trivial de rappeler que le débit d'un grand nombre de bassins-versant est plus ou moins influencé par des retenues d'eau, et que les débits « naturels » doivent être reconstitués, ce qui n'est pas a priori une opération très facile. Signalons aussi le cas des stations de mesure des débits situées en aval de champs d'inondation, qui ne mesurent en fait que l'hydrogramme écrêté.

Les bases de données de débit

La direction de l'Eau du ministère de l'Écologie, du Développement et de l'Aménagement durables gère la base nationale de données hydrométriques, la banque Hydro, qui propose un stockage des données et des logiciels pour extraire des caractéristiques des débits et effectuer des statistiques. Différents services (DIREN, direction régionale de l'environnement, services de prévision des crues, EDF, organismes de recherche, compagnies d'aménagement rural...) participent à l'alimentation de la banque Hydro. Les services ont en charge l'installation et la gestion des stations de mesure en rivière, la collecte et la vérification des données et l'alimentation de la banque. Hydro gère actuellement les chroniques de hauteurs d'eau, à pas de temps variable, de 3 500 stations dont 2 400 sont actuellement en service. À partir de ces chroniques et des courbes de tarage, Hydro calcule les débits instantanés, journaliers, mensuels... La banque Hydro met les données de base à la disposition des utilisateurs sous forme de graphiques et de fichiers informatiques, et propose différentes procédures pour des applications hydrologiques. Par exemple, la procédure Crucal effectue l'échantillonnage des débits journaliers et instantanés, et propose une estimation des quantiles de crue après ajustement d'une loi de Gumbel. Notons que la banque Hydro permet d'accéder librement et gratuitement aux données de débit journalier. Des chroniques, limitées aux pluies journalières, peuvent être achetées à Météo-France à partir de la banque de données Pluvio.

La critique d'une chronique de débits

Il est indispensable, avant d'effectuer des calculs probabilistes sur les débits de crue, de porter un regard critique sur les jeux de données hydrométriques. La qualité des données de crue est en effet très variable d'un site à un autre, suivant la configuration hydraulique du secteur, la disponibilité de mesures de contrôle (jaugeages, topographie, laisses de crue), la gamme de débit et le niveau d'expertise dans l'établissement des courbes de tarage. Certaines stations, implantées pour un objectif de suivi de la qualité des eaux en étiage ou pour un suivi des niveaux de crue, peuvent s'avérer inexploitable pour les débits de crue. Par ailleurs, de nouveaux éléments de contrôle peuvent remettre en cause des estimations de débit de crue faites il y a plusieurs décennies si l'extrapolation de la courbe de tarage de l'époque s'avère erronée.

Nous reprenons ici quelques éléments indicatifs pour la critique des données, issus de l'expérience de services producteurs de données (Forray *et al.*, 1998 ; Armand, 2002) ou d'hydrologues chargés de la reconstitution de données de crue (Naulet, 2002 ; Payrastre,

2005 ; Lang *et al.*, 2006). Dans un premier temps, la simple observation visuelle de la chronique des débits permet de repérer des anomalies grossières (figure 2.2). Il est souvent utile d'agréger les données à un pas de temps assez large du mois, voire de l'année, pour détecter des anomalies flagrantes ou des dérives persistantes de mesure. On peut ensuite raisonner à un pas de temps plus fin (la journée ou l'heure) pour repérer des erreurs temporaires de mesure à l'échelle d'épisodes de crue.

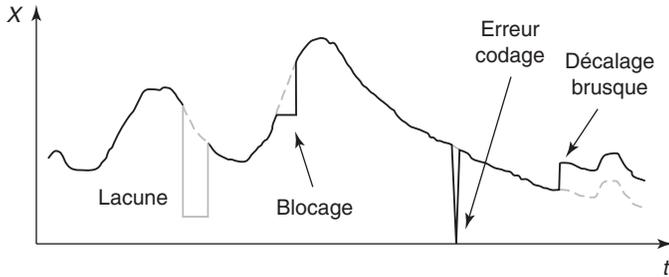


Figure 2.2. Exemples d'anomalies détectées visuellement sur une chronique (Lang, 1995).

Dans un second temps, on doit s'intéresser au dossier hydrométrique de la station afin d'évaluer la qualité de la chaîne de traitement des débits à partir d'une série de critères :

- *Emplacement du site de mesure* : l'ensemble du débit en crue passe au droit de la station, un contrôle hydraulique aval assure une relation univoque entre la hauteur et le débit, la stabilité du fond du lit réduit le risque de détarage de la station, et le débit n'est pas influencé par un ouvrage hydraulique (barrage, prise d'eau, dérivation) ou par des conditions locales (pertes, résurgences).

- *Qualité de la mesure de hauteur* : le dispositif est capable de suivre une variation rapide du niveau, la plage de mesure du capteur est compatible avec les niveaux atteints en crue, le capteur ne présente pas de dérive importante (*cf.* mesures de contrôle), la hauteur mesurée par le capteur est représentative du niveau d'écoulement (contre-exemples : mesure en aval d'un pont en charge, écoulement concentré sur une partie de la section avec une pente transversale, remous au droit du capteur, taux élevé de matières en suspension avec une densité supérieure à 1), et l'échelle limnimétrique et le capteur sont situés à proximité (contre-exemple : en amont et aval d'un pont).

- *Relation hauteur-débit* : stabilité dans le temps (d'après l'historique des courbes de tarage et le nuage de l'ensemble des jaugeages réalisés), sensibilité (à une faible variation de débit doit correspondre une variation suffisamment grande de niveau), dispersion des jaugeages autour de la courbe de tarage, niveau d'extrapolation en crue (plus haut niveau jaugé, plus haut niveau enregistré), méthode d'extrapolation pour les débits de crue (ajustement graphique, relation puissance, modélisation hydraulique).

- *Qualité des jaugeages* : précision (d'après le mode opératoire et la gamme des vitesses), représentativité du débit jaugé par rapport à l'ensemble du débit écoulé au droit du capteur de niveau, variation du niveau pendant le jaugeage.

On portera par ailleurs une attention particulière aux conséquences éventuelles qu'ont pu avoir des événements de la station sur la mesure : changement de matériel, de méthode

de calcul, de barème de tarage, de gestionnaire ; déplacement de l'échelle, du capteur, de la section de jaugeage ; modification de la géométrie de la section de mesure (envasement, incision, prélèvement de matériaux, recalibrage). C'est souvent autour de ces périodes charnières que pourront être constatés des biais systématiques.

Des investigations complémentaires peuvent ensuite être réalisées pour critiquer les données à partir d'informations exogènes. On s'intéressera notamment à la cohérence spatio-temporelle de la série des débits :

– *comparaison des débits entre stations voisines* : comparaison visuelle des chroniques de stations (amont-aval), comparaison du débit spécifique de bassins voisins, test d'homogénéité d'une série par rapport à une autre (double cumul, écarts cumulés, ellipses de Bois, tests de segmentation, Lang *et al.*, 2003), en raisonnant sur les débits mensuels ou les variations de débits journaliers (pour limiter le phénomène d'auto-corrélation).

– *contrôle de cohérence* : bilans hydrauliques d'ouvrages voisins, comparaison des volumes annuels précipités et des volumes annuels écoulés (si besoin est, en année hydrologique), comparaison de chroniques (figure 2.3), reconstitution de données à l'aide d'un modèle hydrologique (modèle GR : pluie-ETP-débit ; modèle Mordor : pluie-température-neige-débit).

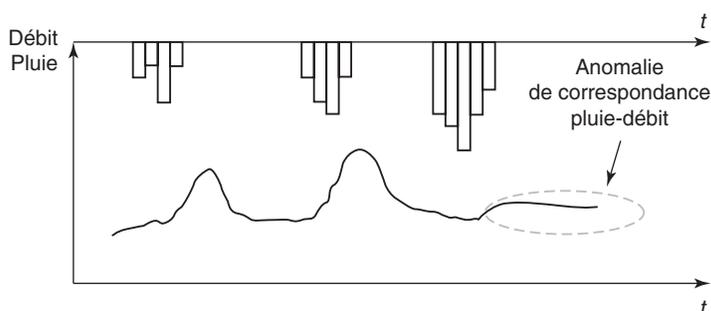


Figure 2.3. Exemple d'anomalie détectée visuellement sur deux chroniques (tiré de Lang, 1995).

Les deux séries d'analyse locale (visualisation de la chronique et qualité de la chaîne de traitement des débits) font partie du travail de validation des données effectué par les producteurs de données, avant archivage dans la banque Hydro. On se rapprochera utilement des gestionnaires hydrométriques pour recueillir ces informations, et évaluer avec eux la qualité des données de débit. Il est rarement inutile de procéder, en liaison avec les gestionnaires, à une critique plus approfondie de certaines données pour corriger l'extrapolation d'une courbe de tarage ou la valeur du débit de crue d'un épisode particulier. Par ailleurs, la cohérence des observations faites à des stations voisines appartenant au même bassin, voire à la même rivière, n'est pas systématiquement assurée, et mérite d'être vérifiée à l'occasion de projets importants.

Représentation d'un échantillon

L'échantillonnage

On désigne par échantillonnage l'opération qui consiste à extraire d'une chronique l'information nécessaire et suffisante pour étudier statistiquement les variables que l'on veut analyser. Dans le cadre d'un PPRI, il s'agit d'évaluer le débit des crues « les plus fortes connues », de sélectionner la plus forte et de constater son occurrence. La valeur du débit de la plus forte crue connue et celles d'autres crues remarquables de la période analysée seront alors comparées à l'estimation de la crue centennale. On s'intéressera prioritairement au débit de pointe de crue pour déterminer la valeur du débit centennal. Nous verrons au chapitre 3 que d'autres variables que la pointe de crue pourront être extraites d'hydrogrammes de référence (débit moyen maximal sur différentes durées) lorsqu'il sera nécessaire de réaliser une simulation hydraulique en régime transitoire (cf. § *Construction d'hydrogrammes de référence*).

L'échantillon des débits de crue sur lequel les statistiques doivent être effectuées peut être construit de différentes façons, en sélectionnant (cf. *infra*) :

- le maximum de chacune des années de la période d'observation ;
- les maximums saisonniers, voire les maximums mensuels ;
- les débits qui excèdent un seuil de sélection.

Le maximum de chacune des années de la période d'observation

Bien que déconseillée, cette méthode est souvent employée. Elle présente l'avantage de la simplicité, mais elle néglige de l'information, et peut conduire à mélanger des crues non homogènes.

On note X^* la valeur maximale annuelle de la variable X étudiée. Le quantile x_p est la valeur ayant la probabilité p de ne pas être dépassée par le maximum annuel :

$$F(x_p) = \text{Prob}(X^* < x_p) = p \quad \text{et} \quad F(x_p) = \int_{-\infty}^{x_p} f(x) dx. \quad (2.2)$$

La période de retour $T_x(x_p)$ correspond à l'intervalle de temps moyen séparant deux années successives, où le maximum annuel est supérieur à la valeur de référence x_p . L'indice x est relatif au mode d'échantillonnage par valeur maximale annuelle. La période de retour T_x est liée à la probabilité au non-dépassement p par la relation :

$$p = 1 - \frac{1}{T_x(x_p)}. \quad (2.3)$$

Afin d'élargir la taille de l'échantillon, et dans le but de ne pas négliger des fortes valeurs observées durant la même année, il est possible de sélectionner les k plus fortes valeurs par an (supposées indépendantes), notées X_k (le plus souvent $k = 2$ ou 3). Si l'on note $G_k(x) = \text{Prob}[X_k < x]$, la distribution de l'échantillon formé par les k maxima annuels, la loi de distribution annuelle est donnée par :

$$F(x) = [G_k(x)]^k. \quad (2.4)$$

Les maximums saisonniers, voire les maximums mensuels

La loi annuelle se déduira du produit des différentes lois saisonnières ou mensuelles. L'information traitée est dans ce cas plus conséquente, mais elle peut contenir des valeurs maximales très faibles (par exemple un échantillonnage mensuel sur un cours d'eau qui présente des étiages sévères) qui ne facilitent pas l'ajustement de lois de probabilité.

Si l'on note $G_{\text{mois } N^0i}(x) = \text{Prob}[(X^*)_{\text{mois } N^0i} < x]$, la loi de distribution de l'échantillon formé par la plus forte valeur du mois N^0i (maxima mensuels), la loi de distribution annuelle est donnée par :

$$F(x) = \prod_{i=1}^{12} G_{\text{mois } N^0i}(x). \quad (2.5)$$

Les débits qui excèdent un seuil de sélection

L'approche par valeurs supérieures à un seuil présente l'intérêt d'une sélection plus rationnelle des valeurs de crue, avec un nombre plus élevé de valeurs et la possibilité de retenir uniquement les plus fortes valeurs. Ainsi, certaines crues maximales annuelles de faible ampleur ne sont pas sélectionnées. En contrepartie, il est nécessaire de choisir le seuil d'extraction S de façon à ce que les crues retenues soient indépendantes, et d'imposer des contraintes d'espacement entre deux épisodes. On trouvera dans le chapitre 5 un exemple d'application de ce mode d'échantillonnage sur la série de la Moselle à Rupt-sur-Moselle, et en annexe 2, § *L'échantillonnage*, des informations sur la façon de constituer l'échantillon des valeurs supérieures au seuil.

On note X_s la valeur maximale d'un épisode supérieur à un seuil. Le début et la fin de chaque épisode correspondent à un franchissement du seuil de référence S (respectivement en montée et en descente). Le quantile x_p est la valeur ayant la probabilité p de ne pas être dépassée au cours d'un épisode supérieur au seuil S :

$$G_s(x_p) = \text{Prob}(X_s < x_p) = p \quad \text{et} \quad G_s(x_p) = \int_{-\infty}^{x_p} g(x) dx. \quad (2.6)$$

La période de retour $T_s(x_p)$ correspond à l'intervalle de temps moyen séparant deux dépassements successifs de la valeur de référence x_p , au cours d'un épisode supérieur au seuil S . L'indice s est relatif au mode d'échantillonnage par valeurs supérieures à un seuil S . La période de retour T_s est liée à la probabilité p par la relation :

$$p = 1 - \frac{1}{[\mu T_s(x_p)]}, \quad (2.7)$$

où μ est le nombre moyen annuel d'épisodes.

La relation entre les distributions issues des deux modes d'échantillonnage (pour passer de la distribution des valeurs supérieures à un seuil à la distribution annuelle) est la suivante (Lang *et al.*, 1999) :

$$F(x) = \exp\{-\mu[1 - G_s(x)]\}, \quad (2.8)$$

Soit, en termes de période de retour :

$$1 - \frac{1}{[T_x(x)]} = \exp\left[-\frac{1}{T_s(x)}\right]. \quad (2.9)$$

Les horsains ou valeurs singulières

Le problème des horsains ou valeurs singulières (*outliers* en anglais) est largement abordé par les mathématiciens et statisticiens. Différents tests ont été développés : régression diagnostic, techniques de *clustering*, M-estimators... Ces techniques recherchent prioritairement la détection d'erreurs de mesure, et ne répondent pas directement à la préoccupation des hydrologues pour lesquels le mot *outlier* n'a pas exactement la même signification.

Il faut d'abord vérifier comment une valeur extrême a été obtenue. Le niveau de la crue a-t-il été observé ou reconstitué ? Est-il compatible avec d'autres laisses de crues ? La courbe de tarage a-t-elle été fortement extrapolée, et si oui, de façon conventionnelle ? (cf. § *Le débit : la variable à étudier*). On doit ensuite apprécier la réalité physique de la valeur extrême. Ce point peut être vérifié par une analyse historique (existence de crues comparables dans le passé), une analyse régionale (ordre de grandeur des records de crue dans la région) ou une analyse hydrologique (caractère exceptionnel des pluies à l'origine de la crue). Si ce débit est physiquement acceptable, le problème réside ensuite dans le traitement statistique de cet horsain. L'hypothèse habituelle consiste à supposer que l'ensemble des valeurs d'un échantillon sont issues d'une même population. L'hydrologue peut envisager deux solutions : exclure la valeur singulière de l'ajustement, auquel cas on aboutit à une estimation souvent excessive de la période de retour de cet événement, ou la garder et retenir une loi de probabilité à plus de deux paramètres, mais dont la détermination est généralement basée sur l'estimation des moments d'ordre 3 ou 4, très sensibles à la présence d'un horsain dans l'échantillon. Cependant, une solution intermédiaire consiste à garder la valeur singulière, mais à retenir une méthode d'estimation des paramètres de la distribution peu sensible aux valeurs extrêmes.

Il est préférable de baser le diagnostic sur l'appartenance d'une valeur singulière à la même population que le reste des valeurs de l'échantillon à partir d'un test statistique donnant une indication sur le caractère plausible de cette hypothèse. Le Water Resources Council (1981) donne une méthode de détection des *outliers* basée sur la valeur du coefficient d'asymétrie $C_s = \frac{m_3}{(m_2)^{3/2}}$ (rapport du moment centré d'ordre 3 au moment centré d'ordre 2 à la puissance 3/2) d'un échantillon de valeurs maximales annuelles de débit. Si le coefficient d'asymétrie est inférieur à 0,4, le test n'est pas activé. Dans le cas contraire, l'équation 2.10 peut être appliquée pour la détection des *outliers* :

$$Y = m_1 + K_n \cdot m_2^{0,5}, \quad (2.10)$$

où :

m_1 et m_2 représentent respectivement la moyenne et la variance du logarithme décimal des débits maxima annuels ;

K_n est tabulé. Pour une taille d'échantillon n respectivement de 10, 20, 30, 50, K_n vaut 2,036, 2,385, 2,563 et 2,768 au niveau de signification de 10 %.

Si le logarithme décimal de la plus forte valeur de l'échantillon est supérieur à Y , cette valeur sera considérée comme un *outlier*.

Dans un premier temps, les valeurs singulières (il peut n'y en avoir qu'une) seront exclues de l'échantillon avant d'en faire l'analyse statistique. On s'intéressera ensuite aux horsains dans le cadre d'un traitement statistique particulier, en les considérant notamment

comme valeur historique (cf. chapitre 3, § *Analyse des crues historiques*). En définitive, on évitera d'éliminer simplement une valeur singulière sous prétexte qu'elle « perturbe » l'analyse statistique (“Those who simply disregard an inconvenient flood, muttering ‘outlier’, should be chained to the river-bank”, extrait de Reed *et al.*, 1999, 1 : 34). Inversement, ce type de valeur ne peut à lui seul remettre en cause l'estimation des crues de fréquence rare (10 ans $< T < 100$ ans), sans investigation détaillée (critique de la valeur estimée, analyse régionale de cohérence, étude historique).

Construction d'un graphique de fréquence cumulée

La construction d'un graphique de fréquence est un mode de représentation privilégié d'un échantillon statistique : il synthétise le comportement de la variable, et il permet de vérifier l'allure globale de la distribution (comportement gaussien, exponentiel, super-exponentiel) pour renseigner sur le type de loi de probabilité bien adapté au comportement de l'échantillon.

La construction d'un graphique de fréquence est incontournable lorsque l'on effectue une étude statistique. Elle consiste à ordonner les valeurs d'un échantillon (x_1, \dots, x_N) de façon croissante :

$$x_1 \leq x_2 \leq \dots \leq x_i \leq \dots \leq x_N, \quad (2.11)$$

et à représenter sur un graphique en ordonnée la valeur x_i , et en abscisse :

$$F_{ND}(i) = \text{Prob}(X < x_i), \text{ la fréquence au non-dépassement,}$$

ou

$$F_D(i) = \text{Prob}(X > x_i), \text{ la fréquence au dépassement,}$$

sachant que :

$$F_{ND}(i) + F_D(i) = 1. \quad (2.12)$$

Théoriquement, la formulation de la fréquence empirique dépend de la loi de probabilité censée représenter l'échantillon étudié (cf. annexe 2, § *Estimation des paramètres des distributions*). Les différentes formulations diffèrent essentiellement pour les fréquences d'apparition rares, à savoir pour les plus fortes valeurs de l'échantillon.

Les figures 2.4 à 2.6 donnent plusieurs représentations graphiques de la distribution de fréquence des débits maxima annuels de crue (*QIX*) du Rhône à Perrache, observés sur la période 1900-1997.

La fréquence empirique de dépassement a été calculée par la formule (cf. annexe 2, § *Estimation des paramètres des distributions, Représentation graphique*) :

$$F_D(i) = 1 - \frac{i - a}{N + 1 - 2a}, \quad (2.13)$$

où i représente le rang de la valeur (échantillon classé par valeurs croissantes) et N la taille de l'échantillon (ici $N = 98$), avec $a = 0,3$. On constate sur la figure 2.4, dans la partie relative aux forts débits, une forte courbure de la distribution de fréquence, qui rend difficile toute extrapolation.

Il est d'usage fréquent de tenter de linéariser et/ou de dilater dans la zone des fréquences rares la courbe, en transformant le repère de projection. La transformation, selon la formulation numérique d'un type de loi de probabilité, permet de constater visuellement

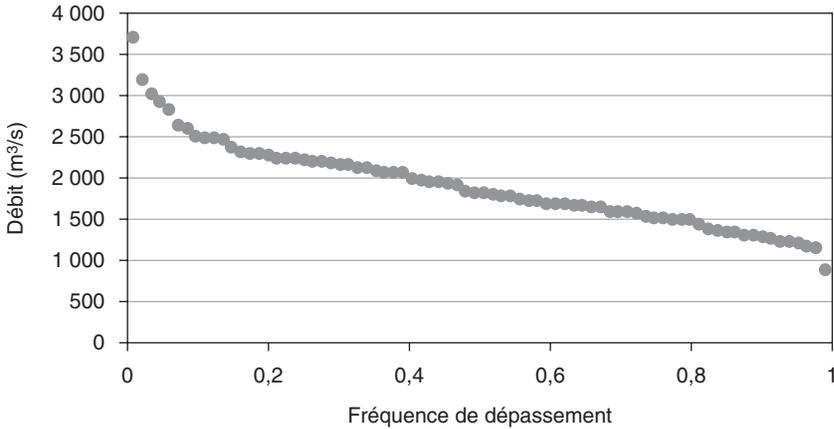


Figure 2.4. Le Rhône à Perrache (1900-1997). Distribution de fréquence empirique des débits maxima annuels de crue (Q_{IX}) (échelle arithmétique des fréquences).

la plus ou moins bonne adéquation avec la loi de probabilité. Toujours pour les débits maxima annuels de crue du Rhône à Perrache, la *figure 2.5* présente la distribution de fréquence après une transformation logarithmique effectuée sur l'axe des fréquences empiriques de dépassement. On remarque la distorsion de l'axe dans la gamme des faibles fréquences (forts débits) et la linéarisation de cette partie de la courbe. Cela permet de mieux appréhender la distribution de fréquence dans la gamme des forts débits.

La transformation logarithmique de l'axe des fréquences de dépassement linéarise un comportement exponentiel. En effet, la fonction de répartition de la loi exponentielle s'écrit :

$$F_{ND}(x) = \text{Prob}(X < x) = 1 - \exp\left[-\frac{(x - x_0)}{a}\right], \quad (2.14)$$

où x_0 et a sont respectivement les paramètres de position et d'échelle de la distribution.

Un changement d'échelle logarithmique sur l'équation 2.14 permet d'obtenir une relation linéaire entre le débit x et $\ln[F_D(x)]$:

$$\ln[F_D(x)] = \ln[1 - F_{ND}(x)] = -\frac{x - x_0}{a}. \quad (2.15)$$

Il est fréquent que les hydrologues construisent les graphiques de fréquence en introduisant la période de retour à la place de la fréquence. Dans notre cas, où nous avons sélectionné un seul maximum par an, l'équation 2.3 conduit à la relation :

$$T(x) = \frac{1}{1 - F_{ND}(x)} = \frac{1}{F_D(x)}, \quad (2.16)$$

où $T(x)$ représente la période de retour exprimée en années.

La transformation logarithmique induit alors la linéarisation de la courbe $x = f(T)$:

$$\ln[T(x)] = -\ln[F_D(x)] = \frac{x - x_0}{a}. \quad (2.17)$$

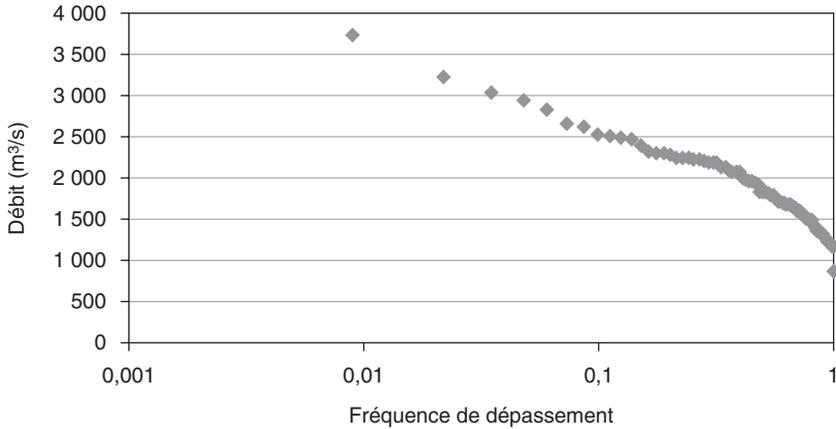


Figure 2.5. Le Rhône à Perrache (1900-1997). Distribution de fréquence empirique des débits maxima annuels de crue (Q_{IX}) (échelle logarithmique des fréquences).

Le mode de représentation de la *figure 2.6* présente les mêmes avantages que ceux de la *figure 2.5*.

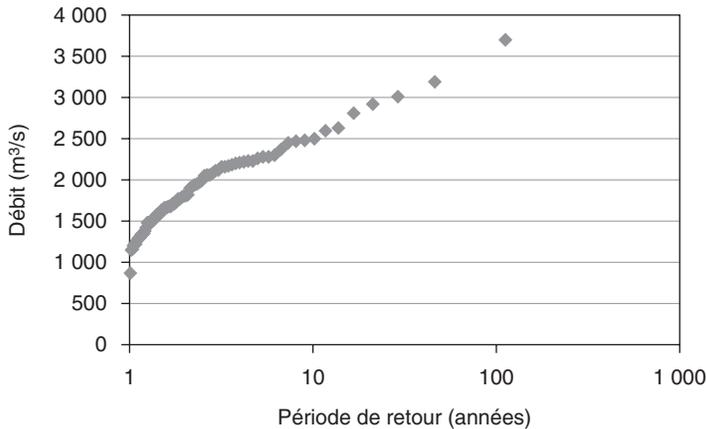


Figure 2.6. Le Rhône à Perrache (1900-1997). Distribution de fréquence empirique des débits maxima annuels de crue (Q_{IX}) (échelle logarithmique des périodes de retour).

Exemples d'ajustement sur la distribution du débit de pointe

Cette section illustre quelques problèmes souvent rencontrés lors de l'ajustement de lois de probabilités à des échantillons. Pour ce faire, nous considérons les échantillons de

débites maxima annuels de différentes stations de mesure des débits, et n'abordons pas ici l'impact de l'échantillonnage. De même, nous ne discutons pas des méthodes d'estimation des paramètres des lois de probabilité et autres difficultés que soulève l'ajustement d'une loi de probabilité à un échantillon (qualité des estimateurs, fréquence empirique, tests d'adéquation, intervalle de confiance...). Ces différents points sont abordés en annexe 2, § *Statistique inférentielle* et § *Applications à l'étude des crues*.

Certaines figures de cette section ont été établies à l'aide du logiciel Hydrolab (Laborde et Mahous, 1998). Notons que d'autres logiciels fréquemment utilisés par les hydrologues francophones sont disponibles, notamment les logiciels Safarhy et Hyfran. Ces logiciels sont rapidement décrits dans l'annexe 1.

Le Rhône à Perrache : un cas d'école

Reprenons l'échantillon des maxima annuels du Rhône à Perrache, sur la période 1900-1997, soit 98 années d'observation. Cet échantillon peut-il être représenté par une loi normale ?

La fonction de répartition de la loi normale, appelée aussi loi de Gauss, s'écrit :

$$N(x_p) = \frac{1}{a\sqrt{2\pi}} \int_{-\infty}^{x_p} \exp\left[-\frac{1}{2}\left(\frac{x-x_0}{a}\right)^2\right] dx = p, \quad (2.18)$$

où x_0 et a sont respectivement les paramètres de position et d'échelle de la distribution, estimés par :

$$\hat{x}_0 = m_1 \quad \text{et} \quad \hat{a} = \sqrt{m_2}, \quad (2.19)$$

avec m_1 et m_2 , respectivement moyenne et variance de l'échantillon.

Le quantile x_p associé à la probabilité p est obtenu par la relation :

$$x_p = x_0 + au_p, \quad (2.20)$$

où u_p est la variable centrée réduite de Gauss, obtenue en inversant la distribution de la loi normale réduite :

$$N_r(u_p) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u_p} \exp\left(-\frac{u^2}{2}\right) du = p \quad \text{et} \quad u_p = N_r^{-1}(p). \quad (2.21)$$

On trouve facilement dans les ouvrages statistiques la table de Gauss, donnant la fonction de répartition $N_r(u)$. On a par ailleurs la correspondance :

$$N(x) = N_r\left(\frac{x-x_0}{a}\right). \quad (2.22)$$

La *figure 2.7* montre la distribution empirique des crues et la distribution normale ajustée sur l'échantillon des 98 crues de la période 1900-1997.

Un changement de l'échelle des abscisses du graphique de la *figure 2.7* ($F(x)$, x) en un graphique ($u = N_r^{-1}[F(x)]$, x) permet de mieux juger de l'adéquation entre :

- la distribution empirique : valeurs $F_{ND}(i)$ remplacées par $N_r^{-1}[F_{ND}(i)]$;
- la distribution théorique : valeurs $N[x(i)]$ remplacées par :

$$N_r^{-1}[N[x(i)]] = N_r^{-1}\left[N_r \cdot \frac{x(i) - x_0}{a}\right] = \frac{x(i) - x_0}{a}.$$

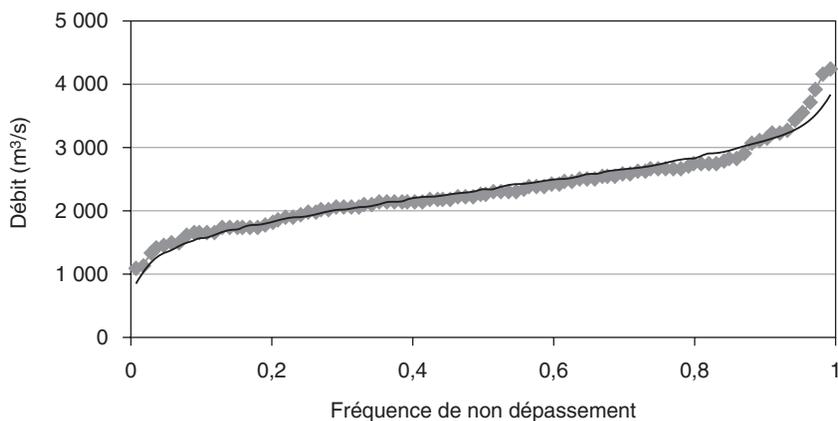


Figure 2.7. Distribution du débit maximal annuel de crue du Rhône à Perrache. Ajustement à une loi normale (échelle arithmétique des fréquences).

La linéarisation de la loi normale est appelée droite de Henry :

$$u = \frac{x - x_0}{a} . \quad (2.23)$$

On constate l'impact de cette transformation entre la *figure 2.7* (coordonnées arithmétiques) et la *figure 2.8* pour laquelle l'axe des fréquences est exprimé en variable centrée réduite de la loi normale (échelle gaussienne des fréquences). La droite de la *figure 2.8* correspond à l'ajustement de la loi normale d'après les équations 2.19 et 2.23. Les intervalles de confiance à 80 % sont aussi reportés. On remarque la mauvaise adéquation de la distribution de fréquence de l'échantillon à la loi normale : courbure de la distribution

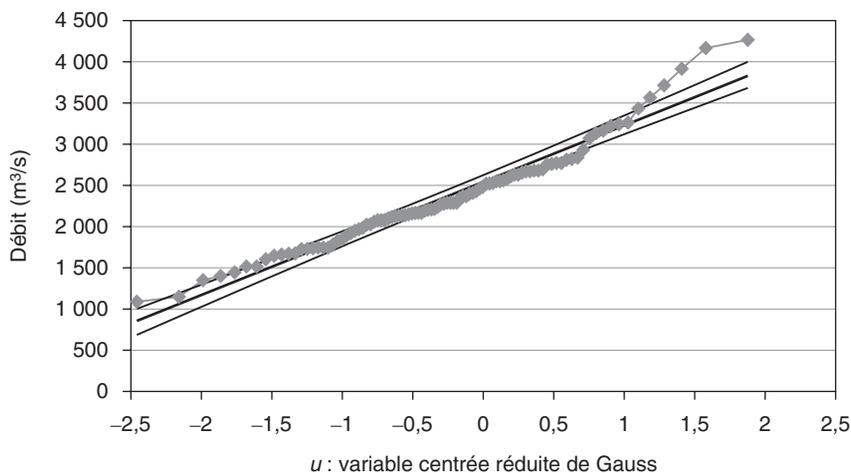


Figure 2.8. Distribution du débit de pointe maximal annuel sur le Rhône à Perrache. Ajustement à une loi normale (échelle gaussienne des fréquences).

expérimentale, nombreux points à l'extérieur des bornes de l'intervalle de confiance. Ce simple examen graphique nous invite à ne pas retenir la loi normale pour représenter les débits de crue du Rhône à Perrache.

En revanche, on notera sur la *figure 2.9* la bonne adéquation de l'échantillon à une loi log-normale. Le calcul de la droite théorique est effectué comme précédemment, mais avec une transformation logarithmique des débits. Pour linéariser la courbe théorique, il convient de construire un repère avec une échelle logarithmique des débits (en ordonnées) et une échelle de Gauss en abscisses (*figure 2.10*). La lecture des estimations de quantiles de crue (débits de différentes fréquences) se fait directement sur ce graphique. Par exemple :

- crue décennale : $\widehat{Q}_{0,9} = 3\,140 \text{ m}^3/\text{s}$ ($u_{0,9} = 1,28$) ;
- crue centennale : $\widehat{Q}_{0,99} = 4\,110 \text{ m}^3/\text{s}$ ($u_{0,99} = 2,33$).

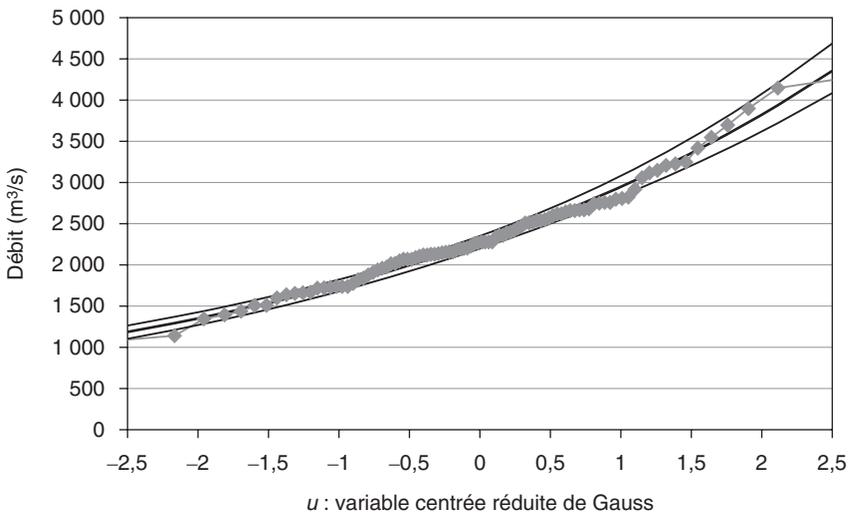


Figure 2.9. Distribution du débit de pointe maximal annuel sur le Rhône à Perrache. Ajustement à une loi log-normale (échelle arithmétique des ordonnées, échelle gaussienne des fréquences).

Le calcul des bornes de l'intervalle de confiance à 80 % donne un encadrement de ces quantiles :

- crue décennale : $\text{Prob}[Q_{0,9} \in [3\,000 ; 3\,280]] = 0,80$;
- crue centennale : $\text{Prob}[Q_{0,99} \in [3\,850 ; 4\,380]] = 0,80$.

Ainsi, nous pouvons dire que, selon l'échantillon traité et en utilisant la loi log-normale, le débit décennal du Rhône à Perrache est estimé à $3\,140 \text{ m}^3/\text{s}$, et que l'on a 80 % de chance pour que la vraie valeur soit comprise entre $3\,000 \text{ m}^3/\text{s}$ et $3\,280 \text{ m}^3/\text{s}$. Si l'on avait pris un intervalle de confiance plus large, par exemple à 90 %, le calcul d'incertitude d'échantillonnage donnerait la fourchette de $2\,960$ à $3\,330 \text{ m}^3/\text{s}$ pour le débit décennal.

Nous pouvons également vérifier (*figure 2.11*) la bonne adéquation de l'échantillon des débits de pointe du Rhône à Perrache avec une loi de Gumbel. La fonction de répartition de la loi de Gumbel s'écrit :

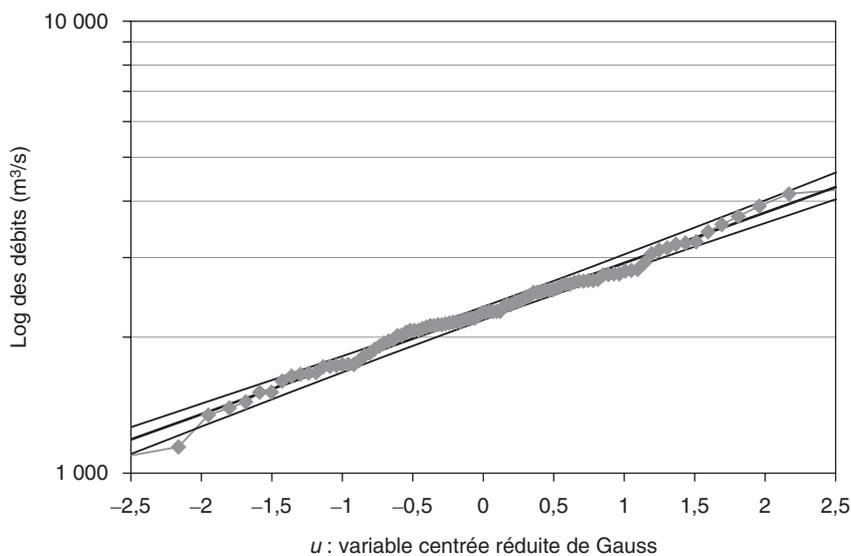


Figure 2.10. Distribution du débit de pointe maximal annuel sur le Rhône à Perrache. Ajustement à une loi log-normale (échelle logarithmique des ordonnées, échelle gaussienne des fréquences).

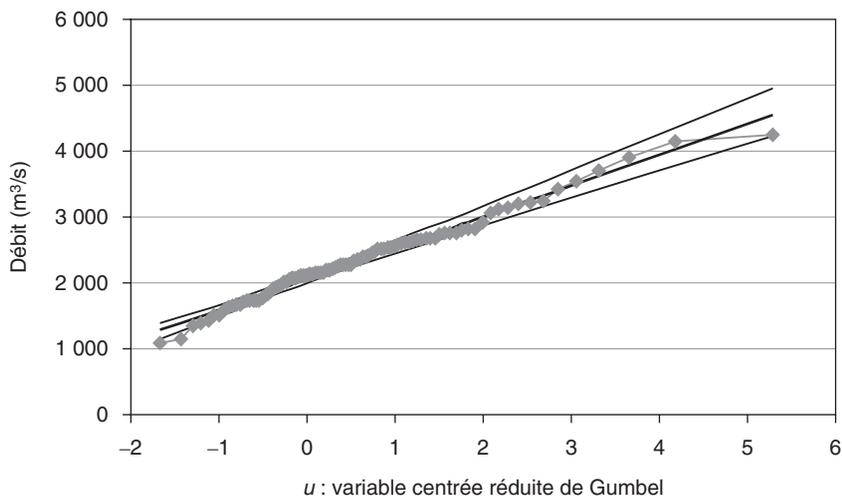


Figure 2.11. Distribution du débit de pointe maximal annuel sur le Rhône à Perrache. Ajustement à une loi de Gumbel.

$$F(x_p) = \text{Prob}(X < x_p) = \exp\left[-\exp\left(-\frac{x_p - x_0}{a}\right)\right] = p, \quad (2.24)$$

où x_0 et a sont respectivement les paramètres de position et d'échelle de la distribution, estimés avec la méthode des moments par :

$$\hat{x}_0 = m_1 - 0,45\sqrt{m_2} \quad \text{et} \quad \hat{a} = \frac{\sqrt{6}}{\pi}\sqrt{m_2}. \quad (2.25)$$

Le quantile x_p associé à la probabilité p est obtenu par la relation :

$$x_p = x_0 + au_p, \quad (2.26)$$

où u_p est la variable réduite de Gumbel, obtenue en inversant la distribution de Gumbel :

$$u_p = \frac{x_p - x_0}{a} = -\ln[-\ln(p)]. \quad (2.27)$$

Ainsi, la double transformation logarithmique des fréquences conduit à une linéarisation du graphique de Gumbel (*figure 2.11*), et des estimations de quantiles de crue voisines de celles obtenues avec la loi log-normale :

– crue décennale :

$$\widehat{Q}_{0,9} = 3\,120 \text{ m}^3/\text{s} \quad (u_{0,9} = 2,25) \quad \text{Prob}[Q_{0,9} \in [2\,975 ; 3\,310]] = 0,80 ;$$

– crue centennale :

$$\widehat{Q}_{0,99} = 4\,230 \text{ m}^3/\text{s} \quad (u_{0,99} = 4,60) \quad \text{Prob}[Q_{0,99} \in [3\,955 ; 4\,580]] = 0,80 .$$

Remarque : la loi de Gumbel montre une tendance linéaire asymptotique sur un graphique dont l'axe des abscisses est exprimé en logarithme de la période de retour.

En effet, d'après les équations 2.24 et 2.27 :

$$F(x) = \exp\{-\exp[-u(x)]\}, \quad (2.28)$$

soit :

$$F(x) \underset{u(x) \rightarrow \infty}{\approx} 1 - \exp[-u(x)]. \quad (2.29)$$

D'après l'équation 2.3 :

$$\ln[T(x)] = -\ln[1 - F(x)], \quad (2.30)$$

ce qui donne :

$$\ln[T(x)] \underset{u(x) \rightarrow \infty}{\approx} u(x) = \frac{x_p - x_0}{a} \quad (2.31)$$

qui traduit bien un comportement asymptotique linéaire entre x et $\ln[T(x)]$. Cette approximation est vérifiée pour les fréquences au-delà de 0,8. Par exemple, pour la crue décennale relative à la période de retour $T = 10$ ans et la probabilité $p = 0,9$:

$$\ln(T) = 2,3 \text{ très peu différent de } u_{0,9} = 2,25 .$$

Non-stationnarité ou fluctuations d'échantillonnage : l'Aude à Belviane

L'Aude à Belviane draine un bassin versant de 692 km². Une chronique d'observations est disponible depuis 1914, avec deux périodes de lacunes (1951 à 1955, et 1983 à 1987). Nous disposons donc de 77 années d'observations. La chronologie des maxima annuels des débits est reportée sur la *figure 2.12*. On constate que quatre crues ont dépassé 300 m³/s sur la seconde partie de la période d'observation, alors que ce débit n'a jamais été atteint de 1914 à 1963.

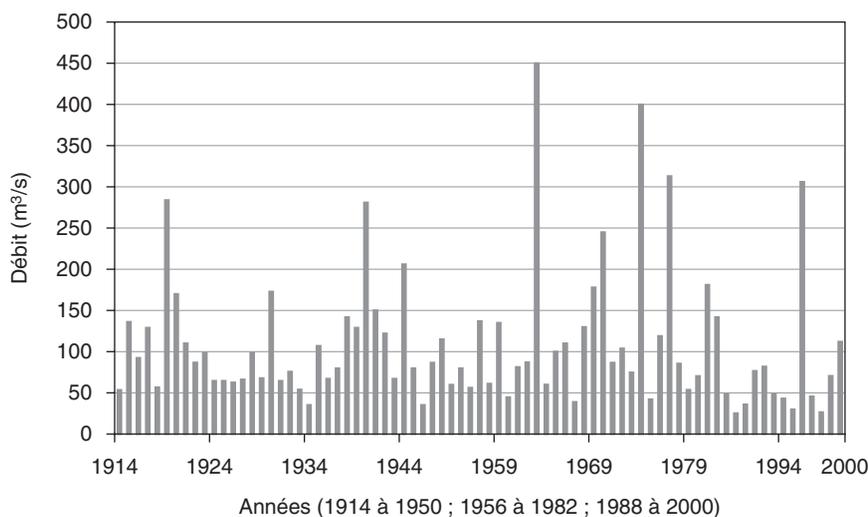


Figure 2.12. L'Aude à Belviane. Chronique des maxima annuels des débits de crue de 1914 à 2000.

Les figures 2.13 et 2.14 montrent que la loi log-normale s'ajuste correctement aux deux échantillons que constituent les périodes 1914-1957 (39 ans) et 1958-2000 (38 ans). Les traits verticaux ($u_{0,9} = 1,28$ et $u_{0,99} = 2,33$) matérialisent la variable centrée réduite associée respectivement à la crue décennale et centennale.

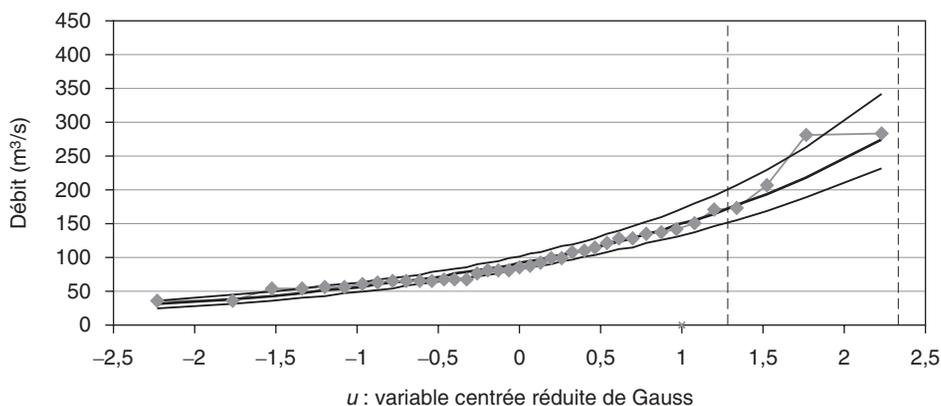


Figure 2.13. L'Aude à Belviane. Ajustement d'une loi log-normale (période 1914-1957).

Cependant, ces deux échantillons conduisent à des estimations quelque peu différentes des quantiles décennaux et centennaux. On constate que les estimations d'une chronique ne sont pas contenues dans les bornes des intervalles de confiance de l'autre chronique (tableau 2.1). Pour l'estimation centennale, les intervalles de confiance ne se recoupent pas ! Simple problème d'échantillonnage ou non-stationnarité de la chronique ?

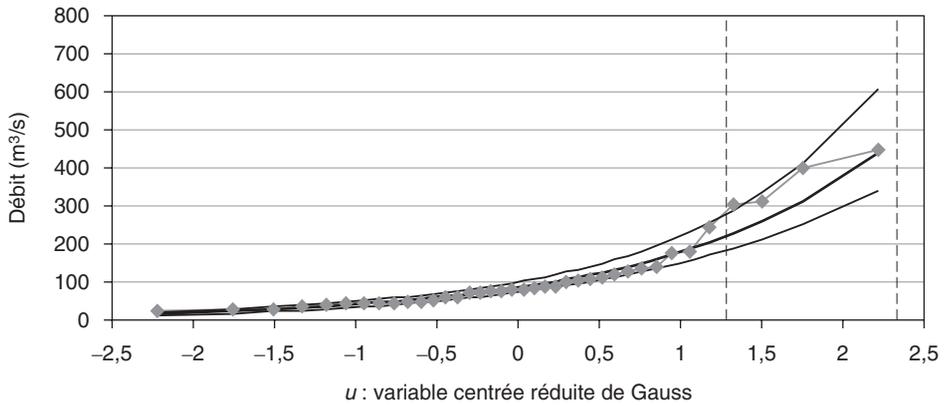


Figure 2.14. L'Aude à Belviane. Ajustement d'une loi log-normale (période 1958-2000).

Tableau 2.1. Quantiles décennaux et centennaux, avec bornes de l'intervalle de confiance à 80 % (en m^3/s).

Période	Estimation décennale	Estimation centennale
1914-1957	[150 ; 170 ; 200]	[240 ; 290 ; 360]
1958-2000	[180 ; 220 ; 275]	[365 ; 475 ; 665]

L'application de quatre tests statistiques (présentation en annexe 2, § *Les tests statistiques, Quelques tests classiques*) ne permet pas d'imputer cette variabilité à une non stationnarité de la chronique (tableau 3.2), quel que soit le risque d'erreur retenu (10, 5 ou 1 %). Par exemple, la statistique du test de Buishand est construite à partir de la somme des écarts à la moyenne. Sous hypothèse de stationnarité, elle doit être proche de zéro, sans atteindre de grandes valeurs. Dans le cas présent, elle vaut 0,6, ce qui est loin d'être une valeur extrême quand on la compare aux différents déciles (respectivement une chance sur 10, sur 20 et sur 100 qu'elle soit supérieure à 1,14, 1,27 et 1,52). En définitive, le problème soulevé par les ajustements sur les deux périodes n'est imputable qu'à l'échantil-

Tableau 2.2. Tests statistiques de détection de non-stationnarité sur la série des débits de pointe (période 1917-2000).

Test	Statistique	Seuil à 10 %	Seuil à 5 %	Seuil à 1 %	Décision sur la stationnarité
Buishand	0,6	1,14	1,27	1,52	Accepté
Kendall	0,99	1,64	1,96	2,58	Accepté
Pettitt	444	480	533	639	Accepté
Spearman	1,23	1,64	1,96	2,58	Accepté

lonnage ! Il suffirait de raisonner sur des intervalles de confiance à 90 % pour avoir un recouvrement des estimations sur les deux périodes (1917-1957 et 1958-2000).

La Berre à Villeneuve-les-Corbières : un événement exceptionnel !

La Berre est un cours d'eau côtier du département de l'Aude. À Villeneuve-les-Corbières, la surface du bassin versant est de 165 km². Une chronique des débits de crue est disponible de 1971 à 2004. La distribution de fréquence empirique des maxima annuels est représentée par la ligne brisée sur la *figure 2.15*. On remarque une valeur remarquable de 750 m³/s (observée le 12 novembre 1999) qui s'éloigne fortement de l'alignement. Cette valeur influence le tracé de la droite théorique de Gumbel, dont les paramètres ont été calculés par la méthode des moments (trait fort) et la méthode du maximum de vraisemblance (trait moyen) (*figure 2.15*). La première méthode montre, dans ce cas, ses limites en proposant un tracé intermédiaire qui ne respecte absolument pas le corps de la distribution.

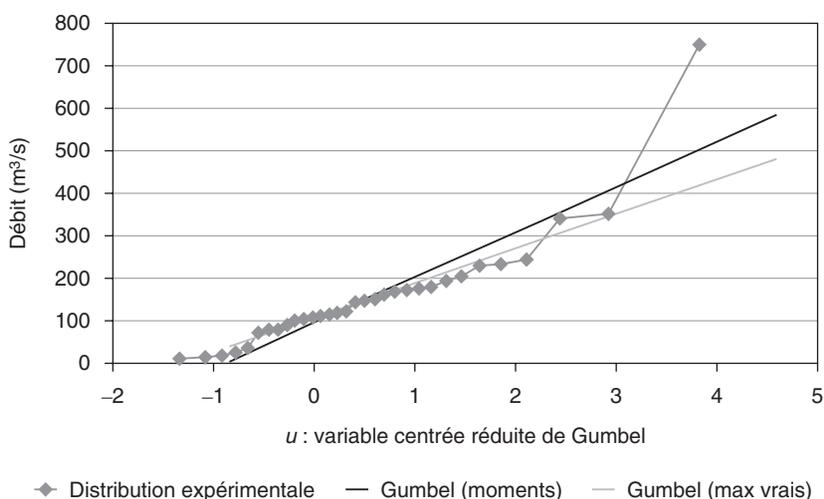


Figure 2.15. La Berre à Villeneuve-les-Corbières. Ajustement d'une loi de Gumbel, échantillon complet.

En excluant la valeur la plus forte, la loi de Gumbel (paramètres ajustés par les moments) représente bien la distribution empirique des débits de pointe (*figure 2.16*). Mais, selon la loi de Gumbel ainsi calée, la période de retour d'un débit de crue de 750 m³/s serait supérieure à 10 000 ans.

Cet exemple illustre deux problèmes souvent rencontrés lors de l'étude statistique des débits.

- *Quelle est la limite d'extrapolation des lois de probabilité sur les débits ?* Étant donné la forte non-linéarité de la relation entre les pluies à l'origine des crues et les débits, les événements courants ne sont que peu informatifs (au sens statistique) pour la prédétermination des débits exceptionnels.

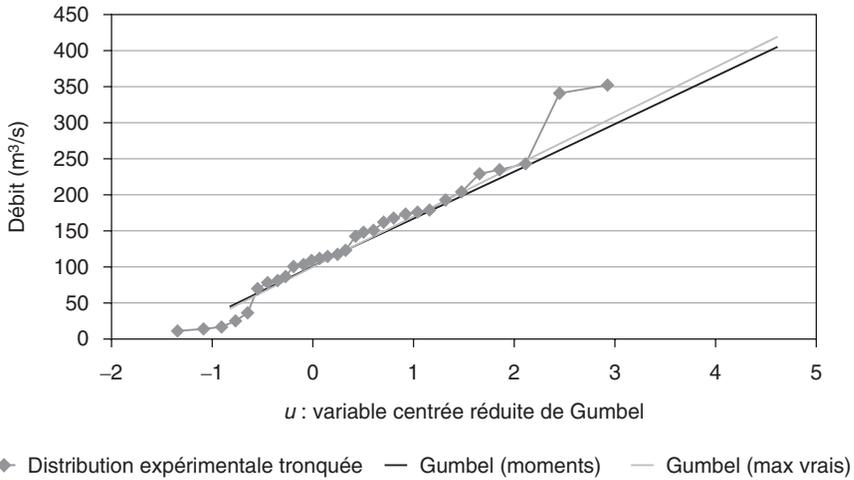


Figure 2.16. La Berre à Villeneuve-les-Corbières. Ajustement d'une loi de Gumbel, échantillon incomplet sans la valeur singulière.

Il est donc vivement conseillé de limiter l'extrapolation des lois de probabilité des débits de crue à la gamme des fréquences moyennes à rares, et certainement pas au-delà de la fréquence de dépassement 10^{-2} , quelle que soit la taille de l'échantillon.

- *Comment traiter le cas d'une valeur a priori exceptionnelle ?* Un événement pluvieux exceptionnel a affecté le département de l'Aude en novembre 1999. Nous ne remettons pas en cause la valeur de ce débit de crue, bien que sa valeur soit certainement entachée d'une forte incertitude.

Trois possibilités :

- écarter la valeur exceptionnelle de l'échantillon ;
- retenir une loi de probabilité à trois ou quatre paramètres qui respectera la fréquence empirique de ce point ;
- retenir une loi de probabilité faiblement paramétrée (deux paramètres), en utilisant une méthode de détermination peu sensible aux valeurs exceptionnelles. On constate sur la *figure 2.15* que la loi de Gumbel avec les paramètres ajustés par la méthode du maximum de vraisemblance respecte bien le corps de la distribution de fréquence, et n'attribue aucun poids particulier à la forte valeur.

Les possibilités (i) et (iii) conduisent sensiblement à la même solution, c'est-à-dire ne pas tenir compte de la valeur singulière. La possibilité (ii) n'est pas une bonne solution : la fréquence empirique de la valeur exceptionnelle est erronée ; retenir une loi qui respecte cette fréquence n'a donc aucun intérêt.

Les lois à plus de deux paramètres doivent être utilisées avec précaution. Confronté à une valeur exceptionnelle, il est souvent édifiant de traiter l'échantillon complet et l'échantillon amputé de la valeur exceptionnelle : les lois surparamétrées conduisent toujours à des tracés très différents pour le calage de ces deux échantillons. La *figure 2.17* illustre bien ce problème.

Le tableau 2.3 récapitule les différentes estimations de la crue décennale et de la crue centennale pour les *figures 2.15, 2.16 et 2.17*.

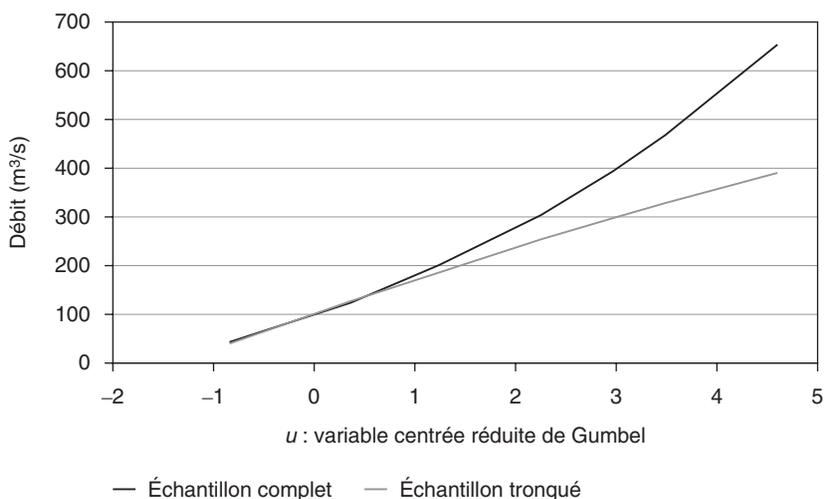


Figure 2.17. La Berre à Villeneuve-les-Corbières. Ajustement d'une loi GEV.

Tableau 2.3. Quantiles décennaux et centennaux (m^3/s), avec ou sans valeur singulière.

Échantillon	Ajustement	Q_{10} ($u_{0,9} = 1,28$)	Q_{100} ($u_{0,99} = 2,33$)	
Complet	Figure 2.15 (Gumbel)	Moments	235	340
		Max. vrais.	210	300
Tronqué	Figure 2.16 (Gumbel)	Moments	185	255
		Max. vrais.	190	260
Complet	Figure 2.17 (GEV/moments pondérés)		210	310
Tronqué			190	260

Recommandations pratiques pour l'ajustement d'une loi de probabilité

Les trois exemples précédents ont fourni des illustrations de graphiques de fréquence qui ont permis d'aborder plusieurs questions relatives à la validation d'une distribution théorique :

- sur le Rhône à Perrache, possibilité de retenir une loi log-normale ou une loi de Gumbel, alors que la loi normale n'est clairement pas adaptée au phénomène étudié (résultat classique pour les crues) ;
- sur l'Aude à Belviane, variabilité des résultats suivant la période de calcul considérée, ce qui renvoie à la notion de distribution d'échantillonnage ;

– sur la Berre à Villeneuve-les-Corbières, difficulté de traitement probabiliste en présence d'une valeur singulière, avec une hétérogénéité des valeurs de l'échantillon.

Face à la variabilité des résultats d'une loi de probabilité à une autre, et au cours du temps avec l'adjonction de nouvelles crues dans l'échantillon d'analyse, le néophyte en retire le plus souvent le sentiment d'une forte imprécision, quitte parfois à préférer n'utiliser que des formules sommaires d'estimation de débit de référence, et ce, sans exploiter des données hydrométriques pourtant disponibles. On observe par ailleurs une confusion dans certaines études « hydrologiques » sur l'interprétation d'un graphe fréquentiel, où l'objectif serait de passer au plus près des points expérimentaux, alors que la position de ceux-ci est imprécise (cf. erreurs de mesure, incertitude d'échantillonnage). On trouvera en annexe 2 des éléments techniques complémentaires permettant de comprendre l'origine des lois de valeurs extrêmes, les méthodes d'estimation des paramètres, le principe de tests statistiques et le calcul de la distribution d'échantillonnage.

Il faut souligner ici que la distribution empirique est considérée à tort comme la référence absolue dans le diagnostic sur l'adéquation d'une loi théorique. Lorsque le graphique de fréquence a été rendu linéaire (cf. exemple figure 2.14), il est tentant de douter de l'adéquation de la loi utilisée dès que trois à quatre points sont au-dessus de la loi ajustée, et ce, même si ces points appartiennent à l'intervalle de confiance. La figure 2.18 donne l'exemple d'une distribution expérimentale qui semble avoir un comportement hyper-exponentiel, alors que les valeurs ont été tirées dans une loi exponentielle. Ce genre de configuration, loin d'être exceptionnel, ne fait que traduire l'élargissement de l'intervalle de confiance pour les faibles fréquences de dépassement. Ainsi, le positionnement des points les plus forts d'un échantillon est sujet à une erreur importante.

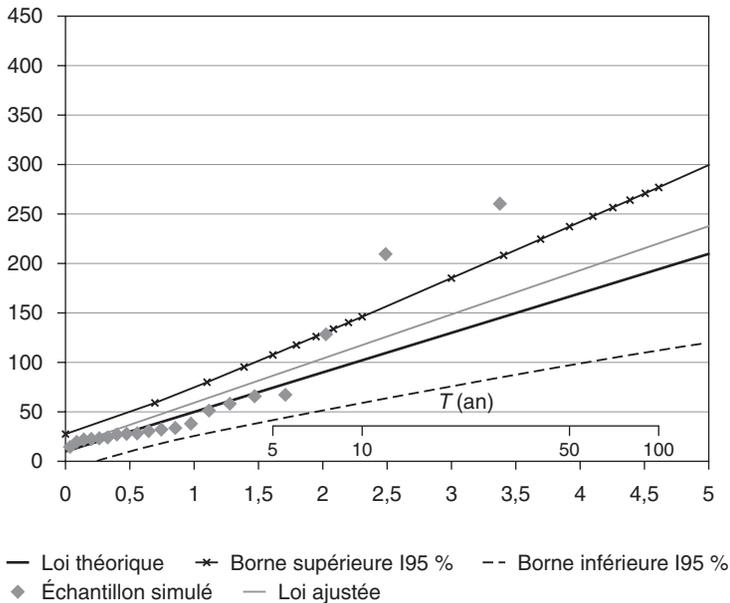


Figure 2.18. Exemple d'échantillon simulé dans une loi exponentielle.

Ce fait est souvent ignoré par les utilisateurs de graphes fréquentiels, qui accordent le même poids, voire davantage, aux valeurs extrêmes qu'au reste de la distribution. Rappelons que le point le plus fort d'un échantillon est entaché d'une erreur sur le débit (cf. incertitudes météorologiques, avec des erreurs relatives de 5 à 10 % pour les débits courants, et des valeurs bien supérieures pour les crues non jaugées) et sur la fréquence empirique. Celle-ci est évaluée à partir du rang de la valeur de l'échantillon et de la taille de l'échantillon (cf. équation 2.13). Or, la crue la plus forte sur 20 ans peut très bien le rester pendant les 80 années qui suivent, auquel cas l'estimation de sa période de retour empirique passe de 29 ans à 143 ans (cf. équations 2.13 et 2.16).

La figure 2.19 récapitule les différentes étapes à suivre lors d'un ajustement d'une loi de probabilité sur les débits de pointe. Rappelons ici que la même procédure pourra être utilisée lors de l'ajustement d'autres variables de débit (courbes QdF, chapitre 3, § *Établissement de courbes débit-durée-fréquence à partir d'observations hydrométriques*), de pluie (méthodes du Gradex et Agregee, chapitre 3, § *Extrapolation de la distribution des crues par la méthode du Gradex et par le modèle Agregee*) ou de valeurs obtenues par simulation (modèle Shypre, § *Extrapolation de la distribution des crues par le modèle Shypre*). Pour chacune des étapes, nous indiquons les tâches minimales à réaliser, sachant qu'il est conseillé d'aborder l'ensemble des points mentionnés, en fonction de l'enjeu de l'étude et du temps disponible.

Lors du *recueil des données* à partir d'une recherche dans la banque Hydro et d'une consultation des services hydrométriques concernés, on s'efforcera d'être le plus exhaustif possible. La *critique des données* permet de détecter des anomalies (lacunes, discontinuités, valeurs aberrantes), et de les corriger quand cela est possible. Les données sont réputées validées (cf. examen de la courbe de tarage). L'*échantillonnage des données* conduit à la sélection de valeurs supposées respecter un certain nombre de propriétés. Dans la mesure du possible :

- (1) raisonner sur des données naturelles reconstituées (dans le cas de valeurs de débit influencées par un ouvrage) ;
- (2) éliminer les valeurs manifestement liées de l'échantillon (même épisode de crue) ;
- (3) tester le caractère saisonnier du régime des crues (si nécessaire, scinder en sous-échantillons de valeurs homogènes) ;
- (4) vérifier la stationnarité du processus (le cas échéant, scinder en deux périodes).

Le cas des valeurs singulières (*outlier* en anglais) est plus délicat, et nécessite un traitement particulier qui peut être abordé par une analyse historique. L'*ajustement fréquentiel* est réalisé en utilisant de préférence une loi de la famille des valeurs extrêmes (respectivement loi généralisée des valeurs extrêmes GEV et loi de Pareto généralisée GPD pour le maximum annuel et les valeurs supérieures à un seuil). Dans le domaine des observations, certaines lois peuvent facilement être éliminées par simple examen visuel du graphique fréquentiel, avec une échelle linéaire (cf. § *Exemples d'ajustement sur la distribution du débit de pointe*). Plusieurs lois peuvent ensuite être retenues, lorsqu'elles donnent des résultats voisins. Le domaine de validité des ajustements des lois de probabilité correspond dans une première approximation aux périodes de retour $T < 2N$, où N est la taille de l'échantillon, sachant que ceci est à moduler en fonction de la qualité de l'ajustement. La *présentation des résultats* sera effectuée avec un nombre de chiffres significatifs compatible avec la précision des résultats. Pour les débits, rappelons qu'une erreur de 5 à 10 % sur la mesure des débits semble un minimum absolu. Pour les périodes de retour T , plutôt

1. Recueil de données

Minimum	<ul style="list-style-type: none"> Chronique de débit sur le site le plus proche Recueil et analyse des études hydrologiques antérieures
<ul style="list-style-type: none"> Recherche de laisses de crue (marques sur le terrain, textes-photos-vidéo, témoignages oraux) Historique des aménagements sur le bassin versant ou le réseau hydrographique susceptibles d'avoir modifié la relation pluie-débit ou la propagation des crues 	

2. Critique des données

Minimum	<ul style="list-style-type: none"> Visualisation de la chronique de débit Examen des courbes de tarage et du nuage des jaugeages Avis du gestionnaire sur la qualité des données
<ul style="list-style-type: none"> Examen du dossier station Visite de terrain Critique des données et validation de l'extrapolation de la courbe de tarage 	

3. Échantillonnage des données

Minimum	<ul style="list-style-type: none"> Typologie des crues (genèse, dynamique spatio-temporelle, caractère saisonnier) Valeurs maximales annuelles ou saisonnières Vérification des propriétés de l'échantillon (caractère aléatoire, indépendance, homogénéité, stationnarité)
<ul style="list-style-type: none"> Valeurs supérieures à un seuil Exploitation de l'information sur les crues historiques 	

4. Ajustement fréquentiel

Minimum	<ul style="list-style-type: none"> Lois Gumbel/GEV sur les valeurs maximales (moment, maximum vraisemblance) Diagnostic à l'aide d'un graphique de fréquence cumulée
<p>Lois exponentielle/GPD sur les valeurs supérieures à un seuil Test et comparaison de différentes lois de probabilité Calcul d'intervalles de confiance</p>	
<p>ATTENTION ! Ne faire ni ajustement « à l'œil », ni par les moindres carrés Ne pas faire d'ajustement sur les hauteurs (sensibles aux changements de courbe de tarage) Ne pas se focaliser sur la (les deux) plus forte(s) valeur(s) Ne pas éliminer sans raison explicite une valeur singulière Ne pas extrapoler trop loin</p>	

5. Présentation des résultats

Minimum	<ul style="list-style-type: none"> Quantiles de crue et ordre de grandeur de la précision du résultat Classe de période de retour des crues historiques les plus fortes connues
<ul style="list-style-type: none"> Analyse de sensibilité aux différentes hypothèses (échantillonnage, loi de probabilité, méthode d'ajustement, extrapolation de la courbe de tarage) Comparaison avec les études antérieures et discussion sur des différences éventuelles 	

Figure 2.19. Étapes à suivre lors d'un ajustement de loi de probabilité sur les débits.

qu'un affichage avec une précision optimiste de 5 ou 10 ans, il est préférable de qualifier la fréquence des crues historiques en distinguant quatre classes : crue fréquente ($T < 2$ ans), crue moyenne ($2 \text{ ans} \leq T < 10$ ans), crue forte ($10 \text{ ans} \leq T < 100$ ans), crue rare ($T \geq 100$ ans) (cf. échelle nationale d'intensité de crue, Lang et Claudet, 2005).

Au-delà du domaine des observations, les incertitudes deviennent importantes, et il est très facile de commettre des erreurs dans un rapport du simple au double (ou davantage) en choisissant une loi ou une procédure mal adaptée au phénomène étudié. La *figure 2.20* donne l'exemple d'une longue série simulée sur 1 000 ans, où l'estimation de la crue centennale ne poserait pas de problème particulier lié à l'échantillonnage (le seuil théorique de la crue centennale a été dépassé sur cet exemple douze fois en 1 000 ans).

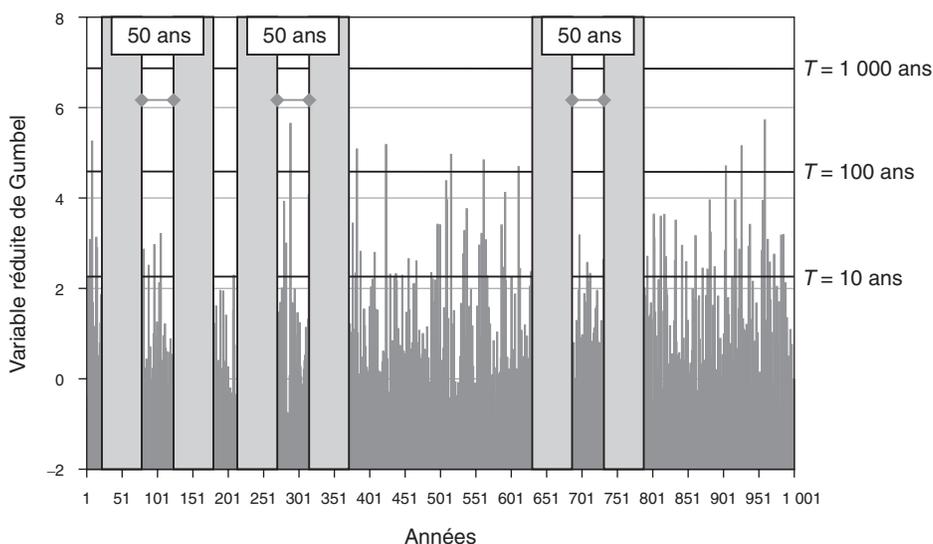


Figure 2.20. Série simulée de valeurs extrêmes dans une loi de Gumbel (1 000 années).

Dans la pratique, on ne dispose le plus souvent que de quelques dizaines d'années d'observation seulement. Suivant la présence ou l'absence dans la fenêtre d'observations de crues exceptionnelles ($T > 100$ ans), le traitement probabiliste classique induira une estimation fort différente de la crue centennale. Cette variabilité peut être approchée par le tracé des intervalles de confiance, qui permet seulement d'appréhender les erreurs d'extrapolation liées à l'échantillonnage, mais plus difficilement celles liées au choix du modèle probabiliste.

Ce problème spécifique de l'extrapolation en fréquence sera abordé au chapitre 3.

Chapitre 3

Méthodes probabilistes spécifiques d'estimation des débits de crue de référence sur un site jaugé

Ce chapitre présente une série d'approches spécifiques à l'hydrologie, en complément du traitement probabiliste des valeurs de pointe de crue (chapitre 2). Il s'agit d'étudier le régime des crues en prenant en compte la dimension spatio-temporelle du phénomène (§ *Établissement de courbes débit-durée-fréquence à partir d'observations hydrométriques*), puis d'exploiter une série d'informations complémentaires pour l'extrapolation aux événements forts et rares (§ *Méthodes d'extrapolation de la distribution des crues*). Une troisième partie aborde la construction d'hydrogrammes de crue de référence, utiles pour la modélisation hydraulique en régime transitoire des inondations (§ *Construction d'hydrogrammes de référence*). Enfin, une série de recommandations pratiques synthétise les points les plus importants à traiter pour l'étude des événements extrêmes et la constitution d'hydrogrammes de crue (§ *Recommandations pratiques pour l'étude des événements extrêmes et la constitution d'hydrogrammes de crue de référence*).

Établissement de courbes débit-durée-fréquence à partir d'observations hydrométriques

Intérêt de l'étude de la durée des crues et des courbes QdF

Le régime des crues est bien souvent réduit à la distribution du débit maximal annuel relatif à la pointe de crue ou au débit journalier le plus fort. Or, cette information ne donne qu'une représentation incomplète de l'aléa hydrologique. À titre d'exemple (Sauquet *et al.*, 2003), la *figure 3.1* présente deux distributions identiques du débit de pointe sur le Volp à Saint-Croix-Volvestre (91 km²), un affluent de la Garonne, et sur la Seille à Nomeny (925 km²), un affluent de la Moselle. L'estimation du débit décennal est dans les deux cas voisine de 100 m³/s. L'examen de quelques hydrogrammes de crue (*figure 3.2*) montre pourtant une forte différence dans la dynamique de crue des deux rivières, avec des événements courts sur le Volp (1 à 2 jours) et des écoulements soutenus pendant une dizaine de jours sur la Seille.

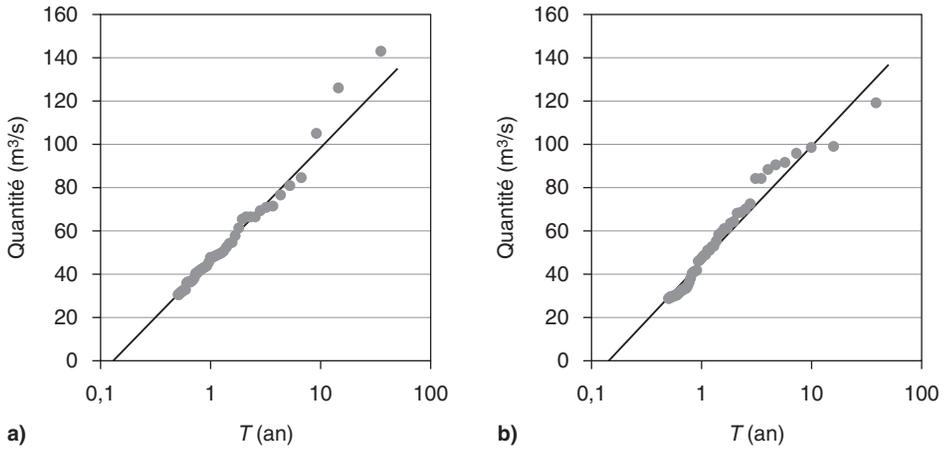


Figure 3.1. Distribution du débit de pointe : (a) le Volp à Saint-Croix-Volvestre ; (b) la Seille à Nomeny.

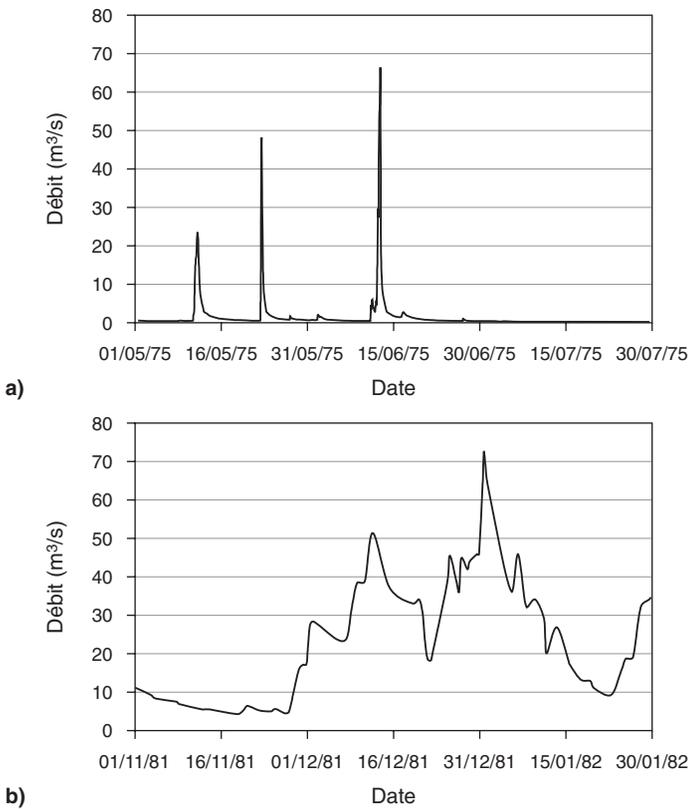


Figure 3.2. Hydrogrammes de crue observées : (a) le Volp à Saint-Croix-Volvestre ; (b) la Seille à Nomeny.

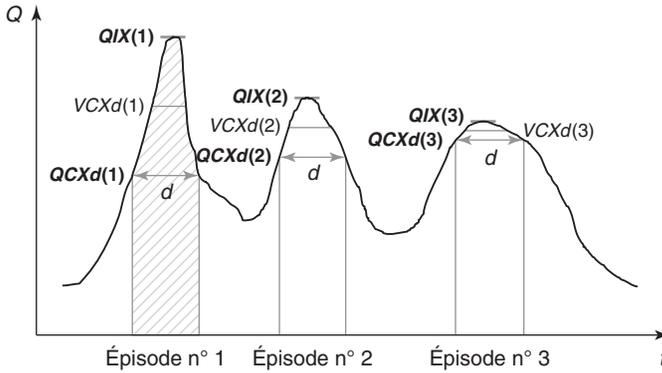


Figure 3.3. Définition des variables de débit de pointe (QIX), de débit moyen maximal ($VCXd$) et de débit seuil maximal ($QCXd$).

Il est préférable de caractériser le régime des crues en un point donné d'une rivière par un ensemble de courbes débit-durée-fréquence, dites courbes Q_dF . Au lieu d'échantillonner uniquement la valeur maximale du débit de pointe (QIX), on s'intéresse maintenant à un débit Q_d (figure 3.3), valeur relative à une durée d , calculée à partir du volume écoulé sur une durée d (débit moyen maximal $VCXd$), ou d'un seuil de débit dépassé de façon continue sur la durée d (débit seuil maximal $QCXd$). Le logiciel Prethy (cf. annexe 1) permet l'extraction d'échantillons multidurées qui peuvent ensuite être exploités pour la construction d'hydrogrammes de référence (cf. § Hydrogramme synthétique monofréquence et § Hydrogramme moyen monofréquence).

Un point quelconque $Q_d(T)$ des courbes Q_dF (figure 3.4) correspond à un quantile de débit maximal calculé sur la durée d , et de période de retour T . La courbe associée à la durée $d = 0$ correspond à la distribution du débit de pointe. La démarche suivie pour établir ces courbes Q_dF consiste à réaliser une analyse fréquentielle à partir d'une série de plusieurs échantillons de valeurs maximales de débits de crue Q_d . Il est possible, pour chaque variable Q_d , d'estimer les paramètres (α , β , γ) de la distribution des crues à partir

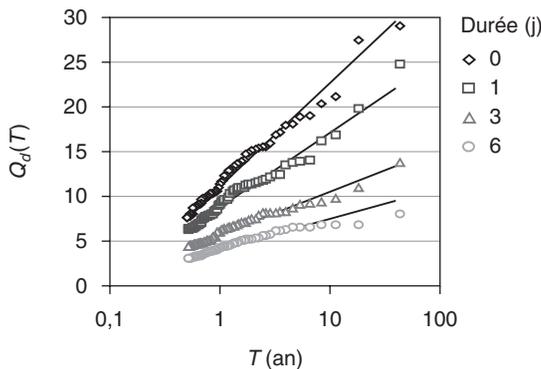


Figure 3.4. Courbes Q_dF : famille de distributions de crues multidurée.

de l'échantillon des valeurs maximales. Le mode de calcul diffère, suivant le mode d'échantillonnage, par valeurs maximales annuelles ou supérieures à un seuil (cf. annexe 2, § L'estimation).

Le modèle QdF convergent

De façon analogue à ce qui est pratiqué pour les courbes intensité-durée-fréquence des pluies, l'étape suivante consiste à lisser les ajustements de façon à ce que l'évolution de la distribution du débit Q_d soit progressive suivant la durée d . Comme pour les pluies (cf. formule de Montana), une formulation analytique va permettre de passer d'une série discrète de quantiles $Q_d(T)$ à une formulation générale à deux variables $Q(d, T)$. Cette consolidation permettra ensuite d'obtenir une cohérence dans la décrue des hydrogrammes de référence (cf. § Construction d'hydrogrammes de référence). Le logiciel Prethy (cf. annexe 1) permet l'ajustement de courbes débit-durée-fréquence d'après le modèle convergent.

Hypothèses et formalisation

Le modèle QdF convergent (Javelle, 2001) est né d'une propriété observée sur un grand nombre de chroniques hydrologiques traitées : les ajustements $Q_d(T)$ tendent tous à se croiser en un même point vers les très faibles périodes de retour. Cette première hypothèse énoncée, une seconde, plus forte, est envisagée, appuyée également par les observations : les lois de probabilité théoriques relatives aux différentes durées se déduisent les unes des autres par affinité orthogonale (figure 3.5), et le point pivot est le point de convergence des distributions théoriques.

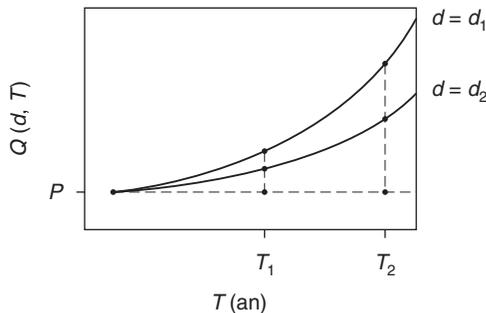


Figure 3.5. Principe de similitude du modèle convergent.

Ces propriétés se traduisent mathématiquement par :

$$\frac{Q(d_1, T_1) - P}{Q(d_1, T_1) - P} = \frac{Q(d_2, T_1) - P}{Q(d_2, T_1) - P}, \text{ ou encore } \frac{Q(d, T) - P}{Q(0, T) - P} = f(d), \quad (3.1)$$

où P est l'ordonnée du point pivot et f une fonction de la seule durée d .

L'examen de nombreux bassins en France métropolitaine (Javelle, 2001), au Canada (Javelle *et al.*, 2002) et à la Martinique (Meunier, 2001) a montré qu'une forme homographique était bien adaptée pour modéliser la relation $f(d)$. Pour les bassins versants où

le module interannuel est négligeable devant les valeurs de crues, le paramètre P peut être fixé à zéro sans compromettre l'efficacité du modèle. D'où la formulation finale :

$$\frac{Q(d, T)}{Q(0, T)} = \frac{1}{1 + d/\Delta}, \quad (3.2)$$

avec Δ homogène à une durée, à définir.

Les quantiles théoriques $Q(d, T)$ se déduisent de la distribution des débits instantanés :

$$Q(d, T) = \frac{Q(0, T)}{1 + d/\Delta}. \quad (3.3)$$

Cette relation qui a été établie pour les quantiles de débit moyen maximal $VCXd(T)$, est exploitée pour obtenir les quantiles de débits seuils $QCX(d, T)$:

$$QCX(d, T) = \frac{Q(0, T)}{(1 + d/\Delta)^2}, \quad (3.4)$$

en supposant que :

$$VCX(d, T) = \frac{1}{d} \int_0^d QCX(\delta, T) d\delta. \quad (3.5)$$

Estimation du paramètre Δ

Le paramètre Δ du modèle QdF convergent est estimé par une méthode itérative (Javelle, 2001) qui consiste à raisonner sur un seul échantillon constitué de l'ensemble des séries expérimentales Q_{d_i} de débit moyen maximal $VCXd$, calculées pour différentes durées $d = d_i$. La relation (3.3) donne une équivalence entre le débit de pointe $Q(0, T)$ et l'expression $Q(d, T) \cdot \left(1 + \frac{d}{\Delta}\right)$. Les distributions expérimentales des N échantillons transformés :

$$x_{d_i}(j) = Q_{d_i}(j) \cdot \left(1 + \frac{d_i}{\delta}\right), \text{ avec } i = 1, \dots, N, \quad (3.6)$$

doivent se confondre avec celles des valeurs instantanées dans le cas où $\delta = \Delta$. La procédure consiste à faire varier δ , et conserver la valeur pour laquelle la dispersion des N distributions pseudo-instantanées x_{d_i} autour de l'échantillon constitués des valeurs :

$$\bar{x}(j) = \frac{1}{N} \sum_{i=1}^N x_{d_i}(j) \quad (3.7)$$

est minimale. La dispersion globale autour de \bar{x} est mesurée par ε , somme des écarts relatifs entre les x_{d_i} et \bar{x} :

$$\varepsilon = \sum_{j=1}^{NV} \sum_{i=1}^N \left(\frac{x_{d_i}(j) - \bar{x}(j)}{\bar{x}(j)} \right)^2, \quad (3.8)$$

où NV correspond au nombre minimal de valeurs sélectionnées par durée. Le critère d'optimisation ne préjuge pas de la loi théorique ajustée, et manipule des écarts relatifs afin de donner autant de poids aux faibles valeurs qu'aux fortes valeurs des échantillons. La der-

nière étape consiste à calculer les paramètres de la distribution du débit de pointe $Q(0, T)$. Ces derniers sont ajustés sur l'échantillon \bar{x} .

Interprétation physique de Δ

La valeur attribuée conditionne la forme des hyperboles dans le plan (Q, d) , selon l'équation :

$$Q(d, T) = \frac{Q(0, T)}{1 + d/\Delta} \quad (3.9)$$

La figure 3.6 permet d'apprécier l'influence du paramètre Δ sur l'allure des courbes QdF de deux bassins versants ayant même distribution du débit instantané $Q(0, T)$, mais deux valeurs distinctes pour Δ . Les quantiles $Q(d, T)$ diminuent rapidement avec la durée lorsque le paramètre Δ est faible, alors que la courbure est peu prononcée quand la durée Δ est élevée. On attend donc que les fortes valeurs du paramètre Δ soient plutôt représentatives de bassins à dynamique lente, pour lesquels les crues durent plusieurs jours, avec des débits peu variables d'un jour à l'autre. En revanche, les bassins qui voient leurs écoulements se tarir rapidement présenteront *a priori* des valeurs faibles de Δ . Cette durée renseigne sur la dynamique du bassin versant en hautes eaux.

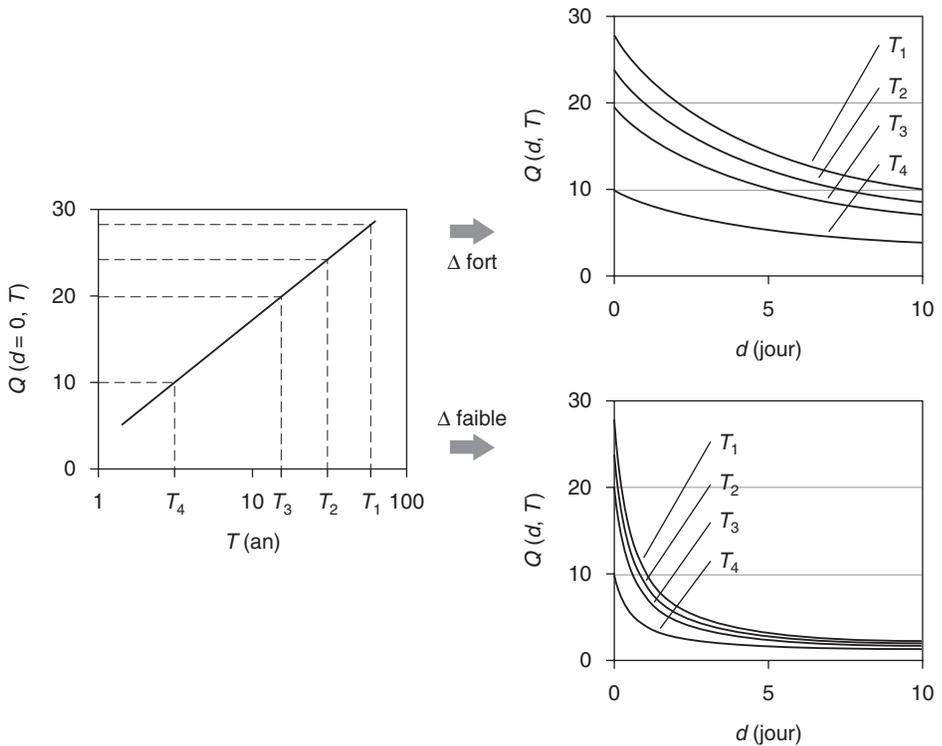


Figure 3.6. Sensibilité des courbes QdF au paramètre Δ .

Reprenons les exemples du Volp et la Seille, qui rappellent la configuration décrite en figure 3.6. Cinq durées sont examinées qui donnent lieu à cinq séries de débits moyens, avec d compris entre 0 et 1 jour pour le Volp, et entre 0 et 18,5 jours. Le modèle local convergent $Q(d, T)$ est appliqué sur les séries de valeurs maximales (figure 3.7). L'examen de la chronique de données brutes de la Seille montre que les débits moyens écoulés sur les durées inférieures à deux jours sont sensiblement équivalents aux valeurs de pointe. Cela se traduit par des distributions des quantiles $Q(0, T)$ et $Q(2,32 \text{ jours}, T)$ relativement proches. En revanche, les ajustements obtenus pour le Volp pour une gamme de durées inférieures à 2 jours sont bien distincts, et révèlent une baisse rapide des volumes écoulés avec la durée. Le quantile $Q(0, T)$ est le double de $Q(1 \text{ jour}, T)$. La durée Δ pour le Volp — à dynamique rapide — est fixée à 24 heures, et est effectivement bien inférieure aux 18,5 jours ajustés pour la Seille à dynamique plus lente. La hiérarchie intuitée dans les valeurs de Δ est ainsi respectée.

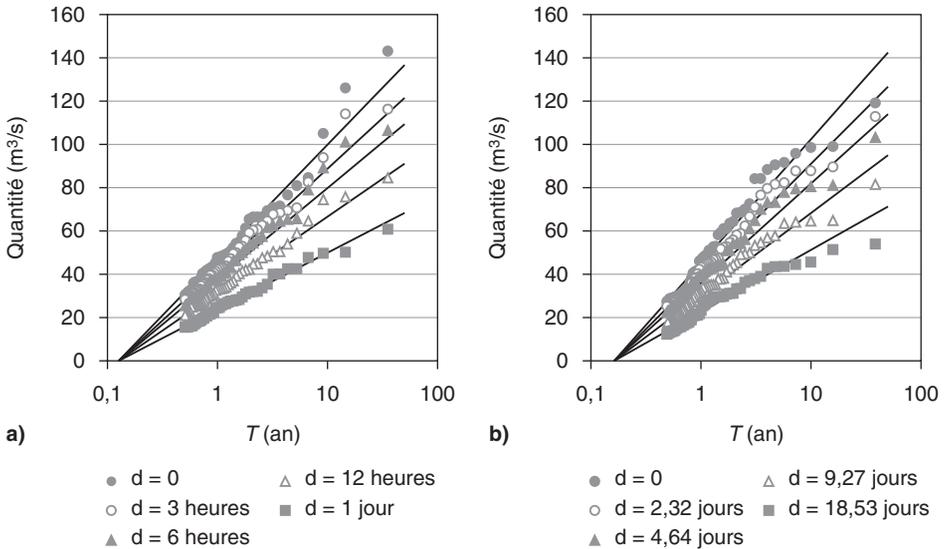


Figure 3.7. Courbes QdF du modèle convergent : (a) le Volp et (b) la Seille.

On note sur les courbes QdF de la Seille un infléchissement au-delà de la période de retour 5 ans, qui conduit à rester prudent dans la validation d'une loi exponentielle pour les crues fortes. Le diagnostic mériterait d'être approfondi, et ce, afin de vérifier si le plafonnement des débits ne vient pas d'un artefact sur l'extrapolation de la courbe de tarage ou d'un phénomène de laminage dû à la présence d'un champ d'inondation.

Méthodes d'extrapolation de la distribution des crues

Nécessité d'informations complémentaires pour l'extrapolation des crues

Nous avons mentionné la question de la distribution d'échantillonnage qui induit une incertitude d'autant plus forte que le quantile de crue x_p à estimer est relatif à une faible probabilité au dépassement ($1 - p$) (cf. fin du chapitre 2). Une seconde incertitude provient du fonctionnement du bassin versant en crue, avec un comportement bien différent entre les crues fréquentes ($T < 2$ ans) et les crues exceptionnelles ($T > 100$ ans). Si l'on excepte les crues strictement thermiques des bassins versants à régime glaciaire, la majorité des crues observées en France sont générées par la pluie, voire par la pluie associée à une fonte du manteau neigeux. Il est classique d'aborder le passage de la pluie en débit, en introduisant la notion de coefficient d'écoulement C qui exprime le rendement de la pluie au cours d'un épisode pluvieux. Le coefficient C , rapport du volume écoulé au volume de pluie précipité sur une même durée, est fortement variable d'une crue à l'autre. L'état hydrique du bassin versant avant la crue, la saison, et la répartition spatiale et temporelle des pluies conditionnent fortement la réponse du bassin versant. Cependant, malgré une dispersion certaine, on note généralement une croissance du coefficient d'écoulement avec le total de pluie de l'épisode. Ce phénomène est parfaitement visible sur la *figure 3.8*, où nous avons reporté le coefficient d'écoulement calculé crue par crue en fonction du total de pluie de l'épisode, pour un petit bassin versant.

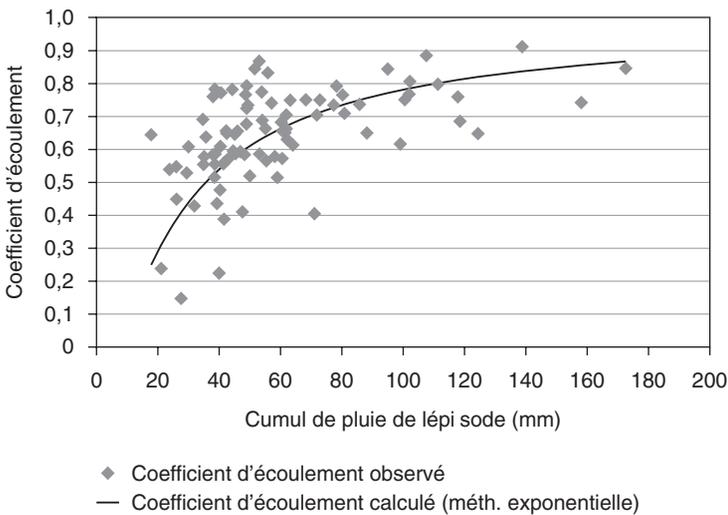


Figure 3.8. Évolution du coefficient d'écoulement avec le total de pluie de l'épisode. Cas d'un petit bassin versant.

Si on utilise la relation entre le coefficient d'écoulement et la pluie pour calculer les volumes écoulés pour une pluie donnée, on aboutit à la tendance qu'illustre la *figure 3.9*. On note le parallélisme asymptotique entre les distributions de pluie et de débit, lié à un

coefficient d'écoulement marginal qui tend vers 1 pour les très fortes intensités de pluie. Par ailleurs, on voit qu'une simple extrapolation des débits dans la gamme de fréquence des observations ($T < 10$ ans) conduirait à une sous-estimation nette des débits de périodes de retour élevées. Cette remarque est fondamentale, et conduit les hydrologues à la plus grande prudence lors de l'utilisation en extrapolation des lois de probabilité calées sur les distributions de fréquence des débits observés, et ce, d'autant plus que nous avons déjà évoqué les problèmes que soulève la métrologie des forts débits, avec le poids abusif de ces derniers quant à la détermination des lois de probabilité... Cette prudence se traduit par la prise en compte d'informations supplémentaires telles que les informations historiques afin de conforter l'extrapolation des crues rares ($10 \text{ ans} < T < 100 \text{ ans}$), ou l'information pluviométrique pour fixer une borne supérieure à la distribution des crues exceptionnelles ($T > 100 \text{ ans}$).

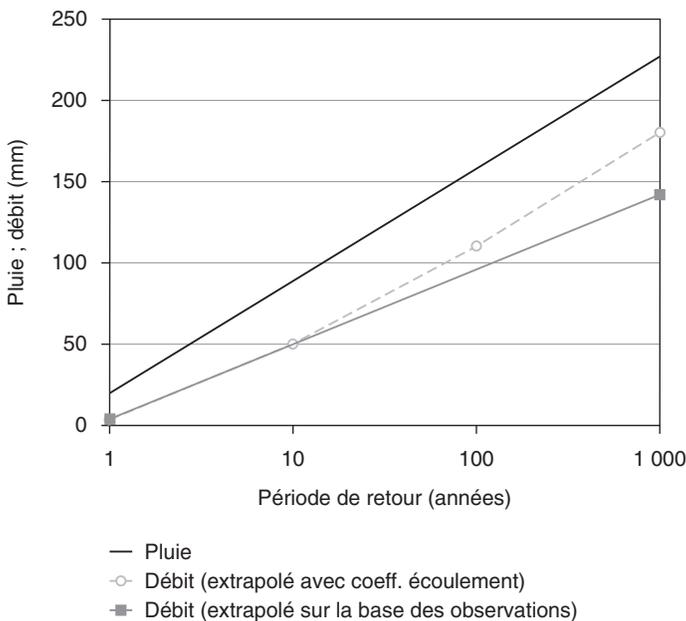


Figure 3.9. Distribution de fréquence des lames précipitées et écoulées (pluies et débits moyens sur une durée de référence) ; cas théorique.

Extrapolation de la distribution des crues par la méthode du Gradex

La méthode du Gradex, développée par EDF (Guillot et Duband, 1967) pour la sécurité des grands barrages, est utilisée couramment en France pour l'estimation des débits de fréquences rare et exceptionnelle. La méthode utilise l'information pluviométrique pour extrapoler la distribution de fréquence des débits. Son succès est en particulier lié à sa (relative) facilité de mise en œuvre qui résulte de la simplification extrême du processus de transformation de la pluie en débit.

Description de la méthode

Cette méthode repose sur trois hypothèses :

– **Hypothèse 1** – La distribution de fréquence des pluies, qui sera utilisée en extrapolation, est à décroissance exponentielle simple. Elle est décrite par une loi de Gumbel ou une loi exponentielle (suivant le type d'échantillonnage). Le gradex $a_p(d)$ est le paramètre d'échelle de la distribution. Il correspond à la pente de la loi de probabilité sur un graphique représentant la pluie en fonction du logarithme de la période de retour. Les valeurs maximales de pluie PXd sont échantillonnées suivant une certaine durée d , égale à la durée moyenne des hydrogrammes.

– **Hypothèse 2** – À partir d'un certain état de saturation du bassin versant, la distribution de la rétention en eau du bassin versant ne dépend plus de la valeur de pluie ou du débit. Cette hypothèse est appliquée sur la durée d : ce qui graphiquement se traduit par une extrapolation, vers les débits de fréquence exceptionnelle, par une parallèle à celle des pluies. Le point de départ de cette extrapolation est fixé à une période de retour T_g comprise entre 10 et 50 ans, selon l'aptitude à l'écoulement en crue du bassin versant.

– **Hypothèse 3** – La forme générale de la fonction de transfert des crues est invariante avec la fréquence, et on peut déduire le débit de pointe de crue du volume écoulé sur la durée d par un simple coefficient d'affinité RXd . Ce rapport est estimé à partir des hydrogrammes observés ; on retient sa valeur moyenne ou sa valeur médiane. On considère que si la durée d a été bien déterminée, le coefficient d'affinité a une valeur de l'ordre de 1,5.

L'expression analytique du quantile de débit de pointe $QIX(T)$ par la méthode du Gradex est la suivante :

$$QIX(T) = QIX(T_g) + RXd \cdot a_p(d)[u(T) - u(T_g)], \quad (3.10)$$

avec :

$$a_p(d) \text{ (m}^3\text{/s)} = \frac{a_p(d) \text{ (mm)} \cdot S \text{ (km}^2\text{)}}{[86,4 \times d \text{ (jours)}]} ; \quad (3.11)$$

$u(T)$ variable réduite de Gumbel ou exponentielle :

– loi de Gumbel :

$$p = \text{Prob}(\text{Max annuel} < x) = 1 - \frac{1}{T} = \exp\left(-\exp\left(-\frac{x - x_0}{a_p(d)}\right)\right),$$

$$u_p = -\ln[-\ln(p)] \text{ et } u(T) = -\ln\left[-\ln\left(1 - \frac{1}{T}\right)\right] ;$$

– loi exponentielle :

$$p = \text{Prob}(\text{Valeur sup-seuil} < x) = 1 - \frac{1}{\mu T} = 1 - \exp\left(-\frac{x - x_0}{a_p(d)}\right),$$

$$u_p = -\ln(1 - p) \text{ et } u(T) = -\ln(\mu T),$$

avec μ le nombre moyen annuel de valeurs sup-seuil.

La méthode est applicable pour des bassins versants dans une large gamme de superficies : de 10 à 10 000 km². Pour les grands bassins versants, il est nécessaire de prendre en compte la variabilité spatiale des pluies. Cela est parfois effectué, en considérant les gradex par poste pluviométrique, et en retenant pour gradex de bassin la moyenne des gradex. Il est cependant préférable de calculer la pluie de bassin par interpolation spatiale

des observations ponctuelles, et de caler ensuite la loi de probabilité sur la chronique de pluie moyenne de bassin ainsi calculée.

Étant donné la forte variabilité du comportement hydrologique des bassins versants ruraux, il est préférable de saisonnaliser la méthode. Les gradex des pluies d'été sont généralement plus importants que ceux de l'hiver, alors que l'on observe l'inverse pour les débits. Pour les bassins montagnards, il est aussi intéressant d'isoler la saison de fonte des neiges. La méthode est appliquée pour les saisons considérées, et la loi annuelle est déduite directement des lois saisonnières.

À titre d'illustration (les valeurs estimées n'étant pas à retenir telles quelles), la méthode du Gradex est appliquée pour le bassin versant de la Garonne à Portet (9 850 km²). Nous disposons de la chronique des débits maxima annuels de 1910 à 1997 et de chroniques de pluies journalières pour dix postes du bassin de 1977 à 1997. La chronique de la pluie de bassin est calculée de 1977 à 1997. Une loi de Gumbel est ajustée à l'échantillon des maxima annuels des pluies journalières et à l'échantillon des maxima annuels de débit (1977-1997). La distribution de fréquence des pluies est alors extrapolée à celle des débits, à partir de la période de retour 10 ans, les pluies et débits étant exprimés dans la même unité (m³/s).

Il semble sur la *figure 3.10* que l'extrapolation par la méthode du Gradex conduise à surestimer la distribution des débits, la pente de la distribution empirique des débits étant nettement inférieure à celle des pluies (gradex).

Nous avons reporté sur la *figure 3.11* l'échantillon complet des débits (1910-1997). On constate que l'extrapolation proposée par la méthode du gradex représente bien les plus forts débits journaliers observés.

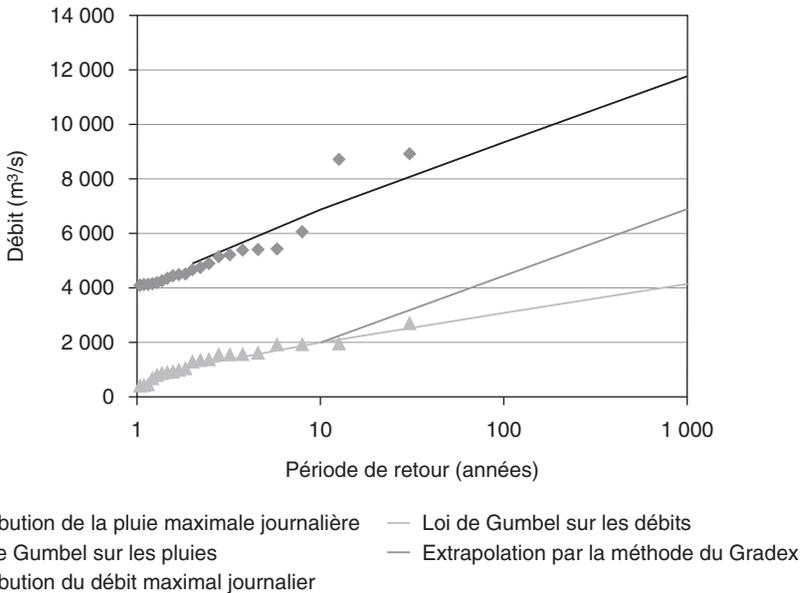


Figure 3.10. La Garonne à Portet. Distribution des valeurs maximales de pluies journalières et de débits journaliers (période 1977-1997).

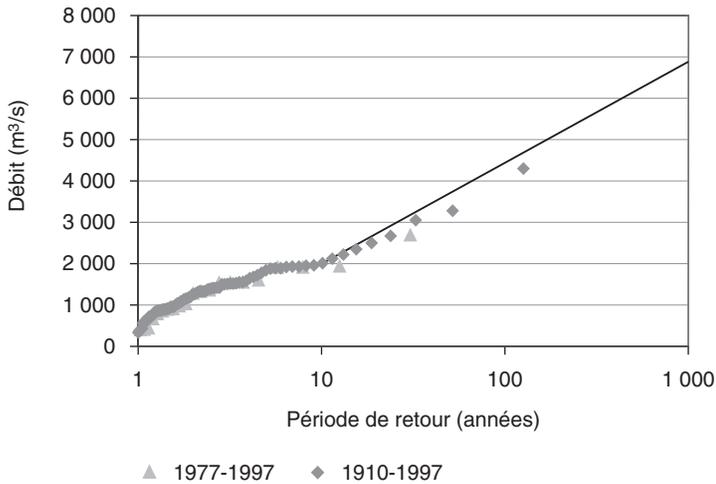


Figure 3.11. La Garonne à Portet. Comparaison des distributions de débit sur les périodes 1977-1997 et 1910-1997.

Dans le cas de petits bassins versants dépourvus de chronique de débit, cette opération n'est pas possible. Une approche régionale basée sur des bassins versants proches et similaires est alors nécessaire. Mais il est aussi admissible d'avoir recours seulement à des synthèses nationales telles que Socose ou Crupedix (cf. chapitre 4). Ces méthodes nécessitent essentiellement des données pluviométriques, et fournissent un ordre de grandeur des débits de pointe de période de retour 10 ans (et aussi 20 ans pour Socose). Il s'avère qu'une erreur même importante sur la crue de retour 10 ans (ou 20 ans) a une influence relativement faible sur la crue de période de retour 1 000 ans ou 10 000 ans, calculée par la méthode du Gradex. Cela permet de qualifier la fréquence d'une crue historique, quand elle est plus forte que la crue centennale.

Dans le cas des bassins versants non jaugés (généralement de superficie réduite), l'évaluation des débits de crue de fréquence rare repose alors quasi exclusivement sur l'information pluviométrique. Celle-ci est fort heureusement plus largement disponible sur une très grande partie du territoire français.

Difficultés d'application de la méthode du Gradex

Hypothèse 1

1) Le choix d'une distribution de pluie à décroissance strictement exponentielle (loi de Gumbel ou exponentielle) conduit à affecter des périodes de retour extrêmement élevées à certains événements effectivement observés en un point quelconque du territoire français. Rappelons que des totaux de précipitations en 24 heures supérieurs à 500 mm ne sont localement pas réellement exceptionnels en France, mais qu'ils restent l'apanage de quelques régions particulières : autour de 1 000 mm dans la région du Canigou (Pyrénées-Orientales) en octobre 1940, 800 mm dans la région de Solenzara (Corse) en octobre 1993, 620 mm à Lézignan-Corbières en novembre 1999, 700 mm dans le Gard en septembre 2002... Cependant, il semble actuellement admis que la méthode n'entraîne pas de biais systéma-

tique par un effet de compensation, la probable sous-estimation des pluies de fréquence exceptionnelle étant compensée par une surestimation du coefficient d'écoulement.

2) Il n'existe pas de règle bien définie pour calculer la durée d sur laquelle l'hypothèse d'égalité des accroissements de pluie et des volumes écoulés est appliquée. Le choix d'un pas de temps trop grand conduit à des coefficients de forme très dispersés, et engendre une incertitude inquiétante sur le coefficient moyen, donc sur le résultat final. Un pas de temps trop petit conduit en revanche à un résultat pessimiste, car l'hypothèse selon laquelle tout volume d'eau précipité supplémentaire sur le pas de temps s'écoulerait à l'exutoire pendant une durée équivalente deviendrait contestable. C'est pourquoi on préconise souvent que le coefficient de forme moyen soit compris entre 1,3 et 2. Seule une analyse détaillée des couples averses-crues permet une estimation « pas trop » hasardeuse. On trouvera au chapitre 4 une présentation de la méthode Socose qui propose une formule régionale d'estimation d'une durée caractéristique de crue.

Remarque : la méthode est très souvent employée au pas de temps journalier, lorsque la superficie du bassin versant est d'une certaine importance, en raison de la plus grande disponibilité des données journalières de pluie et de débit. Le choix de ce pas de temps est à éviter dans le cas de petits bassins versants pour lesquels le temps de concentration est de quelques heures.

Hypothèse 2

3) Bien que d'approche statistique, la méthode sous-tend une relation entre la pluie et le débit. Au-delà du point pivot, le coefficient d'écoulement marginal durant le temps caractéristique est égal à un. De ce fait, la méthode fixe la rétention médiane du bassin versant à la différence entre la pluie et le volume écoulé durant le temps caractéristique du bassin versant. La rétention est donc bornée par la pluie décennale si le point pivot est fixé à la fréquence décennale. On voit sur la *figure 3.12*, établie sur 220 bassins versants des régions méditerranéennes, que la rétention conditionnelle des bassins versants, calculée par différence entre la pluie journalière décennale et la lame journalière décennale, augmente

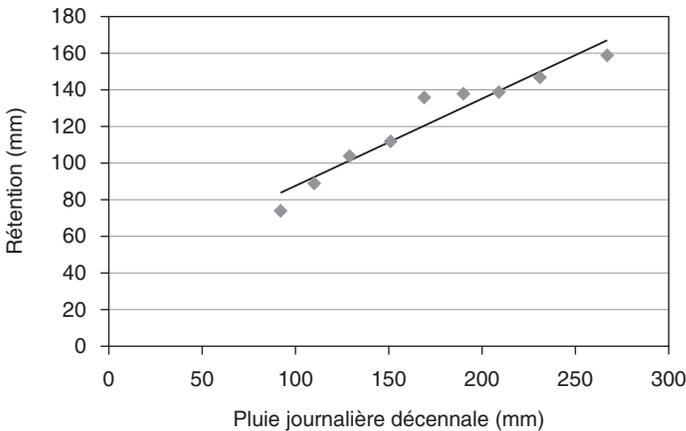


Figure 3.12. Évolution de la rétention « décennale » des bassins versants avec la pluviométrie journalière décennale.

avec la pluviométrie journalière décennale. Les rétentions conditionnelles sont calculées pour des classes de pluie, de 20 mm en 20 mm : 80-100, 100-120... Même si ce graphique n'est pas généralisable à l'ensemble de la France, et s'il est possible que les bassins versants présentent une certaine adaptation à l'infiltration suivant la climatologie, ce type de résultat conduit à émettre deux réserves : le choix du pas de temps journalier n'est pas forcément toujours justifié, et surtout il est vraisemblable que la rétention limite ne soit pas atteinte dès la crue décennale sur certains bassins. Pour les bassins versants à faible pluviométrie, le choix d'une extrapolation directe avec le gradex des pluies, au-delà du point pivot, peut entraîner une sous-estimation de la rétention, et ainsi une surestimation des volumes écoulés et des débits de crue.

4) La brusque cassure qui affecte la loi des débits au point pivot (démarrage de la loi des pluies) conduit à une surestimation des débits de fréquence intermédiaire (périodes de retour de l'ordre de 50 à 100 ans). Comme indiqué *supra*, cette surestimation devient toutefois négligeable en valeur relative pour les crues exceptionnelles (par exemple $T = 1\ 000$ ans), objet initial d'étude de la méthode du Gradex.

Hypothèse 3

5) Le rapport d'affinité pour obtenir le débit de pointe est très variable. La méthode préconise de retenir sa valeur moyenne ou médiane. Elle doit être de l'ordre de 1,5 à 2 si l'on a bien choisi la durée sur laquelle l'accroissement de débit est égal à l'accroissement de pluie. La détermination de ce rapport, notamment en absence d'information sur le bassin versant, reste un point délicat de la méthode, et ce, en raison de son implication directe sur le calcul du débit de pointe.

6) La méthode ne fournit pas une forme d'hydrogramme de projet, ce qui est nécessaire pour la simulation du laminage des crues dans un réservoir ou un champ d'inondation. Une forme bitriangulaire, qui respecte la durée, le débit de pointe et le volume écoulé sur la durée d , peut être retenue. Généralement, ces hydrogrammes de projet entraînent des surestimations de la capacité de laminage, car ils ne permettent pas de déterminer le volume complet de la crue (au-delà de la durée d).

Une évolution récente, appelée « Schadex », a été proposée par EDF/DTG (Paquet *et al.*, 2006) afin d'assouplir les deux premières hypothèses de la méthode du Gradex. Les analyses confirment qu'un ajustement à décroissance exponentielle, sans homogénéisation de la population des pluies (saisons, types de temps), engendre, en France du moins, une sous-estimation des probabilités de pluies extrêmes. Par ailleurs, une modélisation plus détaillée de la production de l'écoulement (avec le modèle Mordor pluie-neige-débit) permet de mieux évaluer le rapport des pentes des distributions des lames d'eau précipitées et écoulées pour les crues rares et extrêmes (rapport de 100 % selon la méthode du Gradex). Dans des cas génériques (bassins d'un millier de km^2 , en zone de reliefs), les auteurs ont trouvé une compensation de ces deux effets, ce qui tend à rassurer sur l'approximation fournie par la méthode originale dans la gamme d'application préconisée par ses auteurs.

En conclusion, cette méthode, de par l'exploitation de l'information pluviométrique qui donne une borne supérieure sur les débits, et de par son usage somme toute assez simple permet d'éviter de grossières erreurs de sous-estimation des crues extrêmes (période de retour supérieure à 1 000 ans), liées à l'extrapolation directe d'une distribution calée sur quelques dizaines d'années de débit. Elle présente toutefois quelques difficultés d'application qui viennent d'être rappelées, notamment pour l'estimation des crues rares (périodes

de retour de 50 à 100 ans) et la construction d'hydrogrammes de crues de référence, pour lesquelles une analyse de la forme des crues est recommandée (cf. § Hydrogramme moyen monofréquence).

Extrapolation de la distribution des crues par le modèle Agregée

Différents développements de la méthode du Gradex ont été proposés (CFGB, 1994), notamment pour élargir son domaine fréquentiel d'application aux événements de période de retour inférieure à 1 000 ans. Parmi ceux-ci on peut mentionner le modèle Agregée (Margoum, 1992 ; Margoum *et al.*, 1994 ; Lang, 1995, 1997), qui respecte le parallélisme entre les distributions de pluie et de débit. Le modèle Agregée comporte trois domaines fréquentiels (figure 3.13) :

- le domaine des crues observables, utilisant l'information des débits observés en continu et éventuellement celle des crues historiques ;
- le domaine des crues rares, issu d'un raccordement progressif entre le premier et le troisième domaine ;
- le domaine des crues extrêmes, piloté par la distribution des pluies.

Au-delà du troisième domaine, on considère que l'extrapolation est trop importante pour pouvoir être réalisée par une approche probabiliste. Comme indiqué au chapitre 1, des approches par maximisation de l'aléa (PMP, courbes enveloppes, géomorphologie et paléohydrologie) peuvent éventuellement prendre le relais.

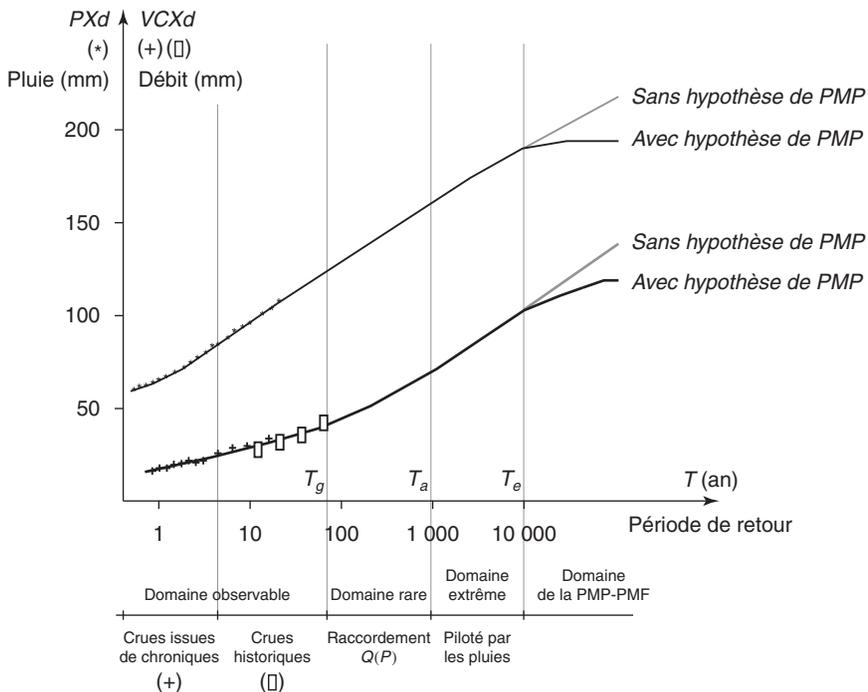


Figure 3.13. Domaines fréquentiels du modèle Agregée.

Au-delà du premier domaine, pour $T > T_g$, la formulation analytique du quantile est la suivante :

$$VCXd(T) = VCXd(T_g) + \frac{a_e(d)}{K_p - K_q} \left[K_p \ln \left(\frac{T + K_p}{T_g + K_p} \right) - K_q \ln \left(\frac{T + K_q}{T_g + K_q} \right) \right], \quad (3.12)$$

avec :

$$\begin{cases} K_p = \left(\frac{a_e(d)}{a_p(T_g, d)} - 1 \right) \cdot T_g & \text{et} & K_q = \left(\frac{a_p(T_g, d)}{a_q(T_g, d)} - 1 \right) \cdot T_g \\ a_e(d) = \lim_{T \rightarrow \infty} a_p(T, d) \\ a_p(T, d) = \frac{\partial PXd(T)}{\partial \ln T} (T = T_g) & \text{et} & a_q(T, d) = \frac{\partial VCXd(T)}{\partial \ln T} (T = T_g) \end{cases} \quad (3.13)$$

Les paramètres $a_p(T, d)$ et $a_q(T, d)$, appelés respectivement pseudo-gradex des pluies et des débits, représentent la pente de la distribution (des pluies ou des débits) dans un graphique de Gumbel pour la période de retour T . Lorsque les pluies suivent une loi de Gumbel (pseudo-gradex $a_p(T, d)$ constant égal à $a_e(d)$), l'expression (3.12) devient (avec $K_p = 0$) :

$$VCXd(T) = VCXd(T_g) + a_e(d) \ln \left(1 + \frac{a_q(T_g, d)}{a_e(d)} \cdot \frac{T - T_g}{T_g} \right), \quad (3.14)$$

expression proposée par Michel (1982), dite du Gradex esthétique.

Le modèle Agregee conserve donc l'hypothèse de base de la méthode du Gradex (sur la loi conditionnelle des pertes), tout en proposant des développements sur l'exploitation de l'information hydrométrique (crues historiques), l'extrapolation progressive de la distribution des crues à partir du seuil T_g , et enfin la généralisation de la loi des pluies à des distributions à caractère asymptotiquement exponentiel (existence d'une limite $a_e(d)$).

La méthode Speed

La méthode Speed (Système probabiliste d'études par événements discrets) a été développée par Sogreah (Cayla, 1993). Cette méthode reprend les hypothèses de la méthode du Gradex : les pluies sont représentées par une loi de Gumbel, et le coefficient d'écoulement marginal est égal à 1 lorsque le bassin versant est saturé. Elle utilise la théorie de l'hydrogramme unitaire pour aboutir à la formulation suivante :

$$Q(T) = \left(\frac{P(T) - P_0}{12} \right) \cdot S^{0,75}, \quad (3.15)$$

où :

$Q(T)$ et $P(T)$ sont respectivement le débit de pointe de crue et la pluie journalière de période de retour T (respectivement en m^3/s et en mm) ;

P_0 est le seuil probabiliste de saturation du bassin versant (en mm) ;

S est la surface du bassin versant (en km^2).

L'estimation de P_0 résulte de l'examen d'un graphique sur lequel on reporte la distribution de fréquence des pluies journalières et la distribution de fréquence des débits

réduits. Les débits réduits sont obtenus par la transformation $\frac{12Q}{S^{0,75}}$, où Q représente la valeur maximale annuelle du débit de pointe en m^3/s . Les auteurs de la méthode détectent une cassure dans la distribution des débits réduits, à partir d'une certaine période de retour T_0 , avec un parallélisme entre les deux distributions de fréquence au-delà de T_0 . Le paramètre P_0 est déterminé comme la différence entre la distribution de fréquence des pluies et celle des débits réduits pour la période de retour T_0 .

Comme dans la méthode du Gradex, l'écart entre la distribution des pluies et des débits est constant. Au-delà d'une période de retour T_0 , le débit journalier, en m^3/s , s'écrit :

$$QJ(T) = \frac{(P(T) - P_0) \cdot S}{86,4} \tag{3.16}$$

si la pluie $P(T)$ est exprimée en mm. La méthode Speed suppose donc un ratio constant entre le débit de pointe et le débit journalier de même fréquence, égal à $\frac{86,4}{12} S^{-0,75}$.

La méthode a été utilisée, à titre d'exemple, pour estimer le débit centennal d'un petit bassin versant méditerranéen (Réal Collobrier, superficie : 70,4 km^2). Sur la *figure 3.14*, la courbe supérieure représente la distribution de fréquence des valeurs maximales annuelles de pluie journalière d'un poste pluviométrique du bassin. On note sur la distribution de fréquence des débits réduits maxima annuels une cassure vers $T_0 = 6$ ans. L'extrapolation au-delà de T_0 par la loi des pluies colle assez bien aux plus forts débits observés.

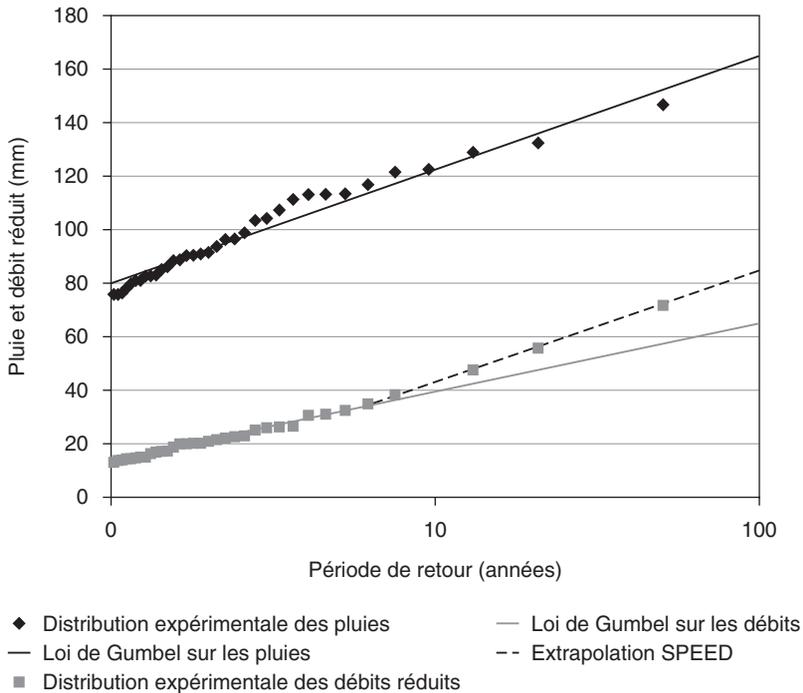


Figure 3.14. Application de la méthode Speed sur le bassin versant du Réal Collobrier.

Par ailleurs, le fait que les équations 3.15 et 3.16 n'exploitent l'information pluviométrique qu'au pas de temps journalier montre que la méthode Speed n'est pas adaptée au cas des bassins rapides dont le temps de réaction n'est que de quelques heures.

Extrapolation de la distribution des crues par le modèle Shypre

Shypre, acronyme de Simulation d'hydrogrammes pour la prédétermination des crues, est une méthode développée au Cemagref (Cernesson *et al.*, 1996 ; Lavabre *et al.*, 1998 ; Arnaud et Lavabre, 2000 et 2002). La méthode associe un générateur aléatoire de pluies horaires et une transformation de la pluie en débit (figure 3.15). Les techniques de simulation permettent de générer de très longues chroniques de pluies horaires (sur 100 000 ans, par exemple) dont on déduit les hydrogrammes de crue par application d'une modélisation simple de la pluie en débit. Le classement des pluies et des débits simulés permet de tracer les distributions de fréquence « empirique » des pluies et des débits de toutes durées, et d'en déduire les quantiles pour une large gamme de périodes de retour, de 2 à 1 000 ans, sans faire appel à ce stade à une loi de probabilité. Le progiciel Gershypre permet d'effectuer l'ensemble des opérations : établissement des couples averses-crues, calage du modèle de simulation des pluies horaires, calage du modèle pluie-débit, simulation des hyétogrammes horaires et des hydrogrammes de crue, échantillonnage des pluies et des débits et tracés des distributions de fréquence.

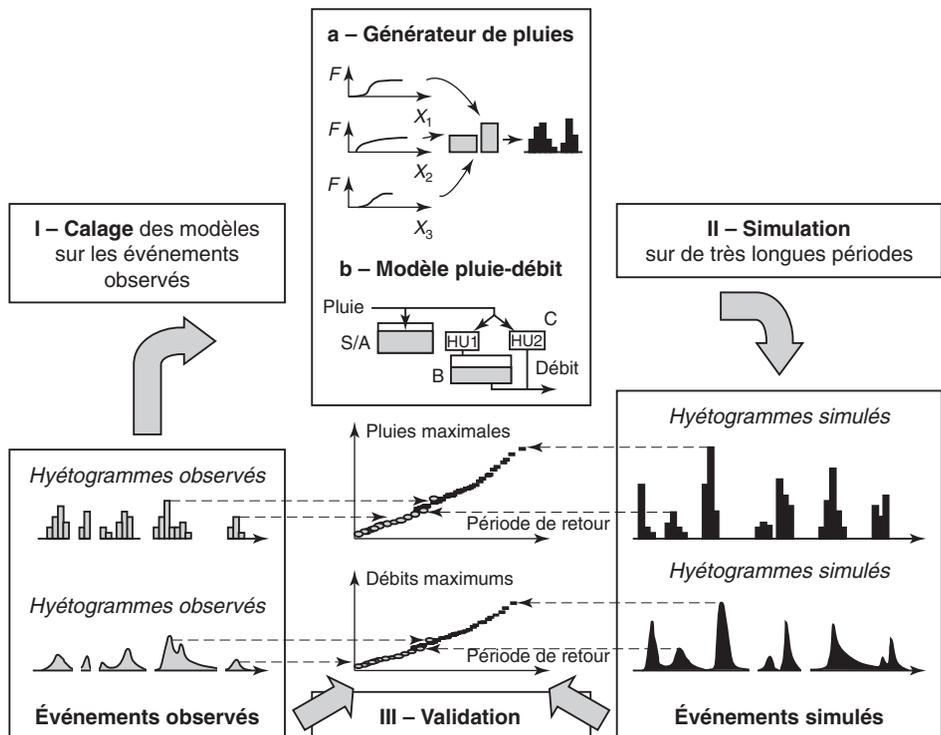


Figure 3.15. Principe de la méthode Shypre.

À partir d'une description géométrique du signal temporel de pluie, on procède à la génération du signal temporel de pluie en deux étapes. Une première phase consiste en l'étude descriptive du phénomène et l'analyse de la structure temporelle interne des événements pluvieux observés au pas de temps horaire. Un événement pluvieux est défini comme une succession de pluies journalières supérieures à 4 mm, et dont au moins un cumul journalier excède 20 mm. Huit variables sont utilisées pour décrire le signal : le nombre d'événements par année, le nombre de périodes pluvieuses par événement (une période pluvieuse est constituée par une séquence de pluies horaires non nulles), le nombre d'averses par période pluvieuse (une averse, dans une période pluvieuse, est constituée par une séquence qui présente un seul maximum), avec pour chaque averse, sa durée, son intensité moyenne, la position relative de l'intensité maximale dans l'averse, le rapport entre l'intensité maximale horaire et l'intensité moyenne de l'averse, et enfin la durée sèche qui sépare deux périodes pluvieuses à l'intérieur d'un événement. Ces variables sont ensuite identifiées par leur loi de probabilité dont les paramètres sont estimés à partir des échantillons constitués d'après les événements pluvieux observés.

Après cette phase d'identification, la seconde étape consiste en la génération de hyéto-grammes horaires à partir des variables descriptives qui sont générées de manière indépendante par un tirage aléatoire dans leur loi de probabilité, et ce, suivant une méthode de Monte-Carlo. On obtient ainsi de multiples scénarios de pluies horaires générées sur de très longues périodes de simulation.

Deux modèles peuvent être mis en œuvre par Shypre pour la transformation pluie-débit :

- le modèle conceptuel GR3h défini par trois paramètres (Edijatno et Michel, 1989) ;
- la combinaison de la fonction de production du Soil Conservation Service (*National Engineering Handbook*, 1985) et d'un hydrogramme unitaire. La fonction de production est définie par un paramètre J qui présente la capacité de rétention du sol pour une crue donnée. Sa formulation est la suivante :

$$R(t) = \frac{(P(t) - 0,2 J)^2}{(P(t) + 0,8 J)}, \quad (3.17)$$

où :

$R(t)$ et $P(t)$ sont respectivement l'écoulement cumulé et la pluie cumulée (en mm) ;
 J le paramètre qui représente la rétention du sol (en mm).

La fonction de transfert est assurée par un hydrogramme unitaire, défini par deux paramètres. Cette fonction s'écrit (Ctgreff *et al.*, 1980-1982) :

$$Q(t) = Q_p \cdot \frac{2\left(\frac{t}{t_m}\right)^\alpha}{1 + \left(\frac{t}{t_m}\right)^{2\alpha}}, \quad (3.18)$$

où :

$Q(t)$ est le débit à l'instant t ;
 Q_p est le débit de pointe de l'hydrogramme unitaire ;
 t_m est le temps de montée de l'hydrogramme unitaire ;
 α est le coefficient de forme de l'hydrogramme unitaire.

Pour la version régionalisée (cf. chapitre 4, § *Modèle Shyreg*), c'est le second modèle qui est utilisé dans une forme simplifiée en le réduisant, comme nous le verrons par la suite, à la seule fonction de production. La méthode Shypre a été appliquée sur le bassin versant du Bevinco. Ce cours d'eau de Haute-Corse draine, à la station de mesure, un bassin de 53,6 km². La période d'observation analysée est de 22 ans, de 1974 à 1995. Un pluviographe est implanté sur le bassin. Sur la *figure 3.16*, nous avons reporté les distributions de fréquence, des pluies et des débits journaliers exprimées en mm. Les points représentent les observations. Les distributions de fréquence des pluies et des débits simulés sont représentées en traits pleins (les valeurs journalières ont été calculées à partir des chroniques de pluies horaires simulées et des hydrogrammes simulés). La distribution des pluies simulées se situe légèrement en dessous des plus fortes observations, tout en attribuant des périodes de retour acceptables aux plus fortes valeurs. En revanche, la distribution de fréquence des débits simulés (échantillon initial) se situe nettement au-dessus des observations, au-delà de la période de retour 2 ans. Après enquête auprès du gestionnaire, il s'est avéré que la série disponible dans la banque de données était tronquée des plus fortes valeurs de crue à cause de dysfonctionnements divers. On remarquera que la seconde simulation, effectuée à partir de l'échantillon complété, est tout à fait comparable à la précédente.

Si l'on reporte maintenant sur la distribution des débits instantanés (*figure 3.17*) les valeurs simulées avec les deux types d'échantillon (initial et complété), on constate que

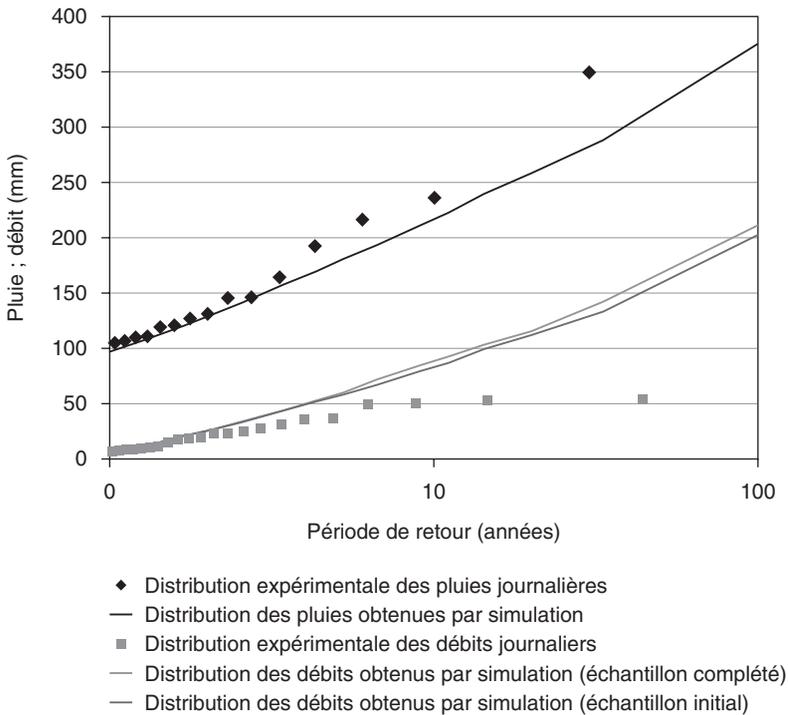


Figure 3.16. Bassin versant de Bevinco. Application de la méthode Shypre. Distribution de fréquence des pluies journalières et des débits journaliers.

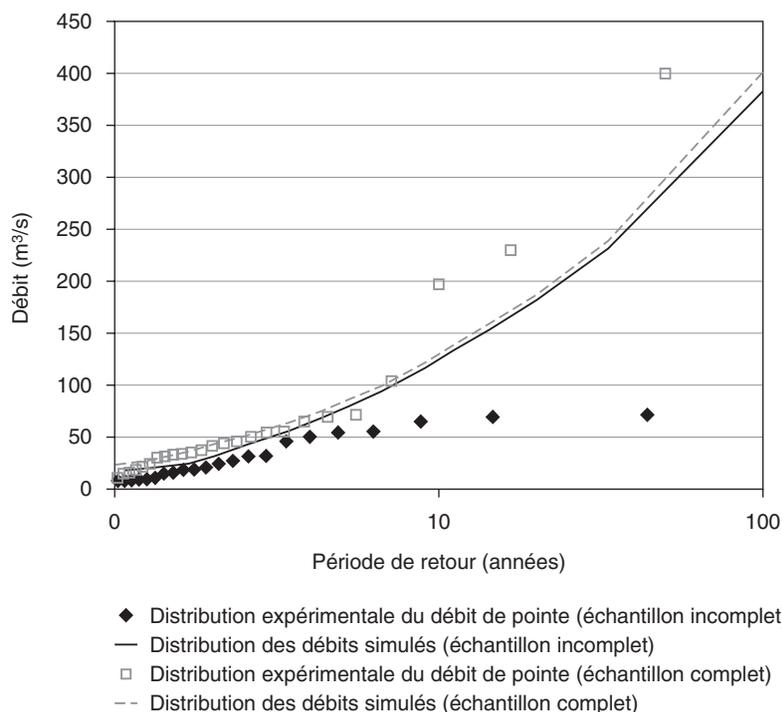


Figure 3.17. Bassin versant de Bevinco. Application de la méthode Shypre. Distribution de la valeur maximale du débit de pointe.

la distribution des valeurs simulées est bien en accord avec les observations (échantillon complété).

La méthode Shypre est plus lourde à mettre en œuvre qu'un simple ajustement à une loi de probabilité. Elle nécessite, en plus des observations de débit, des chroniques de pluies horaires qui ne sont que rarement disponibles. Lorsque cela est le cas, elle permet de mieux exploiter l'information contenue dans les séries pluviométriques (caractéristiques des averses successives au lieu d'un simple cumul). Elle donne la possibilité de générer des scénarios de crue à partir d'une collection d'hydrogrammes simulés de formes diverses. Des travaux récents ont montré (Muller, 2006) que la version régionale du modèle Shypre présente une plus faible sensibilité à l'échantillonnage que les méthodes d'extrapolation des lois de valeurs extrêmes. Les quantiles de crue obtenus à l'aide de la méthode Shypre sont parfois plus forts que ceux issus d'une extrapolation par la méthode du Gradex ou ses variantes (Schadex, Agregée, Speed). Des validations sont encore nécessaires pour évaluer le domaine respectif d'application de chacune de ces méthodes, leur plage d'incertitude et leurs limitations. On retiendra toutefois que les données pluviométriques apportent une information primordiale sur le comportement asymptotique de la distribution des crues, et que ce type d'information doit systématiquement être exploité.

Analyse des crues historiques

L'intérêt d'une recherche historique sur les inondations est d'élargir le champ chronologique d'analyse en recueillant des informations sur les événements majeurs ayant déjà eu lieu dans le passé. Une première source d'information est constituée des services d'annonce des crues qui disposent pour certains de chronologies de crues sur plus d'un siècle. La création du premier service d'annonce des crues date en effet de 1854 sur la Seine à Paris, et de la fin du XIX^e siècle pour les principales grandes rivières françaises. Ces données sont la plupart du temps disponibles sous forme de hauteur d'eau au droit d'une échelle hydrométrique. Elles nécessitent un travail d'analyse hydrologique et hydraulique pour critiquer les données, ainsi que pour reconstituer les débits associés aux crues, à partir des courbes de tarage récentes et d'informations disponibles sur la topographie ancienne de la rivière, et éventuellement de jaugeages anciens. Il est également possible de trouver des informations complémentaires dans les fonds des anciens services des ponts et chaussées (transférés dans les directions départementales de l'Équipement, les services de Prévision des Crues, les directions régionales de l'Environnement ou les archives départementales), dans certains fonds spécialisés (fonds Maurice Pardé à l'Institut de géographie alpine à Grenoble ; École nationale des ponts et chaussées à Paris ; Société hydrotechnique de France à Paris ; EDF pour les anciens services des Grandes forces hydrauliques ; Météo-France pour les données pluviométriques anciennes...) et dans l'inventaire national réalisé par Champion (1858-1864). Le comité consultatif des réseaux d'observation météorologique a rappelé l'importance de préserver les données pluviométriques anciennes (*cf.* journée CCROM du 6 octobre 2006, Météo-France Toulouse), et un programme spécifique d'action va être engagé en ce sens, dans la continuité des saisies de données anciennes déjà réalisées par Météo-France.

La procédure de collecte et de critique des données historiques a fait l'objet de travaux interdisciplinaires entre historiens, hydrologues et hydrauliciens, notamment en ce qui concerne les Alpes (Lang *et al.*, 1998, 2003 ; Cœur et Lang, 2000 ; Brochot *et al.*, 2002), les Cévennes (Lang *et al.*, 2002 ; Naulet *et al.*, 2005 ; Payrastré, 2005) et les Pyrénées (Antoine, 1992). La collecte de données historiques doit se concrétiser, en premier lieu, par un recensement des données suivant une grille de lecture en trois grands ensembles (*tableau 3.1*), avec un historique :

- 1/ des services producteurs de données ;
- 2/ des méthodes et matériels de mesure ;
- 3/ des événements sur le bassin.

En second lieu, il est indispensable de dresser un état général des sources, qui donne l'inventaire des sources d'information disponibles.

La mise au point d'une chronologie qualitative des crues de la rivière permet de repérer dans le temps les événements extrêmes. Ces derniers peuvent ensuite faire l'objet d'une analyse spécifique dans le but de reconstituer les débits de crue. Le traitement probabiliste de cette information soulève plusieurs difficultés liées à la disponibilité et l'exhaustivité des sources d'information, à la qualité de la reconstitution des événements, et enfin à la variabilité climatique et aux changements morphologiques des rivières sur plusieurs siècles. Selon les cas de figure, on pourra exploiter cette information de différentes façons.

– Une des crues historiques est directement retenue comme événement de référence, en supposant que les limites du champ d'inondation de l'époque sont toujours représentatives

Tableau 3.1. Grille de lecture des informations historiques sur les crues.

Grille de lecture	Objectif
1. Historique des services en charge de :	Ciblage des sources éventuelles d'informations
<ul style="list-style-type: none"> – la topographie ; – l'hydrométrie et de l'annonce des crues ; – la météorologie ; – la gestion de l'aménagement du territoire (forêt, travaux publics, énergie hydroélectrique) ; 	
2. Historique des méthodes et matériels de mesure :	Situation spatiale des données historiques Estimation des incertitudes liées aux pratiques de l'époque
<ul style="list-style-type: none"> – la topographie ; – l'hydrométrie : <ul style="list-style-type: none"> • enregistrement des hauteurs d'eau, • estimation des débits (jaugeages), • conversion hauteur-débit (courbe de tarage) ; – la pluviométrie (enregistrement des hauteurs d'eau). 	
3. Historique des événements du bassin :	Modélisation hydraulique des crues Modifications hydrologiques Évolution morphologique du lit Référencement altimétrique des hauteurs de crue Modélisation hydraulique Évolution morphologique du lit Évolution de la prise en compte du risque d'inondation Méthodes d'extrapolation des distributions de crue (méthodes du Gradex et Agregee) Stationnarité
<ul style="list-style-type: none"> – les aménagements (construction, destruction) : <ul style="list-style-type: none"> • barrages, ponts, digues, • forêts, urbanisation ; – les opérations de topographie (vue en plan, profils en travers, profils en long) ; – l'hydrométrie : <ul style="list-style-type: none"> • dossiers des stations hydrométriques, • jaugeages, barème des débits, courbes de tarage ; – les crues : <ul style="list-style-type: none"> • relevés de cotes et de débits (valeur maximale et valeur journalières de hauteur et débit), • limites de champ d'inondation ; – les étiages (relevés de cotes, débits, lignes d'eau) ; – la réglementation ; – la pluviométrie : <ul style="list-style-type: none"> • dossiers des stations pluviométriques, • relevés de précipitations (pluies mensuelles, pluies journalières) ; – la climatologie (sécheresse, périodes froides). 	

des conditions actuelles d'écoulement, ou en effectuant une analyse hydraulique complémentaire pour prendre en compte les changements significatifs intervenus depuis. Il est conseillé de recadrer cet événement historique de référence dans la gamme des crues possibles, en précisant si ladite gamme est représentative des crues fréquentes ($T < 2$ ans), moyennes ($2 \text{ ans} < T < 10$ ans), rares ($10 \text{ ans} < T < 100$ ans) ou exceptionnelles ($T > 100$ ans).

– Un événement singulier récent (cf. chapitre 2, § *Les horsains ou valeurs singulières*) peut être requalifié en fréquence, en exploitant la chronologie des crues anciennes. Ainsi,

une valeur singulière sur la période récente d'observations, qui serait supérieure de plusieurs ordres de grandeur à la seconde valeur de l'échantillon, peut voir sa fréquence empirique réévaluée en considérant qu'elle est la plus forte (deuxième, troisième...) valeur sur la période « historique ». À titre d'exemple, une valeur singulière la plus forte sur la période 1961-2000 et sur le xx^e siècle (1901-2000) voit sa période de retour sensiblement modifiée, en retenant par exemple la formule de fréquence empirique de Chegodayev (cf. annexe 2, § *Applications à l'étude des crues*), pour un échantillon classé en sens décroissant : $x'(1) > x'(2) > \dots > x'(i) > \dots > x'(N)$:

$$F_{ND}'(i) = \text{Prob}[X < x'(i)] = 1 - \frac{i - 0,3}{N + 0,4}, \quad (3.19)$$

et la correspondance avec la période de retour empirique :

$$T[x'(i)] = \frac{1}{1 - F_{ND}'(i)} = \frac{1}{F_D'(i)}. \quad (3.20)$$

- Période 1961-2000 : $T[x'(1)] = \frac{40,4}{0,7} = 58$ ans .

- Période 1901-2000 : $T[x'(1)] = \frac{100,4}{0,7} = 143$ ans .

– Une collection de quelques crues historiques a pu être reconstituée, et exploitée à titre de contrôle pour vérifier si l'extrapolation de la distribution des crues (par une des méthodes précédemment décrite) est en accord avec la distribution empirique des crues historiques. L'utilisation de la formule classique de fréquence empirique peut conduire à des incohérences si on l'applique de façon indépendante pour chacun des deux échantillons de crues. Par exemple, si l'on dispose de l'information suivante :

- 40 crues sur la période 1961-2000 : $x'(1) > x'(2) > \dots > x'(40)$,

- 3 crues historiques sur la période 1901-1960 : $y'(1) > y'(2) > y'(3)$,

- sachant que ces trois crues historiques n'ont pas été dépassées sur la période récente : $y'(3) > x'(1)$, l'équation 3.19 conduit au résultat suivant :

– probabilité au non-dépassement de la plus forte valeur sur 1961-2000 :

$$F_{ND}'(1) = 1 - \frac{0,7}{40,4} = 0,983,$$

– et probabilité au non-dépassement de la troisième plus forte valeur sur 1901-2000 :

$$F_{ND}'(3) = 1 - \frac{2,7}{100,4} = 0,973,$$

soit : $\text{Prob}[X < x'(1)] = 0,983 > \text{Prob}[X < y'(3)] = 0,973$,

alors que : $x'(1) < y'(3)$.

On trouvera en annexe 2, § *Applications à l'étude des crues, Représentation graphique*, la présentation d'une formule de fréquence empirique adaptée au cas d'un échantillon mixte, composé à la fois de valeurs maximales sur un échantillon complet (période récente) et de valeurs supérieures à un seuil sur un échantillon tronqué (période historique).

Les paramètres (α, β, γ) de la distribution des crues sont évalués en exploitant directement l'information disponible sur la période récente (échantillon $x'(1) > x'(2) > \dots$

$> x'(N_1)$ sur N_1 années) et sur la période historique (échantillon $y'(1) > y'(2) > \dots > y'(N_2)$ sur N_2 années). On trouvera dans Miquel (1984) ou Naulet (2002) la présentation de méthodes d'estimation permettant le traitement d'une information incomplète, où les crues historiques sont considérées comme supérieures à un seuil de perception. La figure 3.18 illustre le cas de l'Ardèche à Saint-Martin-d'Ardèche (2 240 km²), pour lequel l'on dispose de :

- relevés automatisés (limnigraphe) sur la période 1955-2000 : valeurs maximales annuelles ;
- relevés manuels (observateur) sur la période 1892-1954 : valeurs maximales annuelles avec des lacunes certaines années ;
- la reconstitution de 15 crues historiques sur la période 1644-1891, dont quatre crues (1890, 1827, 1878, 1857) supérieures à 5 400 m³/s de 1827 à 1891, une crue (1772) supérieure à 5 850 m³/s de 1772 à 1826, et une crue (1644) supérieure à 6 925 m³/s de 1644 à 1771.

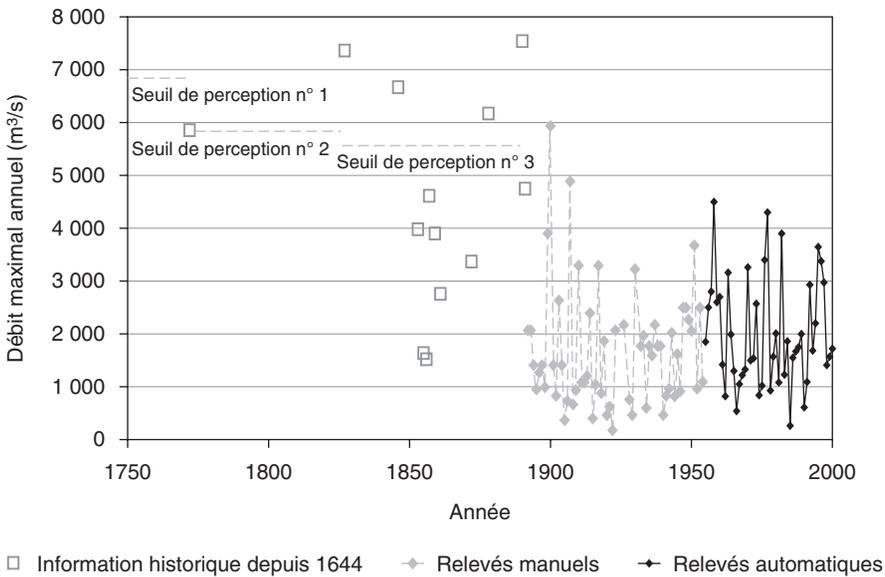


Figure 3.18. Chronologie des crues à Saint-Martin-d'Ardèche.

La figure 3.19 présente une comparaison de la distribution des crues à Saint-Martin-d'Ardèche (Lang *et al.*, 2002) :

- 1/ à partir des données de la période 1955-2000, uniquement avec les débits (courbe 4) ou avec l'information pluviométrique par la méthode du Gradex (courbe 3) ;
- 2/ en exploitant l'information historique depuis 1827 (courbe 1) ou 1644 (courbe 2).

Si les résultats sont concordants jusqu'à la crue décennale, on constate que la simple extrapolation d'une loi à trois paramètres (GEV) sur la période récente 1955-2000 conduit à des valeurs bien inférieures à celles obtenues avec un large recul chronologique. Ainsi, l'estimation de la crue centennale passe de 5 150 m³/s à 6 850-7 300 m³/s lorsque l'on

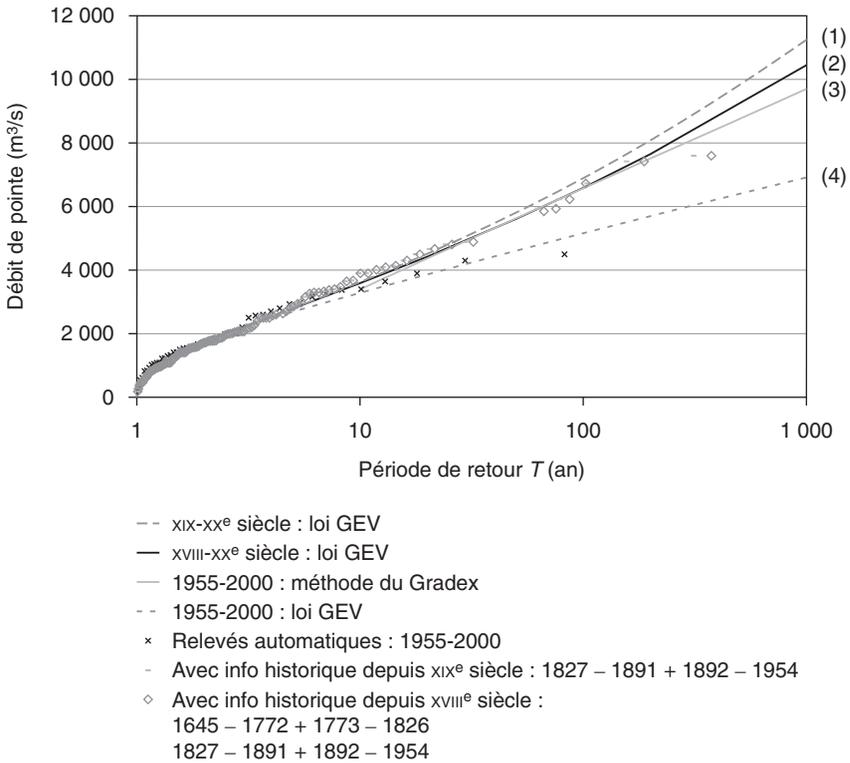


Figure 3.19. Distribution des crues à Saint-Martin-d’Ardèche.

incorpore l’information historique (respectivement sur les périodes 1644-1954 et 1827-1954). Un résultat remarquable est la très bonne concordance des résultats obtenus par l’approche historique (courbes 1 et 2) avec ceux issus de la méthode du Gradex (courbe 3), basée sur l’extrapolation de la distribution des crues, à l’aide de l’information pluviométrique récente.

Construction d’hydrogrammes de référence

Un des objectifs de la simulation hydraulique de scénarios hydrologiques de référence est de caractériser le degré d’exposition d’une zone aux inondations, par l’étude des paramètres fréquence, hauteur, vitesse et durée de submersion. Ce dernier paramètre ne peut être obtenu qu’à l’aide d’une modélisation hydraulique en régime transitoire, à partir d’hydrogrammes de crue injectés sur la branche amont du modèle et les différents affluents. Par ailleurs, la modélisation en régime permanent peut également s’avérer limitée pour l’estimation des hauteurs et des vitesses lorsque des phénomènes de concomitance entre affluents, ou de stockage/laminage dans une retenue ou un champ d’expansion induisent une déformation des hydrogrammes de l’amont vers l’aval.

Le choix d'hydrogrammes de référence peut dans un premier temps reposer sur des hydrogrammes observés, mais il sera alors difficile d'interpréter de façon fréquentielle les résultats obtenus par simulation hydraulique. En effet, un événement historique de référence est rarement homogène au niveau spatial (contribution différenciée des sous-bassins) et temporel (débit de pointe, volume de crue). Une ligne d'eau issue de la propagation de tels hydrogrammes observés ne peut être relative à une seule période de retour.

Nous abordons dans cette section différentes manières de considérer un hydrogramme de référence. Les données manipulées sont des hydrogrammes reconstitués à partir des données de hauteur d'eau (limnigrammes) et de la courbe de tarage, ou obtenus par simulation à l'aide de la méthode Shypre. Dans le second cas, on vérifiera que les propriétés des hydrogrammes simulés sont cohérentes avec celles des hydrogrammes observés.

Formulations analytiques

L'utilisation d'expressions analytiques, comme celle présentée au paragraphe précédent (cf. équation 3.18) — ou à l'aide d'un ratio entre le débit de pointe et le débit moyen maximal sur une durée de référence (cf. méthode du Gradex) —, permet de construire un hydrogramme à partir du débit de pointe et du calage de quelques paramètres de forme. C'est une première approche, mais qui ne permet pas d'associer une seule période de retour à l'hydrogramme. La probabilité rattachée au dépassement du volume de crue de cet hydrogramme peut très bien être différente de celle du débit de pointe. Dans ce cas de figure, il devient difficile d'interpréter une ligne d'eau issue de la propagation d'un tel hydrogramme, et de lui associer une période de retour.

Hydrogramme synthétique monofréquence (HSMF)

Une alternative consiste à raisonner à partir d'hydrogrammes théoriques qui présentent la propriété d'être homogènes en fréquence, quel que soit le débit utilisé pour les caractériser. Les courbes QdF en débit seuil (éventuellement reconstituées par une approche régionale, cf. chapitre 4) vont être exploitées pour la constitution d'hydrogrammes synthétiques monofréquence (HSMF) (figure 3.20) relatifs à une période de retour T :

– La phase de montée de la crue est supposée linéaire.

Le point n° 1 correspond à la pointe de crue. Le temps de montée $t_{\text{montée}}$ peut être estimé à partir d'un échantillon d'hydrogrammes de crue ou de façon approximative à partir d'une estimation de la durée caractéristique de crue D , (cf., chapitre 4, formule Socose) par la relation $t_{\text{montée}} = D$ ou $t_{\text{montée}} = 1,5D$. Le débit de pointe $QIX(T)$ relatif à la période retour T est obtenu par une analyse probabiliste.

– La phase de décrue est constituée pas à pas, à partir des quantiles en débit seuil $QCXd(T)$.

Le point n° i (i varie de 2 à N) est obtenu en reportant en ordonnée la valeur du quantile $QCXd_i(T)$, et en abscisse la durée d_i à partir de la branche de montée de l'hydrogramme. Le débit $QCXd_i(T)$ est en effet le débit dépassé de façon continue pendant la durée d_i .

Galéa et Prudhomme (1994) ont montré la cohérence de l'hydrogramme HSMF avec les courbes QdF en débit moyen maximal $VCXd$. L'intégration de l'hydrogramme HSMF pour une durée d donnée donne un débit $VCXd_{\text{HSMF}}$. Les deux auteurs ont montré que la période de retour de ce débit, déduite des abaques QdF en $VCXd$, était très voisine de la période de retour T de l'hydrogramme HSMF construit à partir des débits seuils $QCXd(T)$.

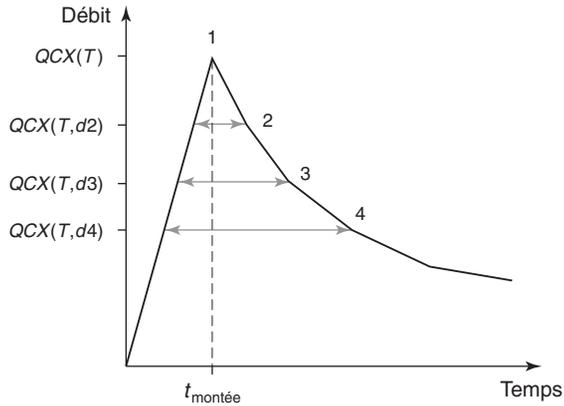


Figure 3.20. Construction d'un hydrogramme synthétique monofréquence HSMF.

Précisons toutefois que la distribution des débits seuils est plus difficile à extrapoler, car elle ne peut être reliée physiquement à la distribution des pluies (cf. *Méthode du Gradex*).

Hydrogramme moyen monofréquence

Un travail complémentaire (Garçon *et al.*, 2002) a permis d'améliorer le mode de construction des hydrogrammes HSMF, en proposant une montée de crue non linéaire qui respecte la forme moyenne des hydrogrammes de crue. La méthode consiste à normaliser les hydrogrammes observés par le débit moyen maximal sur une durée de référence (caractéristique du bassin versant), et à faire coïncider la pointe de crue. La *figure 3.21* présente

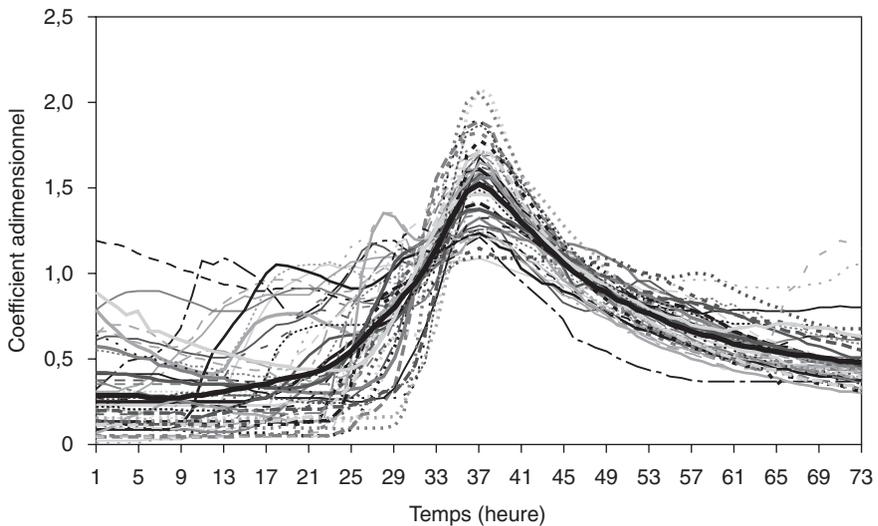


Figure 3.21. Hydrogramme moyen issu des hydrogrammes normalisés sur la Valserine à Lancrans.

un exemple d'hydrogramme moyen normalisé dont on conservera la forme pour la montée de crue. La décrue sera corrigée de façon à ce que les volumes de crue restent homogènes en fréquence pour plusieurs débits $VCXd$.

Des variantes permettent par ailleurs d'introduire des contraintes supplémentaires dans la construction de l'hydrogramme (Le Clerc, 2004), pour respecter les gradients de montée de crue (cas où l'on s'intéresse à la sécurité d'un ouvrage d'écroulement pour lequel le mode de gestion est très sensible à la rapidité de montée des eaux) ou des formes spécifiques d'hydrogrammes de crue (à deux bosses) en classant les événements suivant une typologie de formes (coefficient d'asymétrie et écart-type).

Hydrogramme d'apport par différence

L'injection, sur le linéaire d'une rivière, d'hydrogrammes d'apport HSMF, tous relatifs à la même période de retour, conduit en aval à une dérive en fréquence du fait de la concomitance des événements sur l'ensemble du bassin versant. La méthode par différence (Cemagref, 1993) consiste (figure 3.22) :

- à estimer les hydrogrammes HSMF du cours d'eau principal, de part et d'autre de la confluence avec un affluent, pour une même fréquence ;
- à construire l'hydrogramme d'apport en faisant la différence entre les hydrogrammes HSMF relatifs aux superficies ($S_{\text{amont}} + S_{\text{affluent}}$) et S_{amont} .

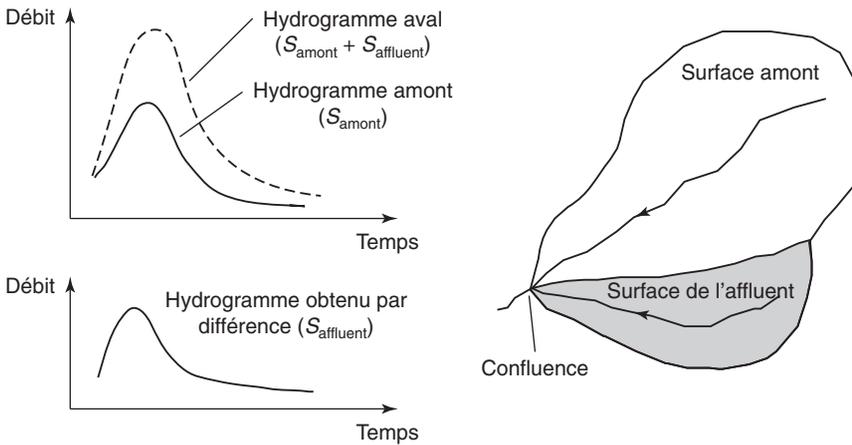


Figure 3.22. Construction d'un hydrogramme d'apport par différence.

Cette procédure permet de respecter la conservation des débits au droit de la confluence et garantit le traitement monofréquence le long du cours d'eau principal. En contrepartie, l'hydrogramme par différence n'est plus homogène en fréquence. C'est un opérateur intermédiaire utilisé pour la modélisation hydraulique le long du cours d'eau principal. Si l'on doit évaluer le risque d'inondation sur un affluent, il faudra alors utiliser un hydrogramme fréquentiel propre au régime de cet affluent, quitte à ce que la sommation de cet hydrogramme avec celui du cours d'eau principal (en amont de la confluence) ne soit pas

la même que celle obtenue avec l'hydrogramme aval déduit du régime du cours d'eau principal.

Recommandations pratiques pour l'étude des événements extrêmes et la constitution d'hydrogrammes de crue de référence

L'étude statistique des événements extrêmes (crue au moins centennale) est par nature délicate, du fait de la longueur réduite des chroniques hydrométriques (en général quelques dizaines d'années seulement) et de la nécessité d'extrapoler la distribution des crues au-delà du domaine des observations. Il est essentiel d'exploiter des informations complémentaires, en mobilisant les données issues du réseau pluviométrique et tout élément sur des crues historiques connues dans la région. Les points sur lesquels il convient d'insister sont les suivants :

Information pluviométrique

- Analyse de la pluie du bassin considéré :
 - analyser la variabilité spatiale des pluies dans le secteur géographique d'étude (au-delà des limites du bassin versant), et éventuellement prendre en compte des phénomènes particuliers : effet orographique, abattement des précipitations en zone d'altitude (limite pluie-neige) ;
 - de préférence raisonner sur une lame d'eau de bassin (obtenue par moyenne arithmétique de plusieurs postes ou par krigeage), sinon retenir un poste représentatif de la pluie de bassin.
- Choix d'un pas de temps et d'un échantillonnage saisonnier adapté à la dynamique des crues du bassin considéré :
 - du fait de la meilleure disponibilité des pluies journalières (plus forte densité du réseau d'observation, longueur des chroniques), retenir un pas de temps supérieur ou égal à la journée, en effectuant, si besoin est, une correction sur le centrage des valeurs (*cf.* correction de Weiss, chapitre 5). Si le temps de réaction du bassin en crue est de quelques heures, il est indispensable d'exploiter directement des données pluviographiques ou d'utiliser des courbes intensité-durée-fréquence ;
 - raisonner sur la saison à plus fort risque plutôt qu'à l'échelle annuelle. Ne pas retenir la saison des plus fortes intensités de pluie si celle-ci ne correspond pas à la saison des plus fortes crues.
- Critique des données et des hypothèses de modélisation pluie-débit :
 - contrôler visuellement la chronique des pluies et comparer les données agrégées à différents pas de temps (journalier → mensuel ; horaire → journalier) ;
 - raisonner en sous-bassins homogènes d'un point de vue climatologique si l'échelle du bassin est trop large ;
 - étudier le rendement des pluies (lame d'eau ruisselée/lame d'eau précipitée) des plus forts événements de crue pour évaluer le seuil d'extrapolation des méthodes statistiques (Gradex, Agregee, Speed) ;

- vérifier si l'hypothèse de correspondance statistique entre les distributions de pluie et débit d'un même pas de temps est pertinente, et le cas échéant utiliser une modélisation plus complète (influence de la neige, présence de karst, crue par remontée de nappe, laminage des crues par une zone d'expansion ou par des ouvrages de retenue, secteur urbanisé...);
- vérifier la qualité du calage du modèle pluie-débit des méthodes par simulation (Shypre) sur les plus forts événements de crue.

Information sur les crues historiques

- Constitution d'un jeu de données :
 - vérifier l'existence de données sur les crues anciennes auprès du service de prévision des crues, de la Diren (atlas des zones inondables), et en consultant les études hydrologiques antérieures et l'inventaire de Maurice Champion (1848-1856) ;
 - se renseigner sur l'existence de marques de crues ;
 - compléter éventuellement ces informations par une enquête historique spécifique (consultation des archives départementales et parcours de fonds plus spécialisés, cf. § *Analyse des crues historiques*).
- Analyse critique des données :
 - vérifier la cohérence des informations (recoupement spatial, analyse des laisses de crue) ;
 - convertir les cotes de crues historiques en débit (vérifier la stabilité du fond du lit et prendre en compte d'éventuels changements dans la géométrie du lit mineur et/ou du lit majeur) ;
 - déterminer un seuil d'exhaustivité au-delà duquel on peut considérer que tous les événements sont connus.
- Analyse spécifique des plus forts événements de crue :
 - faire la synthèse des éléments disponibles qui permettent d'apprécier chaque événement : description météorologique, répartition spatiale des précipitations, date dans l'année, cotes atteintes et zones de fortes vitesses, conditions particulières d'écoulement par rapport à la situation actuelle (changements morphologiques du lit mineur et/ou du lit majeur, endiguements, brèches/ruptures de digues...), dommages recensés.

Hydrogrammes de crue de référence

- Quelle que soit la méthode utilisée pour constituer des hydrogrammes de crue de référence (cf. § *Construction d'hydrogrammes de référence*), il est indispensable de vérifier :
- la cohérence de la forme de crue, en superposant l'hydrogramme de référence avec la ou les plus forte(s) crue(s) connue(s) ;
 - la cohérence fréquentielle de l'hydrogramme, en calculant la période de retour de plusieurs débits moyens de l'hydrogramme de référence dans le référentiel issu des courbes QdF ;
 - l'effet éventuel de concomitance de crue entre affluents et cours d'eau principal, qui pourrait induire une dérive en fréquence vers l'aval.

Le cas échéant, on pourra raisonner sur plusieurs scénarios d'hydrogrammes de référence si le bassin présente une diversité marquée de formes de crue ou une variabilité dans la contribution des différents affluents.

Chapitre 4

Méthodes probabilistes spécifiques d'estimation des débits de crue de référence sur un site non jaugé

Bien qu'il existe en France près de 3 000 chroniques de débits, la plupart du temps, l'hydrologue ne dispose pas d'informations hydrométriques directement exploitables sur le site pour lequel il désire estimer un débit de référence.

Dès lors, trois cas de figure peuvent se présenter à lui :

Cas 1 – Existence d'une station de mesure des débits sur le bassin versant à étudier, et située légèrement en amont ou légèrement en aval

Pour donner un ordre de grandeur, cela peut correspondre au cas où la superficie du bassin jaugé est de l'ordre de la moitié à deux fois la superficie du bassin versant à étudier. Après avoir vérifié que le bassin versant contrôlé peut être considéré comme représentatif du bassin versant à étudier (absence de particularités fortes telles qu'un site urbain important entre le site à étudier et la station de mesure, présence d'un champ d'inondation ayant un effet notable sur les débits de crue, ou apport intermédiaire d'un affluent au régime bien différencié), on effectuera une analyse statistique, telle que précédemment décrite, à partir de la chronique d'observation.

Le transfert d'information entre les deux bassins versants, du site mesuré (*station*) au site à étudier (*projet*), se fera grâce à une relation de la forme :

$$\frac{Q_{\text{projet}}(T)}{Q_{\text{station}}(T)} = \left(\frac{S_{\text{projet}}}{S_{\text{station}}} \right)^{\alpha} \cdot \left(\frac{PJ_{\text{bassin}_{\text{projet}}}(T)}{PJ_{\text{bassin}_{\text{station}}}(T)} \right)^{\beta}, \quad (4.1)$$

où :

$Q(T)$: débit relatif à la période de retour T ;

S : superficie du bassin versant ;

$PJ_{\text{bassin}}(T)$: pluie journalière de période de retour T .

Pour les débits de pointe de crue, on peut retenir pour α une valeur proche de 0,8 et pour β une valeur autour de 2. Ces deux valeurs numériques sont homogènes à la formulation Crupedix (cf. équation 4.6). Pour les débits moyens sur des durées de quelques heures

à un ou plusieurs jours, les valeurs de α et β à retenir tendent vers 1. L'usage de cette formule de transfert spatial doit rester limité à une première estimation de l'ordre de grandeur du débit de référence. Ainsi, dans le cas de petits bassins versants, la pluie journalière n'est pas une référence idéale, et il est possible que d'un site à l'autre le rapport entre les pluies horaires de mêmes durées de retour soit très différent du rapport calculé sur les pluies journalières. De même, le bon sens physique ne permet pas d'extrapoler indéfiniment une formule où β différerait de 1. Il s'agit donc, au mieux, d'une approximation valable pour une certaine gamme de durées de retour. Enfin, α diminue avec les durées de retour : il se rapproche de 1 pour les phénomènes les plus courants, et de 0,5 pour les plus rares (déduit des courbes-enveloppes).

Cas 2 – Existence d'une station de mesure sur le bassin versant, mais éloignée du site étudié

Dans tous les cas, en absence locale d'information près du site à étudier, il est préférable de mettre en œuvre des approches multiples afin de réduire, autant que faire se peut, l'incertitude sur l'estimation des débits. Les formules, dites de régression multiple, présentées ci-après, peuvent être examinées. Notre préférence va vers des formulations qui prennent explicitement la pluie en compte, la variable pluie étant, après la superficie, la variable la plus explicative des débits de crue, hors régime hydrologique glaciaire. En tout état de cause, il est vivement conseillé d'examiner régionalement les performances des formulations sommaires avant d'en retenir les ordres de grandeur.

Cas 3 – Aucune station de mesure n'existe sur le bassin versant

L'information sur les bassins versants proches doit être alors recherchée. Mais, face à l'hétérogénéité spatiale des débits de crue, il est alors préférable de ne pas se limiter à un seul bassin versant voisin, mais d'en examiner au moins deux ou trois pour tester l'homogénéité ou non des quantiles de débits. En cas de non homogénéité (après correction de surface et pluviométrie), des analogies pourront être recherchées, selon les caractéristiques physiographiques et géologiques du bassin, l'occupation du sol...

Régressions multiples

Ces formules empiriques régionalisées sont généralement de la forme :

$$Q(T) = aS^b, \quad (4.2)$$

où :

$Q(T)$: débit de pointe relatif à la période de retour T (en m^3/s) ;

S : superficie du bassin versant (en km^2) ;

a et b : coefficients fonctions de la période de retour T .

Ces formulations découlent de synthèses régionales. Le coefficient a est lié au régime pluviométrique et à l'aptitude au ruissellement du bassin. Le coefficient b est un exposant, généralement compris entre 0,7 et 0,8, qui tient compte de l'abattement des pluies, ce qui résout implicitement le problème de la dimension spatiale des zones de précipitations intenses, et de l'écrêtement du débit dans le réseau hydrographique.

Nous présentons *infra* quelques exemples de formules régionales.

La formulation de Bressand-Golossof (1995)

La direction départementale de l'Équipement du département du Gard propose deux formulations de type $Q = aS^b$, valables pour l'ensemble du département du Gard. Elles ont été établies à partir des courbes enveloppes des plus forts événements observés en milieu méditerranéen (Gard, Hérault et Pyrénées-Orientales). Les formules sont données pour une gamme de surface de bassin versant comprise entre 20 et 400 km² :

$$Q_{\text{rare}} = 30 S^{0,75} \quad \text{et} \quad Q_{\text{exc.}} = 50 S^{0,75}, \quad (4.3)$$

où Q_{rare} et $Q_{\text{exc.}}$ donnent respectivement un ordre de grandeur du débit centennal et millénal.

Les courbes-enveloppes de Francou-Rodier (1967)

À partir des crues maximales observées lors des deux derniers siècles sur 1 400 bassins versants répartis dans le monde, et ayant des superficies dans la fourchette $10^{-2} \times 10^6$ km², Francou et Rodier ont établi la courbe-enveloppe dont la formulation est la suivante :

$$\frac{Q}{Q_0} = \left(\frac{S}{S_0} \right)^{1 - \frac{k}{10}}, \quad (4.4)$$

où :

Q : débit de pointe de la crue (en m³/s) ;

S : superficie du bassin versant (en km²) ;

$Q_0 = 10^6$ et $S_0 = 10^8$.

Le paramètre k est un paramètre régionalisé : il varie en France de 5,5 (zone méditerranéenne) à 3,5 (zone océanique du Nord de la France). S'agissant de l'enveloppe des crues maximales observées, ces estimations de débit ne sont pas affectées d'une fréquence d'apparition. Les auteurs indiquent cependant qu'une bonne partie des points qui ont servi à paramétrer k , correspondent à une période de retour de 100 ans.

Synthèse des débits de période de retour 1 000 ans calculés par la méthode du Gradex (Duband *et al.*, 1988)

La méthode du Gradex a été appliquée par EDF sur de nombreux bassins versants français dont la superficie varie de quelques km² à quelques milliers de km². La régression établie sur 170 bassins versants du débit de pointe de période de retour 1 000 ans en fonction de la superficie du bassin versant s'écrit :

$$Q = \lambda S^{0,72}, \quad (4.5)$$

où S est la surface du bassin versant en km² et λ un paramètre donné par le *tableau 4.1* pour les trois zones suivantes :

- **zone I** : les bassins affluents de la Loire-Inférieure (Vienne, Creuse...), situés au nord du Massif central, ceux de la Saône, de la Moselle et de la Bretagne ;
- **zone II** : les bassins des Pyrénées occidentales et centrales, de l'Aude et de l'Ariège, de la Dordogne et du Lot, les bassins de la Durance, du Fier et de l'Arve, des Dranses, de l'Isère ;

– **zone III** : les bassins de la Haute-Loire, des Cévennes, du Tarn, des affluents rive droite du Rhône à l'aval de Lyon (Eyrieux, Ardèche...) des Alpes-Maritimes, de la Corse.

Tableau 4.1. Valeurs régionales du coefficient λ (régionalisation de la méthode du Gradex).

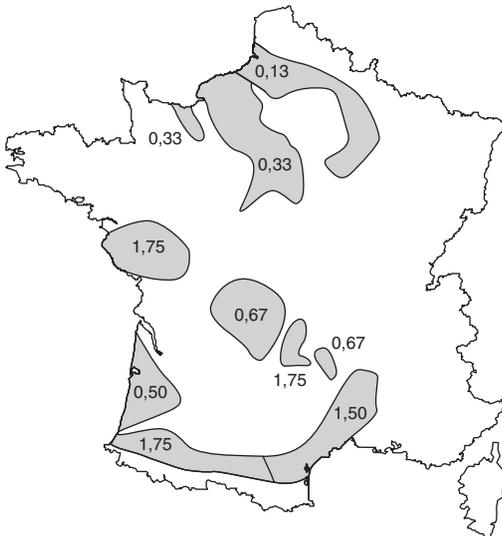
Zone	I	Fourchette à 90 %	Fourchette à 70 %
I	4,05	3,07-5,36	3,4-4,8
II	7,4	5,2-10,4	5,9-9,2
III	16,4	9,1-29,7	11,3-23,9

L'équation 4.5 n'est pas applicable aux bassins versants inférieurs à quelques km². Elle ne donne qu'un ordre de grandeur, qu'il faut toujours affiner par une étude locale. Cet ordre de grandeur est seulement à considérer pour porter un premier jugement sur la capacité de l'évacuateur de crue. Il ne dispensera jamais d'une étude plus complète.

Formule Crupedix (Ctgreff *et al.*, 1980-1982)

$$Q(10) = S^{0,8} \left(\frac{PJ(10)}{80} \right)^2 R, \quad (4.6)$$

où $PJ(10)$ est la pluie journalière décennale en mm. Le paramètre R varie en France dans la fourchette $[0,1 ; 1,75]$. La *figure 4.1* présente une cartographie des valeurs de R sur la métropole.



Estimation de R

R est un coefficient régional établi empiriquement à partir du rapport entre la valeur de la crue décennale observée et la valeur obtenue par la relation : $Q = S^{0,8} \cdot (P/80)^2$.

Il consiste donc en une estimation du résidu moyen dans un secteur géographique donné. La carte ci-contre présente les valeurs du coefficient R .

Figure 4.1. Coefficient régional R de la méthode Crupedix.

Les auteurs, après analyse des écarts obtenus lors de la comparaison entre les valeurs de Crupedix et les quantiles estimés à la station, donnent un intervalle de confiance sur la crue décennale. Ainsi, la probabilité que la vraie valeur de débit décennal se situe dans l'intervalle $\left[\frac{Q}{2}; 2Q\right]$ est de l'ordre de 90 %. Elle est voisine de 70 % pour l'intervalle $\left[\frac{2Q}{3}; \frac{3Q}{2}\right]$.

Formule Socose (Ctgreg *et al.*, 1980-1982)

$$Q(10) = \frac{kS}{(1,25D)^b} \cdot \frac{\rho^2}{15 - 12\rho}, \quad (4.7)$$

avec :

$$\ln D = -0,69 + 0,32 \ln S + 2,2 \sqrt{\frac{PA}{PJ(10)t_A}}, \quad (4.8)$$

$$k = \frac{24^b PJ(10)}{21 \times \left(1 + \sqrt{\frac{S(30D)}{D}}\right)}, \quad (4.9)$$

$$\rho = 1 - \frac{0,2J}{k(1,25D)^{1-b}}, \quad (4.10)$$

$$J = 260 + 21 \ln\left(\frac{S}{L}\right) - 54 \sqrt{\frac{PA}{PJ(10)}}, \quad (4.11)$$

où :

k et ρ : coefficients intermédiaires de calcul ;

S : superficie du bassin versant (en km^2) ;

D : durée caractéristique de crue (en h), définie comme la valeur médiane de la durée de dépassement de la moitié du débit de pointe. En l'absence de données hydrométriques sur le site, la formule régionalisée 4.8 peut être utilisée ;

PA : pluie annuelle moyenne (en mm) ;

$PJ(10)$: pluie journalière de fréquence décennale (en mm) ;

t_A : température annuelle moyenne réduite au niveau de la mer (en °C) (*cf. figure 4.2*) ;

b : deuxième coefficient de Montana des courbes *Idf* des pluies pour la période de retour 10 ans : ad^{-b} (intensité I en mm/h, durée d en h) ;

J : interception potentielle du bassin (en mm) ;

L : longueur du talweg le plus long (en km).

On notera que la température t_A n'est pas une variable explicative de la durée des crues, mais plutôt, en France, la latitude où les températures comme les pluies courtes augmentent du sud au nord (à ne pas employer dans un autre pays). Le logiciel Crueproj met en œuvre directement la méthode Socose, et permet aussi de générer l'hydrogramme de crue en réponse à un hyétogramme. Il est mis gratuitement à disposition par le Cemagref, groupement d'Aix-en-Provence.

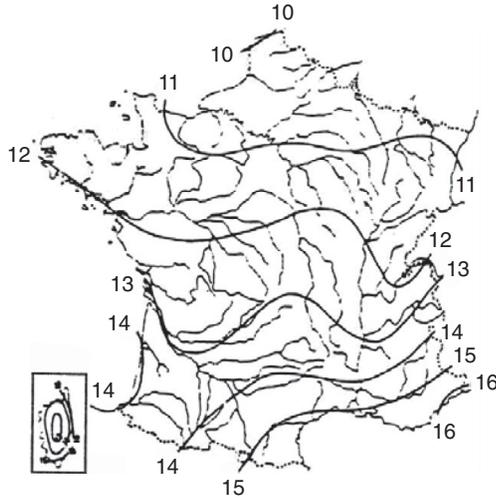


Figure 4.2. Température annuelle moyenne réduite au niveau de la mer.

Remarque : en faisant l'hypothèse que J et D sont invariants avec la fréquence, il est possible d'utiliser la formulation Socose pour l'estimation de quantiles de débits de fréquence plus rares. Il suffit de retenir la valeur b de la loi de Montana des pluies relative à la fréquence désirée. Les auteurs de la méthode limitent cependant l'application à une gamme de période de retour 5 à 50 ans.

Modèles préparamétrés

La modélisation de la pluie en débit est un domaine extrêmement vaste. Nous nous limitons ici à exposer deux méthodes simples, d'utilisation courante en ingénierie. Ces méthodes sont généralement dédiées à l'estimation du débit de pointe de crue d'une fréquence donnée. La fréquence du débit est identique à celle de la pluie, bien que cette hypothèse soit rarement vérifiée en hydrologie rurale. Ces méthodes conduisent aussi à une forme très simple des hydrogrammes de crue, qui découle de l'estimation du volume écoulé pendant un certain temps.

La modélisation de la pluie en débit repose sur la définition d'une fonction de production et d'une fonction de transfert. La fonction de production détermine la part de la pluie qui s'écoule effectivement à l'exutoire du bassin versant pendant la crue. Le transfert hydraulique des volumes de pluie sur le bassin jusqu'à l'exutoire est assuré par la fonction de transfert. *A minima*, cette fonction est définie par une durée caractéristique du bassin versant et une forme d'hydrogramme. Les deux méthodes qui vont être présentées reposent sur une formulation très simplifiée de la relation pluie-débit, avec des paramètres d'écoulement qui restent difficiles à caler en l'absence d'information hydro-pluviométrique.

La méthode rationnelle (Stifel W.J. et Mc Pherson M.B., 1964)

Cette méthode, d'utilisation courante, repose d'une part sur la définition d'un coefficient d'écoulement C_e (fonction de production), et d'autre part sur le calcul de la durée caractéristique du bassin qui est généralement définie par le temps de concentration T_c du bassin versant. La méthode fait l'hypothèse que la pluie cumulée sur la durée caractéristique T_c et de période T , notée $P(T_c, T)$, est constante et homogène dans l'espace sur l'ensemble du bassin versant de superficie S . Le volume écoulé est $V = P(T_c, T)C_e S$. Si l'on assimile la fonction de transfert à un triangle isocèle, de base $2T_c$, on peut écrire que $V = q(T)T_c$, soit :

$$q(T) = \frac{1}{3,6} C_e I(T_c, T) S, \quad (4.12)$$

où :

$q(T)$: débit de pointe de la crue (en m^3/s) ;

C_e : coefficient d'écoulement ;

$I(T_c, T)$: intensité de la pluie (en mm/h) pendant la durée T_c .

T : période de retour de référence ;

S : superficie du bassin versant (en km^2).

L'hydrogramme ainsi défini est un triangle isocèle de volume V et de sommet $q(T)$.

À titre d'exemple, la *figure 4.3* représente l'hydrogramme généré par la méthode rationnelle pour 1 mm de pluie nette, et avec comme fonction de transfert un triangle isocèle de base $2T_c$. Pour un bassin de $1 km^2$, le débit de pointe est de $0,278 m^3/s$. Toujours pour un bassin versant de $1 km^2$, la *figure 4.4* représente les hydrogrammes obtenus, selon les conditions suivantes : $C_e = 0,5$; $I(T_c, T) = 50 T_c^{-0,5}$; $T_c = 0,5$ et $1,5$ h. Les débits de pointe obtenus pour ces deux valeurs de T_c , sont respectivement de $9,8$ et $5,6 m^3/s$.

Remarque : une confusion est souvent effectuée. Si l'hydrogramme de projet est construit avec un temps de montée égal à T_c et un temps de décrue de $2T_c$, le temps de base est alors

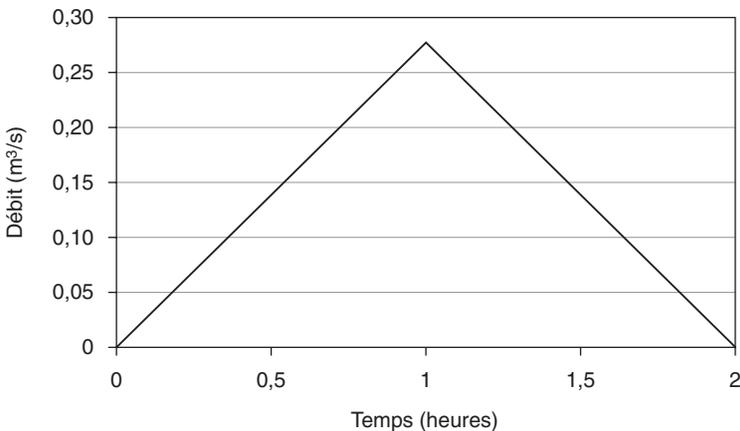


Figure 4.3. Fonction de transfert de la méthode rationnelle. Exemple d'hydrogramme pour $T_c = 1$ heure et une pluie nette de 1 mm (*pluie nette* = *pluie* \times C_e) pour un bassin versant de $1 km^2$.

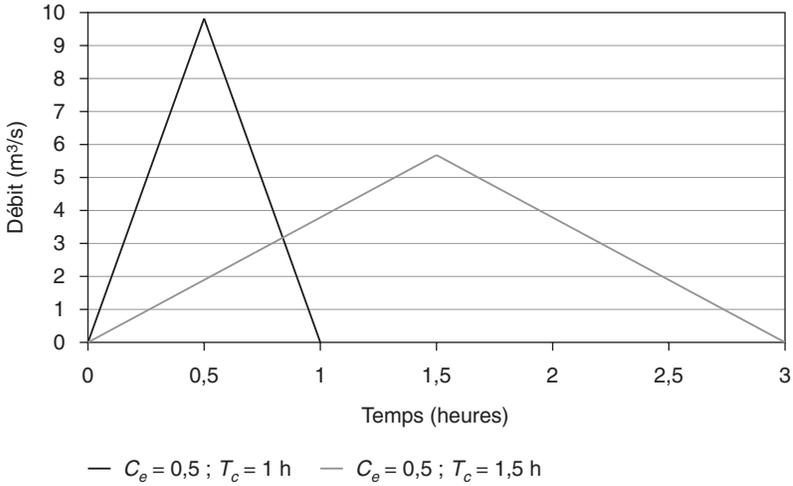


Figure 4.4. Exemples d'hydrogrammes générés par la formule rationnelle, pour un bassin versant de 1 km².

$3T_c$, ce qui entraîne une majoration du coefficient C_e dans un rapport 3/2 pour aboutir à la même valeur $q(T)$ du débit de pointe.

La méthode implique, outre le choix de la forme de la fonction de transfert, la définition de $C_e(T)$ et de T_c . Le premier coefficient $C_e(T)$, compris entre 0 et 1, croît avec la période de retour T . Il existe de nombreuses tables qui permettent de déterminer la valeur de ce coefficient en fonction des caractéristiques physiques du bassin versant. Van Te Chow *et al.* (1988) proposent une table où $C_e(T)$ est fonction de l'occupation des sols et de la pente moyenne du bassin versant. Il croît de 0,81 à 0,95 pour des surfaces imperméables (respectivement période de retour 10 à 100 ans). Pour les autres types d'occupation du sol (lande, prairie, terres agricoles, forêt...), on obtient les valeurs suivantes :

$$\begin{cases} C_e(T = 10 \text{ ans}) = 0,35, \text{ avec une fourchette de variation } [0,25 ; 0,45] \\ C_e(T = 100 \text{ ans}) = 0,50, \text{ avec une fourchette de variation } [0,36 ; 0,55] . \end{cases} \quad (4.13)$$

Pour la région méditerranéenne, un groupe de travail, constitué pour la détermination des débits à retenir pour la sécurité hydraulique de la ligne TGV, a proposé la formulation suivante du coefficient d'écoulement de fréquence centennale (Astier *et al.*, 1993) :

$$C_e(T = 100 \text{ ans}) = 0,80 \times \left(1 - \frac{P_0}{PJ(100)} \right), \quad (4.14)$$

où :

$PJ(100)$: pluie journalière de fréquence centennale ;

P_0 : coefficient tabulé qui varie entre 0 (zones cultivées en région montagneuse) et 90 (zones boisées avec un relief plat).

Dans la formule rationnelle, l'intensité de la pluie $I(T_c, T)$ est calculée sur la durée du temps de concentration T_c . La formulation générale pour déterminer le temps de concentration T_c est de la forme :

$$T_c = KL^n Pe^{-cI^{-d}}, \quad (4.15)$$

où :

L : longueur du drain hydraulique le plus long (parfois la longueur totale des drains, ou la longueur du rectangle équivalent du bassin versant...);

n : coefficient de rugosité de Manning-Strickler ;

Pe : pente du drain le plus long (parfois la pente des versants) ;

I : intensité de la pluie nette ou pluie nette (pluie qui participe effectivement à l'écoulement), à déterminer par itérations successives, $I = I(T_c, T)$;

K : constante qui dépend des unités utilisées ;

a, b, c et d : exposants dont certains peuvent être nuls.

Il convient d'être prudent lors du calcul du temps de concentration, les différentes formulations aboutissant à des estimations dans un rapport de 1 à 10 ! La formule rationnelle aboutit à des estimations d'autant plus fortes sur le débit de pointe et faibles sur le volume écoulé que le temps T_c est faible. Inversement, une valeur forte de T_c entraîne une estimation faible du débit de pointe et un volume écoulé plus fort.

À titre d'exemple, nous donnons les coefficients de la formulation de T_c proposée par Lefort, in Astier *et al.* (1993), que nous utiliserons au chapitre 5, § *Applications de méthodes de régionalisation en zone méditerranéenne française, Modèles préparamétrés* :

– $K = 1,8$;

– $a = 0,6, b = 0, c = 0,33, d = 0,23$;

– $I = Pn$: pluie nette = $0,8 (PJ(100) - P_0)$ pour la fréquence centennale.

Comme suite aux hypothèses de la méthode rationnelle, son utilisation est limitée aux petits bassins versants (jusqu'à quelques dizaines de km²). En cas de non-homogénéité de l'occupation de l'espace, il est possible de calculer un coefficient d'écoulement pondéré.

L'instruction interministérielle de 1977 relative aux réseaux d'assainissement d'agglomérations préconise l'utilisation de la formule rationnelle pour calculer les débits pluvieux en zone urbaine. Le calcul des débits est réalisé pour des zones de coefficient d'écoulement C_e homogène. Une règle d'addition des débits de l'ensemble des zones permet de déterminer le débit de pointe à l'exutoire.

La méthode SCS (Soil Conservation Service, 2003)

Cette méthode du Soil Conservation Service de l'USDA, quoique moins utilisée que la formule rationnelle, présente l'avantage d'être relativement simple d'emploi tout en étant susceptible de reproduire des événements observés, ce qui n'est pas le cas de la formule rationnelle. Nous reprenons la formulation déjà exposée avec la méthode Shypre. La fonction de production du Soil Conservation Service (*National Engineering Handbook*, 1985) est définie par un paramètre J qui représente la capacité de rétention du sol pour une crue donnée. Sa formulation est la suivante :

$$R(t) = \frac{(P(t) - 0,2J)^2}{P(t) - 0,8J}, \quad (4.16)$$

où :

$R(t)$: écoulement cumulé (en mm) ;

$P(t)$: pluie cumulée (en mm) ;

J : paramètre qui représente la rétention du sol (en mm).

Il existe une table pour calculer le paramètre J en fonction de l'occupation de l'espace. La confrontation de la méthode avec les données hydrométriques de bassins ruraux aboutit à des valeurs de J dans l'intervalle [100 ; 250 mm].

Chaque millimètre de pluie nette, déterminé par la fonction de production, est transformé en débit à l'exutoire par un hydrogramme unitaire synthétique. Cet hydrogramme est de forme triangulaire, avec un temps de montée T_p et un temps de décrue de $\frac{5}{3}T_p$.

Le débit pour 1 mm de pluie nette est alors donné par :

$$q(t) = \frac{0,208S}{T_p}, \quad (4.17)$$

où :

$q(t)$: débit de pointe (en $\text{m}^3/\text{s} \cdot \text{mm}$) ;

S : surface du bassin versant (en km^2) ;

T_p : temps de montée (en h).

T_p est défini en fonction de la durée de la pluie T_r et du *lag time* t_{lag} (délai de réaction du bassin versant). La méthode du SCS retient $t_{lag} = 0,6T_c$, avec T_c le temps de concentration, soit :

$$T_p = \frac{T_r}{2} + t_{lag} = \frac{T_r}{2} + 0,6T_c. \quad (4.18)$$

La sommation des hydrogrammes obtenus pour chaque pas de temps T_r de pluie permet d'obtenir l'hydrogramme final.

Remarque : si l'on considère que la pluie de référence est homogène sur le temps de concentration T_c , le paramètre T_p vaut $\frac{T_c}{2} + 0,6T_c = 1,1T_c$. On est ramené à quelque chose de semblable à la fonction de transfert de la formule rationnelle, avec un temps de base égal à $\left(1 + \frac{5}{3}\right) \times (1,1T_c) = 2,93T_c$, alors que le temps est de $2T_c$ lorsque l'on retient un triangle isocèle comme fonction de transfert, et de $3T_c$ si l'on considère une fonction de transfert avec un temps de décrue de $2T_c$.

Modèles QdF de référence

Un nombre restreint de modèles régionaux débit-durée-fréquence, notés QdF, ont été retenus pour la reconstitution du régime des crues en tout site non jaugé (Prudhomme, 1995 ; Galéa et Prudhomme, 1997). La procédure d'estimation des quantiles de crue est réalisée selon les trois étapes décrites *infra*.

Estimation de trois paramètres locaux représentatifs du bassin versant

On s'intéresse à la durée caractéristique de crue D , au débit décennal de pointe $Q(10)$ et au gradex des pluies $a_p(d)$. Les deux premiers paramètres, respectivement des descripteurs de la fonction de transfert et de production des crues, sont estimés directement à partir de mesures hydrologiques lorsqu'elles existent, ou autrement à l'aide de régressions linéaires établies à l'échelle nationale ou régionale, et utilisant des variables explicatives

facilement estimables comme la superficie du bassin versant et des valeurs caractéristiques de pluie (formules Socose et Crupedix, cf. équations 4.6 et 4.8). Le gradex des pluies est estimé pour différents pas de temps d à partir de postes pluviographiques situés sur le bassin versant, ou à partir de documents présentant la cartographie du gradex.

Choix d'un modèle de référence QdF adimensionnel

Pour différentes durées d , la valeur du rapport $\frac{a_p(d)}{Q(10)}$ obtenu en première étape est comparée à deux limites $L_1(d^* = \frac{d}{D})$ et $L_2(d^*)$.

Suivant la position de ce rapport — inférieur à $L_1 = \frac{1}{0,768d^* + 2,332}$; compris entre L_1 et L_2 ; supérieur à $L_2 = \frac{1}{0,419d^* + 1,580}$ —, on choisira respectivement l'un des trois modèles de référence, appelés Vandenesse, Florac et Soyans, d'après le nom de chaque bassin utilisé pour les constituer (figure 4.5).

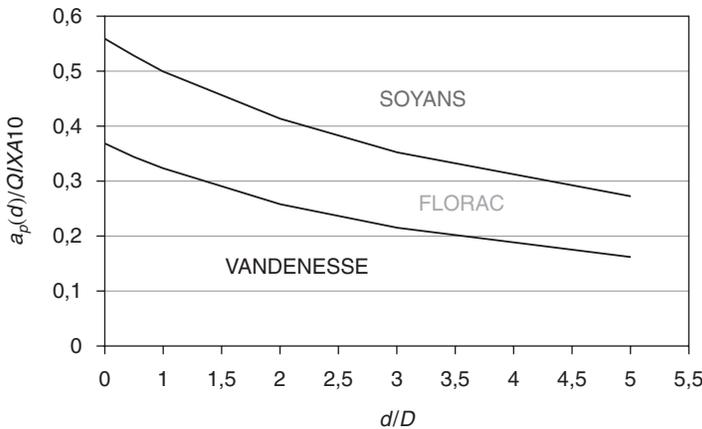


Figure 4.5. Critères de choix d'un modèle QdF de référence.

Chaque modèle QdF de référence se présente sous la forme d'une famille de courbes adimensionnelles :

$$\left\{ \begin{array}{l} T \leq 10 \text{ ans : } \frac{Q(T, d)}{Q(10)} = A\left(\frac{d}{D}\right) + B\left(\frac{d}{D}\right) \ln T \\ T > 10 \text{ ans : } \frac{Q(T, d)}{Q(10)} = \frac{Q(10, d)}{Q(10)} + C\left(\frac{d}{D}\right) \ln \left(1 + \frac{B\left(\frac{d}{D}\right)}{C\left(\frac{d}{D}\right)} \cdot \frac{T-10}{10} \right) \end{array} \right. \quad (4.19)$$

Les fonctions A , B et C de $d^* = \frac{d}{D}$ ont une forme homographique :

$$f(d^*) = \frac{1}{ad^* + b} + c,$$

où les valeurs des paramètres a , b et c dépendent des fonctions A , B et C des trois modèles de référence (cf. tableau 4.2).

Tableau 4.2. Valeurs des paramètres des trois modèles QdF de référence.

Modèle QdF	Fonction	Modèle en débit moyen $VCXd$			Modèle en débit seuil $QCXd$		
		a	b	c	a	b	c
Vandenesse	A	2,635	6,19	0,016	3,97	6,48	1,91
	B	1,045	2,385	0,172	1,91	1,91	0,097
	C	1,083	1,75	0	3,674	1,774	0,013
Florac	A	1,12	3,56	0	3,05	3,53	0
	B	0,95	3,18	0,039	2,13	2,96	0,01
	C	1,56	1,92	0,085	2,78	1,77	0,04
Soyans	A	0,87	4,6	0	2,57	4,86	0
	B	1,07	2,50	0,099	2,10	2,10	0,05
	C	0,569	0,69	0,046	1,49	0,66	0,017

Reconstitution du régime des crues par dénormalisation du modèle de référence

Le quantile local de crue est alors estimé en dénормant la distribution adimensionnelle du modèle de référence :

$$\begin{cases} Q(T, d')_{\text{local}} = \left(\frac{Q(T, d)}{Q(10)} \right)_{\text{modèle de référence}} \cdot Q(10)_{\text{local}} \\ \text{avec } \frac{d'}{D_{\text{local}}} = \frac{d}{D_{\text{modèle}}} \end{cases} \quad (4.20)$$

On voit par exemple sur la figure 4.6, à gauche les courbes QdF adimensionnelles du modèle de Vandenesse (équation 4.19 et tableau 4.2), et à droite les courbes QdF reconstituées à l'aide de l'équation 4.20.

Cette procédure de régionalisation comporte plusieurs aspects originaux.

– L'objectif de la méthode est de reconstituer la distribution des crues, pas seulement pour le débit de pointe mais également pour des débits calculés sur une durée d quelconque (cf. quantile $Q(T, d)$ des équations 4.19 et 4.20). La variable Q du débit étudié est relative,

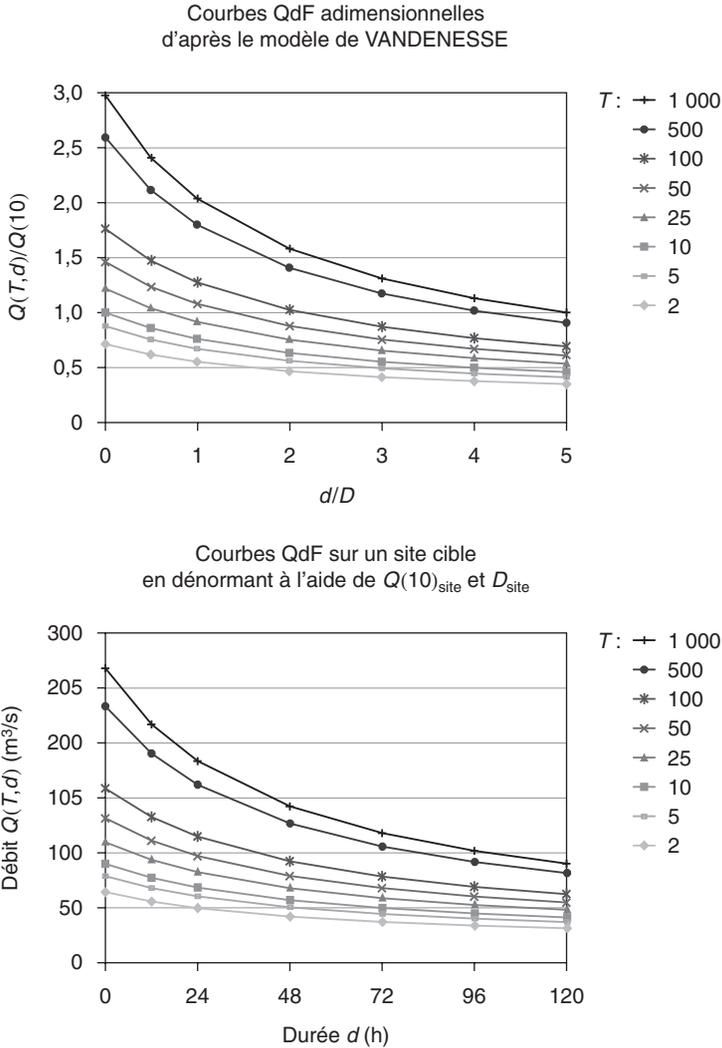


Figure 4.6. Estimation sur site non jaugé à partir d'un modèle QdF de référence.

soit à un débit moyen maximum sur la durée d , débit moyen $VCXd(T)$, soit à un débit dépassé de façon continue pendant la durée d , débit seuil maximum $QCXd(T)$.

– La norme de débit $Q(10)$ utilisée pour travailler sur une distribution adimensionnelle des crues est relative à une période de retour de 10 ans, bien plus élevée que celle utilisée par la méthode de l'indice de crue (cf. § *Modèle QdF régional convergent*), où la moyenne (ou la médiane) des valeurs maximales annuelles de débits de crue a une période de retour voisine de 2 ans. Ce choix conduit à donner davantage de poids, dans la procédure de régionalisation, aux descripteurs locaux. En contrepartie, les modèles de référence ont un champ d'application beaucoup plus vaste. Actuellement, de nombreux régimes de crue

en France et en Europe ont pu être reconstitués à l'aide d'un des trois modèles de référence adimensionnel et des deux descripteurs locaux.

– La procédure de régionalisation intègre une information pluviométrique, par le gradex des pluies. Cette information est exploitée à deux niveaux : dans le choix du modèle de

référence par le critère $\frac{a_p(d)}{Q(10)}$, et dans la forme des courbes adimensionnelles de référence

où la deuxième partie de l'équation 4.19 ($T > 10$ ans) a une forme issue du modèle Agregee.

Modèle QdF régional convergent

La méthode d'estimation précédente se raisonne sur trois distributions régionales adimensionnelles (chacune ayant trois paramètres), dont l'une est censée présenter une forte similarité hydrologique avec le site non jaugé à étudier. Les courbes QdF adimensionnelles de référence sont alors transférées d'un des trois modèles au site projet.

Une approche inverse consiste à d'abord retenir les sites hydrométriques « voisins » du site à étudier, puis à construire une distribution régionale adimensionnelle en agglomérant toutes les données de ce voisinage. L'intérêt est non seulement de transférer l'information d'un site mesuré à un site quelconque non jaugé mais aussi de consolider à l'échelle régionale toute l'information disponible.

Une première application de cette approche, connue sous le nom de la méthode de l'indice de crue (Dalrymple, 1960), consiste à supposer que la distribution des crues de tous les sites d'une même région hydrologique est identique, à un facteur d'échelle près, fonction des caractéristiques physiographiques de chaque bassin. Cette méthode a été reprise depuis dans de nombreux pays, comme le Royaume-Uni (Reed *et al.*, 1999), en suivant trois étapes.

• Définition d'une région hydrologiquement homogène

Initialement les régions étaient définies suivant des critères géographiques, avec une partition de l'espace en régions. D'autres critères ont ensuite été introduits pour aboutir à la notion de voisinage hydrologique, basée sur une distance hydrologique (Ouarda *et al.*, 1999). Ainsi, pourront être considérés comme voisins des bassins pas forcément contigus, mais aux caractéristiques physiographiques et climatologiques similaires.

• Construction d'une distribution régionale $Q^*(T)$

Les données de m stations similaires au site cible sont agglomérées en un seul échantillon, après avoir transformé les valeurs de chaque échantillon de crue $x_k(1) < \dots < x_k(i) < \dots < x_k(N_k)$ de la station n° k par la relation :

$$x_k^*(i) = \frac{x_k(i)}{\mu_k}, \quad (4.21)$$

où μ_k est une statistique de l'échantillon, appelée indice de crue (souvent pris égal à la moyenne de l'échantillon). Des techniques de pondération ont été introduites par Hosking *et al.* (1985) pour tenir compte du fait que les m stations du voisinage ne disposent pas forcément d'échantillons de crue de taille identique.

• Estimation de la distribution des crues en un site quelconque

Le quantile de crue au droit du site à étudier (projet) est obtenu en multipliant la distribution régionale par l'indice de crue :

$$Q_{\text{projet}}(T) = Q^*(T) \cdot \mu_{\text{projet}} \quad (4.22)$$

Une formule régionalisée, comme celles présentées au chapitre 4, § *Régressions multiples*, permet d'obtenir la valeur de l'indice de crue μ_{projet} à partir de la superficie, de la pluviométrie...

Ce type d'approche de régionalisation a été repris sur le bassin de la Moselle par Javelle (2001) et également au Québec par Javelle *et al.* (2002), en combinant la méthode de l'indice de crue avec le modèle QdF convergent. Il est ainsi possible de reconstituer le quantile $Q(T, d)$ en un point quelconque à partir de l'estimation de deux paramètres locaux, la durée Δ et le quantile de pointe de crue décennal $Q(10)$, et l'estimation des paramètres de la distribution régionale, obtenue par agglomération des données des stations du voisinage hydrologique.

• **Lien entre le paramètre Δ et la durée de Socose**

Pour faciliter l'interprétation physique et l'estimation du paramètre Δ , Javelle (2001) a comparé cette variable à la durée de Socose (D_{Socose}), médiane conditionnelle de la durée de dépassement de la moitié du débit de pointe pour une crue décennale. L'application a concerné 103 bassins versants français, de surface comprise entre 7 et 9 400 km², et de durée d'enregistrement comprise entre 9 et 33 ans. Dans ce contexte, il existe une solide relation entre les deux durées (figure 4.7), qui renforce l'interprétation de Δ en termes de dynamique de crue :

$$\Delta = 3,44D_{\text{Socose}}, \text{ avec } r^2 = 0,85 \quad (4.23)$$

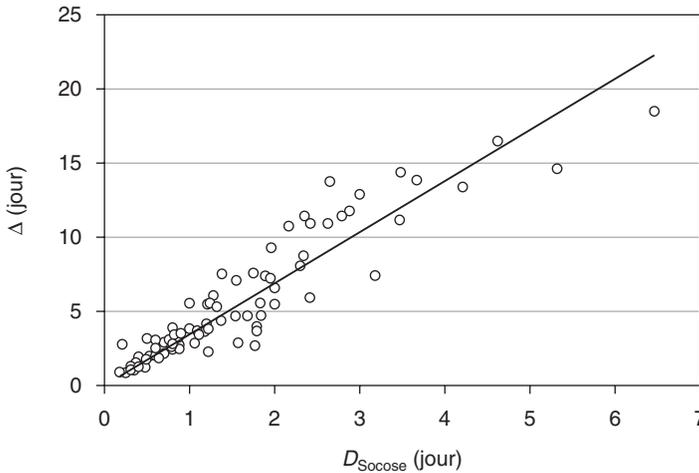


Figure 4.7. Relation entre la durée Δ et la durée de Socose.

Modèle Shyreg : une méthode régionale pour l'estimation des débits de crue

Cette méthode a été développée pour permettre une première estimation rapide des débits de référence dans le cadre d'élaboration des PPRI. La méthode a été calée sur les quantiles de débits de pointe et journalier des bassins versants jaugés, dans une gamme de superficie de 10 à 1 000 km². Son domaine d'application doit être limité à cette gamme de superficies et, bien sûr, aux cours d'eau qui ne présentent pas de spécificités hydrologiques et hydrauliques marquées : bassins excessivement karstiques, bassins avec des champs d'inondation importants... Shyreg est une application régionale de la méthode Shypre, qui associe un générateur de pluies horaires et une modélisation de la pluie en débit pour générer des scénarios de crue. Le développement de Shyreg a nécessité la régionalisation du générateur de pluies horaires et une approche régionale de la modélisation de la pluie en débit. Shyreg aboutit à l'élaboration de bases de données pixellisées au km² des pluies et des débits qui sont utilisées pour estimer, en tout point du réseau hydrographique, les débits de différentes durées et de différentes fréquences. Un logiciel a été développé, PRD'crue, pour exploiter les bases de données générées par l'application Shyreg. En *tableaux 4.3 et 4.4*, nous avons reporté des sorties du logiciel sur un pixel de l'espace. Les pluies et les débits sont définis pour des périodes de retour entre 2 et 100 ans et pour des durées de 1 à 72 heures.

Les informations contenues dans les *tableaux 4.3 et 4.4* concernent la région méditerranéenne mais elles sont également disponibles pour l'ensemble de la France métropolitaine et sur les départements d'outre-mer.

Les débits spécifiques permettent de calculer les débits de crue en tout point du réseau hydrographique des zones étudiées (actuellement, les 17 départements de la France méditerranéenne et l'île de la Réunion). Il est prévu de calculer systématiquement les débits de crue sur les bassins versants de la zone étudiée et de créer un site Web pour les mettre à disposition des utilisateurs.

Le générateur de hyétogrammes

Dans sa conception initiale — calage local sur une chronique de pluies horaires — le modèle de génération de pluies est défini par huit variables. La définition de ces variables pour les deux saisons considérées par le modèle (hiver de décembre à mai, été de juin à novembre) nécessite le calage de 21 paramètres. La régionalisation du modèle comporte trois étapes.

– *La simplification du modèle.* Différents tests de sensibilité ont permis de réduire, pour chaque saison, le nombre de variables du modèle à cinq sans altérer sensiblement ses performances.

– *La recherche de variables exogènes explicatives des paramètres du modèle de génération.* L'information de la pluie journalière a pu être utilisée, ce qui est particulièrement intéressant au regard de la bonne disponibilité de cette information dans l'espace. Les grandeurs caractéristiques de la pluie au pas de temps journalier, déterminées en chaque poste pour les deux saisons, sont le nombre N_e moyen d'événements pluvieux par saison, la moyenne des pluies journalières maximales de chacun des épisodes $\mu PJMAX$ (en mm) et la moyenne des durées des épisodes pluvieux $\mu DTOT$ (en jours).

Tableau 4.3. Hauteurs $PXd(T)$ de pluie maximale (en mm), de différentes durées d et périodes de retour T , d'un pixel de la base des pluies Shyreg (région méditerranéenne).

T	2 ans	5 ans	10 ans	20 ans	50 ans	100 ans
$PXd(1\text{ h})$	36,5	47,4	58,2	69,2	84,8	97,1
$PXd(2\text{ h})$	49,2	62,0	74,9	88,4	106,2	120,1
$PXd(3\text{ h})$	61,4	77,0	91,4	106,6	126,0	141,8
$PXd(4\text{ h})$	70,5	88,7	104,3	120,9	142,7	161,3
$PXd(6\text{ h})$	83,5	106,0	124,8	144,8	174,2	199,8
$PXd(12\text{ h})$	109,2	142,5	170,8	206,1	253,2	287,9
$PXd(24\text{ h})$	131,8	178,3	220,3	265,7	331,4	377,9
$PXd(48\text{ h})$	150,0	206,4	257,4	315,0	386,4	435,5
$PXd(72\text{ h})$	159,4	217,8	270,6	332,1	405,7	457,3

Tableau 4.4. Débits spécifiques de crue $Q_d(T)$ (en $\text{m}^3/\text{s} \cdot \text{km}^2$), de différentes durées d et périodes de retour T , d'un pixel de la base des débits Shyreg (région méditerranéenne).

T	2 ans	5 ans	10 ans	20 ans	50 ans	100 ans
$Q_{\text{pointe}}(d = 0)$	1,3	2,3	3,3	4,6	6,5	8,0
$Q_d(1\text{ h})$	1,2	2,1	2,9	4,1	5,9	7,2
$Q_d(2\text{ h})$	1,1	2,0	2,8	4,0	5,7	6,9
$Q_d(3\text{ h})$	1,1	1,9	2,7	3,7	5,4	6,6
$Q_d(4\text{ h})$	1,0	1,8	2,5	3,5	5,1	6,3
$Q_d(6\text{ h})$	0,9	1,6	2,2	3,1	4,5	5,6
$Q_d(12\text{ h})$	0,7	1,2	1,7	2,3	3,4	4,3
$Q_d(24\text{ h})$	0,4	0,7	1,1	1,4	2,1	2,7
$Q_d(48\text{ h})$	0,3	0,4	0,6	0,8	1,2	1,5
$Q_d(72\text{ h})$	0,2	0,3	0,4	0,6	0,8	1,0

– La régionalisation proprement dite des trois paramètres N_e , $\mu PJMAX$ et $\mu DTOT$. L'objectif de cette étape est de mettre en évidence une structure spatiale afin de disposer d'une estimation des paramètres étudiés en tout point de la zone étudiée. Une approche semblable à celle de la méthode Aurelhy (Benichou *et al.*, 1986) est utilisée, sur la base d'une grille d'altitude plus fine (la maille de la grille est ici de 9 km^2 , contre 100 km^2 dans la méthode Aurelhy). Un traitement en composantes principales permet de dégager les

éléments structurants du relief, qui sont utilisés comme variables explicatives. Des équations de régression linéaire sont ensuite établies avec les composantes principales du relief et avec des paramètres ayant une influence sur la climatologie locale, tels que la proximité de la mer ou de reliefs importants, la présence de grandes vallées... La *figure 4.8* présente les cartographies des paramètres du modèle, et, à titre d'exemple, nous avons reporté les quantiles décennaux de la pluie journalière et des pluies de durée 1 heure de la France métropolitaine (*figure 4.9*).

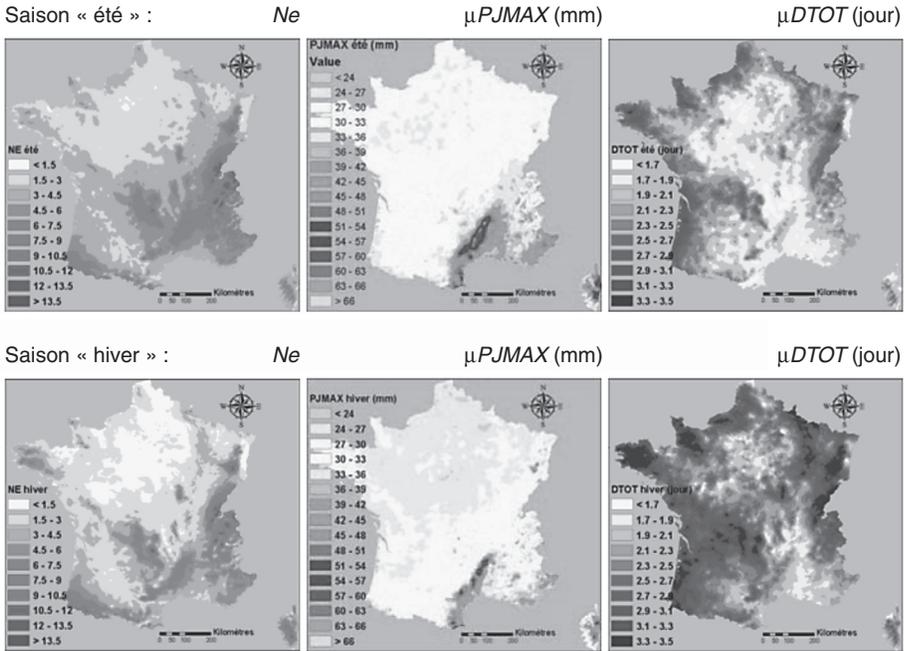


Figure 4.8. Cartographie au pixel du km² des paramètres journaliers du modèle Shyreg.

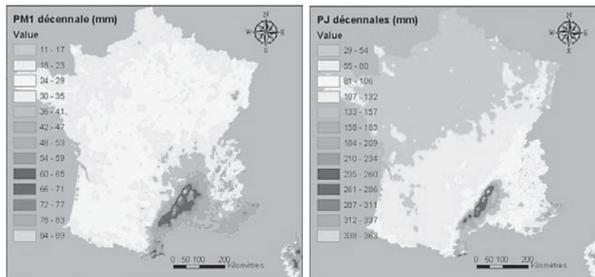


Figure 4.9. Exemples de cartes de quantiles de pluies obtenues par simulations du modèle Shyreg.

La modélisation de la pluie en débit

Lors des premiers développements de Shyreg, le modèle SCS a été utilisé, bien qu'il ne soit pas le meilleur outil de modélisation pluie-débit utilisable. Mais sa robustesse est reconnue, et son faible paramétrage est un atout pour une approche régionale. Une modélisation de la pluie en débit par un modèle GR simplifié à un paramètre unique est en cours de développement. Par ailleurs, l'objectif de construire une méthodologie susceptible d'estimer les débits de crue en tout point du réseau hydrographique, donc pour des superficies de bassin versant très variables, soulève, entre autres, le problème de la connaissance spatiale de la pluie. Quelques formulations de l'abattement des pluies en fonction de la superficie des bassins versants sont disponibles (Neppel *et al.*, 2003), mais elles sont limitées à des quantiles de pluie. Rien d'opérationnel n'est disponible pour une approche spatio-temporelle à faible pas de temps.

À ce stade de l'étude, et ce, afin d'évacuer le problème de l'abattement des pluies, nous appliquons le modèle de transformation de la pluie en débit à un pas d'espace fin, en faisant l'hypothèse que l'abattement des pluies devient négligeable. En fait, l'espace d'étude est approché par une grille régulière de maille 1×1 km, chaque maille constituant un bassin versant virtuel de 1 km^2 sur lequel la transformation pluie-débit est effectuée. Afin de simplifier au maximum la modélisation, une fonction de transfert unique (caractérisée par une valeur de α et de t_m , cf. équation 3.16) est retenue pour l'ensemble des pixels de 1 km^2 . Les bassins versants virtuels sont donc caractérisés par le seul paramètre J de la fonction de production (équation 4.16).

La régionalisation du modèle pluie-débit

Les paramètres α et t_m sont constants pour l'ensemble des pixels. La régionalisation porte essentiellement sur le paramètre J de la fonction de production, qui représente la rétention maximale du bassin versant. Il peut être déterminé à partir de tables publiées par le Soil Conservation Service (*National Engineering Handbook*, 1985). Dans le cadre du développement du modèle Shyreg, il a été choisi de déduire le paramètre J d'informations locales issues des bases de données disponibles : altimétrie, occupation de l'espace, géologie. Le croisement de ces couches d'information permet d'établir une cartographie du paramètre de production du modèle à la maille du km^2 . À cette étape, il est possible, pour chacune des mailles, de simuler des chroniques horaires de pluie et de les transformer en hydrogrammes de crue. De ces chroniques de crue sont déduits les quantiles de débit de différentes durées et de différentes fréquences qu'il est possible de cartographier (*tableau 4.4 et figure 4.10*).

Agglomération de l'information à l'échelle du bassin versant

Pour obtenir le débit à l'exutoire d'un bassin versant, il est nécessaire de définir une règle d'agglomération des débits des pixels qui le composent. L'agglomération des pixels d'un quantile de débit donné pour estimer le même quantile de débit à l'exutoire du bassin versant n'est pas une pratique classique. Elle a conduit à proposer la notion de Fonction de transfert statistique (FTS). Des développements méthodologiques pour améliorer l'estimation de cette FTS sont encore nécessaires. Nous nous limitons ici à la définition d'une fonction FTS globale qui dépend uniquement de la superficie du bassin versant. Elle s'écrit :

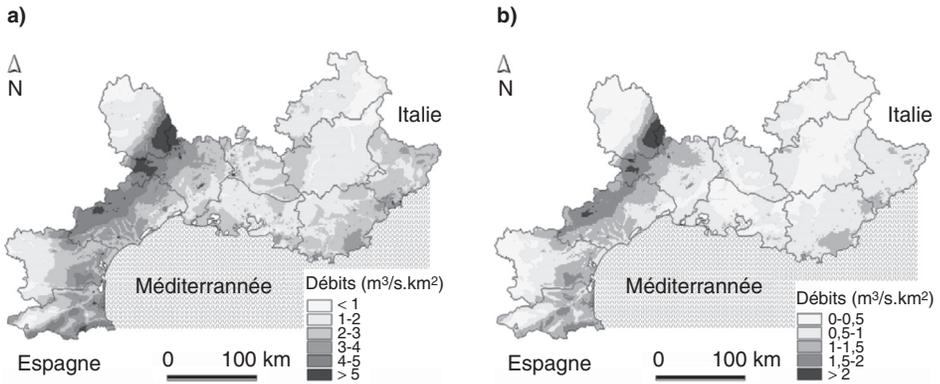


Figure 4.10. Cartographie par le modèle Shyreg des quantiles décennaux des débits de pointe et des débits journaliers de la zone méditerranéenne française : (a) débit de pointe de crue ; (b) débit journalier.

$$Q(d, T) = \left(\sum_{i=1}^n (q(d, T))_i \right) \cdot S^{1-\beta(1-S^{-\gamma})}, \quad (4.24)$$

avec :

$Q(d, T)$: débit de durée d et de période de retour T à l'exutoire du bassin versant ;

$[q(d, T)]_i$: débit de durée d et de période de retour T pour un pixel donné i à l'intérieur des limites du bassin versant ;

N : nombre de pixels à l'intérieur du bassin versant ;

S : surface du bassin versant en km^2 (très peu différente de n) ;

β : paramètre fonction de la durée d des débits (débit de pointe, débit moyen en 1, 2, ..., 72 h). β décroît avec la durée ;

γ : paramètre qui limite l'abattement pour les petites surfaces. Il est fonction de la période de retour T .

Cette formulation a été utilisée pour estimer les débits de crue pour l'ensemble des bassins versants instrumentés de la zone géographique étudiée, ce qui a permis de contrôler la pertinence de l'approche.

Contrôle des performances du modèle régional

Le modèle est évalué sur la reconstitution des quantiles décennaux en débit de pointe de crue et en débit journalier, avec un coefficient de Nash respectivement égal à 86 % et 87 %. Malgré ces valeurs élevées, on remarque sur la *figure 4.11* une tendance assez marquée à la surestimation des débits obtenus par simulation, notamment pour les valeurs faibles de débits. Dans 25 % des cas, la surestimation des quantiles décennaux de débit atteint un ratio de 2, et elle peut atteindre des valeurs plus importantes pour certains bassins versants, généralement de type karstique. Une idée de l'incertitude de la méthode peut être avancée : dans 75 % des cas, l'estimation régionale est comprise dans la fourchette du double et de la moitié de la valeur déduite de l'observation.

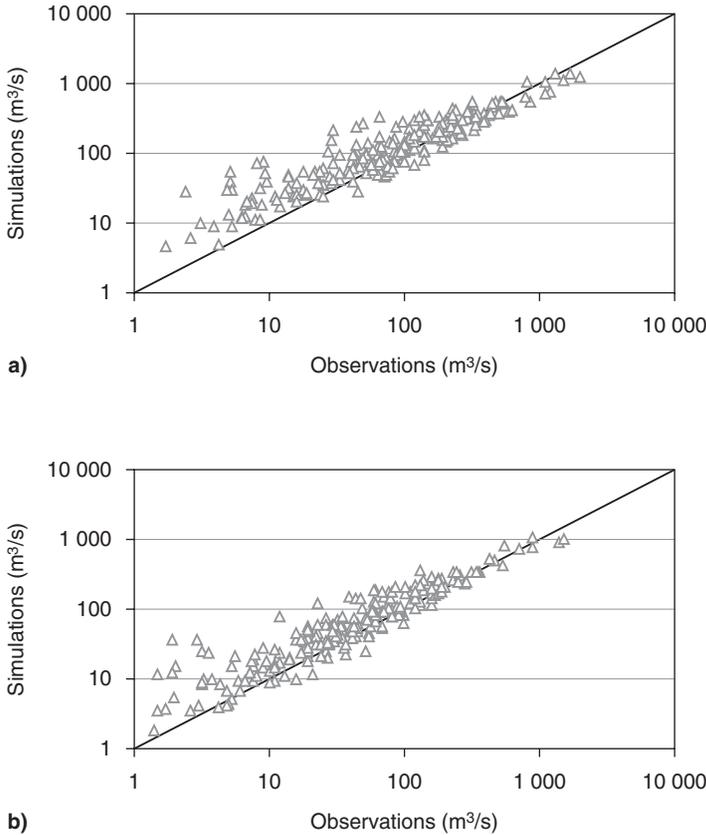


Figure 4.11. Comparaison du débit décennal estimé à partir des observations ou simulations issues de Shyreg : (a) débit de pointe de crue ; (b) débit journalier.

La méthode est entre autres destinée à l'estimation des quantiles centennaux. À des fins de vérification, nous avons comparé les résultats de la simulation aux résultats obtenus par la méthode du Gradex qui, rappelons-le, est d'utilisation courante en France pour l'estimation des débits rares et exceptionnels.

Pour l'appliquer rapidement sur l'ensemble des 220 bassins versants étudiés, les hypothèses suivantes ont été posées :

- la méthode est utilisée au pas de temps journalier. Le gradex des pluies journalières est estimé par l'expression suivante :

$$\text{gradex} = \frac{PJ(100) - PJ(10)}{2,3}, \quad (4.25)$$

où $PJ(10)$ et $PJ(100)$ représentent respectivement les pluies journalières de période de retour 10 et 100 ans, estimées par Shyreg et exprimées en mm.

- le rapport d'affinité entre les débits de pointe et les débits journaliers est unique et égal, pour un bassin versant donné, au ratio entre les quantiles décennaux des débits de pointe de crue et des débits journaliers.

Sur l'ensemble des bassins versants, le gradex varie entre 16 mm/jour et 60 mm/jour ; les ratios entre les quantiles décennaux appartiennent à la fourchette [1,1;4,7]. On constate sur la *figure 4.12* une bonne liaison entre les estimations centennales proposées par les deux méthodes, avec un coefficient de Nash égal à 90 %. On remarque aussi la tendance de la simulation à la surestimation des faibles débits. Cependant, les faibles débits sont principalement observés sur des bassins versants de faible superficie, et peuvent être sous-estimés par la méthode du Gradex, ici appliquée à un pas de temps journalier pour une vérification rapide — mais qui mériterait d'être complétée.

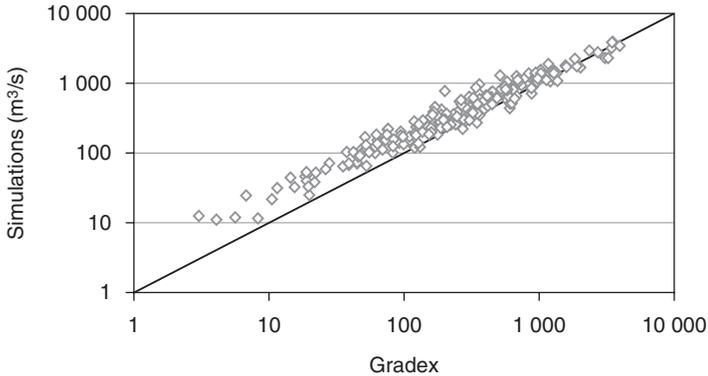


Figure 4.12. Comparaison du quantile centennial (en m^3/s) des débits de pointe de crue estimés par la méthode du Gradex, au pas de temps journalier, et par simulation du modèle régional Shyreg.

Recommandations pratiques pour l'estimation de débits de crue de référence en site non jaugé

Comme indiqué en introduction, la densité du réseau hydrométrique fait que la plupart du temps il est nécessaire de transférer une estimation de quantile de crue d'un site mesuré à un site non jaugé. Il faut garder à l'esprit que les formules sommaires d'estimation (régressions multiples et modèles préparamétrés) restent très imprécises, et que leurs résultats ne doivent être considérés que comme un premier ordre de grandeur. Des approches plus élaborées (modèles QdF et Shyreg) sont conseillées, sachant qu'en tout état de cause il convient toujours de vérifier les résultats annoncés par les modèles régionaux sur des stations hydrométriques du secteur étudié. Il faut ainsi systématiquement les comparer avec ceux obtenus par une approche locale, tant sur le domaine des observations (chapitre 2) que sur le domaine des extrapolations (chapitre 3).

Une action nationale d'inter-comparaison des méthodes d'estimation de débits de crue de référence devrait être soutenue à partir de 2008 par le ministère de l'Écologie, du Développement et de l'Aménagement durables. Un des volets portera sur la validation croisée des méthodes par test des approches régionales sur des sites jaugés non utilisés pour le calage des modèles.

Chapitre 5

Exemples d'application

Construction d'un hydrogramme de projet relatif à la crue centennale. Cas du bassin versant de la Moselle à Rupt-sur-Moselle

L'exemple traité est celui de la Moselle à Rupt-sur-Moselle (A4050620), sur la période 1980-1997. Le bassin se situe dans le Nord-est de la France et draine 153 km^2 . Les valeurs extraites de la chronique au pas de temps variable sont pour partie issues des travaux de Javelle (2001). Elles ont fait l'objet d'une critique rapide, et donc mériteraient une réactualisation et une étude plus détaillée (en partie sur la qualité de la courbe de tarage). Les quantiles obtenus dans cet exemple d'application ne peuvent être exploités au voisinage immédiat de la station pour des besoins opérationnels.

Notre objectif est l'obtention d'un hydrogramme de projet centennal. Pour y parvenir, nous exploiterons successivement les éléments sur la dynamique de crue (cf. chapitre 3, § *Hydrogramme moyen monofréquence*), les courbes QdF et IdF sur la base des volumes écoulés et précipités à l'échelle du bassin (cf. chapitre 3, § *Établissement de courbes débit-durée-fréquence à partir d'observations hydrométriques*) dans le domaine des observations, et l'extrapolation de la distribution des crues selon la méthode du Gradex esthétique (cf. chapitre 3, § *Extrapolation de la distribution des crues par le modèle Agregée*). L'hydrogramme de projet issu de cette étude sera comparé à celui obtenu par une approche sommaire basée uniquement sur l'ajustement d'une loi de Gumbel sur les débits de pointe et le choix d'un hydrogramme de crue triangulaire.

Examen de la dynamique de crue

Cette durée est appréciée à partir d'une sélection d'hydrogrammes sur une fenêtre autour des pics de crues majeures. Pour chaque station, nous avons extrait l'événement majeur de chaque année, l'année hydrologique débutant en septembre. Puis, nous avons rendu adimensionnels les hydrogrammes en divisant chacun d'eux par le débit de pointe Q_{\max} . Tous les hydrogrammes normés ont donc un débit maximal de 1. À partir de cet échantillon de crues normées, nous construisons un hydrogramme supposé représentatif

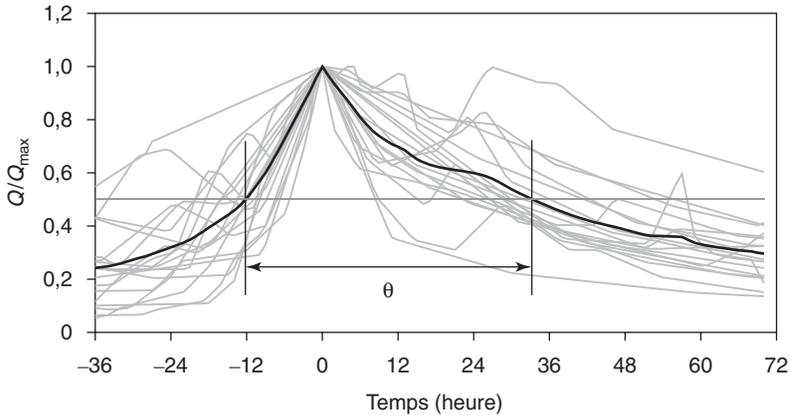


Figure 5.1. Hydrogrammes adimensionnels des crues sélectionnées par max annuel.

du régime de crue par simple moyenne arithmétique. Finalement la durée θ correspond au temps de dépassement de la moitié du débit de pointe de l'hydrogramme normé moyen ($Q = 0,5$). Les éléments de la *figure 5.1* illustrent la procédure de calcul. L'hydrogramme moyen est tracé en noir, et en gris l'ensemble des événements qui ont conduit à son calcul.

Les crues ont une durée de l'ordre de 4,5 jours, et le cœur de la crue s'évacue sur $\theta = 2$ jours.

Les volumes générés sur différentes durées

Compte tenu de la durée des épisodes, l'étude des volumes générés concerne six durées multiples de la durée θ : 0, 6, 12, 24, 48 et 108 heures (0, $\theta/8$, $\theta/4$, $\theta/2$, θ et $2,5\theta$). Les valeurs de référence nécessaires à l'élaboration des courbes QdF sont obtenues avec un échantillonnage des valeurs maximales d'un ensemble d'événements, ceux-ci commençant et finissant lorsque le débit passe au-dessus, puis au-dessous, d'un seuil fixé. Nous retiendrons la valeur maximale et la date d'apparition de chaque événement.

L'estimation des quantiles de crue repose sur l'étude conjointe :

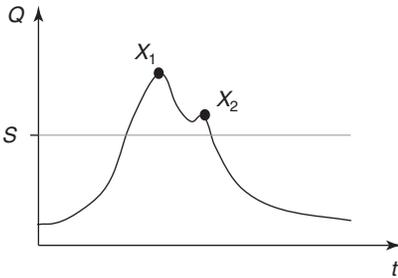
- du processus d'occurrence (méthode du renouvellement), à partir du nombre $m(i)$ d'événements sélectionnés chaque année i ;
- de la distribution fréquentielle du maximum des événements X_S supérieurs au seuil S , permettant de définir la fréquence théorique de chaque épisode, exprimée en termes de période de retour T .

L'extraction des valeurs supérieures à un seuil est plus délicate que son homologue par maximum annuel. Il faut notamment veiller à respecter l'indépendance des événements — pour ne pas introduire une information redondante — et assurer la robustesse des estimations des paramètres des lois ajustées, et ainsi la robustesse des valeurs des quantiles. Trois paramètres S , α et δ_{\min} ont été définis pour y parvenir (Lang *et al.*, 1999), qui alimentent une procédure d'extraction automatique des événements de la chronique analysée.

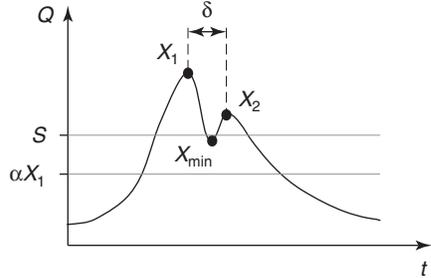
Considérons dans la chronique un épisode de maximum $X_1 (> S)$, observé en t_1 , et un maximum local X_2 , observé en t_2 , avec $t_2 > t_1$ et $X_1 \geq X_2 \geq S$. L'épisode de crue associé à X_2 rejoint la sélection des événements analysés si :

- le débit entre t_1 et t_2 est redescendu sous le seuil S et sous un certain pourcentage α de la valeur maximale de l'épisode précédent, c.-à-d. $Q(t) < \min(\alpha X_1; S)$, avec $\alpha \leq 1$;
- une durée minimale δ_{\min} est respectée entre deux valeurs échantillonnées, c.-à-d. $t_2 - t_1 > \delta_{\min}$.

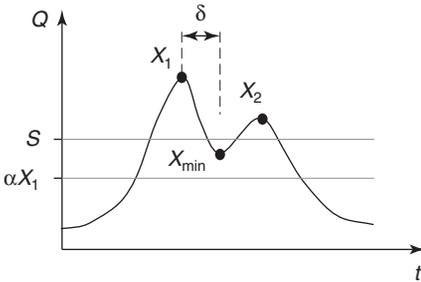
La figure 5.2 présente plusieurs configurations types.



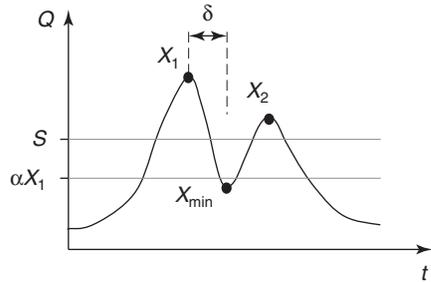
(1) : X_2 est un maximal local, le débit n'est pas redescendu en dessous du seuil S , seule la valeur X_2 est conservée.



(2) : entre X_1 et X_2 , le débit est redescendu sous le seuil S (deux événements sont identifiés), mais reste supérieur à αX_1 ; de plus l'espacement temporel requis n'est pas vérifié ; seule la valeur X_1 est conservée.



(3) : entre X_1 et X_2 , le débit est redescendu en dessous du seuil S , deux événements sont donc identifiés ; la contrainte temporelle est respectée mais le débit n'est pas redescendu en dessous de αX_1 ; seule la valeur X_1 est conservée.



(4) : entre X_1 et X_2 , le débit est redescendu en dessous du seuil S , deux événements sont donc identifiés ; les deux contraintes d'indépendance sont respectées ; X_1 et X_2 font partie de l'échantillon.

Figure 5.2. Modalités d'extraction des valeurs supérieures à un seuil.

La difficulté majeure est d'attribuer des valeurs pertinentes aux paramètres d'échantillonnage. On comprend en effet qu'il n'est pas judicieux d'imposer une contrainte temporelle très courte — une heure, par exemple — pour un bassin versant de temps de réponse de l'ordre de plusieurs jours. De même, trop contraindre le critère de redescente perturbe la sélection des événements lorsque le régime présente une forte composante souterraine (c'est-à-dire un fort débit de base).

Nous suggérons des valeurs par défaut inspirées de la physique des phénomènes :

- le critère d'espacement δ_{\min} a un lien étroit avec la dynamique de crue, et une valeur par défaut peut être obtenue à partir d'une sélection d'hydrogrammes de crue dont on extrait un hydrogramme sans dimension représentatif du régime de crue par simple moyenne arithmétique. La valeur δ_{\min} est prise égale à la durée θ de dépassement de 1/2 pour l'hydrogramme normé moyen (figure 5.1) ;
- le critère de redescente α est conditionné par le débit de base : sa valeur évolue avec la durée analysée entre 0,05 et 0,50 (compte tenu du lissage croissant avec la durée d'intégration pour les moyennes mobiles). Une étude du rapport débit de pointe/débit de base permet de fixer une valeur pour les débits instantanés ;
- le seuil S doit être suffisamment élevé pour ne piéger que les crues, et suffisamment bas pour obtenir un nombre suffisant de valeurs. Si la rivière présente un comportement spécifique au-delà d'un certain seuil (transparence d'ouvrages en amont), celui-ci pourra être retenu pour garantir une homogénéité des types de crue.

Une démarche itérative permet ensuite de confirmer le choix des valeurs par défaut des trois paramètres d'échantillonnage (δ_{\min} , α et S). Un premier test consiste à vérifier que le processus d'occurrence des crues suit une loi de Poisson qui suppose implicitement la stationnarité du processus et l'indépendance des valeurs de l'échantillon. On peut utiliser

le test de Cunnane (1979), basé sur l'indice de dispersion $I = \frac{\text{var}(m)}{E(m)}$ du processus d'occur-

rence, où $m(i)$ est le nombre d'événements sélectionnés chaque année i . Dans le cas d'un processus d'occurrence qui suit une loi de Poisson, l'indice de dispersion est théoriquement égal à 1. Si sur la série analysée, l'indice de dispersion s'écarte de manière significative de 1 (c.-à-d. la valeur se situe en dehors de l'intervalle de confiance à 90 % centré sur 1), il faut soit modifier les critères d'échantillonnage, soit retenir une autre loi pour modéliser le processus d'occurrence. En règle générale, pour les crues, il suffit de vérifier la stationnarité du processus (cf. annexe 2, § *Les tests statistiques*) et de choisir des paramètres d'échantillonnage permettant de respecter la contrainte sur l'indice de dispersion (à un risque d'erreur près).

Un second test consiste à représenter l'évolution de la moyenne des dépassements au-dessus du seuil en fonction du seuil : $E(X_S) - S = f(S)$. On peut montrer d'un point de vue théorique (Davison et Smith, 1990 ; Naden et Bayliss, 1993) que la fonction f présente un comportement linéaire lorsque la distribution des valeurs X_S suit une loi de Pareto généralisée, et ne change pas avec le seuil avec une loi exponentielle. Dans ce dernier cas, la moyenne des dépassements au-dessus du seuil correspond au paramètre d'échelle de la loi exponentielle. Le test est alors équivalent à celui d'une analyse de sensibilité où l'estimation du paramètre d'échelle varie peu avec le choix du seuil.

Un troisième critère consiste à retenir un seuil avec suffisamment de valeurs, mais pas trop pour ne pas sélectionner des valeurs trop basses, qui pourraient ne pas suivre la même distribution que les valeurs les plus fortes (cf. hypothèse d'homogénéité des valeurs de l'échantillon). Cunnane (1973) a montré, dans le cas d'une loi exponentielle pour la distribution des dépassements au-dessus du seuil, que pour compenser l'introduction du paramètre de la loi de Poisson il fallait sélectionner un nombre moyen annuel de valeurs $\mu > 1,65$ pour que l'incertitude d'échantillonnage soit moins forte que celle obtenue à partir de valeurs maximales annuelles. Dans le cas de lois à trois paramètres (cf. loi de Pareto

généralisée avec un paramètre de forme $k \neq 0$), Madsen *et al.* (1997) ont montré que la condition sur le nombre μ dépendait de la valeur du paramètre k . Dans la pratique, on prendra un seuil avec $\mu \in [2 ; 4 \text{ à } 5]$.

Un dernier test (Lang, 1995) consiste enfin à vérifier la stationnarité du processus d'occurrence depuis le début des observations, à partir du graphique du nombre cumulé d'événements m_t . Si l'on suspecte un changement à une date donnée t^* , on vérifie que le nombre m_{t^*} est compris dans l'intervalle de tolérance à 90 %, avec un risque d'erreur $\alpha = 10$ %. S'il n'y a pas a priori sur un changement quelconque, on compte le nombre de points de la courbe expérimentale sortant de l'intervalle de tolérance. Lang *et al.* (2004) ont mis au point un tableau (*cf.* annexe 2, § *Les tests statistiques*) qui donne le pourcentage minimum de points devant sortir de l'intervalle de tolérance $I_{1-\alpha}$ pour rejeter l'hypothèse de stationnarité avec le risque d'erreur α . Ce pourcentage dépend de la taille de l'échantillon, du nombre moyen annuel de valeurs μ et du risque α choisi. Si la taille de l'échantillon est supérieure à 25, une règle simplifiée donne un pourcentage limite de points hors intervalle de respectivement 29 %, 26 % et 21 %, pour les risques $\alpha = 10$ %, 5 % et 1 %.

Les valeurs considérées sur cet exemple sont :

- $\delta_{\min} = \theta = 2$ jours pour les cinq premières durées, et 4,5 jours pour l'ultime durée examinée (pour éviter le recouvrement des périodes glissantes utilisées pour calculer Vd) ;
- α , de 0,2 pour l'extraction des valeurs instantanées à 0,5 pour la durée 4,5 jours, pour rendre compte du lissage croissant avec la durée d'intégration.

Les deux premiers tests (a) et (b) de la *figure 5.3* nous incitent à choisir un seuil dans la gamme $[20 ; 60 \text{ m}^3/\text{s}]$. Le troisième test (c) conduit à retenir un seuil $S = 54,5 \text{ m}^3/\text{s}$, de façon à avoir en moyenne deux événements par an. On vérifie enfin sur le quatrième test (d) que, pour ce dernier seuil, le processus peut être considéré comme stationnaire avec un nombre cumulé d'événements depuis le début des observations qui reste compris dans les bornes de l'intervalle de confiance à 90 %. Il y a quatre valeurs en dehors de l'intervalle, soit un pourcentage $4/40 = 10$ % qui est inférieur au pourcentage limite de 29 % (*cf.* annexe 2, § *Les tests statistiques*).

Nous avons enfin ajusté les paramètres d'une loi exponentielle aux échantillons constitués pour les différents seuils, et nous avons extrait les quantiles de période de retour 2, 5 et 10 ans, sans examiner la qualité des ajustements. Si les quantiles de période de retour élevée sont les plus sensibles au choix du seuil, on note sur la *figure 5.4* que les ajustements donnent des résultats assez stables.

Finalement, le seuil a été fixé de telle sorte qu'en moyenne deux événements par an soient retenus, quelle que soit la durée considérée. Du point de vue de la saisonnalité, les événements extraits se concentrent sur la période octobre-avril.

Courbes QdF dans le domaine des observations

Nous procédons à l'ajustement d'une loi exponentielle sur chaque échantillon de données Vd . Cette loi semble décrire correctement les échantillons jusqu'à une période de retour d'environ 10 ans (*figure 5.5a*). Le faisceau de courbes *QdF* décrit les variations de Vd en fonction de la durée d et de la période de retour T . Les débits de pointe correspondent à $d = 0$. Nous avons ensuite calé le modèle convergent, et la durée Delta a été fixée à 107 heures (*figure 5.5b*). Le *tableau 5.1* récapitule les quantiles obtenus par ajustement.

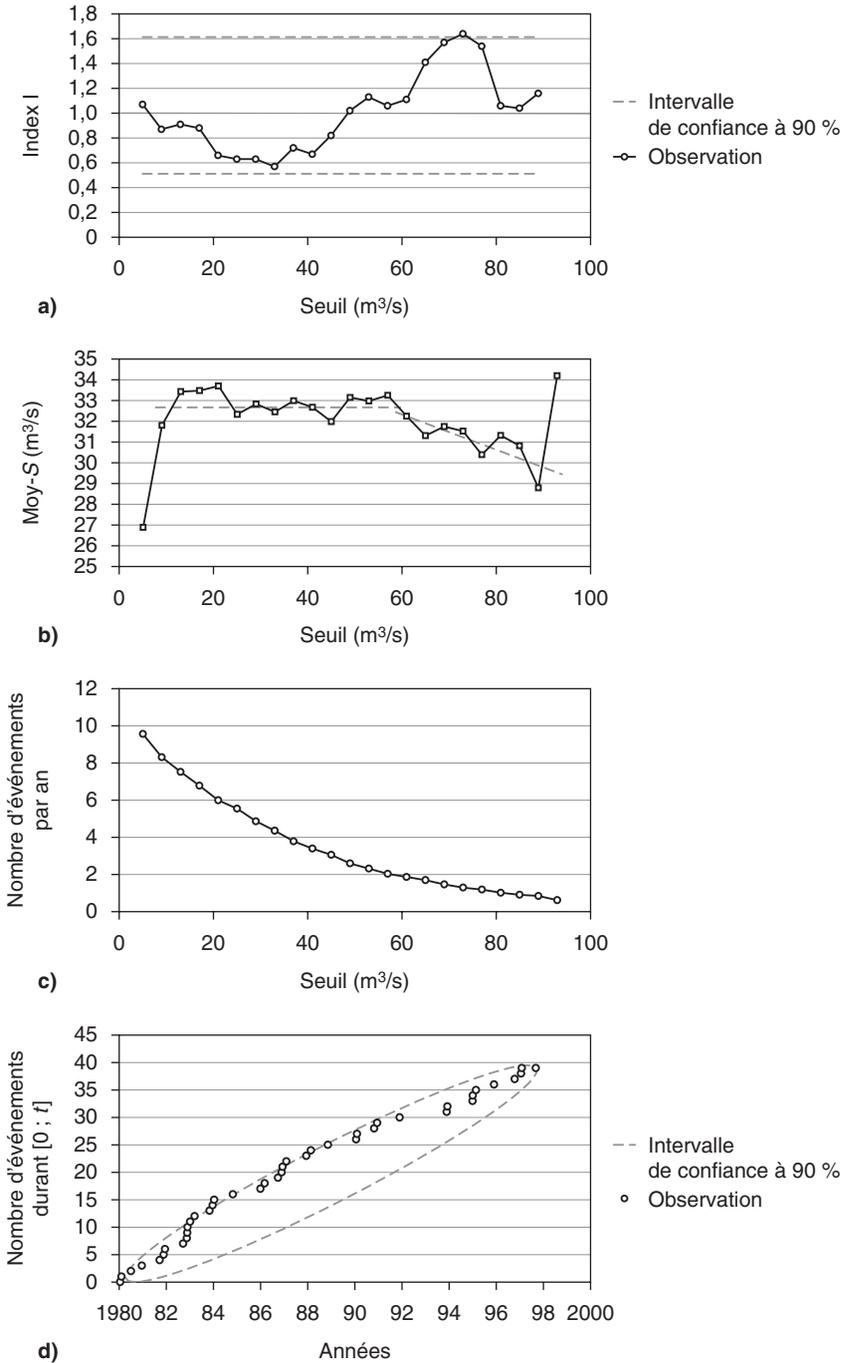


Figure 5.3. Tests utilisés pour la sélection des événements supérieurs au seuil S : (a) indice de dispersion ; (b) dépassement moyen au-dessus du seuil ; (c) nombre moyen d'événements par an ; (d) stationnarité du processus d'occurrence.

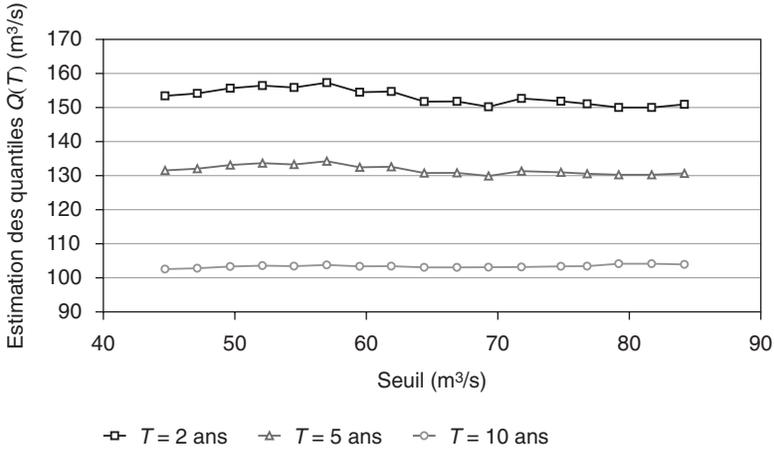


Figure 5.4. Analyse de sensibilité de l'estimation des quantiles de débit de pointe $Q(T)$ au choix du seuil S .

tement d'une loi exponentielle pour chaque échantillon Vd , et le *tableau 5.2* donne ceux issus du modèle convergent pour l'ensemble des courbes QdF .

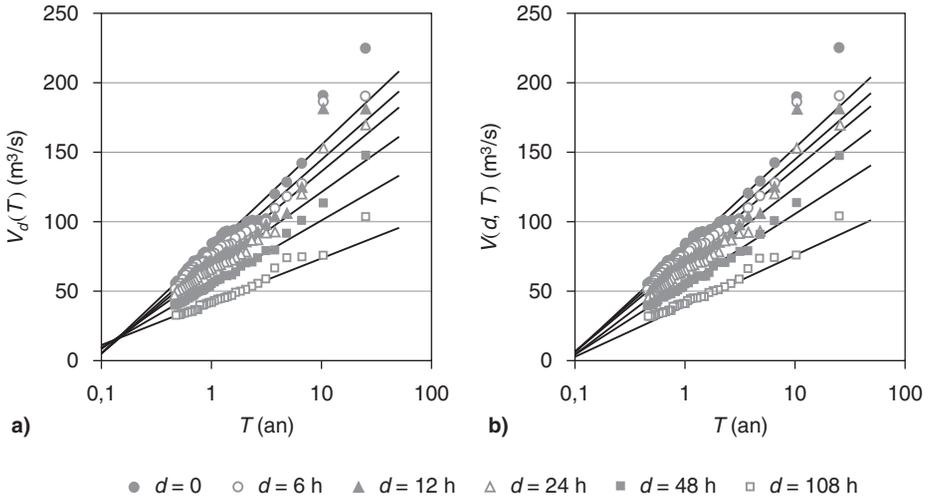


Figure 5.5. Courbes QdF – ajustement loi exponentielle, obtenues : (a) individuellement pour chaque durée ; (b) obtenues par application du modèle convergent.

Tableau 5.1. Synthèse des quantiles de débits $Vd(T)$ pour différentes durées (la première colonne correspond au débit de pointe).

T (an)	d (jour)					
	0	0,25	0,5	1	2	4,5
1	80,9	74,8	71,2	65,4	55,8	42,8
2	103,4	95,9	90,9	82,4	69,5	52,1
5	133,3	123,8	116,9	105,0	87,7	64,5
10	155,8	144,9	136,6	122,0	101,5	73,8
20	178,4	166,0	156,3	139,1	115,2	83,1

Tableau 5.2. Synthèse des quantiles de débits $V(d,T)$ pour différentes durées (la première colonne correspond au débit de pointe).

T (an)	d (jour)					
	0	0,25	0,5	1	2	4,5
1	79,9	75,6	71,8	65,2	55,1	39,7
2	101,9	96,4	91,6	83,2	70,2	50,6
5	131,0	124,0	117,7	106,9	90,3	65,1
10	153,0	144,9	137,5	124,9	105,5	76,0
20	175,0	165,7	157,3	142,9	120,7	86,9

Extrapolation des courbes QdF aux fréquences rares

L'information pluviométrique recueillie aux différents postes est agrégée à l'échelle du bassin versant pour considérer l'ensemble des pluies participant aux phénomènes, et obtenir *in fine* des pluies de bassin.

Étude des pluies extrêmes

L'extraction des intensités moyennes maximales PXd sur différentes durées d est réalisée par valeurs supérieures à un seuil, avec pour objectif en moyenne deux événements par an. Seules les portions entre octobre et avril ont été considérées pour rester cohérent avec la saisonnalité des crues. Les courbes intensité-durée-fréquence (*figure 5.6*) ont été établies sur cinq durées (de 1 à 5 jours, la durée maximale est de l'ordre de deux fois la durée caractéristique du bassin). Elles représentent les quantiles d'intensité de pluie pour différentes durées en fonction de la période de retour. La loi exponentielle a été ajustée pour chaque durée. La correction de Weiss a été appliquée pour passer de l'évaluation d'une intensité moyenne sur une durée d calendaire, notée PXd , à une intensité moyenne calculée

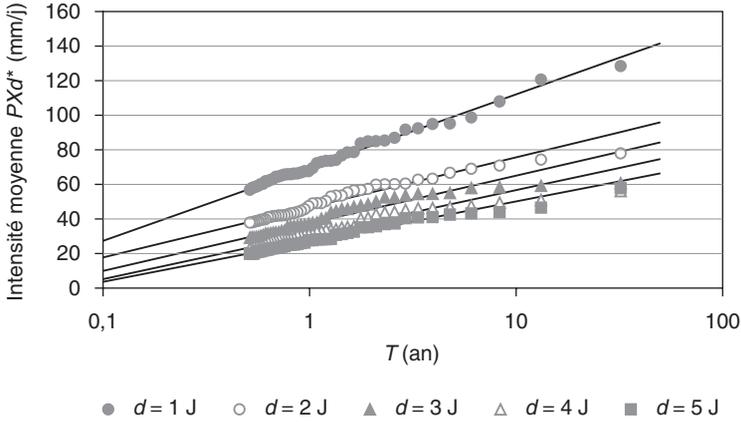


Figure 5.6. Courbes IdF – Loi exponentielle pour la pluie de bassin.

à partir des valeurs maximales centrées, notée PXd^* . Dans le cas de données échantillonnées sur une durée δ , cette correction de Weiss s'exprime par :

$$PXd^*(T) = K \cdot PXd(T), \text{ avec } K = \frac{d}{d - \frac{\delta}{8}}. \quad (5.1)$$

Un modèle de type Montana a été ajusté pour représenter l'évolution du gradex ou paramètre d'échelle $a_p(d)$ de la loi exponentielle calée avec la durée d exprimée en jour (représenté en mm/jour en *figure 5.7*) : $a_p(d) = 31,1d^{-0,35}$ (en m^3/s), avec une surface égale à 153 km^2 .

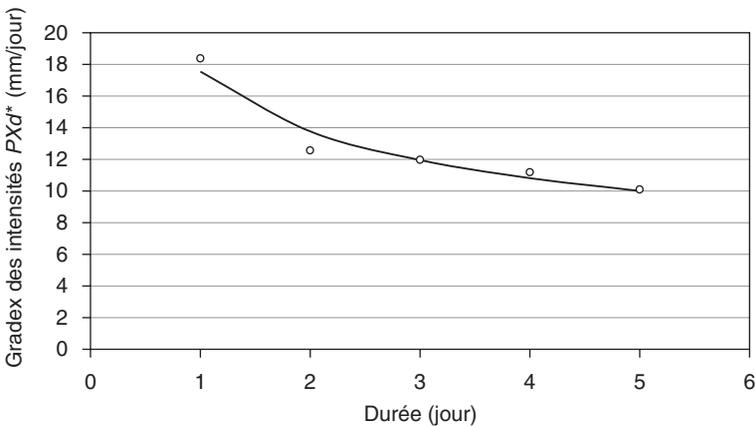


Figure 5.7. Loi de Montana ajustée sur les couples durée-gradex de la pluie de bassin.

Extrapolation aux fréquences rares selon la méthode du Gradex esthétique

La distribution des débits moyens $V(d,T)$ est extrapolée suivant le modèle d'extrapolation du Gradex esthétique, pour des périodes de retour T supérieures à T_g (en général, de l'ordre de 10 ans) :

$$\text{pour } d \in \left[\frac{D}{2} ; 5D \right] : V(d, T) = V(d, T_g) + C(d) \cdot \ln \left(1 + \frac{A(d)}{C(d)} \cdot \frac{T - T_g}{T_g} \right), \quad (5.2)$$

$$\text{pour } d \in \left[0 ; \frac{D}{2} \right] : V(d, T) = \bar{r}(d, D) \cdot V(D, T), \quad (5.3)$$

avec :

T_g : seuil exprimé en période de retour à partir duquel l'information pluviométrique prend progressivement le relais pour l'estimation des quantiles ;

$V(d, T_g)$: issu des courbes QdF ;

$A(d)$: gradex des débits (pente de la courbe QdF relative à la durée d et la période de retour T_g) ;

$C(d)$: gradex des pluies, exprimé en m^3/s ;

D : durée caractéristique des crues à définir sur la chronique ;

$\bar{r}(d, D)$: valeur moyenne du rapport de pointe de crue $r(d, D) = \frac{Vd}{VD}$.

La durée caractéristique D est fixée à $\theta = 2$ jours. Le gradex des pluies $C(d)$ est donné par la formule de Montana ajustée plus haut : $C(d) = a_p(d) = 31,1d^{-0,35}$. Le gradex des débits $A(d)$ est issu de la formulation QdF convergente calée sur les observations. Le *tableau 5.3* récapitule les valeurs obtenues pour quatre durées. On voit sur cet exemple que les valeurs de gradex des pluies sont légèrement supérieures à celles du débit.

L'hypothèse du gradex suppose implicitement un pilotage par les pluies. Dans ce cas d'application, les écoulements en crue peuvent être dus à des phénomènes de fonte couplés à des précipitations liquides (par exemple, la crue de décembre 1947 sur ce bassin). Nous retenons ici une extrapolation de type gradex esthétique avec un seuil établi à 10 ans. Nous nous plaçons délibérément dans une démarche sécuritaire.

Tableau 5.3. Valeurs du gradex des débits et des pluies (en m^3/s).

Durée d (en heures)	Gradex des débits $A(d)$	Gradex des pluies $C(d)$
0	31,8	–
$\frac{D}{2} = 24$	26,0	31,1
$D = 48$	22,0	24,4
$2,5D = 108$	15,8	18,4

Le calcul du rapport de passage du débit moyen maximal Vd , pour $d = D$, au débit de pointe ($d = 0$), pour $T > T_g$, s'appuie sur l'étude des coefficients $\frac{QIX}{Vd}$. Pour cela, on exa-

mine le ratio $\frac{QIX}{Vd}$ pour chaque rang (i -ème plus forte valeur de QIX divisée par la i -ème plus forte valeur de Vd des échantillons). Nous adoptons la valeur médiane des rapports pour les valeurs dépassant le quantile biennal associé à chaque durée, ce qui conduit à conserver les neuf premières valeurs des échantillons de chaque durée (tableau 5.4).

Tableau 5.4. Estimation du coefficient multiplicatif r du gradex.

Durée d (en heures)	Min	Médiane	Max	Moyenne
0	1,30	1,42	1,68	1,45

Le calcul des statistiques sur les rapports de pointe permet d'apprécier les incertitudes liées à l'hypothèse d'invariance de forme des crues. Cette variabilité de rapport est peut-être due à des crues d'origines différentes (une saisonalisation peut être utile), ou au fait que l'incertitude d'échantillonnage reste forte.

Le modèle d'extrapolation couplé à la formulation convergente fournit les quantiles suivants. L'extrapolation a été réalisée au-delà de $T = 10$ ans, à partir de l'équation 5.2 pour les durées $d \geq D = 12$ heures, sinon à partir de l'équation 5.3.

On voit sur la figure 5.8 que l'extrapolation de la distribution des crues avec la méthode du Gradex n'induit, sur cet exemple, qu'une faible inflexion vers le haut pour les fortes crues. Comme indiqué précédemment, cela est dû à la faible différence entre les gradex des pluies et de débits. Le débit de pointe de la crue centennale est estimé ici à 225 m³/s.

On obtient généralement des différences significatives sur des bassins à crues rapides. On se reportera à l'exemple du bassin de l'Ardèche à Saint-Martin-d'Ardèche (cf. figure 3.19), où l'estimation de la crue centennale passe de 5 150 à 6 500 m³/s lorsque l'on extrapole la distribution des crues avec une loi GEV ou à l'aide du gradex des pluies.

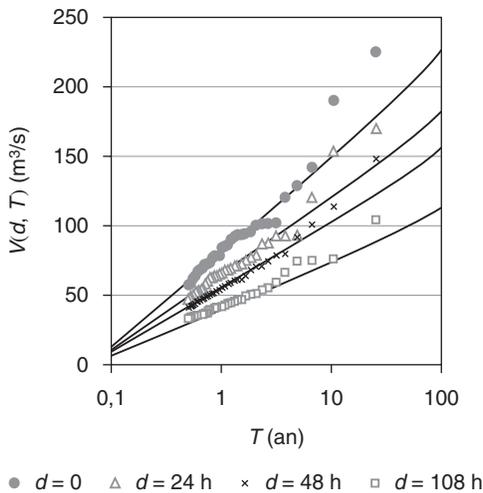


Figure 5.8. Courbes QdF extrapolées par la méthode du Gradex esthétique.

Hydrogramme de projet centennial (HSMF)

Nous reprenons l'échantillon des hydrogrammes sélectionnés par maximum annuel, et nous exploitons la montée de l'hydrogramme moyen tracé en trait épais noir sur la *figure 5.1*. En multipliant cette portion d'hydrogramme adimensionnel par le quantile de période de retour T , nous obtenons la phase de montée de l'hydrogramme de projet relatif à cette période de retour, la décrue étant obtenue sous contrainte de débit moyen maximal sur la durée d , qui est égal au quantile $V(d, T)$.

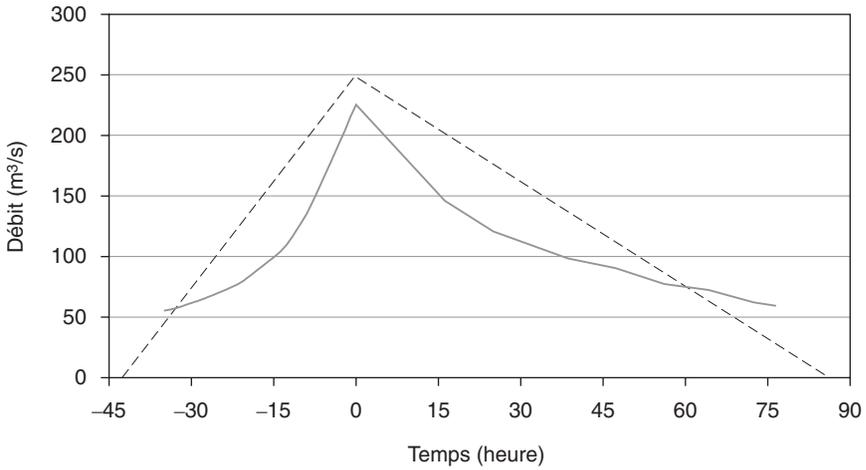


Figure 5.9. Hydrogrammes de la crue centennale issus de l'étude complète en trait continu, et de l'étude sommaire en trait pointillé.

Comparaison avec un hydrogramme de projet centennial issu d'une méthode sommaire

Nous allons comparer l'hydrogramme précédent avec celui obtenu par une approche sommaire qui consiste à retenir un hydrogramme triangulaire, avec un temps de montée égal à la durée caractéristique de crue, une descente de crue deux fois moins rapide, et un débit de pointe centennial estimé directement à partir d'un ajustement fréquentiel sur les débits.

Un échantillonnage par valeurs maximales annuelles a été réalisé en ne conservant que les années complètes à 90 %. L'année considérée débute le 1^{er} septembre et se termine le 31 août. L'ajustement d'une loi de Gumbel sur les débits de pointe (*figure 5.10*) conduit aux estimations suivantes : $Q(10) = 165 \text{ m}^3/\text{s}$ et $Q(100) = 250 \text{ m}^3/\text{s}$, avec une forte extrapolation de l'ajustement pour la crue centennale, réalisée ici uniquement à titre de comparaison (dans la pratique, à proscrire). Ces valeurs diffèrent assez peu de celles obtenues précédemment : $Q(10) = 153 \text{ m}^3/\text{s}$ (cf. courbes QdF , *tableau 5.2*) et $Q(100) = 225 \text{ m}^3/\text{s}$ (extrapolation par le Gradex esthétique, *figure 5.8*), avec un écart relatif de 8 et 11 %. Cela n'est généralement pas le cas, l'extrapolation avec le gradex des pluies conduisant le plus souvent à des valeurs de crue centennale plus élevées. Il est vraisemblable ici qu'un

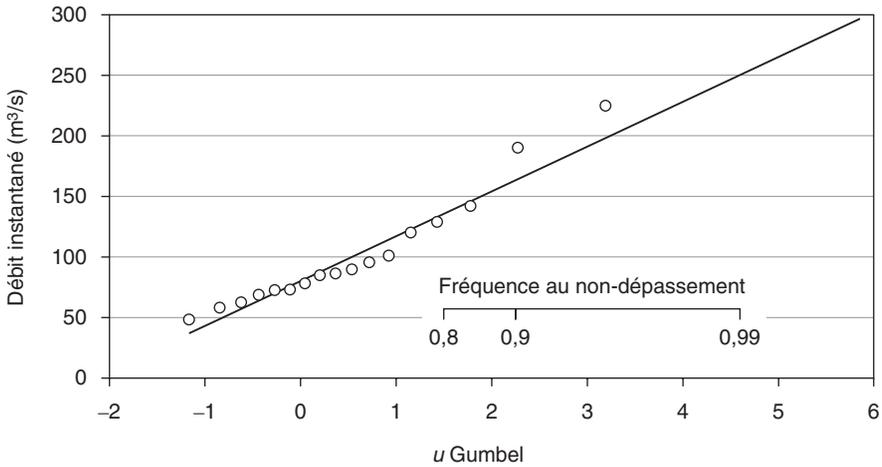


Figure 5.10. Ajustement d'une loi de Gumbel sur les maxima annuels de la Moselle à Rupt-sur-Moselle.

laminage des crues dû au lit majeur de la Moselle ait réduit les valeurs que l'on obtiendrait directement avec un ruissellement standard.

La *figure 5.11* présente, pour tous les hydrogrammes de crue issus de l'échantillonnage des maxima annuels, la durée de dépassement de la moitié de la pointe de crue en fonction du débit de pointe. On retient ici la durée médiane $D_{\text{Socose}} = 43$ heures, en conservant les cinq plus fortes crues.

On voit sur la *figure 5.9* que l'hydrogramme obtenu par cette approche sommaire surestime le volume de la crue si l'objectif est de rester homogène en fréquence. Le volume généré sur 108 heures équivaut à un débit moyen de $145 \text{ m}^3/\text{s}$. Dans le référentiel issu du

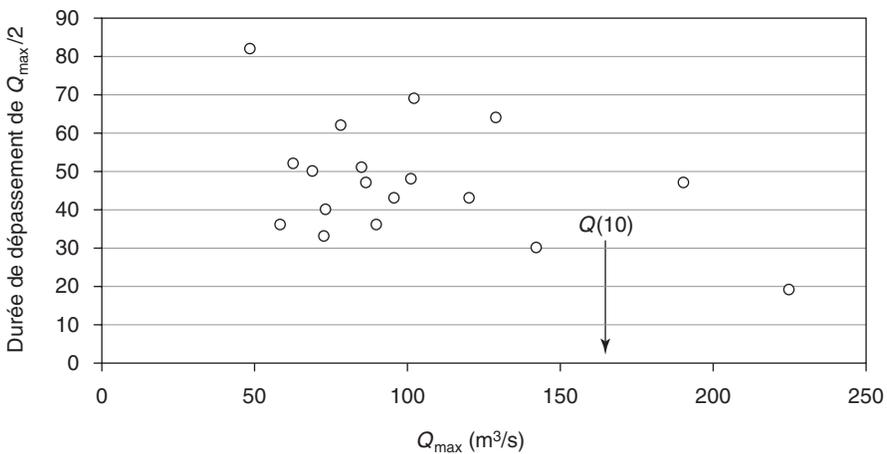


Figure 5.11. Détermination du temps caractéristique de crue.

modèle convergent et de l'extrapolation par le modèle du Gradex, cela équivaut à un volume de période de retour 200 ans. Par ailleurs, la forme triangulaire de la crue reste assez éloignée de celle des crues observées (cf. figure 5.1).

Construction d'un hydrogramme de projet relatif à la crue centennale. Cas d'un bassin versant à crues lentes du Nord de la France

Le bassin étudié est situé dans le département du Pas-de-Calais, entre Calais et Lille, juste en amont de Saint-Omer. La chronique examinée est la série des débits au pas de temps variable de l'Aa à Wizernes (code Hydro E4035710). Elle débute en janvier 1970 pour s'achever en décembre 2002. Le bassin versant défini par la station de jaugeage draine 392 km² (figure 5.12).

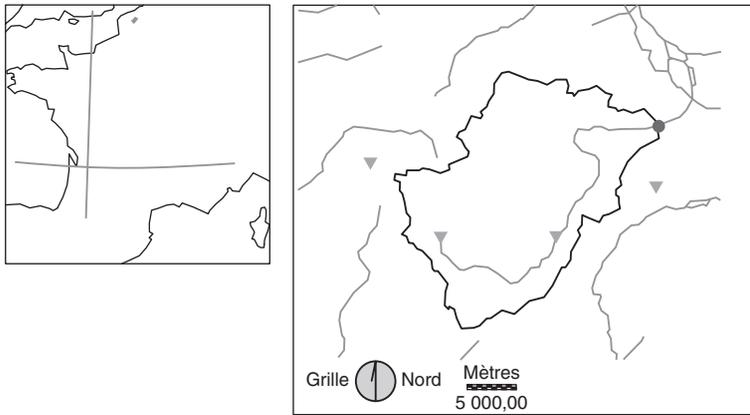


Figure 5.12. Bassin versant de l'Aa à Wizernes (Pas-de-Calais). La station hydrométrique est représentée par le point grisé, les triangles identifiant les stations pluviométriques exploitées *infra*.

L'objectif est de déterminer la pointe de crue de période de retour 100 ans et l'hydrogramme de projet associé. Nous examinerons successivement :

- 1/ les données hydrométriques à traiter ;
- 2/ la distribution des crues fréquentes à moyennes ;
- 3/ l'extrapolation de la distribution des crues pour les événements rares ;
- 4/ la construction d'hydrogrammes de référence.

Les données hydrométriques

La station de mesure est implantée juste en amont d'un seuil déversant (ancien moulin) qui contribue à la bonne stabilité dans le temps de la relation hauteur-débit. La figure 5.13 présente la courbe de tarage de la station et les jaugeages réalisés depuis 1970. On peut constater d'une part la faible dispersion des jaugeages autour de la courbe de tarage, et d'autre part l'existence de mesures de débits dans la partie haute de la courbe (hauteur

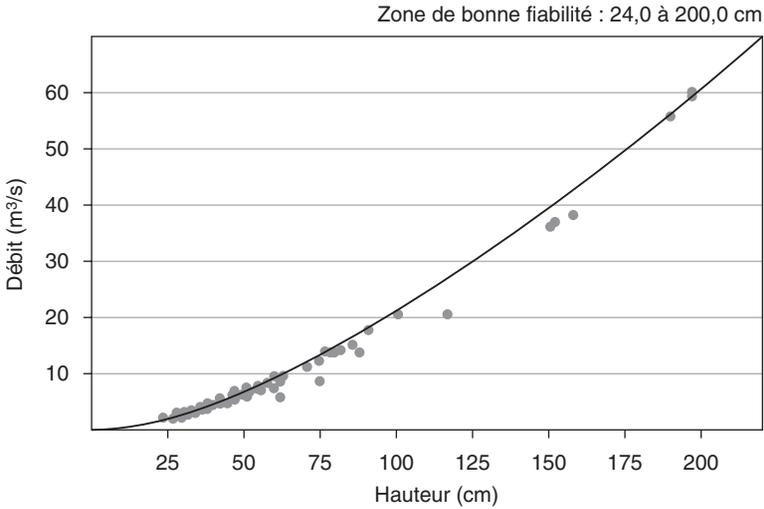


Figure 5.13. Courbe de tarage de l'Aa à Wizernes (source : Diren NPC).

maximale jaugée : 1,90 m ; hauteur de la crue maximale en 2002 : 1,93 m), bien au-delà de la crue décennale (environ 40 m³/s). Cette station, du fait de sa configuration hydraulique et du bon suivi hydrométrique, peut être considérée comme de très bonne qualité, avec une faible extrapolation de la courbe de tarage en période de crue.

Il n'y a pas d'erreur notable et détectable à la seule visualisation de la chronique (figure 5.14). Les événements les plus forts se concentrent en fin de série, d'où une non-

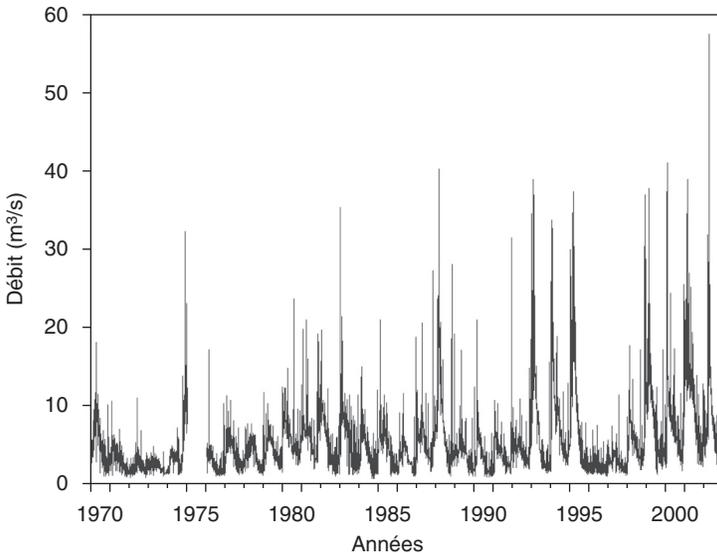


Figure 5.14. Chronique des débits de l'Aa à Wizernes (1970-2002).

stationnarité détectée par le test basé sur le coefficient de corrélation ρ de Spearman appliqués aux maxima annuels. Ce test permet de vérifier le caractère aléatoire d'une série, donc d'identifier une tendance. On calcule la série r_i des rangs attribués aux valeurs de l'échantillon classées par ordre croissant. ρ est le coefficient de corrélation entre i et r_i . Pour $n > 30$, ρ suit approximativement la loi normale de moyenne nulle et de variance $\frac{1}{n-1}$.

Une tendance est détectée si ρ est significativement différent de 0, c.-à-d. :

$$|\rho| \sqrt{n-1} > u_{1-\alpha/2}, \quad (5.4)$$

avec $u_{1-\alpha/2}$ quantile de la loi normale centrée réduite de fréquence au non-dépassement $1 - \frac{\alpha}{2}$. Ainsi, sur l'Aa, nous obtenons $\rho = 0,49$, et l'hypothèse de stationnarité est rejetée au seuil de $\alpha = 1\%$ ($\frac{u_{1-\alpha/2}}{\sqrt{n-1}} = 0,46$), avec $n = 32$. Sur la *figure 5.15*, on constate effective-

ment que les dernières crues sont classées de manière quasi systématique parmi les plus fortes (rangement par ordre croissant).

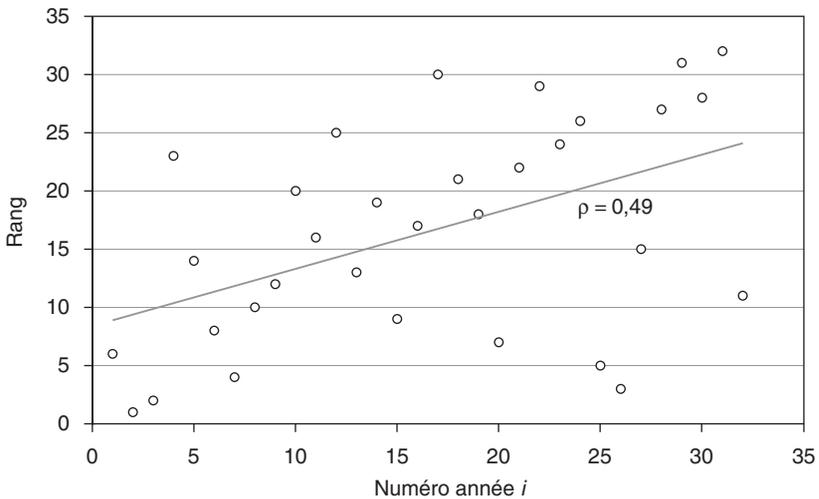


Figure 5.15. Éléments de calcul du coefficient de corrélation sur les rangs pour la série des maxima annuels.

À une échelle plus fine, on constate une forte variabilité interannuelle :

- certaines années se caractérisent par l'absence de crue significative, et même de réponse claire aux impulsions. C'est le cas de l'hiver 1996-1997 (*figure 5.16*), où le bassin ne réagit que faiblement aux précipitations pourtant fréquentes, mais peu intenses : la chronique ressemble plus à un bruit blanc qui vient s'ajouter à une constante (débit de base) ;
- pour d'autres années (hiver 1992-1993 et suivant), on pourrait assimiler le fonctionnement du bassin versant à un réservoir qui se vidange du début de l'année à l'automne

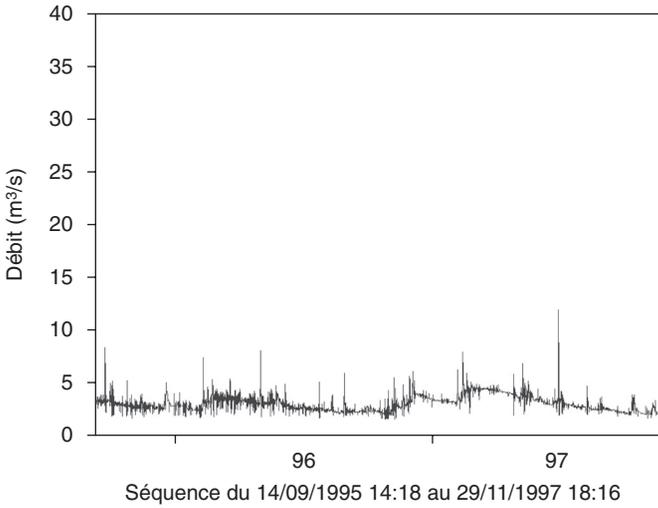


Figure 5.16. Extrait de la chronique des débits de l'Aa à Wizernes (septembre 1995-novembre 1997).

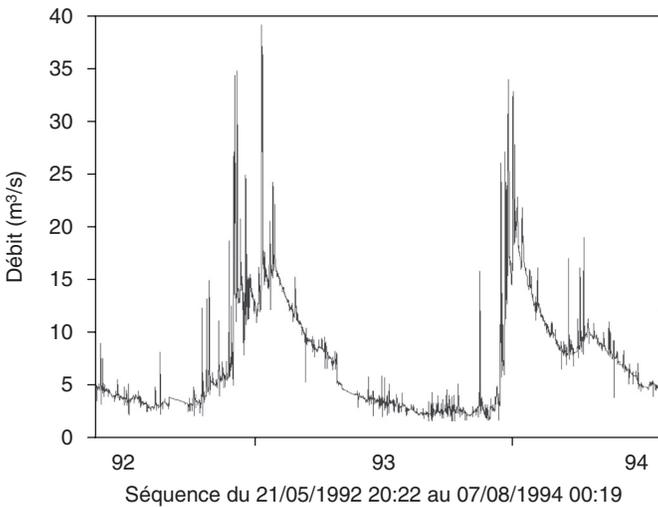


Figure 5.17. Extrait de la chronique des débits de l'Aa à Wizernes (mai 1992-août 1994).

(figure 5.17). Le cours d'eau est relativement insensible aux précipitations qui peuvent se produire dans cette phase de récession.

On note par ailleurs sur ces deux figures (5.16 et 5.17) de fortes fluctuations du débit instantané. Renseignements pris auprès de la Diren Nord-Pas-de-Calais, celles-ci proviennent de lâchures et de manœuvres de vannes qui ne remettent pas en cause les pointes de crue principales.

La non-stationnarité des crues mise en évidence sur la *figure 5.15* est liée au comportement du bassin versant. À un signal de base gouverné par le souterrain s'ajoute la réponse atténuée à la pluie. Le bassin versant se comporte comme une « éponge » qui libère son eau après saturation des réservoirs. On pourrait même considérer qu'il n'y a qu'une seule crue au cours de l'année hydrologique comme pour un bassin de fonte nivale pour lequel les événements majeurs correspondent à la fonte combinée aux pluies. Ici, les événements rares résulteraient d'une vidange couplée à un ruissellement direct engendré par des précipitations intenses (comme l'épisode majeur de mars 2002). Ces processus ont principalement lieu entre l'automne et le printemps. Cette période de l'année sera privilégiée dans notre analyse.

Ces considérations déduites d'une simple visualisation de la série sont renforcées par le descriptif du bassin disponible sur le site de l'agence de l'eau Artois-Picardie :

L'Aa est un cours d'eau qui prend sa source à Bourthes à l'altitude de 121 mètres (dans l'arrondissement de Montreuil/Mer – Pas-de-Calais) et se jette dans la mer du Nord à Gravelines (Nord). Le bassin versant de l'Aa rivière est limité au Nord par le marais de Saint-Omer, sa superficie est de 398 km². La vallée est étroite et tortueuse, à versants très inclinés ; ses affluents sont à forte pente. La déclivité de l'Aa dans cette section dépasse 2 mètres par kilomètre, ce qui correspond à un régime rapide pour la région.

La source, située au lieu-dit « Le Crocq » à Bourthes, est temporaire et ne coule que lorsque la nappe a été bien alimentée par des pluies d'hiver. Généralement, au printemps, les sources se trouvent déplacées vers l'aval pour se situer au pied du « Mont Timet ».

Le débit de cette rivière est en moyenne de l'ordre de 5,6 m³/s, et en étiage sévère de l'ordre de 1,8 m³/s.

Source :

http://www.eau-artois-picardie.fr/DPE/DPE_MILIEU/HTML/L'Audomarois.html.

Remarque : sur cette même page Web, il est fait mention d'une zone d'étang, située en aval du bassin versant étudié (marais Audomarois), qui n'interfère pas sur les écoulements à Wizernes.

Distribution des crues fréquentes à moyennes

Choix d'un mode d'échantillonnage et d'une distribution pour les débits de pointe

Dans les conditions énoncées *supra*, il paraît délicat d'engager un échantillonnage par valeurs supérieures à un seuil. On peut s'en convaincre sur quelques essais (impossibilité de vérifier les hypothèses du modèle de renouvellement utilisé habituellement, avec un processus de Poisson), sauf à retenir une condition très stricte d'espacement temporel entre événements, ce qui revient en fait à saisonnaliser l'échantillonnage.

Nous avons donc procédé à un échantillonnage par sélection de la valeur maximale annuelle sur l'automne-hiver-printemps, en cohérence avec le fonctionnement du bassin et dans le souci de ne pas mélanger les processus de genèse des crues. En effet, il est préfé-

nable pour la suite de l'étude (cf. § *Extrapolation de la distribution des crues pour les événements rares*) de raisonner sur la même saison de référence entre les pluies et les débits. La période considérée va du mois de septembre au mois d'avril suivant. Trente-deux valeurs de débits instantanés maximaux QIX ont été retenues, et un ajustement à la loi de Gumbel a été réalisé (figure 5.18). Il est tout juste acceptable, avec un tassement apparent des valeurs observées autour de $40 \text{ m}^3/\text{s}$. L'interprétation reste délicate : est-ce attribuable au pouvoir tampon de la nappe ou uniquement au hasard de l'échantillonnage ? Nous avons testé l'utilisation de la loi GEV (3 paramètres) qui n'apporte pas d'amélioration notable sur la qualité de l'ajustement. Par la suite, nous retiendrons un échantillonnage par valeurs maximales saisonnières (septembre-avril) et une loi de Gumbel.

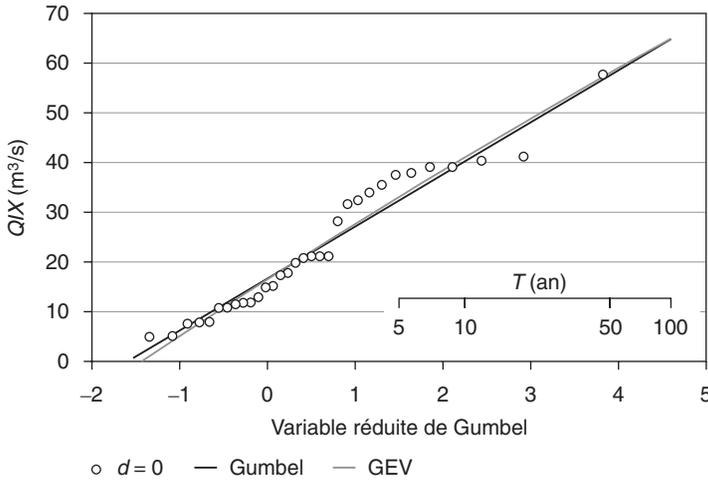


Figure 5.18. Distribution du débit de pointe de l'Aa à Wizernes (1970-2002).

Estimation de la durée caractéristique de crue du bassin versant

Afin d'évaluer le temps de réaction en crue du bassin à la pluviométrie, nous avons retenu les vingt plus forts épisodes parmi les trente-deux disponibles, avec l'échantillonnage maximal saisonnier. Compte tenu du comportement du bassin versant, certaines années ne présentent pas de crue marquée ; nous avons donc choisi de restreindre l'analyse aux événements les plus intenses et les plus manifestes. Le débit de pointe de ces événements a dépassé le seuil de $15 \text{ m}^3/\text{s}$. On peut noter sur la figure 5.19 un fort étalement des crues associé à une réaction lente du bassin, et la grande variabilité du débit de base. Ce dernier point est confirmé par la figure 5.20, sur laquelle on constate que le débit de base varie fortement entre 4 et $25 \text{ m}^3/\text{s}$ dans la gamme de débit de pointe de 30 à $40 \text{ m}^3/\text{s}$.

Pour filtrer la contribution des eaux souterraines et isoler la réponse du bassin à l'impulsion pluviométrique, nous examinerons les hydrogrammes de crue $Q'(t) = Q(t) - Q_{\text{base}}$ définis par le débit privé de la composante du débit de base Q_{base} (c.-à-d. au minimum sur la fenêtre temporelle définissant l'épisode). On peut comparer sur la figure 5.21 les durées $d_{1/2}$ de dépassement de la moitié du débit de pointe, calculées à partir des hydrogrammes

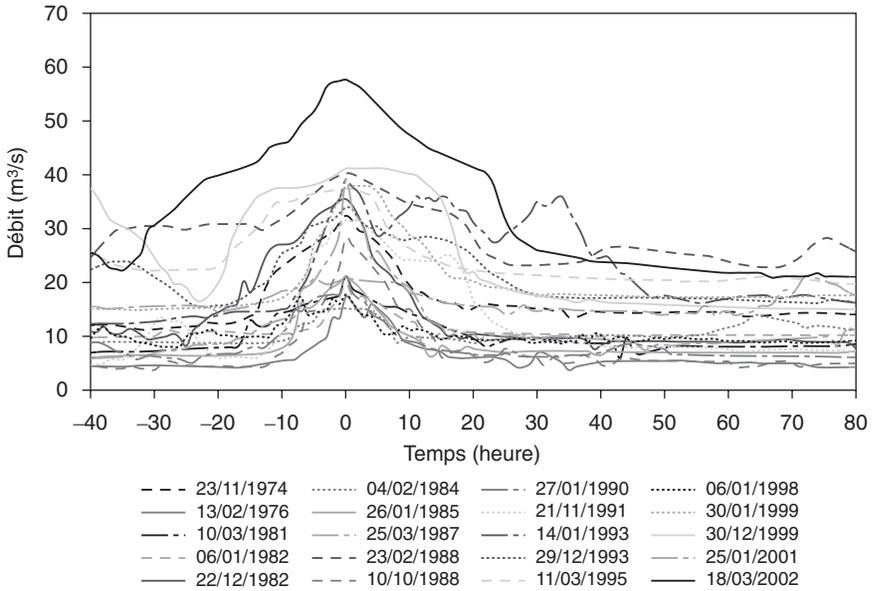


Figure 5.19. Collection d'hydrogrammes de crue de l'Aa à Wizernes, synchronisés sur le maximum.

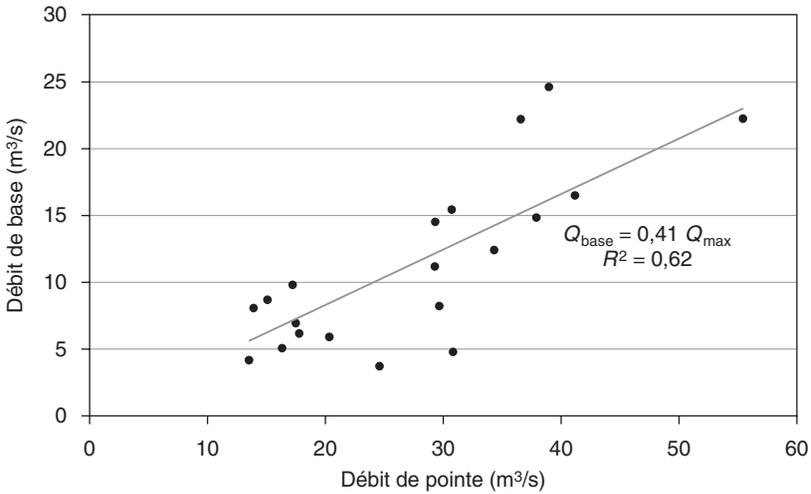


Figure 5.20. Relation débit de base-débit de pointe de crue (l'Aa à Wizernes).

$Q(t)$ ou $Q'(t)$. On constate une réduction de la dispersion des durées lorsque l'on raisonne uniquement sur les dépassements au-dessus du débit de base. Dans ce dernier cas, la durée de Socose (valeur médiane des durées $d_{1/2}$ pour un débit de pointe décennal) est à considérer dans l'intervalle [30 h ; 40 h].

Une autre approche consiste à raisonner sur la forme moyenne des hydrogrammes de crue, en ayant au préalable effectué une normalisation par le débit de pointe. Dans le cas

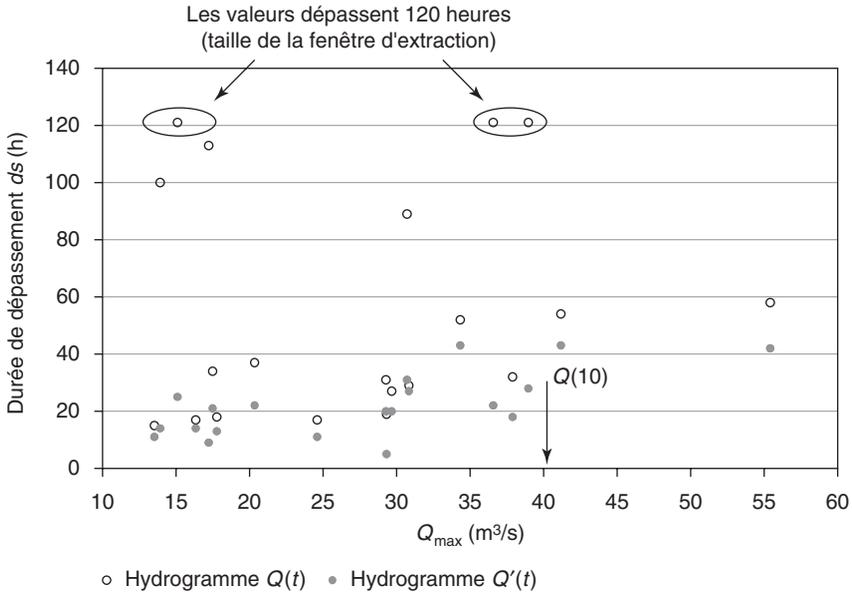


Figure 5.21. Étude des durées de dépassement de crue (l'Aa à Wizernes).

des hydrogrammes $Q'(t)$ normalisés, on peut identifier un temps de réponse du bassin D de l'ordre de 36 heures (1,5 jour) (figure 5.22) pour un événement décennal.

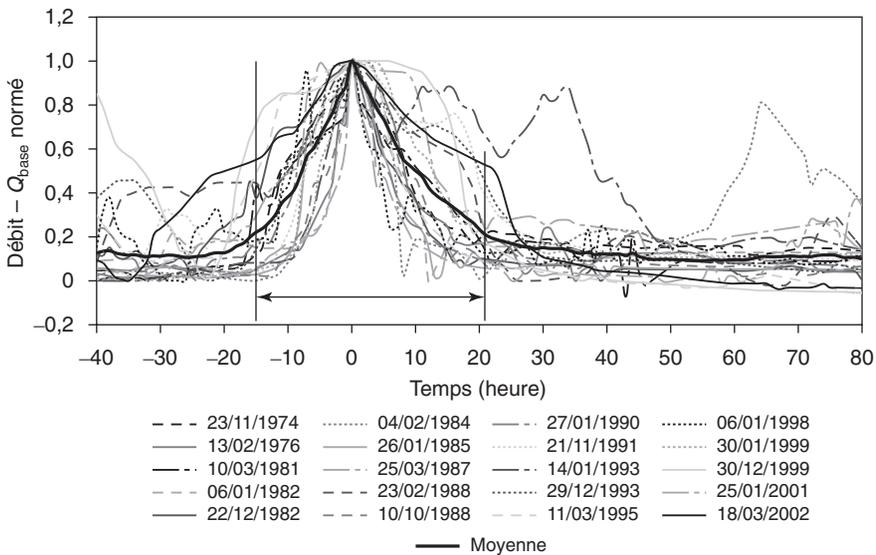


Figure 5.22. Hydrogramme moyen de crue au-dessus du débit de base (l'Aa à Wizernes).

Établissement des courbes QdF

Nous avons étendu l'analyse précédemment réalisée sur les débits de pointe à la variable $VCXd$, débit moyen sur la durée continue d , en retenant la valeur maximale saisonnière. Quatre durées ont été examinées : $d = 1, 1,5, 2,5$ et 4 jours. Nous avons exclu les durées supérieures à quatre jours, car le pas de temps n'est plus réellement compatible avec la dynamique de crue ($D = 1,5$ jour). La loi de Gumbel a été calée sur les séries de maxima annuels. Les ajustements des courbes QdF restent de qualité moyenne, de façon analogue à ce qui a été constaté pour la distribution du débit de pointe (figure 5.23). Le tableau 5.5 donne la valeur des quantiles de crue issus des cinq ajustements (la durée $d = 0$ correspond au débit de pointe).

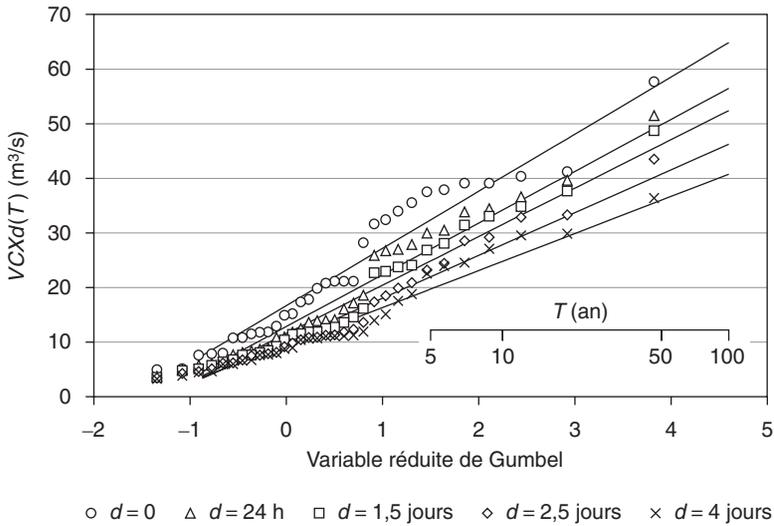


Figure 5.23. Courbes QdF (l'Aa à Wizernes).

Tableau 5.5. Quantiles de débits de crue $VCXd(T)$ (en m^3/s) issus des courbes QdF (l'Aa à Wizernes).

d (jour)	T (an)					
	1,1	2	5	10	20	50
0	7,5	20,5	32,4	40,2	47,8	57,5
1	4,5	16,3	27,0	34,2	41,0	49,8
1,5	3,8	14,8	24,9	31,5	37,9	46,2
2,5	3,3	13,1	21,9	27,8	33,5	40,8
4	3,6	12,0	19,7	24,8	29,7	36,0

Consolidation des courbes QdF

Le modèle QdF convergent permet de synthétiser dans une formulation analytique unique les quantiles de crue relatifs à différentes durées :

$$\frac{VCX(d, T)}{VCX(0, T)} = \frac{1}{1 + d/\Delta}, \quad (5.5)$$

avec :

$VCX(0, T)$: distribution théorique du débit de pointe, obtenue par consolidation des courbes QdF ;

Δ : paramètre à caler, homogène à un temps et relié à la dynamique de crue.

Le calage des paramètres à partir des cinq échantillons de débit moyen maximal $VCXd$ ($d = 0, 1, 1,5, 2,5$ et 4 jours) conduit à l'expression suivante :

$$VCX(d, T) = \frac{15,3 + 10,95 u(T)}{1 + d/6}, \quad \text{avec } u(T) = -\ln\left[-\ln\left(1 - \frac{1}{T}\right)\right]. \quad (5.6)$$

Le *tableau 5.6* donne la valeur des quantiles de crue obtenus après consolidation par le modèle QdF convergent des cinq échantillons. On peut noter sur la *figure 5.24* que la valeur $\Delta = 6$ jours obtenue pour le modèle QdF convergent est compatible avec la régression présentée au chapitre 2, § Modèle QdF régional convergent :

$$\Delta = 3,44 D_{\text{Socose}}, \quad \text{avec } r^2 = 0,85, \quad \text{et l'intervalle de confiance à 80 \% : } \Delta \pm 2 \text{ jours}, \quad (5.7)$$

soit :

$$\Delta(D_{\text{Socose}}) = 5,16 \text{ jours}, \quad \text{et l'intervalle de confiance à 80 \% : } [3,16 \text{ jours} ; 7,16 \text{ jours}]. \quad (5.8)$$

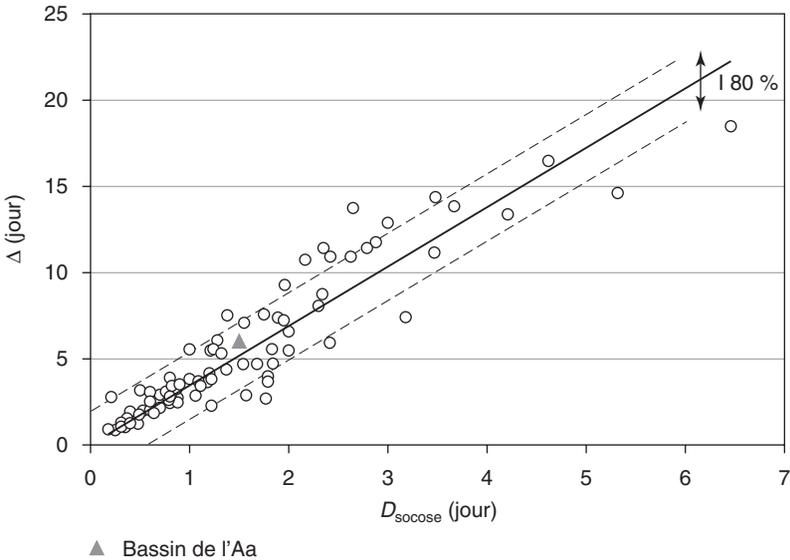


Figure 5.24. Position du paramètre Δ estimé par le modèle QdF convergent dans la régression $\Delta = f(D_{\text{Socose}})$ (d'après Sauquet *et al.*, 2003).

Tableau 5.6. Quantiles de débits de crue $VCXd(T)$ (en m^3/s) issus du modèle QdF convergent (l'Aa à Wizernes).

d (jour)	T (an)					
	1,1	2	5	10	20	50
0	5,7	19,3	31,7	39,9	47,8	58,0
1	4,9	16,6	27,2	34,2	41,0	49,7
1,5	4,6	15,5	25,4	31,9	38,3	46,4
2,5	4,0	13,6	22,4	28,2	33,8	41,0
4	3,4	11,6	19,0	24,0	28,7	34,8

Extrapolation de la distribution des crues pour les événements rares

Établissement des courbes intensité-durée-fréquence (IdF)

Quatre chroniques de pluie au pas de temps journalier sont disponibles sur la période 1961-2002 (Fauquembergues, Herbelles, Bourthes, Desvres) à proximité du bassin versant. La région n'ayant pas un relief accidenté, nous avons construit une chronique de pluie de bassin, par simple moyenne arithmétique, sur la période d'enregistrement commune aux quatre stations, et, en cohérence avec l'analyse des débits, nous avons extrait les maxima annuels des pluies moyennes sur plusieurs durées. L'échantillonnage s'appuie sur la période septembre-avril de chaque année pour rester en cohérence avec l'extraction des débits de crue.

Les courbes IdF (figure 5.25) ont été établies sur six durées (de 1 à 6 jours), et un modèle de type Montana a été ajusté pour représenter l'évolution du gradex avec la durée (figure 5.26). La correction de Weiss (cf. équation 5.1) a été appliquée pour passer de l'évaluation du quantile d'intensité moyenne sur une durée d calendaire, noté $PXd(T)$, à un quantile calculé à partir des valeurs maximales centrées, noté $PXd^*(T)$.

Extrapolation par la méthode du Gradex esthétique

Pour prédéterminer les quantiles de fréquence rare, on introduit une information pluviométrique qui prend progressivement le relais pour l'estimation des quantiles à partir de T_g , seuil fréquentiel d'extrapolation. Le modèle d'extrapolation des distributions choisi est une variante de la méthode du Gradex (formulation du Gradex esthétique, cf. chapitre 3) :

$$VCXd(T) = VCXd(T_g) + a_p(d) \ln \left(1 + \frac{a_p(T_g, d)}{a_p(d)} \cdot \frac{T - T_g}{T_g} \right), \text{ pour } \frac{D}{2} \leq d \leq 5D, \quad (5.9)$$

$$QIX(T) = r \cdot VCXd(T), \text{ pour } d = D, \quad (5.10)$$

avec :

$VCXd(T_g)$: quantile de débit moyen sur la durée d relatif à la période de retour T_g , seuil d'extrapolation ;

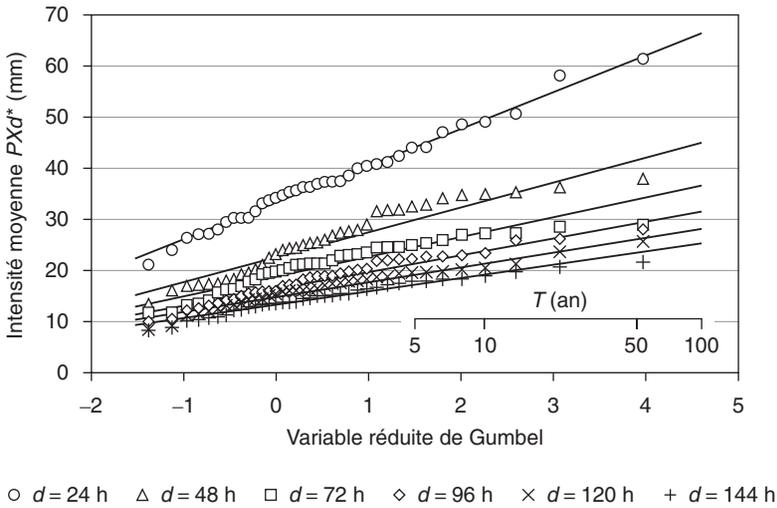


Figure 5.25. Courbes IdF de la pluie moyenne de bassin (l'Aa à Wizernes).

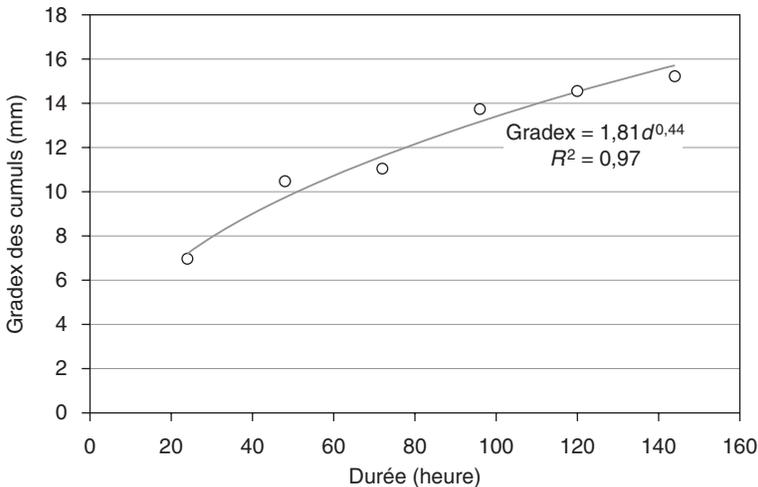


Figure 5.26. Loi de Montana sur le gradex de la pluie de bassin (l'Aa à Wizernes).

$a_q(T_g, d)$: gradex des débits relatif à la durée d et la période de retour T_g ;
 $a_p(d)$: gradex des pluies maximales pour les intensités relatives à la durée d .

Ce modèle d'extrapolation est applicable aux quantiles de débits moyens sur des durées proches de la durée caractéristique du bassin (cf. équation 5.9). L'extrapolation pour les quantiles instantanés est ensuite réalisée selon un rapport d'homothétie entre la pointe et le volume sur la durée caractéristique (cf. équation 5.10).

On calcule dans un premier temps le gradex des pluies pour la durée $d = 36$ heures à l'aide de la formule de puissance calée sur les données journalières. La valeur du gradex $a_p(d)$ (en m^3/s) est celle issue de l'expression de Montana :

$$C(d) = 197 d^{-0,56}, \quad (5.11)$$

avec d en heure (cf. expression de la figure 5.26, convertie de mm en m^3/s).

Le deuxième paramètre est le point d'inflexion T_g de la courbe d'ajustement. Au-delà de ce seuil T_g , on considère que tout nouvel accroissement de précipitation induit un accroissement équivalent en terme de lame d'eau écoulée. La valeur admise par défaut est usuellement 10 ans pour un bassin versant avec une forte aptitude au ruissellement, et 20 ans pour un bassin versant réagissant moins rapidement. Ici, compte tenu des grandes capacités d'infiltration du bassin, l'information pluviométrique prend le relais plus tardivement que dans un bassin au comportement plus « classique ». Pour nous aider dans le choix final du seuil T_g , nous avons examiné les coefficients d'écoulement, ainsi que le ratio $\frac{VCXd}{PXd}$, pour les dix valeurs de débit $VCXd$ les plus fortes dans les échantillons. On compare la pluie maximale sur la durée d (avec correction de Weiss) au débit maximal moyen sur la même durée d pour les dix épisodes majeurs. On obtient de faibles valeurs du coefficient d'écoulement (cf. tableau 5.7) pour les trois durées $d = 1, 3$ et 6 jours compatibles avec la durée de Socose. On atteint au maximum la valeur de 50 % pour certains épisodes qui suivent une période particulièrement arrosée. À titre de comparaison, nous avons tracé les chroniques de moyenne mobile des pluies centrées et débits pour les trois durées sur l'épisode le plus fort, de février-mars 2002 (cf. figure 5.27). Il demeure que nous sommes loin d'atteindre la saturation du bassin versant, même pour un événement auquel on attribue une fréquence empirique de 46 ans. Nous retiendrons ici une extrapolation de type Gradex esthétique avec un seuil établi à 50 ans.

Tableau 5.7. Coefficients d'écoulement pour la durée $d = 1, 3$ et 6 jours.

Date $VCXd$	$VCXd$ (en m^3/s)	PXd			Coefficient d'écoulement
		brut (en mm)	corrigé Weiss (en mm)	corrigé (en m^3/s)	
01/03/2002	51 / 40 / 33	44 / 26 / 21	51 / 27 / 22	230 / 124 / 98	0,22 / 0,33 / 0,34
05/01/2001	27 / 20 / 18	21 / 13 / 9	24 / 14 / 9	109 / 62 / 43	0,25 / 0,32 / 0,41
27/12/1999	40 / 33 / 25	26 / 20 / 11	30 / 20 / 11	137 / 92 / 51	0,29 / 0,36 / 0,49
16/01/1999	30 / 22 / 20	15 / 10 / 19	18 / 10 / 19	80 / 45 / 86	0,38 / 0,49 / 0,23
30/01/1995	34 / 28 / 26	26 / 18 / 16	30 / 19 / 17	136 / 84 / 75	0,25 / 0,33 / 0,34
11/01/1993	34 / 25 / 23	25 / 14 / 10	29 / 15 / 10	130 / 68 / 47	0,26 / 0,36 / 0,48
31/12/1993	30 / 27 / 22	22 / 19 / 13	25 / 20 / 13	112 / 92 / 60	0,27 / 0,30 / 0,36
20/11/1991	27 / 31 / 18	43 / 20 / 17	49 / 21 / 18	222 / 94 / 81	0,12 / 0,33 / 0,35
06/02/1988	37 / 17 / 13	27 / 20 / 12	31 / 21 / 12	141 / 97 / 54	0,26 / 0,18 / 0,24
21/12/1982	28 / 19 / 16	28 / 20 / 8	32 / 21 / 8	146 / 97 / 36	0,19 / 0,19 / 0,46

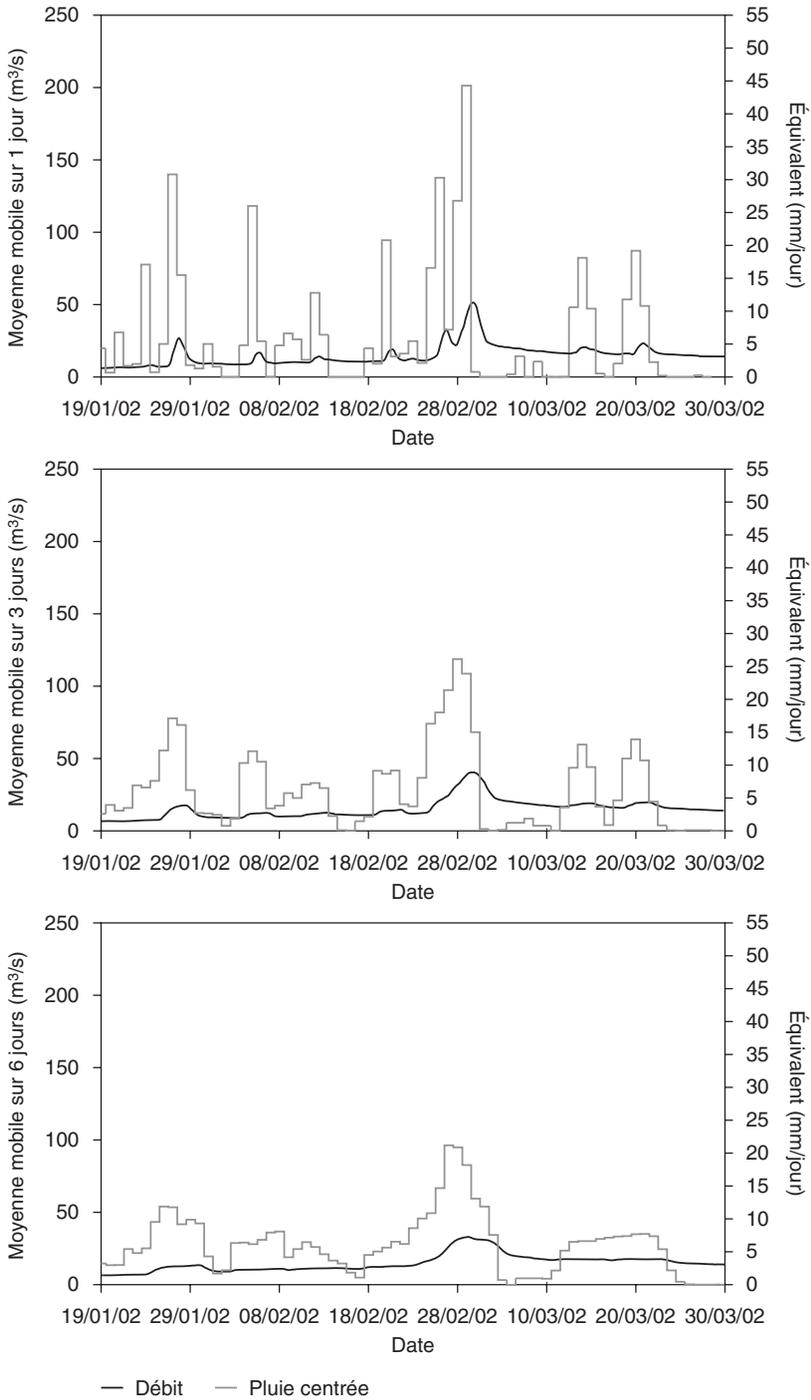


Figure 5.27. Comparaison des lames d'eau précipitées et ruisselées sur l'Aa à Wizernes (crue de février-mars 2002).

Le calcul du rapport de pointe de crue, ce dernier étant égal au ratio du débit moyen maximal $VCXd$ pour $d = 36$ heures, avec le débit de pointe QIX pour $T > 50$ ans, s'appuie sur les huit plus fortes valeurs issues de l'échantillonnage par valeurs maximales annuelles.

On calcule pour chaque rang le coefficient $\frac{QIX}{VCXd}$ qui varie entre 1,09 et 1,47 (cf. tableau 5.8). On en conserve la valeur médiane égale à 1,22 (cf. figure 5.28).

Tableau 5.8. Valeur du rapport de pointe de crue pour les huit plus forts événements (l'Aa à Wizernes).

QIX	$VCXd = 36 \text{ h}$	r
57,7	48,7	1,18
41,2	37,6	1,09
40,3	34,8	1,16
39,1	33,0	1,18
39,1	31,4	1,24
37,9	28,1	1,35
37,5	26,8	1,40
35,5	24,1	1,47

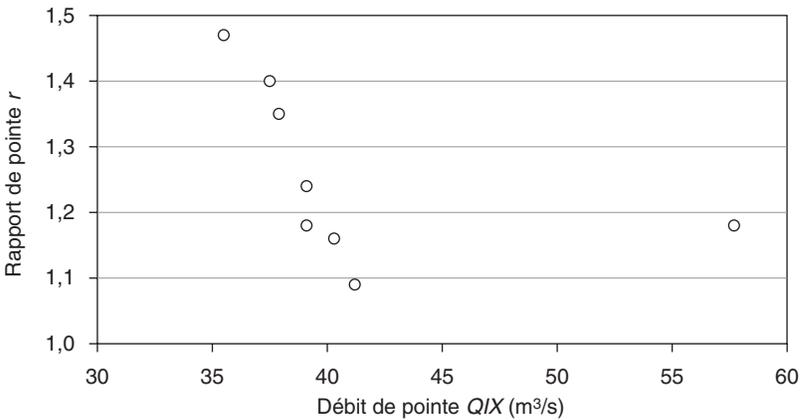


Figure 5.28. Détermination du rapport de pointe de crue pour l'extrapolation aux grandes périodes de retour (l'Aa à Wizernes).

Le modèle d'extrapolation couplé à la formulation convergente fournit les quantiles suivants (cf. tableau 5.9 et figure 5.29).

Tableau 5.9. Quantiles de débits de crue $VCX(d,T)$ (en m^3/s) issus du modèle QdF convergent et de la formulation du Gradex esthétique (l'Aa à Wizernes).

d (jour)	Modèle QdF convergent						Gradex esthétique		
	$T = 1,1$ an	2	5	10	20	50	100	500	1 000
0	5,7	19,3	31,7	39,9	47,8	57,2	64,9	99,6	118,8
1	4,9	16,6	27,2	34,2	41,0	49,0	57,1	90,3	109,5
1,5	4,6	15,5	25,4	31,9	38,3	45,7	53,2	81,7	97,4
2,5	4,0	13,6	22,4	28,2	33,8	40,4	46,8	69,7	81,8
4	3,4	11,6	19,0	24,0	28,7	34,3	39,7	58,0	67,5

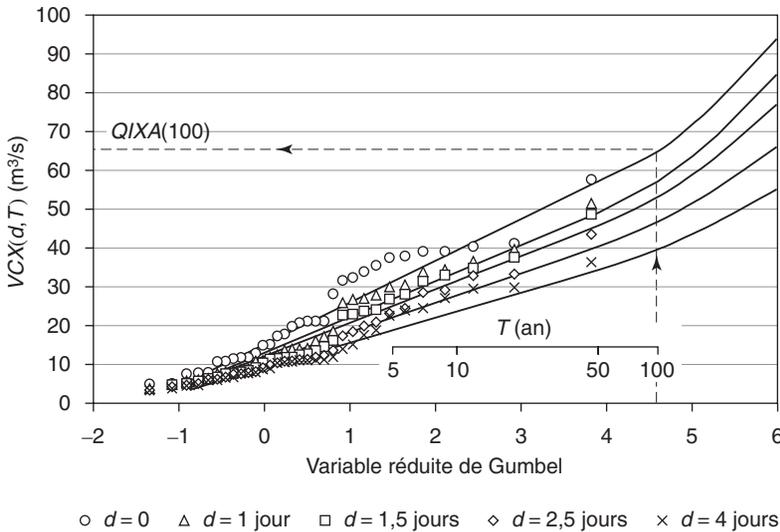


Figure 5.29. Courbes QdF sur l'Aa à Wizernes (combinaison du modèle local convergent et du gradex des pluies).

L'extrapolation a été réalisée au-delà de $T = 50$ ans, à partir :

- de l'équation 5.9 pour les durées $d = 1, 1,5, 2,5$ et 4 jours ;
- de l'équation 5.10 pour le débit de pointe ($d = 0$).

On pourra comparer ces résultats aux quantiles de débit qui apparaissent sur le site de l'agence de l'eau Artois-Picardie (figure 5.30) :

Superficie du bassin versant	392 km ²
Période de référence	1964-1994
QMNA5 ans en m ³ /s	2,16
QMNA5 ans en l/s/km ²	5,51

Écoulements inter-mensuels et inter-annuels

Paramètres	janv.	fév.	mars	avril	mai	juin	juil.	août	sept.	oct.	nov.	déc.	année
Débit (en m ³ /s)	6,80	7,20	6,60	6,30	5,40	4,45	3,76	3,19	2,99	3,30	4,08	5,75	4,83
Débit (en l/s/km ²)	17,35	18,37	16,84	16,07	13,78	11,35	9,59	8,14	7,63	8,42	10,41	14,67	12,32
Lame d'eau (en mm)	46,46	49,20	45,10	43,05	36,90	30,41	25,69	21,80	20,43	22,55	27,88	39,29	408,73

Coefficients de pondération

Module annuel maxi en m ³ /s	7,74 en 1994
Module annuel mini en m ³ /s	2,36 en 1973
$Q_{\text{mensuel mini}} / Q_{\text{mensuel maxi}}$	0,08

Crues

Temps de retour (en années)	QIX (en m ³ /s)	QJX (en m ³ /s)	VCX3 (en m ³ /s)	VCX10 (en m ³ /s)
2	20,30	14,20	11,50	9,50
5	28,90	22,00	18,00	14,70
10	34,60	27,20	22,30	18,10
20	40,10	32,10	26,50	21,30
50	47,20	38,50	31,80	25,60
100				
Maxi connu pour l'année	40,50 1988	36,00 1988	31,10 1988	26,10 1988

Étiages

QCN3 (en m ³ /s)	QCM10 (en m ³ /s)	VCN30 (en m ³ /s)	
2,25	2,30	2,45	
1,90	1,95	2,10	
1,75	1,80	1,90	
Mini connu pour l'année	1,54 1973	1,56 1973	1,61 1973

Bilan hydrologique

Module pluviométrique P	850 mm
Lame d'eau équivalente L	409 mm
Déficit d'écoulement $L - P$	-441 mm
Coefficient d'écoulement L/P	0,48

Observations

Figure 5.30. Débits caractéristiques de l'Aa à Wizernes (d'après Diren Nord-Pas-de-Calais, Services de l'Eau et des Milieux Aquatiques).

L'incorporation des épisodes postérieurs à 1994 entraîne une révision à la hausse des quantiles. Cette hausse est relativement sensible et généralisée pour les débits de pointe QIX. L'augmentation est ainsi de 10 m³/s pour la crue cinquantennale, soit +25 %.

Construction d'hydrogrammes de référence

Hydrogrammes de projet HSMF

L'Hydrogramme synthétique monofréquence $HSMF(T)$ est construit à partir de l'estimation des quantiles de débit. La pointe de crue est définie par le débit $QIX(T)$ instantané maximal de la période de retour T d'intérêt. La montée de l'hydrogramme est supposée en croissance linéaire avec le temps. Pour différentes durées d , on construit la courbe de récession de sorte que le débit moyen maximal sur la durée d soit égal à $V CX(d, T)$. L'avantage de l'hydrogramme $HSMF(T)$ est de garantir que toutes les caractéristiques de l'hydrogramme sont rattachées à une même fréquence de dépassement $\frac{1}{T}$. Outre les quantiles de débit $V CX(d, T)$, deux paramètres sont utilisés pour construire l'hydrogramme $HSMF(T)$: le débit de base et le temps de montée (par défaut égal à la durée de Socose). Pour fixer le premier paramètre de l'hydrogramme, nous avons retenu le résultat de la régression, présentée en *figure 5.20* (coefficient de détermination de 0,62), entre le débit de base (approximativement le débit minimal sur les 40 heures qui précèdent la pointe) et la pointe de crue, soit :

$$Q_{\text{base}}(T) = 0,41 QIX(T) . \quad (5.12)$$

Le second paramètre est pris égal à la durée de Socose ($t_m = D = 36$ h). Cette valeur est compatible avec l'échantillon du temps de montée des vingt plus forts hydrogrammes (*figure 5.31*), mais elle est dans la gamme des valeurs fortes. Une seconde possibilité est offerte en calculant la médiane des durées de dépassement de la moitié de la différence entre le débit de pointe et le débit de base sur la totalité de l'échantillon de la *figure 5.21*. On obtient une deuxième valeur représentative du débit égal à 20 heures. Ces deux valeurs seront testées dans la construction d'hydrogrammes de projet.

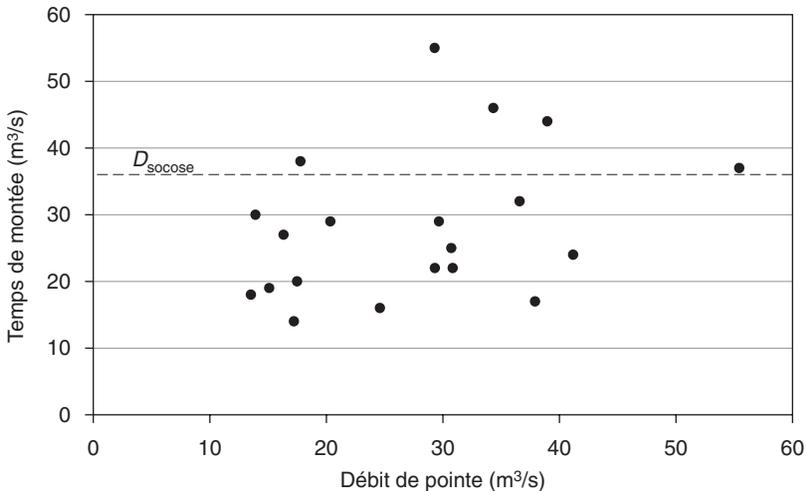


Figure 5.31. Liaison entre le temps de montée et le débit de pointe de crue (l'Aa à Wizernes).

À titre d'illustration, les figures 5.32 et 5.33 présentent les hydrogrammes $HSMF(T)$ pour $T = 10$ et 100 ans, construits de façon à être homogène en fréquence à la fois pour le débit de pointe et pour les volumes de crue écoulés. On peut noter que les hydrogrammes construits avec un temps de montée de 36 heures présentent une allure peu réaliste, avec une décrue plus rapide que la montée de crue ! Le choix d'un temps de montée de 20 heures est ici plus adapté, avec une forme de crue plus en rapport avec celle des huit plus forts hydrogrammes de crue.

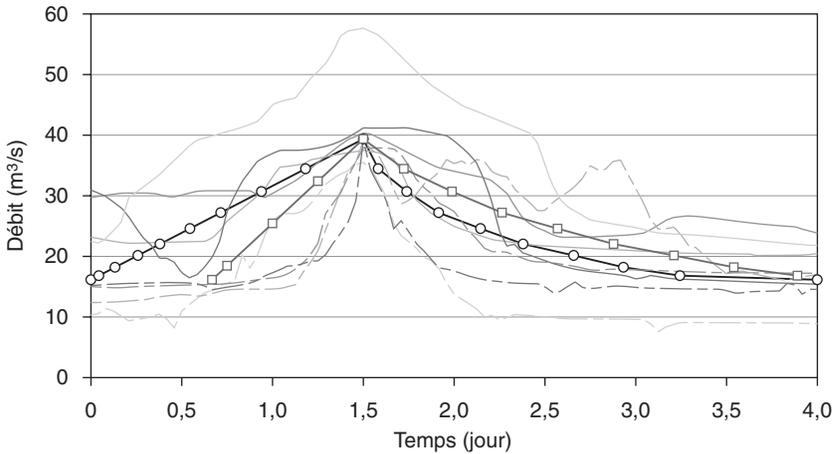


Figure 5.32. $HSMF(T = 10$ ans) sur l'Aa à Wizernes : symbole (O) avec un temps de montée de 36 heures (durée de Socose dans sa déclinaison classique μ) ; symbole (□) avec un temps de montée de 20 heures (susceptible de représenter la durée de ruissellement θ).

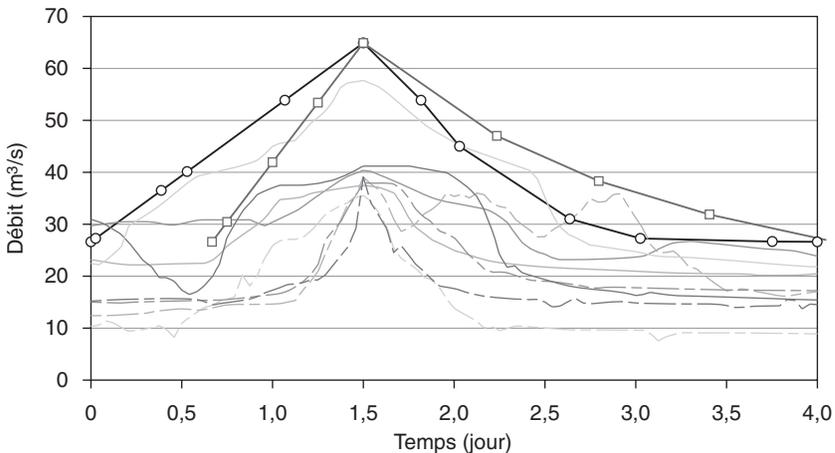


Figure 5.33. $HSMF(T = 100$ ans) sur l'Aa à Wizernes : symbole (O) avec un temps de montée de 36 heures (durée de Socose dans sa déclinaison classique μ) ; symbole (□) avec un temps de montée de 20 heures (susceptible de représenter la durée de ruissellement θ).

Hydrogrammes de projet avec prise en compte de la forme des crues

Les hydrogrammes précédents (figures 5.32 et 5.33) ont été construits en supposant une montée linéaire des crues. Lorsque la notion de gradient de montée est importante (cf., par exemple, une étude de diagnostic du comportement d'un ouvrage évacuateur de crue), il est préférable d'examiner les propriétés statistiques des formes d'hydrogramme afin de déterminer si l'hypothèse simplificatrice d'une montée linéaire peut être acceptée.

L'examen de la figure 5.22 montre une certaine courbure des hydrogrammes. Nous avons représenté sur les figures 5.34 et 5.35 les hydrogrammes *HSMF* (symbole \square) et des hydrogrammes corrigés (symbole \triangle) de la manière suivante :

- la montée de la crue est déterminée à partir de l'hydrogramme moyen (figure 5.22) ;

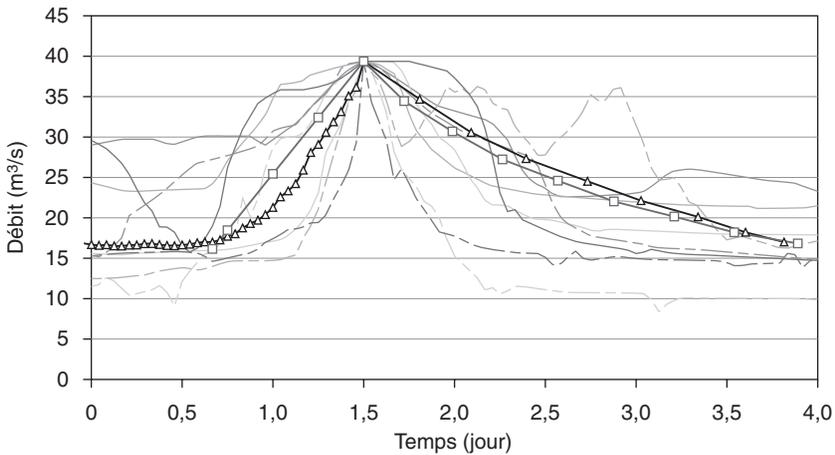


Figure 5.34. Hydrogrammes de référence ($T = 10$ ans) sur l'Aa à Wizernes.

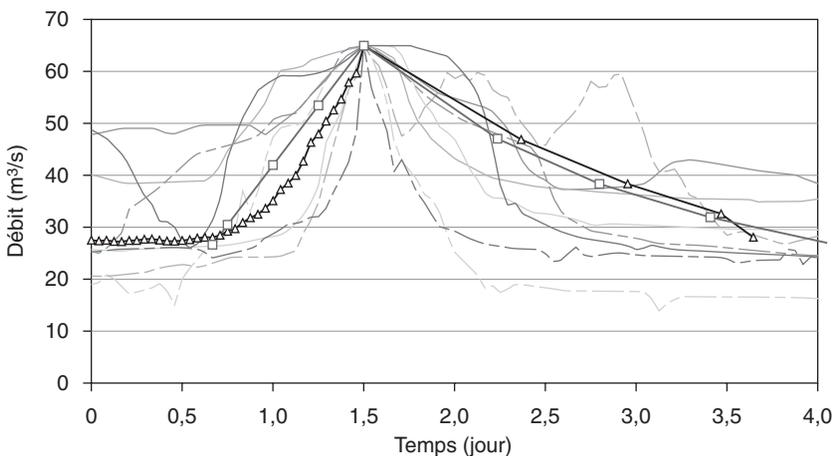


Figure 5.35. Hydrogrammes de référence ($T = 100$ ans) sur l'Aa à Wizernes.

– la décrue est construite de manière à garder des volumes de crue cohérents avec les quantiles de débit $VCXd(T)$ relatifs à la période de retour T choisie.

Dans les deux cas de figure ($T = 10$ ou 100 ans), la correction conduit à un hydrogramme avec un plus fort gradient de montée de crue et une décrue moins rapide. L'hydrogramme *HSMF* et sa version corrigée constituent deux scénarios possibles pour la caractérisation des crues, sachant qu'il existe une forte variabilité des formes d'hydrogrammes de crue.

Conclusion

Cet exemple constitue un cas limite d'application des approches probabilistes présentées aux chapitres 2 et 3. En effet, le fonctionnement du bassin versant, fortement influencé par la nappe de la craie, fait qu'il existe une forte auto-corrélation des crues d'une même année, voire de plusieurs années successives. L'hypothèse d'un échantillon de crues composé de valeurs indépendantes est donc sujette à caution. Une alternative consisterait à ne prendre par exemple qu'une crue tous les 2 à 3 ans, mais la taille de la chronique (environ 30 ans) fait que l'échantillon résultant serait trop limité (une dizaine de valeurs) pour estimer correctement la crue décennale.

Par ailleurs, l'hypothèse d'un coefficient de ruissellement marginal voisin de 1 à partir d'une période de retour prise ici à 50 ans reste très discutable. Il est vraisemblable que les plus fortes crues soient la conjonction d'un niveau de nappe élevé (suite à plusieurs années de pluviométrie excédentaire) et de volumes importants de précipitations sur plusieurs jours. Il serait préférable de raisonner sur plusieurs variables : le débit de base avant la crue est lié au fonctionnement de la nappe et à l'historique de la pluviométrie à une échelle pluri-annuelle, et la réponse du bassin versant à une sollicitation pluviométrique, compte tenu de l'état de saturation du bassin. L'apport d'un outil de modélisation hydrologique incluant les différents compartiments concernés (atmosphère, sol, nappe) semble ici incontournable pour affiner l'estimation des crues de référence.

Applications de méthodes de régionalisation en zone méditerranéenne française

Présentation des approches testées

Les différentes méthodes de régionalisation exposées au chapitre 4 vont être appliquées sur trois sous-bassins du Réal Collobrier. Le *tableau 5.10* récapitule leurs principales caractéristiques géométriques et climatologiques.

Les paramètres sur les pluies maximales annuelles ont été obtenus à partir de l'ajustement des paramètres d'une loi de Gumbel sur la chronique du poste des Bonnaud (période 1966-2001), puis d'un lissage par une formule de Montana. La *figure 5.36* présente la série des courbes de cumul pluviométrique-durée-fréquence (de 1 à 48 heures) et la relation du gradex des précipitations avec la durée.

Tableau 5.10. Caractéristiques hydrologiques des trois bassins étudiés.

Rivière	Valescure	Réal Collobrier	
Station	Collobrières	Collobrières	Pont de Fer
Chronique hydrométrique	1967-2001	1972-2002	1966-2001
Superficie S (en km^2)	9,4	29	70,4
Longueur du talweg principal L (en km)	7	13	22
Pluie annuelle PA (en mm)	1 000	1 000	1 000
Pluie journalière centennale $PJ(100)$ (en mm)	185	185	185
Pluie journalière décennale $PJ(10)$ (en mm)	130	130	130
Coefficient de Montana = b	0,30	0,30	0,30
Température annuelle t_A (en $^{\circ}\text{C}$)	15	15	15
Plus fort débit de la chronique (en m^3/s)	19	36	145

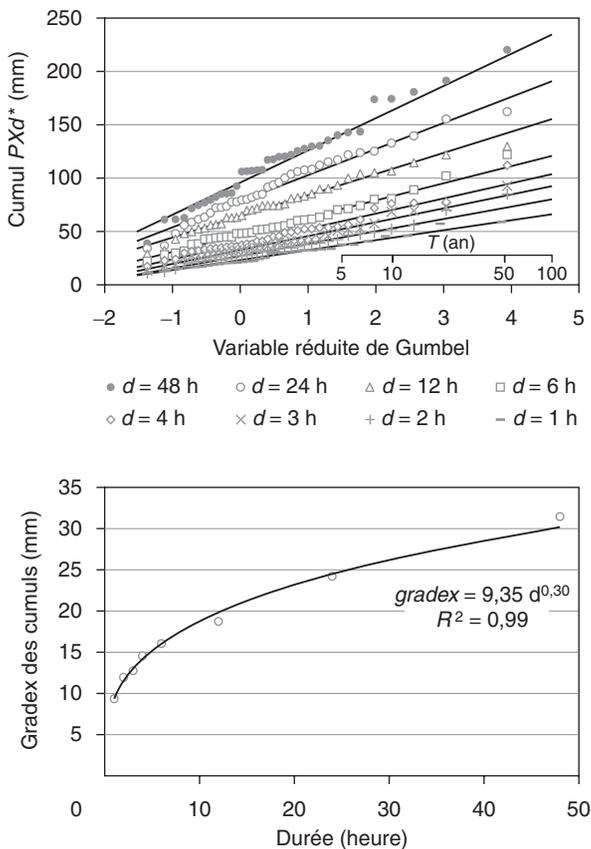


Figure 5.36. Courbes lame précipitée-durée-fréquence et valeurs du gradex (poste des Bonnaud).

La figure 5.37 présente les courbes QdF ajustées à partir d'une loi exponentielle, puis consolidées à l'aide d'une formulation du type « modèle QdF convergent » :

$$V(d, T) = \frac{V(0, T)}{1 + \frac{d}{\Delta}}, \text{ avec } V(0, T) = x_0 + a_0 \ln T. \quad (5.13)$$

Les quantiles de débit $V(d, T)$ ont été calculés pour chacune des cinq durées d (avec $\frac{d}{D^*} = 0, 0,5, 1, 2$ et 4 ; D^* est la durée moyenne de dépassement de la moitié du débit de pointe), à partir d'un échantillonnage par valeurs supérieures à un seuil (en moyenne 2 épisodes par an).

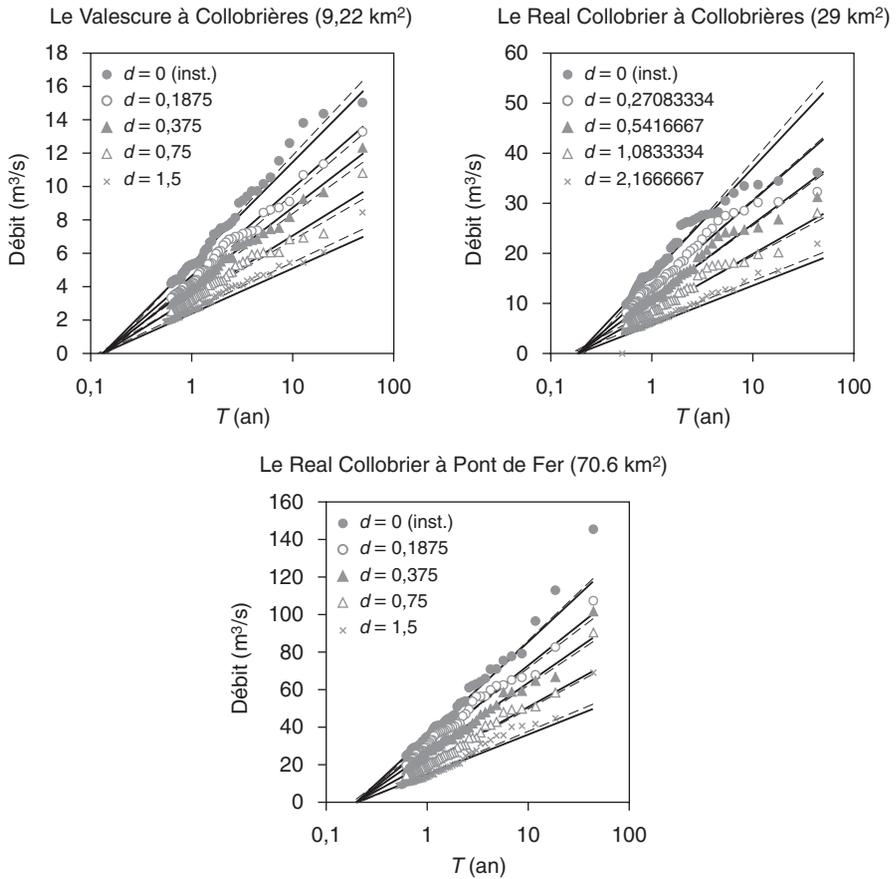


Figure 5.37. Courbes QdF de référence, obtenues à partir des données hydrométriques.

Les paramètres de ces courbes QdF sont donnés dans le *tableau 5.11*. On remarque que les durées Δ et D^* varient peu avec la superficie. On constate en effet que les crues du Réal Collobrier à Pont de fer (en aval) sont souvent en avance sur celles de Collobrières

(en amont). Quant au Valescure, ce bassin plus boisé réagit plus lentement aux précipitations que celui du Réal Collobrier.

Tableau 5.11. Caractéristiques de crue des trois bassins étudiés.

Rivière	Valescure		Réal Collobrier
Station	Collobrières	Collobrières	Pont de Fer
x_0 (en m ³ /s)	5,38	15,7	34,1
a_0 (en m ³ /s)	2,65	9,3	21,3
Durée Δ (en jour)	1,2	1,25	1,1
Durée D^* (en jour)	0,375	0,542	0,375

Régressions multiples

Le tableau 5.12 donne les valeurs de débit de référence obtenues à l'aide de régressions du type :

$$Q(T) = aS^b. \quad (5.14)$$

Les résultats sont systématiquement arrondis, compte tenu des incertitudes de ce type d'approches.

Tableau 5.12. Débits de référence (en m³/s) obtenus à partir de régressions multiples.

Rivière	Valescure		Réal Collobrier
Station	Collobrières	Collobrières	Pont de Fer
Crue maximale (Stanescu)	350	900	1 800
Crue maximale (Francou-Rodier)	700	1 100	1 700
Crue millénale (Gradex régionalisé)	80	200	350
Crue millénale (Bressand-Golossof)	270	620	1 200
Crue centennale (Bressand-Golossof)	160	370	730
Crue d'ordre centennal (Francou-Rodier)	190-300	340-540	550-840
Crue décennale (Crupedix)	15	40	85
Crue décennale (Socose)	5	15	40

Régressions sans variable explicative liée à la pluie

La crue maximale a été évaluée à partir des courbes enveloppes de Stanescu (cf. courbe n° 1 de la figure 1.1) et de la formule de Francou-Rodier, avec un coefficient $k = 5,5$.

$$\widehat{Q}_{\max} = 10^6 \left(\frac{S}{10^8} \right)^{0,45}. \quad (5.15)$$

La crue millénaire a été estimée à partir d'une analyse régionale du gradex, avec un coefficient $\lambda = 16,4$:

$$\widehat{Q}(1\ 000) = 16,4 S^{0,72}. \quad (5.16)$$

Rappelons que les bornes de l'intervalle de confiance à 90 % situent l'estimation entre 0,55 et 1,8 fois la valeur trouvée avec $\lambda = 16,4$ (tableau 4.1).

Les crues centennales et millénales ont été obtenues à partir de la formule de Bressand-Golossof associée aux crues rares et exceptionnelles :

$$\widehat{Q}_{\text{rare}} = 30 S^{0,75} \quad \text{et} \quad \widehat{Q}_{\text{exc.}} = 50 S^{0,75}. \quad (5.17)$$

La crue d'ordre centennale est estimée par Francou-Rodier, avec pour la zone méditerranéenne, hors Pyrénées-Orientales et Cévennes, une valeur de k comprise entre 4,7 et 5 :

$$10^6 \left(\frac{S}{10^8} \right)^{0,53} \leq \widehat{Q}(100) \leq 10^6 \left(\frac{S}{10^8} \right)^{0,5}. \quad (5.18)$$

Régressions avec une variable explicative liée à la pluie

La crue décennale a été obtenue par la formule de Crupedix, avec une pluie journalière décennale $PJ(10) = 130$ mm et un coefficient régional $R = 1$:

$$\widehat{Q}(10) = S^{0,8} \left(\frac{130}{80} \right)^2. \quad (5.19)$$

La crue décennale par la formule de Socose a été obtenue à partir de la relation qui suit et des résultats intermédiaires du tableau 5.13 :

$$\widehat{Q}(10) = \frac{kS}{(1,25D)^b} \cdot \frac{\rho^2}{15 - 12\rho}, \quad (5.20)$$

avec :

$$k = \frac{24^b PJ(10)}{21 \times \left(1 + \frac{\sqrt{S/D}}{30D} \right)}; \quad (5.21)$$

$$\ln D = -0,69 + 0,32 \ln S + 2,2 \sqrt{\frac{PA}{PJ(10)t_A}}; \quad (5.22)$$

$$\rho = 1 - \frac{0,2J}{k(1,25D)^{1-b}}; \quad (5.23)$$

$$J = 260 + 21 \ln \left(\frac{S}{L} \right) - 54 \sqrt{\frac{PA}{PJ(10)}}. \quad (5.24)$$

Tableau 5.13. Calcul de la crue décennale par la formule Socose.

Rivière	Valescure		
Station	Collobrières	Collobrières	Pont de Fer
Interception potentielle J (en mm)	116	127	135
Durée caractéristique de crue D (en h)	5,0	7,1	9,5
Coefficient k	15,9	15,9	15,9
Coefficient ρ	0,59	0,65	0,70
Crue décennale (en m ³ /s)	4	14	40

Modèles préparamétrés

Formule rationnelle

La formule rationnelle attribue au débit la même fréquence que celle de la pluie. Il n'y a pas de limite fréquentielle à son utilisation, bien qu'elle soit généralement utilisée pour estimer les débits de crue de périodes de retour comprises entre 10 et 100 ans.

La formule rationnelle a pour expression :

$$\widehat{Q}(T) = 0,278 C_e(T) I(t_c, T) S. \quad (5.25)$$

Le coefficient d'écoulement $C_e(T)$ est estimé à partir d'une étude régionale en zone méditerranéenne (Astier *et al.*, 1993) :

$$C_e(T) = \alpha(T) \left(1 - \frac{Po}{PJ(T)} \right). \quad (5.26)$$

Tableau 5.14. Détermination du coefficient d'écoulement $C_e(T)$.

Période de retour T (en années)	10	100	1 000
$\alpha(T)$	0,6	0,8	0,85
$PJ(T)$ (en mm)	130	185	240
Po^*	60	60	60
$C_e(T)$	0,32	0,54	0,64

* La valeur $Po = 60$ mm correspond à une zone montagneuse et boisée avec un sol sableux.

L'intensité de la pluie nette $I(t_c, T)$ sur le temps de concentration t_c est obtenue à l'aide des courbes pluie cumulée-durée-fréquence (figure 5.36). Le temps de concentration t_c est calculé d'après la formule suivante (cf. chapitre 4, § Modèles préparamétrés) :

$$t_c = 1,8L^{0,6} P e^{-0,33} P n^{-0,23}. \quad (5.27)$$

On peut noter, dans le tableau 5.15, que la rétention des bassins versants ($PJ - Pn$) est de l'ordre de 90 mm pour les trois périodes de retour étudiées. Ces rétentions sont

plus faibles que celles du *tableau 5.13* estimées par la méthode Socose. L'application de la formule rationnelle permet d'estimer les débits de crue (*tableau 5.16*). Les intensités de la pluie sont calculées d'après la formulation de Montana (*tableau 5.17*).

Tableau 5.15. Détermination du temps de concentration t_c (en heures).

Rivière		Valescure	Réal Collobrier	
Station		Collobrières	Collobrières	Pont de Fer
Longueur L du talweg principal (en km)		4,1	6,3	13
Pente Pe des versants (en %)		38	35	31
Pn : pluie nette (en mm)	$T = 10$ ans	42	42	42
	$T = 100$ ans	100	100	100
	$T = 1\ 000$ ans	154	154	154
Temps de concentration t_c (en heures)	$T = 10$ ans	2,4	3,2	5,2
	$T = 100$ ans	2,0	2,6	4,2
	$T = 1\ 000$ ans	1,8	2,4	3,9

Tableau 5.16. Débit de référence de crue par application de la formule rationnelle.

Rivière		Valescure	Réal Collobrier	
Station		Collobrières	Collobrières	Pont de Fer
Intensité de pluie $I(t_c, T)$ (en mm)	$T = 10$ ans	27	22	16
	$T = 100$ ans	44	36	26
	$T = 1\ 000$ ans	61	50	36
Débit de crue $Q(T)$ (en m^3/s)	$T = 10$ ans	25	60	100
	$T = 100$ ans	60	155	270
	$T = 1\ 000$ ans	100	260	450

Tableau 5.17. Formulation de Montana pour les courbes intensité-durée-fréquence :
 IdF (en mm/h) = at^{b-1} , (t en heures).

Période de retour T (en années)	10	100	1 000
a	50	71	93
b	0,3	0,3	0,3

Méthode SCS

La méthode décrite au chapitre 4, § *Modèles préparamétrés*, associe une fonction de production définie par un paramètre J comparable au coefficient d'écoulement de la formule rationnelle, et une fonction de transfert triangulaire, tout comme la formule rationnelle. Alors que cette dernière considère un cumul de pluie sur une durée prise égale au temps de concentration t_c , notre expérience dans la mise en œuvre de la méthode SCS nous conduit à l'appliquer sur des durées de pluie sensiblement très supérieures à t_c (de l'ordre de $5t_c$). En effet, la formulation de la méthode implique que l'écoulement est nul tant que le total de pluie est inférieur à $0,2J$, soit 20 mm pour une faible valeur de $J = 100$ mm, et 50 mm pour une valeur de $J = 250$ mm.

L'hydrogramme résultant est obtenu par sommation des hydrogrammes successifs, obtenus à chaque pas de temps élémentaire d'une durée de l'ordre de $0,15t_c$.

Pour l'application aux trois bassins versants, nous retenons les valeurs suivantes :

- capacité de rétention du sol $J = 120$ mm (*tableau 5.13*) ;
- hydrogramme unitaire triangulaire de temps de montée Tp , de temps de descente $1,67Tp$,

avec $Tp = \frac{Tr}{2} + 0,6t_c$;

- pas de temps élémentaire $Tr = 1$ heure ;
- les temps de concentration t_c sont repris du *tableau 5.15* ;
- durée de l'hydrogramme $D = 12$ heures ;
- le hyétogramme est calculé à partir des valeurs a et b de la formule de Montana (*tableau 5.17*). La répartition de la pluie dans le hyétogramme est reprise de la méthode Socose (Ctgreff *et al.*, 1980-1982). Le cumul $P(t)$ des pluies sur une durée t est égal à :

$$P(t) = a \frac{t^b}{\left[\left(\frac{t}{D} \right)^2 + \frac{2D}{t} - 2 \right]^{\frac{3b}{7}}}, \quad (5.28)$$

où :

a et b sont les coefficients de Montana,

D est la durée du hyétogramme.

On vérifie que pour $t = D$, on retrouve $P(D) = aD^b$.

Le hyétogramme est construit pas par pas, pour $t = 1, 2, 3, \dots, 12$ heures. La pluie maximale en 1 heure est donc obtenue pour $t = 1$ heure. La pluie du pas de temps élémentaire suivant est obtenue par la différence entre $P(2 \text{ h})$ et $P(1 \text{ h})$, et ainsi de suite jusqu'à 12 heures. Les pluies de chaque pas de temps sont ensuite réorganisées autour du maximum de pluie du pas de temps élémentaire, que l'on situe au centre du hyétogramme.

Les calculs du *tableau 5.18* ont été faits par le logiciel Cruproj, développé au Cemagref pour mettre en œuvre la méthode SCS.

La *figure 5.38* présente les sorties du logiciel Cruproj : fiche récapitulative des résultats et hydrogramme de projet pour la fréquence au non-dépassement 0,9 pour la station de Collobrières sur le Réal Collobrier. Notons que la méthode permet de construire un hydrogramme de projet utilisable pour le dimensionnement d'ouvrage (volume de retenue, écoulement transitoire...).

Tableau 5.18. Débits de référence de crue par application de la méthode SCS (en m³/s).

Rivière		Valescure	Réal Collobrier	
Station		Collobrières	Collobrières	Pont de Fer
Temps de montée T_p (en h) *	$T = 10$ ans	1,9	2,5	3,6
	$T = 100$ ans	1,6	2,1	3,0
	$T = 1\ 000$ ans	1,6	1,9	2,8
Débit de crue $Q(T)$ (en m ³ /s)	$T = 10$ ans	12	41	76
	$T = 100$ ans	36	96	200
	$T = 1\ 000$ ans	60	190	360

* Les T_p sont calculés selon la formulation SCS ($T_p = Tr/2 + 0,6T_c$), avec les valeurs de T_c du tableau V.4.6, et $Tr = 1$ heure.

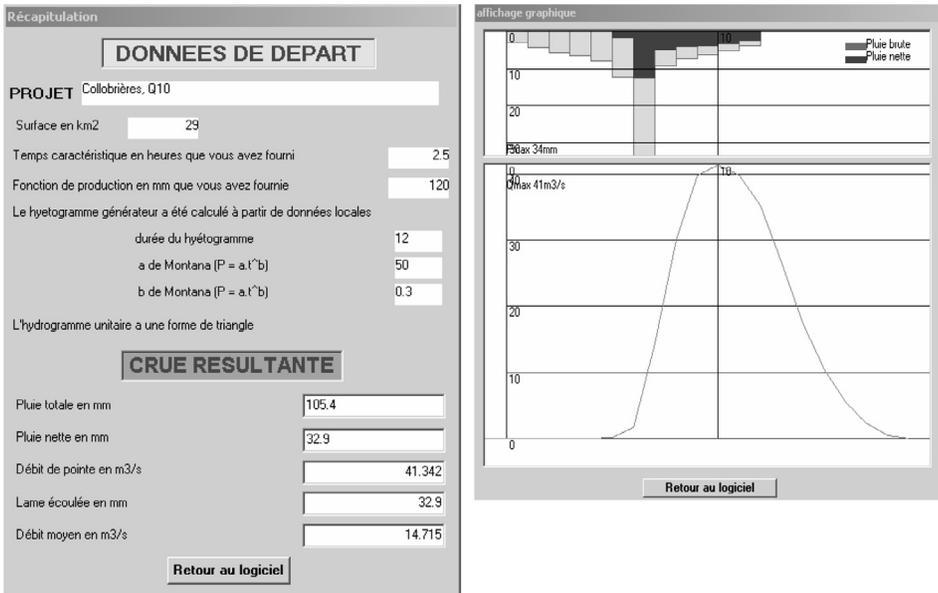


Figure 5.38. Résultats du logiciel Cruproj sur la méthode SCS : fiche de résultats, hyétogramme et hydrogramme de crue.

Modèle QdF de référence

Le choix d'un modèle QdF de référence est réalisé en comparant le ratio $\frac{a_p(d)}{Q(10)}$ (gradex des pluies sur la durée d par le débit décennal de pointe), avec deux limites :

$$L_1\left(\frac{d}{D}\right) = \frac{1}{0,419\frac{d}{D} + 1,58} \quad \text{et} \quad L_2\left(\frac{d}{D}\right) = \frac{1}{0,768\frac{d}{D} + 2,332}, \quad (5.29)$$

respectivement pour Florac-Soyans et Vandenesse-Florac. Le modèle de Florac est choisi pour les trois postes étudiés. Le *tableau 5.19* et la *figure 5.39* illustrent l'application du critère de choix pour le poste de Pont de Fer.

Tableau 5.19. Comparaison du ratio $a_p(d)/Q(10)$, avec les limites $L_1(d/D)$ et $L_2(d/D)$ sur le poste de Pont de Fer.

Durée d (en heures)	4,5	9	18	36
Gradex $a_p(d)$ (en mm)	15	18	22	27
Gradex $a_p(d)$ (en m ³ /s)	64	39	24	15
d/D	0,64	1,29	2,57	5,14
$a_p(d)/Q(10)$	0,76	0,47	0,29	0,18
Limite Florac-Soyans	0,60	0,56	0,51	0,43
Limite Vandenesse-Florac	0,35	0,30	0,23	0,16

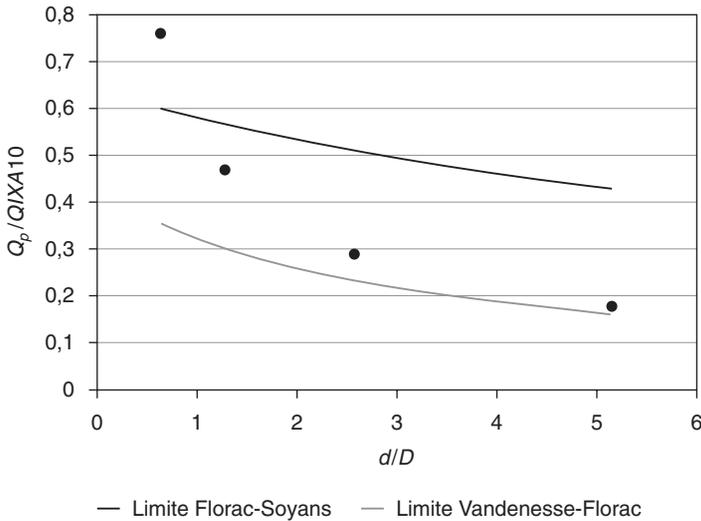


Figure 5.39. Choix d'un modèle QdF de référence sur le poste de Pont de Fer.

On voit sur la *figure 5.40* que, par construction, l'adéquation est bonne en ce qui concerne le débit instantané et la fréquence décennale. Elle le reste sur les deux durées suivantes (0,375 et 0,75 jour), mais le modèle sous-estime ensuite les quantiles de durée plus importante ($d = 1,5$ jour).

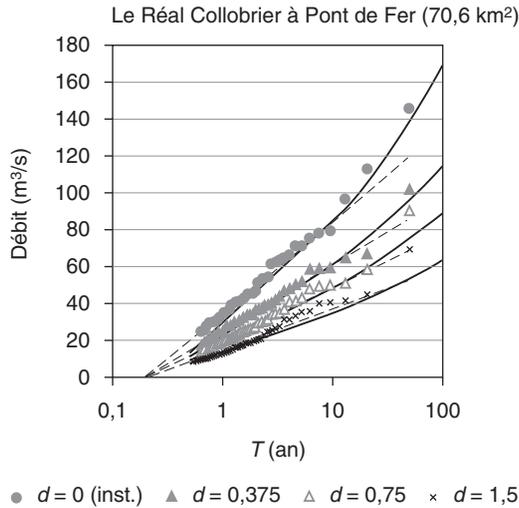


Figure 5.40. Comparaison sur le poste de Pont de Fer des courbes QdF obtenues à l'aide du modèle de Florac (trait plein), le modèle convergent (pointillé) et les observations (points).

Modèle QdF régional convergent

Une alternative à l'approche précédente est de bâtir un modèle QdF régional $V_R(d, T)$ adimensionnel, issu de la formulation du modèle QdF convergent :

$$V(d, T) = Q(10) \frac{V_R(d, T)}{1 + d/\Delta} \quad (5.30)$$

Dans le cas où on se limite aux crues courantes ($T < 10$ ans), on peut retenir une distribution exponentielle du type :

$$V_R(d, T) = x_R + a_R \ln T \quad (5.31)$$

Les deux paramètres x_R et a_R sont obtenus par moyenne de l'ensemble des paramètres $x_i(j)$ et $a_i(j)$ relatifs à l'ajustement des échantillons de crue normés : débit moyen maximal du poste j , calculé sur la durée d_i , multiplié par $\left(1 + \frac{d_j}{\Delta_j}\right)$ et divisé par $Q(10)_{\text{poste } j}$.

Dans le cas des trois postes du bassin du Réal Collobrier, la distribution résultante conduit à l'expression suivante, avec des coefficients régionaux $x_R = 0,426$ et $a_R = 0,24$:

$$V(d, T) = \widehat{Q}(10) \frac{0,426 + 0,24 \ln T}{1 + d/\widehat{\Delta}} \quad (5.32)$$

où $\widehat{Q}(10)$ et $\widehat{\Delta}$ sont estimés localement sur le site choisi.

Le *tableau 5.20* présente une comparaison des quantiles de débit $V(d, T)$ au poste de Pont de Fer, d'après :

- a/ les données hydrométriques du poste de Pont de Fer ;
- b/ le modèle QdF de référence de Florac ;
- c/ le modèle régional convergent calé sur les trois postes du bassin du Réal Collobrier.

Tableau 5.20. Comparaison des quantiles de crue à Pont de Fer.

d (en h)	$V_d(T)$ (en m^3/s)					
	$T = 0,5$	$T = 1$	$T = 5$	$T = 10$	$T = 50$	$T = 100$
0	19,5	34,4	69,1	84,0	134,7	165,4
9	14,3	25,1	50,0	60,7	97,9	120,2
18	11,4	20,0	39,8	48,3	75,2	89,9
36	8,2	14,8	30,1	36,7	55,5	65,1

d (en h)	$V(d,T)$ (en m^3/s) – Modèle de référence Florac					
	$T = 0,5$	$T = 1$	$T = 5$	$T = 10$	$T = 50$	$T = 100$
0	13,3	29,7	67,7	84,1	137,6	167,9
9	10,7	22,4	49,4	61,1	95,7	113,7
18	9,2	18,2	39,2	48,3	74,7	88,2
36	7,4	13,7	28,2	34,5	53,1	62,9

d (en h)	$VCX(d,T)$ (en m^3/s) – Modèle régional convergent					
	$T = 0,5$	$T = 1$	$T = 5$	$T = 10$	$T = 50$	$T = 100$
0	21,8	35,8	68,3	82,2	–	–
9	16,3	26,7	50,9	61,3	–	–
18	13,0	21,3	40,6	48,9	–	–
36	9,2	15,1	28,9	34,8	–	–

La *figure 5.41* permet d'apprécier la pertinence de la forme du modèle régional convergent (cf. équation 5.30), avec une variation du quantile $V(d,T)$ en fonction de la durée d d'après une forme homographique en $\frac{1}{1+d/\Delta}$. Les valeurs $\frac{x_i}{Q(10)}$ et $\frac{x_R}{1+d/\Delta}$, et les valeurs $\frac{a_i}{Q(10)}$ et $\frac{a_R}{1+d/\Delta}$ suivent bien une forme homographique en fonction de $\frac{d}{\Delta}$.

Application de la méthode Shyreg à la zone méditerranéenne française

Le logiciel PRD'crue est utilisé pour calculer les débits de crue sur les trois bassins versants étudiés. Il permet de mettre très facilement en œuvre la méthode à partir des bases de données pluie et débit qui lui sont associées. Les limites des trois bassins versants, sur la *figure 5.42*, ont été tracées automatiquement à partir de la base altimétrique au pas de 50 m de l'Institut géographique national (IGN).

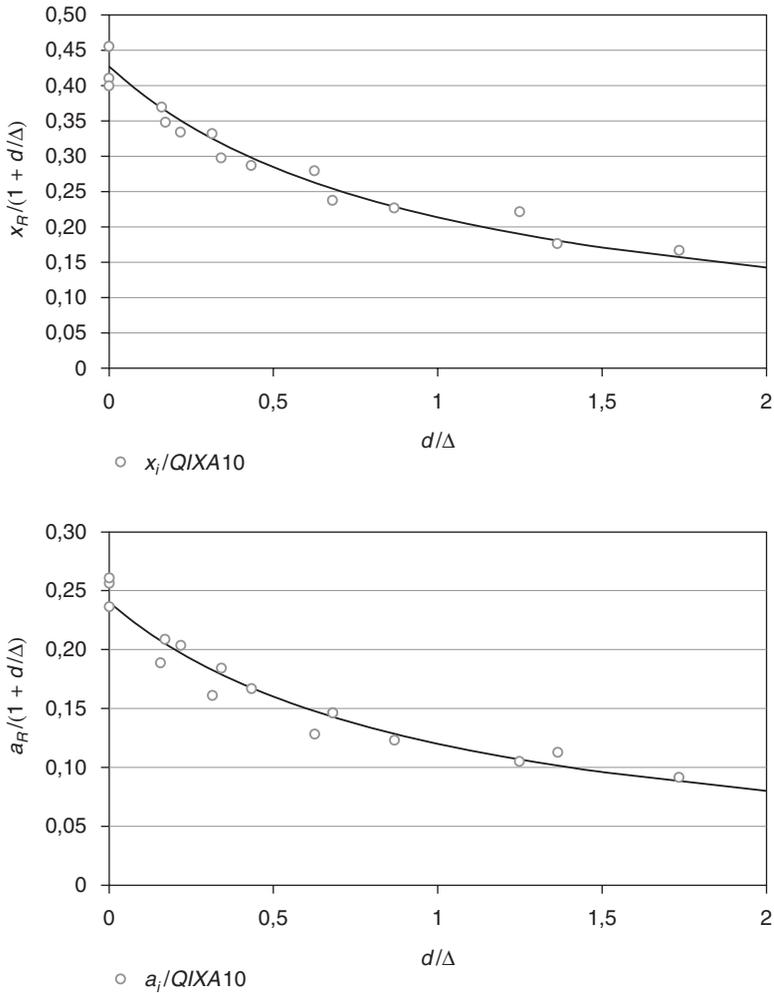


Figure 5.41. Pertinence de la forme du modèle régional convergent sur le poste de Pont de Fer.

Le *tableau 5.21* présente les caractéristiques moyennes de la pluviométrie du bassin du Pont de Fer. Notons que la pluie journalière maximale « moyenne » *PJX* de fréquence décennale est tout à fait comparable à la valeur déduite de la distribution de fréquence (les estimations maximum et minimum de ce quantile sont respectivement de 150 et 120 mm sur le bassin). En revanche, la pluie journalière centennale est sensiblement supérieure, 256 mm à comparer à l'estimation de 185 mm obtenue par une extrapolation strictement exponentielle des observations.

Le logiciel PRD'crue propose ensuite, pour différentes fréquences et différentes durées, les débits de référence de crue des bassins versants. Le *tableau 5.22* présente, à titre d'illustration, les résultats obtenus pour la station du Pont de Fer. Dans le cadre de cet exemple, le calcul du débit de pointe de fréquence décennale a été effectué à l'aide du logiciel

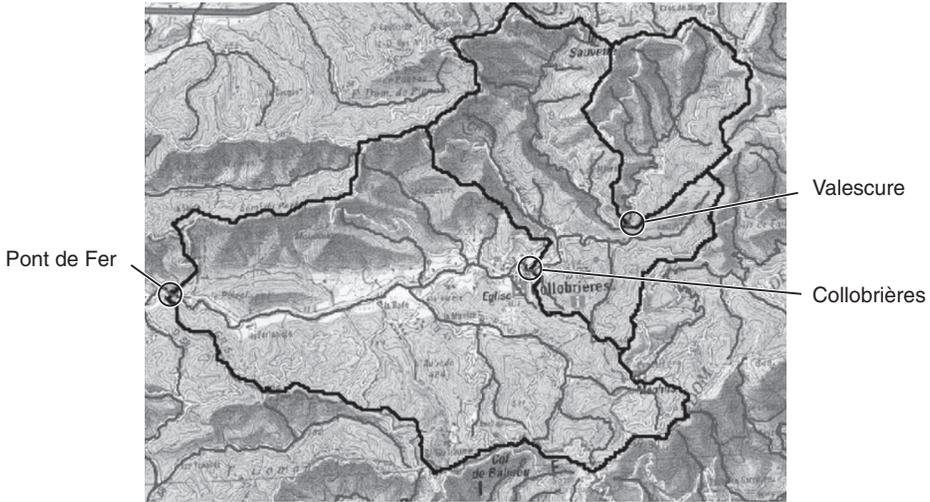


Figure 5.42. Limites des trois bassins versants sur le Réal Collobrier.

Tableau 5.21. Caractéristiques de la pluviométrie simulée sur le poste de Pont de Fer (en mm).

Période de retour	2 ans	5 ans	10 ans	20 ans	50 ans	100 ans
<i>PJX</i>	83	116	140	170	215	256
<i>PX(1 h)</i>	31	41	48	57	73	86
<i>PX(2 h)</i>	40	53	62	75	98	117
<i>PX(3 h)</i>	48	63	74	91	122	144
<i>PX(4 h)</i>	53	71	83	101	140	167
<i>PX(6 h)</i>	60	80	95	115	161	197
<i>PX(12 heures)</i>	76	104	123	148	196	240
<i>PX(24 h)</i>	95	131	159	192	240	283
<i>PX(48 heures)</i>	117	155	187	225	278	322
<i>PX(72 h)</i>	129	170	202	239	292	337

PRD'crue, selon les données suivantes, extraites automatiquement de la base des débits pixellisés :

- débit du pixel le plus fort : $2,6 \text{ m}^3/\text{s} \cdot \text{km}^2$;
- débit du pixel le plus faible : $2,2 \text{ m}^3/\text{s} \cdot \text{km}^2$;
- somme des pixels à l'intérieur des limites du bassin versant : $166 \text{ m}^3/\text{s} \cdot \text{km}^2$;
- coefficient d'abattement : $S^{-0,18} = 0,47$, avec $S = 70,4 \text{ km}^2$;
- débit de pointe décennal : $Q(10) = 166 \times 0,47 = 78 \text{ m}^3/\text{s}$.

Tableau 5.22. Débits de référence de crue du Réal Collobrier (en m³/s) (poste de Pont de Fer).

Période de retour	2 ans	5 ans	10 ans	20 ans	50 ans	100 ans
Q_{pointe}	34	57	78	110	178	267
$QX(1 \text{ h})$	31	51	70	99	159	239
$QX(2 \text{ h})$	30	50	69	97	155	233
$QX(3 \text{ h})$	29	48	67	94	148	222
$QX(4 \text{ h})$	28	47	64	90	142	212
$QX(6 \text{ h})$	25	43	60	83	129	191
$QX(12 \text{ heures})$	21	36	50	70	104	151
$QX(24 \text{ h})$	17	29	40	55	81	110
$QX(48 \text{ heures})$	13	22	29	39	55	71
$QX(72 \text{ h})$	5	9	12	17	24	34

Les estimations des débits de pointe de crue pour les trois bassins versants sont données dans le *tableau 5.23*.

Tableau 5.23. Estimation des débits de crue par la méthode Shyreg (en m³/s).

Rivière	Valescure		
Station	Collobrières	Collobrières	Pont de Fer
$Q(10)$	17	40	78
$Q(100)$	55	130	270
$Q(1\ 000)^*$	100	250	520

* Valeurs fournies à titre indicatif, la plage validée de période de retour étant comprise entre 2 et 100 ans.

Comparaison des différents résultats

Le *tableau 5.24* récapitule les estimations de débit, obtenues par les différentes approches régionales, pour les trois stations du Réal Collobrier. Il a été complété par les estimations issues d'approches locales (ajustement statistique d'une distribution sur les débits, extrapolation par la méthode du Gradex).

L'estimation de la crue maximale varie dans un rapport de 1 à 2, selon que l'on utilise les courbes enveloppes données par Stanescu ou celles de Francou-Rodier. Les valeurs, qui sont largement supérieures à la crue millénaire, peuvent difficilement être validées.

Tableau 5.24. Comparaison des méthodes d'estimation des débits de référence.

	Rivière	Valescure	Réal Collobrier	
	Station	Collobrières	Collobrières	Pont de Fer
Crue maximale (courbes enveloppe)	Stanescu	350	900	1 800
	Francou-Rodier	700	1 100	1 700
Crue millénaire	Ajustement statistique	25	85	185
	Méthode du Gradex	35	120	285
	Méthode Shyreg	100	250	520
	Modèle QdF Florac	40	125	280
	Gradex régionalisé	80	200	350
	Bressand-Golossof	270	620	1 200
	Formule rationnelle	130	330	590
	Méthode SCS	60	190	360
Crue centennale	Ajustement statistique	18	60	135
	Méthode du Gradex	20	75	165
	Méthode Shyreg	55	130	270
	Modèle QdF Florac	25	75	170
	Bressand-Golossof	160	370	730
	Francou-Rodier	190-300	340-540	550-840
	Formule rationnelle	60	150	270
	Méthode SCS	35	100	200
Crue décennale	Ajustement statistique	12	40	85
	Modèle QdF Florac	15	40	85
	Méthode QdF convergent	12	40	80
	Méthode Shyreg	17	40	78
	Formule Crupedix	16	40	80
	Formule rationnelle	20	50	80
	Méthode SCS	12	40	80
	Méthode Socose	4	14	40

L'estimation de la crue millénaire et centennale est à notre sens la mieux approchée par les méthodes basées sur l'exploitation de l'information pluviométrique (méthode du Gradex, modèles QdF, méthode Shyreg, méthode rationnelle). Le rapport du simple au

double entre ces différentes estimations vient du fait que la méthode du gradex et les modèles QdF sont basés sur un comportement strictement exponentiel de la distribution des pluies extrêmes, alors que la méthode Shyreg aboutit à un comportement hyperexponentiel des précipitations (quantiles plus élevés). Des travaux de recherche français sont en cours actuellement pour approfondir cette question, en travaillant sur de longues séries expérimentales ou sur des échantillons régionaux de valeurs extrêmes. La formule rationnelle appliquée hors de son domaine d'application conduit à une estimation plutôt élevée, en accord avec les hypothèses retenues d'augmentation du coefficient d'écoulement et de diminution du temps de concentration.

On notera toutefois que ces approches basées sur les pluies restent préférables à celles basées uniquement sur les débits, qui sous-estiment généralement les quantiles. Les estimations centennales obtenues par les régressions multiples (Bressand-Golossof, Francou-Rodier) sont largement supérieures aux estimations des autres méthodes. Il est acquis que ces méthodes de courbes enveloppes présentent un biais systématique qui conduit à la surestimation des débits, sur un bassin versant donné. Par construction, ces méthodes englobent les plus forts débits connus sur une région donnée, et une confusion existe entre la période de retour locale (sur un bassin versant) et la période de retour régionale. De plus, la formule de Bressand-Golossof, bien que souvent appliquée en zone méditerranéenne, a été développée pour le département du Gard dont les caractéristiques pluviométriques sont maximales. Cette formule a été calée sur les stations hydrométriques disponibles, situées en pied de relief cévenol où les précipitations présentent des gradients élevés avec l'altitude. La formule est donc difficilement transposable en dehors de son contexte, alors qu'il est malheureusement courant de constater que de telles formules sont appliquées sans vérification particulière hors de leur zone de calage.

L'estimation de la crue décennale peut être recoupée l'aide des observations de débit. La dispersion des résultats est assez faible, inférieure à 10 %, à l'exception de la méthode Socose qui sous-estime la valeur locale, sur cet exemple, de près de 50 %. Les méthodes Crupedix et Shypre utilisées en aveugle, à savoir sans connaissance de l'estimation statistique à partir de la chronique d'observation, donnent des estimations tout à fait comparables à l'estimation locale. La formule rationnelle, avec un paramétrage également en aveugle des coefficients C_e et de T_c , conduit à une estimation légèrement plus forte que l'estimation locale, mais d'un ordre de grandeur tout de même comparable. Remarquons que, très souvent, la formule rationnelle est utilisée avec de faibles valeurs du temps de concentration T_c , ce qui conduit à des estimations élevées. Notre connaissance des bassins nous a conduits à retenir une valeur de rétention maximale de 120 mm pour appliquer la méthode SCS. L'estimation SCS est alors identique à l'estimation locale. Bien entendu, une autre valeur du paramètre J , une autre durée de hétéogramme et un autre pas de temps élémentaire conduiraient à des estimations de $Q(10)$ différentes.

On retiendra ici qu'il est fortement recommandé de tester plusieurs jeux de paramètres des modèles pour avancer un ordre de grandeur de l'incertitude de l'estimation, et de systématiquement vérifier les performances des méthodes « sommaires » avant de les utiliser sur des bassins non jaugés. Le bon comportement des modèles QdF n'est pas ici significatif puisque ces modèles exploitent directement l'information sur la valeur du débit de pointe de la crue décennale. Il faudrait les vérifier sur d'autres durées de débit de crue, en comparant les courbes débit-durée-fréquence issues des analyses locales et celles reconstituées par les modèles QdF.

ANNEXES

- 1 – Principaux logiciels d'ajustements de lois de probabilité en hydrologie**
- 2 – Éléments de probabilités et statistiques appliquées à l'hydrologie**
- 3 – Références bibliographiques**

Annexe 1

Principaux logiciels d'ajustements de lois de probabilité en hydrologie

Logiciel *Extrêmes*

Description

Le logiciel est utilisable sous Matlab via une interface graphique. Il est téléchargeable gratuitement sur Internet. Ses fonctionnalités sont regroupées en deux catégories :

- Fonctionnalités statistiques classiques :
 - simulations de variables aléatoires de lois normale, log-normale (deux paramètres), exponentielle (un paramètre), gamma (deux paramètres), Weibull, Chi2, Student, Pareto, Beta, Uniforme et Pareto généralisée ou GPD ;
 - graphique des densités, fonctions de répartition, fonction de survie, fonctions quantiles des lois précitées ;
 - estimation des paramètres des lois précitées ;
 - estimation non-paramétrique de la densité (méthode du noyau, histogramme) ;
 - estimation paramétrique des quantiles ;
 - tests d'adéquation aux lois précitées par les tests d'Anderson-Darling et Cramer-Von-Mises.
- Fonctionnalités sur les valeurs extrêmes :
 - vérification de l'exponentialité des excès : il s'agit de s'assurer que la fonction de répartition des données se situe dans le domaine d'attraction de la loi Gumbel, et que le nombre d'excès est convenablement choisi. Cette vérification s'effectue graphiquement avec un QQ-plot et un test d'exponentialité sur les excès ;
 - estimation des paramètres de la loi GPD, via les méthodes de Hill, Hill généralisé, moments pondérés d'Hosking et Wallis, maximum de vraisemblance et Zipf ;
 - estimation des quantiles extrêmes avec la méthode POT ;
 - test ET et test GPD : deux tests d'adéquation pour la queue de distribution. Ils sélectionnent, par comparaison avec la méthode POT, les modèles centraux produisant de bonnes estimations de la queue de distribution. Lorsqu'on souhaite reconstituer la loi des observations aussi bien dans la région centrale qu'extrême, on applique d'abord à un

ensemble de modèles un test usuel (Anderson-Darling ou Cramer-Von-Mises), puis un test d'adéquation de la queue de distribution (ET ou GPD) ;

– si aucune loi n'est acceptée à la fois par un test central et par un test extrême (ET), la procédure de régulation bayésienne permet, à partir d'un modèle adapté aux valeurs les plus probables, d'améliorer l'adéquation extrême grâce à un avis d'expert sur la queue de distribution.

Contact

Garrido M., ENAC, Département MI 7, Avenue Édouard Belin, BP 4005, 31055 Toulouse cedex 04, France, garrido@recherche.enac.fr

Ouvrage technique

Thèse de Myriam Garrido, disponible sur le site web de l'Inrialpes :
<http://www.inrialpes.fr>

Logiciel Hydro

Description

Base nationale de données hydrométriques, qui, outre l'archivage des données, permet l'extraction de variables hydrologiques (données brutes ou données de synthèse) via une série de traitements :

QTVAR	Réalisation d'un hydrogramme, et calcul d'un volume d'écoulement.
VOIR, VOIR-AN	Examen graphique simultané de plusieurs séries de débits journaliers.
ENTRE2	Graphe de l'écoulement sur l'année en cours, comparé au passé.
INTERA	Écoulements moyens interannuels, mensuels et annuel (module).
TOUS-MOIS	Tous les écoulements mensuels disponibles sur une station.
PERIODE	Chronique et fréquence des débits moyens sur une période donnée.
VCN-QCN	Débits d'étiage, débits de base, débits minimaux sur N jours consécutifs.
QMNA	Débits mensuels minimaux, chronique et statistiques.
VCX-QCX	Hautes eaux : débits maximaux sur N jours consécutifs.
CRUCAL	Crues : historique des débits et estimation des fréquences de crue.
TOUS-QX	Tous les débits maximaux instantanés mensuels disponibles sur une station.
CORJOU	Corrélation des débits journaliers entre deux stations.
CORMEN	Corrélation des débits mensuels entre deux stations.
CUMUL	Test de l'homogénéité de deux séries de débits mensuels.
QTFIX	Tableau de débits à pas de temps fixe (débits horaires...).
DEBCLA	Débits classés : débits journaliers classés sur toute une période.
SYNTHESE	Synthèse hydrologique sommaire sur une station.
ZONAGE	Graphe de synthèse sur un ensemble de stations.
CARTES	Visualisation de cartes nationales sur l'hydrologie et ses facteurs.

Contact

Équipe d'administration de la banque HYDRO, Météopole/SCHAPI, 42, avenue Gaspard Coriolis, 31057 Toulouse cedex 01, hydro@schapi.ecologie.gouv.fr.

Accès à la base

<http://www.hydro.eaufrance.fr>

Logiciel *Hydrolab*

Description

Hydrolab est un outil très simple d'utilisation (environnement Microsoft Excel) et diffusé gratuitement. Il a pour principales fonctionnalités :

- ajustements par la méthode des moments de lois statistiques : Gauss, Galton (log-normale), Gauss avec transformation racine carrée, Gumbel, Fuller, Weibull, Poisson ;
- régressions multiples ;
- analyse en composantes principales ;
- diverses tâches hydrologiques : comblement de lacunes dans les séries d'observation, détection d'anomalies...

Hydrolab propose le graphique de fréquence sur lequel sont reportés les points expérimentaux, le tracé de la loi théorique, les intervalles de confiance. *Hydrolab* propose aussi une table des quantiles avec les bornes de l'intervalle de confiance, et offre la possibilité de calculer un quantile pour une probabilité donnée ou, inversement, la probabilité théorique d'une valeur donnée de la variable.

Contact

Laborde J.-P, Université Sophia Antipolis, Équipe Gestion et Valorisation de l'Environnement de l'UMR 5651 du CNRS, 98, Boulevard Édouard Herriot, BP 209, 06204 Nice cedex 3.

Ouvrage technique

Laborde et Mahous (1998).

Logiciel *Hyfran*

Description

Hyfran est un logiciel d'ajustement de lois statistiques, téléchargeable sur Internet, qui possède les fonctionnalités suivantes :

- vérification d'hypothèses : indépendance, homogénéité, présence de tendances ou de valeurs singulières, adéquation des lois théoriques aux données observées ;
- ajustement à 16 lois (exponentielle, GEV, Gumbel, normale, log-normale (deux ou trois paramètres), Weibull, gamma (ou Pearson, deux ou trois paramètres), gamma généralisée, gamma inverse, log-Pearson type 3, Halphen (types A, B et B-1) par trois méthodes (maximum de vraisemblance, moments, moments pondérés) ;

- calcul de quantiles avec intervalles de confiance (avec les quantiles de loi normale, ou par Bootstrap paramétrique) ;
- interfaçage avec tableurs.

Contact

Bobée B., Institut national de la recherche scientifique/Eau, Terre et Environnement, 2800 Einstein, CP 7500, Sainte-Foy (Québec), Canada G1V4C7.

Téléchargement

Sur le site web de l'INRS (rubrique logiciels), <http://www.ete.inrs.ca>

Logiciel *PRD'crue*

Description

PRD'crue fonctionne dans les environnements Mapinfo (écriture CETE Méditerranée) et Arc View (écriture Cemagref). Une version Microsoft Windows sera prochainement disponible. L'utilisation du logiciel est actuellement limitée aux services de l'État. *PRD'crue* a été écrit pour exploiter les informations des bases de données pluviométriques et hydrologiques générées par la méthode SHYPRE.

La base de données pluviométriques est pixellisée au km². Pour chacun des pixels, les quantiles de pluie de différentes durées entre 1 et 72 heures, de période de retour entre 2 et 100 ans, sont disponibles. La base des quantiles de débits — débits de pointe et débits moyens de différentes durées entre 1 et 72 heures, de période de retour entre 2 et 100 ans — est aussi définie au km². Ces débits ont été obtenus par une modélisation de la pluie en débit, en considérant chaque pixel de l'espace comme un bassin versant virtuel.

Ces bases sont actuellement disponibles pour les régions Corse, Provence-Alpes-Côte d'Azur et Languedoc-Roussillon.

PRD'crue propose trois menus :

- Menu pluie : cartographie des différents quantiles de pluie, courbes intensité-durée-fréquence des pluies pour un pixel donné.
- Menu débit : cartographie des différents quantiles de débits spécifiques (au km²), courbes débits-durées-fréquence pour un pixel donné.
- Menu bassin versant : *PRD'crue* active le logiciel Hydrokit (développé par la société Strategis à Montpellier) pour tracer les limites du bassin versant. Si Hydrokit n'est pas disponible, il est possible de tracer les limites du bassin versant à l'écran manuellement à partir des scans. *PRD'crue* permet alors d'éditer les informations de pluie et des débits spécifiques sur le bassin (pixels minimum, maximum, valeur moyenne), et agglomère l'estimation pixellisée pour proposer une estimation des débits, durée et fréquence à l'exutoire du bassin versant. La gamme d'utilisation est de 10 à 1 000 km².

Contact

Lavabre J., UR OHAX, Cemagref, BP 31, 13612 Aix-en-Provence cedex 1.

Ouvrage technique

Cemagref *et al.* (2002).

Logiciel *Prethy*

Description

Prethy est l'acronyme de programme d'échantillonnage et de traitements hydrologiques.

Ce logiciel permet à l'utilisateur :

- de visualiser les chroniques hydrologiques ;
- de critiquer les valeurs ;
- de créer les chroniques de débits moyens ou de débits seuil par moyenne mobile et minima mobiles sur différentes durées ;
- d'effectuer un échantillonnage de valeurs sur les données brutes ou transformées, selon deux modalités : échantillonnage par valeurs supérieures à un seuil ou par valeurs maximales annuelles pour ensuite réaliser un ajustement à deux types de loi (loi exponentielle et loi de Gumbel, selon type d'échantillonnage) sur une ou plusieurs variables, puis fournir les quantiles associés.

Prethy offre la possibilité de caler un modèle de synthèse QdF (débit Q , durée d , fréquence F) selon l'approche locale convergente ou l'approche par bassin de référence. *In fine*, le logiciel fournit des hydrogrammes de projet HSMF (Hydrogramme synthétique monofréquence), en lien avec les courbes QdF.

Le logiciel ne traite que les chroniques au format conventionnel HYDRO (pas de temps variable ou journalier) ou PLUVIO (pas de temps journalier).

Ergonomie (IHM) : il s'agit d'un logiciel avec interface de type Microsoft Windows.

Contact

Sauquet E., UR Hydrologie-Hydraulique, Cemagref Lyon, 3bis quai Chauveau, 69336 Lyon cedex 09.

Ouvrage technique

Galéa et Ramez (1995) ; Javelle *et al.* (2000) ; Sauquet *et al.* (2003).

Logiciel *Safarhy*

Description

Safarhy est un logiciel de calculs statistiques et d'analyse fréquentielle, adapté à l'évaluation du risque en hydrologie. Il comprend les fonctionnalités suivantes :

- gestion d'échantillons de valeurs observées (saisie d'échantillons, avec ou sans date, exportation et importation d'échantillons dans des fichiers textes) ;
- calcul des principales caractéristiques de l'échantillon (effectif, moments, etc.) ;
- établissement d'une table de conseils orientant l'utilisateur dans son choix de distributions théoriques pour un échantillon ;
- ajustement de lois de probabilités (lois discrètes : géométrique, binomiale négative tronquée ; continues : normale, racine carrée normale, log-normale (deux ou trois paramètres), exponentielle (un ou deux paramètres), Gumbel, log-Gumbel, Jenkinson (ou GEV), Weibull (deux ou trois paramètres), Gamma (ou Pearson) (deux ou trois paramètres), log-gamma (deux ou trois paramètres), loi des fuites), graphique des ajustements, avec des échelles

adaptées à la loi ajustée de manière à ce que la loi théorique soit une droite. L'ajustement peut être effectué sur un lot d'échantillons ;

- tests d'ajustement (tests du χ^2 et d'Anderson-Darling) ;
- évaluation des intervalles de confiance, représentés graphiquement ;
- estimation des probabilités de non-dépassement et de quantiles fixés, et inversement ;
- estimation des périodes de retour de quantiles fixés, et inversement ;

Contact

Raous P., IRD Montpellier, raous@ird.fr

Ouvrage technique

Manuel d'utilisation de Safarhy ; Lubes *et al.* (1994).

Logiciel *Winfap-Feh*

Description

Winfap-Feh est un logiciel dédié au traitement statistique des données hydrologiques (huit distributions possibles) par une approche locale ou régionale. Les procédures s'appuient sur des échantillons de maxima annuels mais également des échantillons de valeurs supérieures à un seuil (possibilité réduite de traitement). Il permet l'exploitation de la base nationale de données hydrologiques (installation simultanée). *Winfap-Feh* est un utilitaire qui reprend les méthodes décrites dans le volume 3 du *Flood Estimation Handbook* (FEH), et qui permet le calcul de quantiles avec intervalles de confiance. Il offre la possibilité de former des groupes de stations, d'en étudier les propriétés statistiques (un regroupement forme-t-il un ensemble homogène ?) sur la base des maxima annuels à des fins de régionalisation (estimation en site non jaugé). Il propose de stocker des données autres qu'anglaises (entrées manuellement, extraites d'un fichier), d'afficher les statistiques du jeu de données traité, d'ajuster une distribution aux séries de maxima annuels (en précisant les intervalles de confiance), et de représenter les traitements sous forme de tableaux et de graphiques. Les mêmes options sont offertes pour le traitement d'un groupe de stations.

Le logiciel fonctionne sous environnement Microsoft Windows.

Contact

CEH Institute of Hydrology, Wallingford, Oxfordshire, OX10 8BB, UK,
<http://www.nwl.ac.uk/ih/feh/>

Ouvrage technique

Reed *et al.* (1999).

Annexe 2

Éléments de probabilités et statistiques appliquées à l'hydrologie¹

Introduction

Cette annexe présente des éléments de compréhension sur les probabilités et statistiques utilisées en hydrologie. Il reprend pour une grande part le cours d'hydrologie statistique dispensé dans le cadre du master professionnel « Mécanique appliquée au transport et à l'environnement » de l'université Lyon I, et à l'ENTPE. Elle a été complétée d'éléments techniques mentionnés dans le corps du présent *Guide méthodologique pour l'estimation de la crue centennale*.

Elle est décomposée en quatre parties :

- Dans la première partie, nous donnerons quelques résultats et techniques permettant de résumer ou d'extraire de l'information à partir d'un échantillon de données. L'ensemble de ces méthodes se rapporte au domaine des statistiques descriptives.
- La théorie des probabilités fournit un ensemble d'outils pour quantifier le hasard. Nous allons en fait étudier une version « allégée » de la théorie probabiliste qui nécessite pour être abordée en toute rigueur des notions mathématiques assez sophistiquées.
- Ces deux premières parties seront ensuite utilisées conjointement dans le cadre de la statistique inférentielle qui permet de quantifier l'influence du hasard dans les processus générateurs de données.
- Enfin, nous étudierons l'application de ces méthodes à l'analyse fréquentielle des crues, qui constitue un des outils de base de l'hydrologue.

Aucun prérequis particulier n'est nécessaire pour comprendre ce cours ; les notions mathématiques utilisées devraient vous être plus ou moins familières.

Les ouvrages et sites suivants constituent d'excellentes sources d'information :

- Saporta, 1990 ; Wonnacott *et al.*, 1991.
- <http://www.math-info.univ-paris5.fr/smel/> ;
- <http://www.agro-montpellier.fr/cnam-lr/statnet/>

1. Annexe rédigée par Benjamin Renard et Michel Lang, non examinée par le comité de relecture.

- <http://www.hds.utc.fr/~ggovaert/sy02/documents/poly.pdf> (cours en ligne).
- <http://www.lsp.ups-tlse.fr/Besse/pub/sdm1.pdf> (statistiques descriptives).

Statistiques descriptives

Dans cette section, nous allons nous intéresser à la description d'un ensemble de données, en résumant l'information qu'il contient grâce à quelques grandeurs caractéristiques.

Quelques définitions

Le plus simple est de considérer l'exemple suivant : on souhaite étudier la taille des Français. On dispose pour cela de valeurs observées sur 35 personnes, que nous noterons (x_1, \dots, x_{35}) . Les Français constituent la **population** à laquelle on s'intéresse, via un **échantillon d'effectif 35**. Chaque personne de cet échantillon est un **individu** sur lequel on a mesuré la **variable quantitative continue** « taille ». Chacun de ces termes peut être utilisé dans toutes les situations que nous rencontrerons, il ne faut donc pas s'attacher à leur signification « démographique ». Ainsi, si on étudie la taille des gouttes d'eau à la sortie d'un mécanisme d'irrigation (<http://afeid.montpellier.cemagref.fr/Montreal/Molle.PDF>), alors l'ensemble des gouttes constitue notre population, celles ayant transité par l'appareil de mesure (un spectro-pluviomètre optique) étant les individus de notre échantillon sur lequel on a relevé la variable quantitative continue « diamètre de la goutte ».

Remarque : la constitution d'un échantillon est la première étape de toute analyse statistique ; c'est également l'une des plus importantes, et parfois l'une des plus difficiles. Il faut en particulier s'assurer de la représentativité de l'échantillon, et rester attentif aux biais pouvant être « cachés » dans les données (la taille de la goutte est sûrement fonction de la distance à l'asperseur).

Il est possible de mesurer différents types de variables sur les individus :

- **Variables quantitatives continues**, comme dans les exemples précédents. Il s'agit de variables dont les valeurs varient théoriquement de manière continue (entre deux valeurs, il en existe toujours une troisième possible). Dans la pratique, la continuité est souvent limitée par la précision des instruments de mesure.
- **Variables quantitatives discrètes**, dont les valeurs possibles sont discontinues (les données issues de comptage, par exemple).
- **Variables qualitatives ordonnées**, comme l'intensité d'une douleur, par exemple : faible, moyenne, forte.
- **Variables qualitatives non ordonnées** telles que, dans notre exemple, la couleur des cheveux de notre échantillon de 35 Français.

Dans la suite de cette annexe, nous nous intéresserons principalement aux variables quantitatives, continues ou non. Nous noterons (x_1, \dots, x_n) les valeurs de telles variables observées sur un échantillon de taille n .

Comment résumer un échantillon ?

Représentations graphiques d'un échantillon

Il existe une infinité de façons de représenter graphiquement un jeu de données : « camemberts », diagrammes en bâtons, courbes, nuages de points... Tout dépend de la structure des données mais également de l'information que l'on souhaite faire passer. Nous allons nous intéresser ici à deux représentations fondamentales pour les données quantitatives : l'**histogramme** et la **courbe des fréquences cumulées**.

L'histogramme consiste à représenter la **fréquence** (ou parfois l'effectif) de chaque valeur présente dans l'échantillon. Notre variable étant supposée ordonnée, il est possible de cumuler ces fréquences par ordre croissant des valeurs rencontrées : on peut ainsi tracer la courbe des fréquences cumulées, qui représente à chaque abscisse la proportion de valeurs de l'échantillon inférieures ou égales à cette abscisse.

Ces deux représentations sont primordiales, car elles peuvent être reliées aux notions probabilistes de densité et de fonction de répartition dont nous parlerons plus loin.

Exemple 1 – Nombre de crues supérieures à 24,7 m³/s pour L'Ubaye à Barcelonnette.

Les individus sont ici des années, et la variable « nombre de crues » est quantitative discrète.

Les calculs des effectifs, des fréquences et des fréquences cumulées sont consignés dans le *tableau A2.2*.

Tableau A2.1. Tableau de données.

Année	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
Nombre	1	1	3	1	5	2	1	3	3	5	0	3	5	2

Tableau A2.2. Effectifs, fréquences et fréquences cumulées.

Valeur	Effectif	Fréquence	Fréquence cumulée
0	1	0,07	0,07
1	4	0,29	0,36
2	2	0,14	0,5
3	4	0,29	0,79
4	0	0,00	0,79
5	3	0,21	1

Pour des données quantitatives continues, il y a de fortes chances pour que chaque valeur n'apparaisse qu'une seule fois. Cela ne pose pas de problème pour le tracé de la courbe des fréquences cumulées ; en revanche, un histogramme où toutes les ordonnées seraient égales à 1 ne serait pas très informatif. Pour y remédier, on effectue un regroupement des

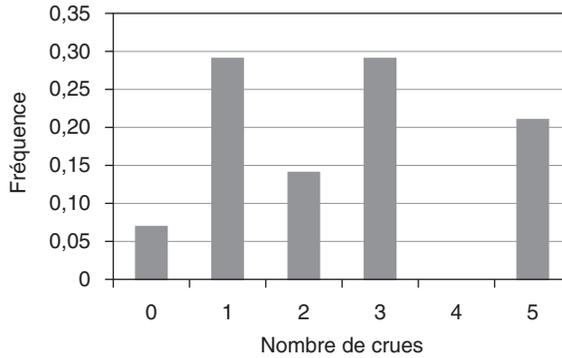


Figure A2.1. Histogramme des fréquences.

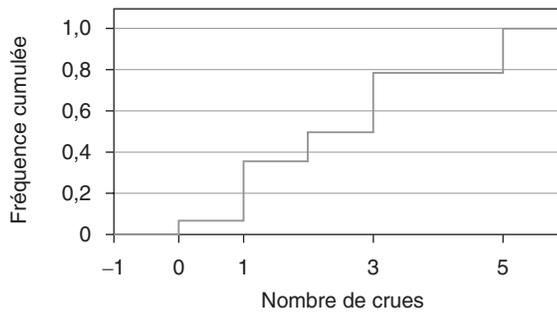


Figure A2.2. Courbe des fréquences cumulées.

individus en classes, ce qui revient à « discrétiser » notre variable continue. Le choix des classes reste arbitraire ; on essaiera en général de créer entre 5 et 15 classes de même étendue.

Exemple 2 – Lors des 14 dernières crues de La Zorn à Waltenheim, les débits journaliers maximaux ont été de 28,9, 45,8, 67,6, 60,8, 53,6, 33,5, 49,9, 58,1, 35,9, 33,3, 28,4, 28,3, 49,5 et 25,9 m³/s. L’histogramme des fréquences pour des classes d’étendue de 10 m³/s, en partant de 25 m³/s, est représenté en figure A2.3.

La différence de représentation par rapport à l’histogramme de la figure A2.1 est à noter : des rectangles pleins ont été représentés pour bien signifier qu’il s’agit de la fréquence d’une classe, et non d’une valeur ponctuelle discrète. Pour être tout à fait précis, seul ce dernier graphique est un histogramme, le terme de « diagramme en bâtons » étant plutôt utilisé pour le cas discret.

Les indicateurs de localisation

La moyenne, que nous noterons \bar{x} :
$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$$

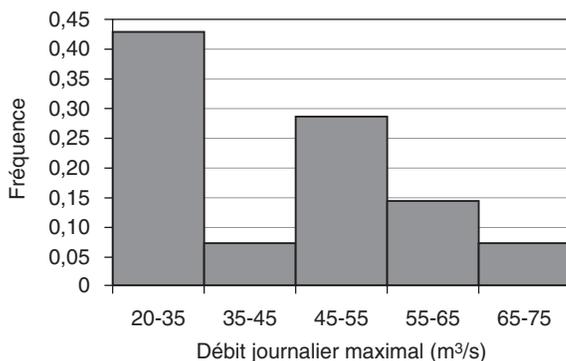


Figure A2.3. Histogramme des fréquences.

La médiane : c'est la valeur qui sépare l'échantillon ordonné en deux sous-parties de même effectif. Par exemple, la médiane des valeurs 10, 15, 18, 19, 20 vaut 18. Si le nombre de valeurs est pair, on calculera la demi-somme entre les deux valeurs du milieu, soit une médiane de 16,5 pour l'échantillon des quatre premières valeurs ci-avant. La médiane peut également être lue sur la courbe des fréquences cumulées ; il s'agit de la valeur correspondant à une fréquence cumulée de 0,5.

L'exemple classique pour bien comprendre la différence entre ces deux indicateurs est le salaire des Français : la moyenne de ce dernier vaut environ 1 750 €, sa médiane 1 400 €... (Salaires nets en 2001 dans le secteur privé, source Insee).

Un troisième indicateur est parfois utilisé, surtout pour les données discrètes ; il s'agit du **mode** qui est la valeur — pas forcément unique — la plus fréquente dans l'échantillon. Pour des données quantitatives continues, il faut (comme pour l'histogramme) procéder à un regroupement des individus. L'inconvénient est que le mode devient dépendant du regroupement arbitraire effectué.

Il est également fréquent de s'intéresser aux valeurs extrêmes d'un échantillon. On utilisera pour cela le **quantile** d'ordre p qui est la valeur de l'échantillon dont la fréquence cumulée vaut p . Le quantile peut donc être lu sur la courbe des fréquences cumulées, en faisant éventuellement une interpolation linéaire entre deux points. La médiane est ainsi le quantile d'ordre 0,5.

Exemple 3 – Pour les données de l'exemple 1, la moyenne vaut 2,5, la médiane 2 et les modes (distribution bimodale) 2 et 4. Pour l'exemple 2, le quantile d'ordre 0,9 peut se déduire de la figure A2.4, soit ici une valeur d'environ 60 m³/s.

Les indicateurs de dispersion

La **variance** permet de mesurer la façon dont les données se dispersent autour de la moyenne :

$$\text{Var} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2.$$

On trouve parfois, notamment dans les calculettes ou les tableurs de type Microsoft Excel, une autre définition qui n'est pas recommandée pour les petits échantillons :

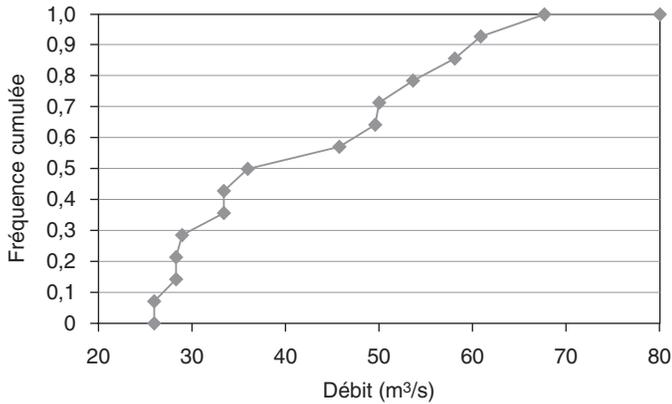


Figure A2.4. Courbe des fréquences cumulées.

$$\text{Var}^* = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2.$$

Ces deux définitions deviennent équivalentes lorsque n est grand. Nous verrons ultérieurement la raison pour laquelle ces deux définitions coexistent.

On définit également l'**écart-type** qui a l'avantage d'avoir la même dimension que les données :

$$\sigma = \sqrt{\text{Var}} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}.$$

Le **coefficient de variation** est également utilisé pour comparer la variabilité de plusieurs séries de données dont les ordres de grandeurs ne sont pas comparables :

$$CV = \frac{\sigma}{\bar{x}}.$$

Il n'est évidemment pas défini pour les données dont la moyenne est nulle.

Autres indicateurs

Le **moment d'ordre k** est défini de la manière suivante :

$$m_k = \frac{1}{n} \sum_{i=1}^n x_i^k.$$

Pour $k = 1$, on reconnaît la moyenne. Les moments n'ont pas de signification aussi claire que les autres indicateurs, mais ils seront centraux lorsque nous nous intéresserons à la théorie de l'estimation.

On définit de même le **moment centré d'ordre k** :

$$m'_k = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^k.$$

On peut déduire de ces moments centrés deux indicateurs relatifs à la forme de l'histogramme :

– asymétrie : $\beta_1 = \frac{m'_3}{\sigma^3}$;

– aplatissement : $\beta_2 = \frac{m'_4}{\sigma^4}$.

Étude de deux variables

Plaçons-nous dans la situation suivante : sur un échantillon, nous avons mesuré deux variables quantitatives que nous noterons X et Y . Nous obtenons donc deux séries de valeurs $\mathbf{x} = (x_1, \dots, x_n)$ et $\mathbf{y} = (y_1, \dots, y_n)$. Remarquez les notations que nous conserverons tout au long du cours : les variables, que nous appellerons plus tard variables aléatoires, sont en majuscules, et les observations de ces variables sur l'échantillon sont en minuscules (et en gras pour une notation vectorielle). Les deux séries peuvent bien entendu être décrites indépendamment l'une de l'autre (cf. § *Comment résumer un échantillon ?*), mais nous allons ici nous intéresser à la liaison qui peut exister entre ces variables.

La régression linéaire

Exemple 4 – Sur un échantillon de 10 petits cours d'eau, on a calculé la superficie du bassin versant (X), ainsi que la crue décennale (Y) (tableau A2.3).

Tableau A2.3. Tableau de données.

BV	Superficie	$Q(10)$	BV	Superficie	$Q(10)$
1	16,2	5,77	6	62,1	25,31
2	21,4	5,37	7	64	20,18
3	38,7	7,29	8	68	39,27
4	42,3	12,66	9	73	39,59
5	55,8	11,09	10	94	28,30

La manière la plus simple de représenter ces données est de projeter les couples (x_i, y_i) dans le plan : on obtient ainsi un **nuage de points** (figure A2.5).

En observant la figure A2.5, il apparaît que les deux variables ont tendance à évoluer conjointement : les « grands » bassins versants ont des crues décennales plus importantes. C'est cette coévolution que nous allons tenter d'étudier.

Commençons par définir la **covariance** entre \mathbf{x} et \mathbf{y} :

$$\text{cov}(x, y) = \frac{1}{n-1} \sum_{i=1}^n [(x_i - \bar{x})(y_i - \bar{y})].$$

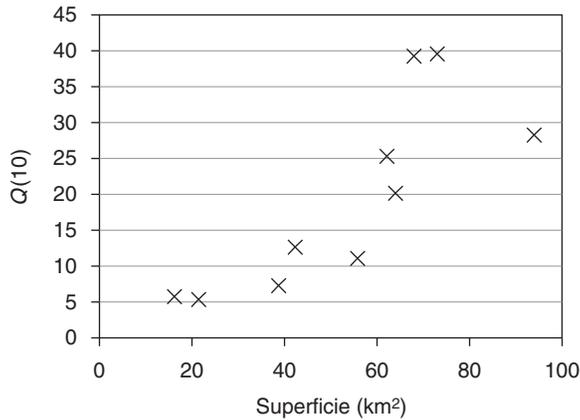


Figure A2.5. Nuage de points.

Cette quantité mesure la manière dont X et Y varient « ensemble ». Comme pour la variance, on trouve parfois une définition alternative avec $\frac{1}{n}$. Le **coefficient de corrélation linéaire** est directement dérivé de la covariance :

$$r = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y}.$$

Il s'agit d'un coefficient adimensionnel compris entre -1 et 1 : une valeur absolue proche de 1 sera la signature de deux variables liées linéairement, tandis qu'une valeur proche de zéro signifiera l'absence de relation linéaire. Il est important de noter que cela n'interdit pas que les variables soient liées par un autre type de relation (polynomiale, sinusoïdale...). Dans le cadre de l'exemple 4, nous obtenons les valeurs suivantes :

$$\sigma_x = 24, \quad \sigma_y = 13,17, \quad \text{cov}(x, y) = 255,01 \quad \text{et} \quad r = 0,81.$$

Ces chiffres confirment la liaison entre les variables. Nous pouvons essayer d'aller plus loin, en cherchant la droite la plus pertinente qui s'ajusterait au nuage de point, c.-à-d. en évaluant une relation du type $Y = aX + b$. Bien entendu, cette relation ne peut pas être parfaite (tous les points ne sont pas alignés) ; nous introduisons donc des termes d'erreurs, ce qui nous donne la relation :

$$y_i = ax_i + b + e_i, \quad \forall i = 1, \dots, n.$$

Une « bonne » droite permettrait de minimiser ces erreurs. Nous allons donc définir un critère, dit des moindres carrés, construit à partir de la somme des carrés des erreurs :

$$S = \sum_{i=1}^n e_i^2.$$

Il s'agit donc à présent de trouver les valeurs de a et b minimisant S .

Exercice

- 1) Exprimez S en fonction de a et b .
- 2) Calculez les valeurs de a et b minimisant S , que l'on notera \hat{a} et \hat{b} .
- 3) Vérifiez qu'avec ces valeurs la somme — donc la moyenne — des erreurs est nulle.

4) Démontrez que $\text{Var}(y) = \text{Var}(\hat{y}) + \text{Var}(e)$, où $\hat{y}_i = \hat{a}x_i + \hat{b}$ est la valeur du débit décennal estimée par notre modèle.

5) On définit un critère pour mesurer la qualité de notre modèle, égal à $c = 1 - \frac{\text{Var}(e)}{\text{Var}(y)}$. Interprétez cet indicateur et démontrez que $c = r^2$.

La droite de régression, dite des moindres carrés, est donc définie par les paramètres \hat{a} et \hat{b} suivants :

$$\hat{a} = \frac{\text{cov}(X, Y)}{\text{Var}(X)} = r \frac{\sigma_Y}{\sigma_X},$$

$$\hat{b} = \bar{y} - \hat{a}\bar{x}.$$

La qualité de cet ajustement est mesurée par r^2 qui mesure la part de variance expliquée par notre modèle linéaire, et qui varie dans $[0 ; 1]$. L'ajustement pour les données précédentes est donné en *figure A2.6*.

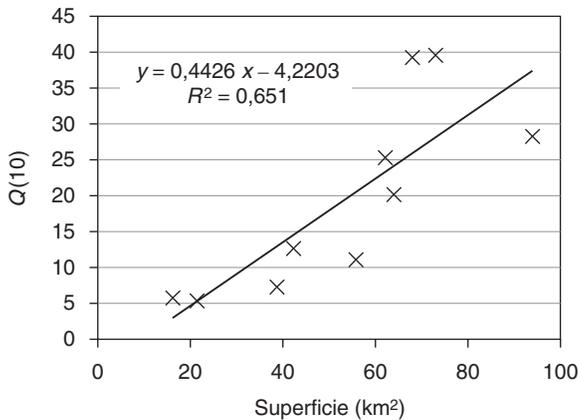


Figure A2.6. Ajustement de la droite de régression.

La démarche présentée dans ce cadre de recherche de relation linéaire entre variables peut être généralisée à toute forme de dépendance du type $Y = f(X)$:

- choix d'une fonction f ;
- calcul des paramètres optimaux de cette fonction, au sens d'un certain critère (méthode des moindres carrés, par exemple) ;
- évaluation de la qualité de l'ajustement par un indicateur du type $1 - \frac{\text{Var}(e)}{\text{Var}(y)}$ (qui n'est plus égal à r^2 si f n'est pas affine).

Bien entendu, la plupart des calculs se compliquent par rapport au cas linéaire, le recours à des méthodes d'optimisation numérique étant souvent indispensable.

Une autre extension de la méthode consiste à intégrer plusieurs variables pour en « expliquer » une autre : dans le cas précédent, il pourrait ainsi être bénéfique d'intégrer la pluviométrie ou la nature et l'occupation du sol pour améliorer l'explication du débit décennal. En guise d'exemple, la méthode dite Crupedix est présentée dans le chapitre 4 ;

il s'agit d'une régression non linéaire multiple visant à estimer le débit décennal sur des sites non jaugés.

Représentation fréquentielle d'un couple de variables

Nous allons à présent généraliser la notion d'histogramme au cas de deux variables, et en profiter pour proposer quelques définitions. Nous nous appuyerons sur un exemple utilisant des données quantitatives discrètes ; pour des données continues, le même type de représentation est possible moyennant un regroupement des couples de valeurs en classes.

Exemple 5 – On se propose d'étudier le nombre annuel de crues entre les années 1958 et 2002 pour deux stations S_1 et S_2 . Les données sont fournies dans le *tableau A2.4*.

Tableau A2.4. Tableau de données.

Année	N_1 : nombre de crues en S_1	N_2 : nombre de crues en S_2
1958	2	1
1959	1	3
...
2002	0	4

Il est fréquent de représenter ce type de données bivariées dans un **tableau de contingence** qui permet de dénombrer l'effectif ou la fréquence de chaque couple de valeurs possible. Dans le cadre de l'exemple 5, nous obtenons le *tableau A2.5* de contingence (en fréquence).

Tableau A2.5. Tableau de contingence.

$N_2 \backslash N_1$	0	1	2	3	4	5	6	Cumul
0	0,044	0,022	0,044	0,022	0,022	0,000	0,000	0,156
1	0,044	0,067	0,022	0,022	0,022	0,000	0,000	0,178
2	0,044	0,089	0,111	0,000	0,000	0,000	0,000	0,244
3	0,000	0,000	0,067	0,022	0,044	0,000	0,000	0,133
4	0,022	0,044	0,022	0,044	0,044	0,000	0,000	0,178
5	0,000	0,022	0,000	0,044	0,000	0,000	0,022	0,089
6	0,000	0,022	0,000	0,000	0,000	0,000	0,000	0,022
Cumul	0,156	0,267	0,267	0,156	0,133	0,000	0,022	1,000

Le *tableau A2.5* se lit de la manière suivante : la cellule grisée nous informe que pour 2,2 % des années étudiées, on a observé deux crues en S_1 et une crue en S_2 . La dernière colonne et la dernière ligne contiennent les cumuls de fréquence respectivement par ligne et par colonne. Enfin, la dernière case, en bas à droite, contient le « cumul des cumuls » qui vaut 1... Fort heureusement pour nous ! Le tableau équivalent en effectif peut se déduire aisément en multipliant chaque case par l'effectif de l'échantillon (ici, 45 années).

Intéressons-nous à une ligne en particulier ; la première, par exemple : y figurent les fréquences associées aux différentes valeurs de N_1 , dans le cas particulier où $N_2 = 0$. Nous pouvons identifier cette première ligne à la **distribution conditionnelle de N_1 , sachant que $N_2 = 0$** . Cette identification est abusive ; il faudrait en fait diviser toutes les valeurs par le cumul de leur fréquence (0,156), qui représente la fréquence de la valeur 0 pour N_2 , en considérant toutes les associations possibles avec N_1 . En raisonnant de même ligne par ligne, on voit que la dernière colonne représente les fréquences des différentes valeurs prises par N_2 compte tenu des différentes associations observées avec N_1 : nous parlerons de la **distribution marginale** de la variable N_2 . De même, la dernière ligne est la distribution marginale de N_1 . Le tableau de contingence représente quant à lui la **distribution jointe** des variables N_1 et N_2 . Il est possible de représenter cette distribution par un histogramme en trois dimensions (*figure A2.7*).

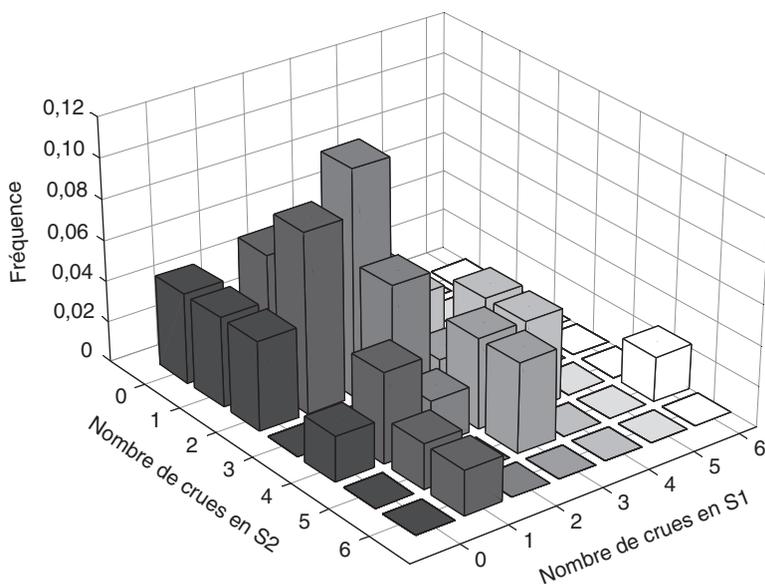


Figure A2.7. Histogramme multidimensionnel.

Sur la *figure A2.7*, l'histogramme représenté en noir est proportionnel à la distribution de N_2 , conditionnellement à $N_1 = 0$.

Le modèle probabiliste

Nous avons exposé dans la section *Statistiques descriptives* un certain nombre de méthodes permettant de décrire une série d'observations et d'en extraire de l'information. La théorie des probabilités fournit un cadre pour représenter de manière plus abstraite le comportement de variables « génératrices » de données. La puissance de cette théorie tient à sa généralité ; cependant, la définition même du terme « probabilité » faisant appel à des notions mathématiques trop élaborées pour être présentées dans cette annexe, nous allons y étudier une version « allégée » de la théorie probabiliste, qui sera cependant valable dans l'immense majorité des cas rencontrés.

Quelques rappels de théorie des ensembles

Sans rentrer dans les détails, nous définirons un ensemble E comme une collection d'objets bien définis. *Infra*, figurent quelques définitions et notations qui nous seront utiles par la suite :

- $x \in E$ signifie que l'élément x **appartient** à E ;
- \emptyset est l'**ensemble vide** ;
- $A \subset E$ signifie que l'ensemble A est **inclus** dans E : $x \in A \Rightarrow x \in E$;
- $A \cup B$ est la **réunion** des ensembles A et B : $x \in A \cup B \Rightarrow x \in A$ ou $x \in B$;
- $A \cap B$ est l'**intersection** des ensembles A et B : $x \in A \cap B \Rightarrow x \in A$ et $x \in B$;
- si $A \cap B = \emptyset$, on dit que les ensembles A et B sont **disjoints** ;
- si $A \cap B = \emptyset$ et $A \cup B = E$, on dit que A et B forment une **partition** de l'ensemble E .

Il en est de même si $A_i \cap A_j = \emptyset, \forall i \neq j$ et $\bigcup_i A_i = E$;

- C_A ou A^c ou \bar{A} est le **complémentaire** de A : $x \in A^c \Leftrightarrow x \notin A$;
- $\wp(E)$ est l'**ensemble des parties** de E , c.-à-d. l'ensemble de tous les sous-ensembles de E : $A \in \wp(E) \Leftrightarrow A \subset E$.

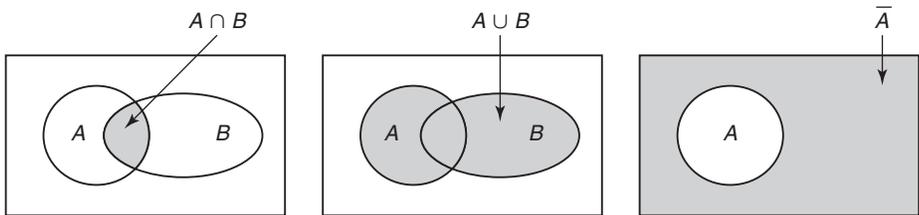


Figure A2.8. Réunion, intersection et complémentaire.

Premières définitions : probabilité et variables aléatoires.

Une **probabilité** sur un ensemble Ω est une fonction $P : \wp(\Omega) \mapsto [0 ; 1]$ qui vérifie les trois propriétés suivantes :

- $P(A) \geq 0, \forall A \subset \Omega$;
- $P(\Omega) = 1$;

– soit (A_i) une suite de sous-ensembles de Ω deux à deux disjoints, c.-à-d. $A_i \cap A_j = \emptyset$,

$$\forall i \neq j. \text{ Alors } P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

Ces trois axiomes conduisent aux propriétés suivantes :

- $P(A) \in [0 ; 1], \forall A \subset \Omega$;
- $P(\emptyset) = 0$;
- $A \subset B \Rightarrow P(A) \leq P(B)$;
- $P(A^c) = 1 - P(A)$;
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Traduction intuitive : supposons que nous ayons réalisé une **expérience aléatoire**, ou effectué des mesures sur un ensemble d'individus. Ω représente l'ensemble des issues possibles de l'expérience ; il est appelé **univers** ou **ensemble fondamental**, tandis qu'un sous-ensemble A de Ω est appelé un **événement**. Une probabilité est ainsi une fonction permettant de mesurer la « vraisemblance » d'un événement.

Remarque : la définition donnée *supra* est abusive, car il peut exister (cas pathologiques) des sous-ensembles de Ω pour lesquels la probabilité n'est pas définie. Il faut alors définir un ensemble d'événements mesurables, ce qui requiert des outils mathématiques trop sophistiqués pour être évoqués dans cette annexe.

Une **variable aléatoire réelle** est une fonction $X : \Omega \mapsto]-\infty ; +\infty[$. Elle permet de traduire un événement en nombre réel. Supposons par exemple que nous lançons deux dés, et que nous nous intéressons à la somme des deux faces observées. L'ensemble Ω est ici égal à l'ensemble des couples (a, b) , où a et b sont des entiers entre 1 et 6. X sera la fonction somme :

$$X : (1, \dots, 6) \times (1, \dots, 6) \mapsto (2, \dots, 12)$$

$$(a, b) \mapsto a + b.$$

Ainsi, l'événement « la somme des deux dés vaut 4 » sera noté $X = 4$. Il s'agit d'une notation simplifiée, car, en toute rigueur, nous nous intéressons à l'ensemble des couples (a, b) pour lesquels $a + b = 4$, soit $X^{-1}(\{4\}) = \{(1,3), (3,1), (2,2)\}$. De même, $X < 6$ représentera l'événement « La somme des deux dés est strictement inférieure à 6 ». Il est possible de mesurer la probabilité de tels événements, que nous noterons $P(X = 4)$ ou $P(X < 6)$. On voit dans ce cas que X ne prend pas ses valeurs dans l'ensemble de tous les nombres réels, mais seulement dans un sous-ensemble de l'ensemble des entiers : il s'agit d'une **variable aléatoire discrète**.

En des termes plus intuitifs, la variable aléatoire X représente tout simplement la variable que nous souhaitons étudier. Étant donné que cette variable est fréquemment mesurée sur des individus, sa définition est en général très naturelle, comme par exemple « Nombre de crues observées en une année » pour l'exemple 1. Il est en revanche important de bien faire la différence entre des observations (x_1, \dots, x_n) , qui sont des valeurs numériques mesurées, et la variable aléatoire X qui génère ces données, qui est un objet abstrait.

Loi de probabilité d'une variable aléatoire

Variable aléatoire discrète

Soit X une variable aléatoire discrète à valeurs dans un ensemble dénombrable ordonné $\{x_i, i = 1, 2, \dots\}$.

La **loi de probabilité** de X est la donnée des quantités $f(x_i) = P(X = x_i), \forall i = 1, 2, \dots$. Cette loi de probabilité peut être décrite par un diagramme en bâtons (figure A2.9).

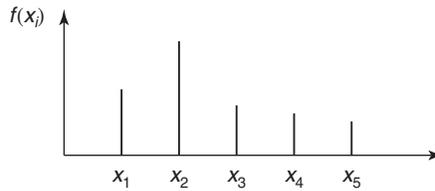


Figure A2.9. Loi de probabilité discrète.

Il existe évidemment une analogie entre cette représentation et celle exposée à la section *Statistiques descriptives*, où nous avons reporté les fréquences observées en ordonnées : nous aurons l'occasion de revenir à plusieurs reprises sur cette analogie fréquence/probabilité.

Nous définissons également la **fonction de répartition** F de la variable aléatoire X de la manière suivante : $F(x_i) = P(X \leq x_i)$ (figure A2.10).

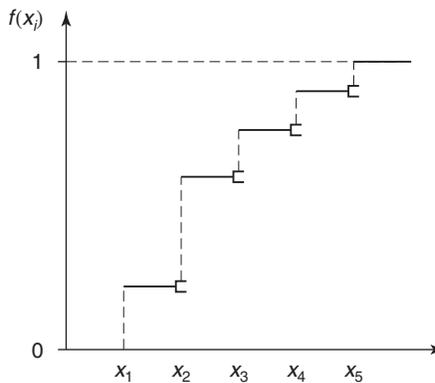


Figure A2.10. Fonction de répartition discrète.

Il est aisé de démontrer la relation suivante :

$$F(x_i) = \sum_{k=1}^i f(x_k).$$

En corollaire, F est une fonction croissante qui part de 0 pour tendre vers 1 en l'infini.

De ces deux définitions découlent quelques grandeurs caractéristiques de la variable aléatoire X (notez encore une fois l'analogie avec la section *Statistiques descriptives*) :

- l'**espérance** $E(X) = \sum_{i=1}^{\infty} x_i f(x_i)$;
- la **variance** $\text{Var}(X) = \sum_{i=1}^{\infty} (x_i - E(X))^2 f(x_i)$ et l'**écart-type** $\sigma(X) = \sqrt{\text{Var}(X)}$;
- les **moments d'ordre k** $\mu_k = \sum_{i=1}^{\infty} x_i^k f(x_i)$.

Si $k = 1$, on reconnaît l'espérance ;

- les **moments centrés d'ordre k** $\mu'_k = \sum_{i=1}^{\infty} (x_i - E(X))^k f(x_i)$.

Si $k = 2$, on reconnaît la variance ;

- le **quantile d'ordre p** , noté x_p , qui vérifie $F(x_p) = p$ ($\Leftrightarrow x_p = F^{-1}(p)$).

Remarque : les quantités faisant intervenir des sommes infinies peuvent ne pas exister.

Généralisation au cas continu

Cette fois, nous allons commencer par définir la loi de probabilité d'une variable aléatoire continue X comme la donnée d'une **fonction de répartition** F , telle que $F(x) = P(X \leq x)$. La **densité de probabilité** de X sera alors la fonction f telle que

$F(x) = \int_{-\infty}^x f(t)dt$. En d'autres termes, f est la dérivée de la fonction de répartition F (figure A2.11).

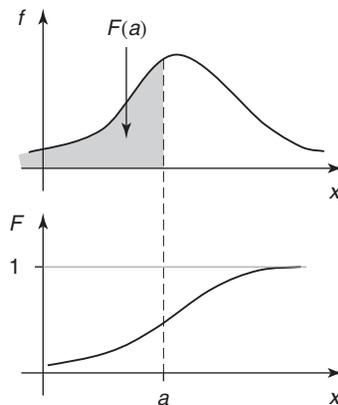


Figure A2.11. Densité et fonction de répartition continues.

Propriété : si f est continue, alors $P(a < X \leq b) = \int_a^b f(t)dt = F(b) - F(a)$.

Dans le cas continu, on voit ainsi apparaître une analogie entre la probabilité d'un événement et l'aire sous la courbe de la densité. Cette analogie implique en particulier que $P(X = x_0) = 0, \forall x_0 \in \mathbb{R}$. C'est la raison pour laquelle il n'est pas possible de définir la loi de probabilité directement à partir des probabilités des éléments de Ω , ce qui conduit à raisonner sur des intervalles. On remarque encore une fois l'analogie avec les histogrammes dans le cas continu de la section *Statistiques descriptives*, où nous nous retrouvons dans l'obligation de procéder à des regroupements.

À partir de la densité de probabilité, il est possible de définir les mêmes grandeurs caractéristiques que dans le cas discret, et ce, en remplaçant les sommes par des intégrales :

– l'**espérance** $E(X) = \int_{-\infty}^{+\infty} xf(x)dx$;

– la **variance** $\text{Var}(X) = \int_{-\infty}^{+\infty} (x - E(X))^2 f(x)dx$ et l'**écart type** $\sigma(X) = \sqrt{\text{Var}(X)}$;

– les **moments d'ordre k** $\mu_k = \int_{-\infty}^{+\infty} x^k f(x)dx$.

Si $k = 1$, on reconnaît l'espérance ;

– les **moments centrés d'ordre k** $\mu'_k = \int_{-\infty}^{+\infty} (x - E(X))^k f(x)dx$.

Si $k = 2$, on reconnaît la variance ;

– le **quantile d'ordre p** , noté x_p , qui vérifie $F(x_p) = p$ ($\Leftrightarrow x_p = F^{-1}(p)$).

Remarque : encore une fois, rien ne garantit l'existence des intégrales infinies. De plus, certaines lois n'étant pas définies sur \mathbb{R} tout entier, ces intégrales doivent alors être réduites aux supports de ces lois.

Quelques lois de probabilités classiques

Nous donnons *infra* quelques lois de probabilités qui sont régulièrement utilisées, notamment en hydrologie. Ces lois dépendent d'un certain nombre de **paramètres** afin de pouvoir s'adapter à une plus grande variété de situations.

Lois discrètes

■ Loi de Bernoulli $Be(p)$

Elle est utilisée pour des variables aléatoires binaires, de type succès-échec ou présence-absence, qui prendront donc pour valeurs 0 ou 1.

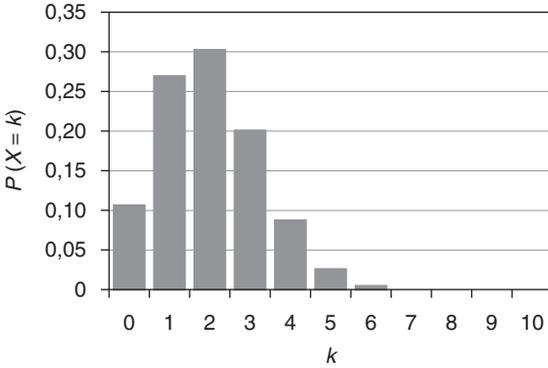
$$P(X = 1) = p, \quad P(X = 0) = 1 - p,$$

$$E(X) = p,$$

$$\text{Var}(X) = p(1 - p).$$

■ Loi binomiale $B(n, p)$

Si l'on répète n fois indépendamment une épreuve de type Bernoulli, alors la variable aléatoire X mesurant le nombre de succès suit une loi binomiale.



$$P(X = k) = C_n^k p^k (1 - p)^{n - k},$$

$$\text{avec : } C_n^k = \frac{n!}{k!(n - k)!},$$

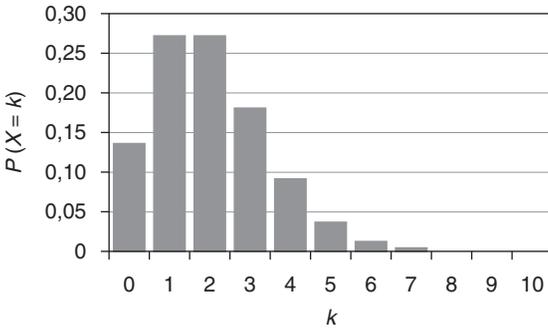
$$E(X) = np,$$

$$\text{Var}(X) = np(1 - p).$$

Figure A2.12. Loi binomiale $B(10 ; 0,2)$.

■ Loi de Poisson $P(\mu)$

Cette loi est fréquemment utilisée pour décrire des données issues de comptages d'événements rares (processus d'occurrences de crues, par exemple), et respectant quelques propriétés supplémentaires (stationnarité, indépendance, absence de regroupements).



$$P(X = k) = e^{-\mu} \frac{\mu^k}{k!},$$

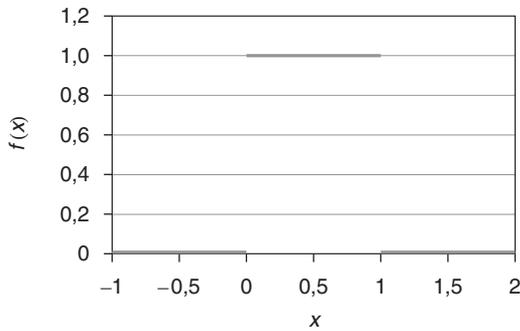
$$E(X) = \mu,$$

$$\text{Var}(X) = \mu.$$

Figure A2.13. Loi de Poisson(2).

Lois continues

■ *Loi uniforme* $U(a, b)$



$$f(x) = \begin{cases} \frac{1}{b-a} & \text{si } x \in [a, b], \\ 0 & \text{sinon} \end{cases}$$

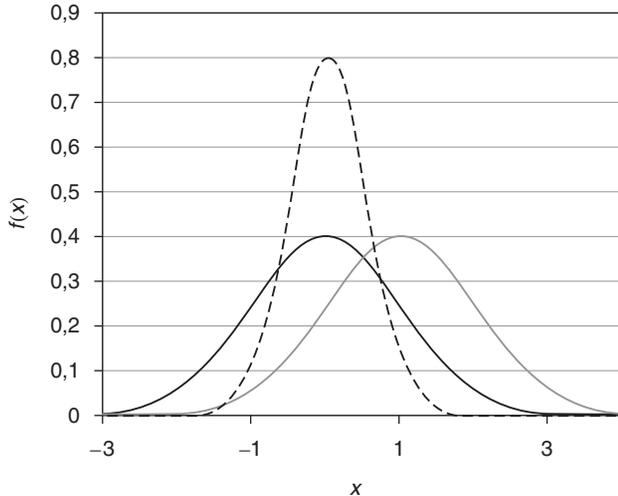
$$E(X) = \frac{a+b}{2},$$

$$\text{Var}(X) = \frac{(b-a)^2}{12}.$$

Figure A2.14. Loi uniforme $U(0;1)$.

■ *Loi de Gauss ou loi normale* $N(\mu, \sigma)$

Cette loi est centrale en statistiques puisqu'un grand nombre de phénomènes ont un comportement approximativement gaussien. De plus, un théorème que nous verrons ultérieurement affirme qu'une somme de variables aléatoires (*cf.* débit moyen annuel) tend (presque) toujours en distribution vers une loi normale.



$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right],$$

$$E(X) = \mu,$$

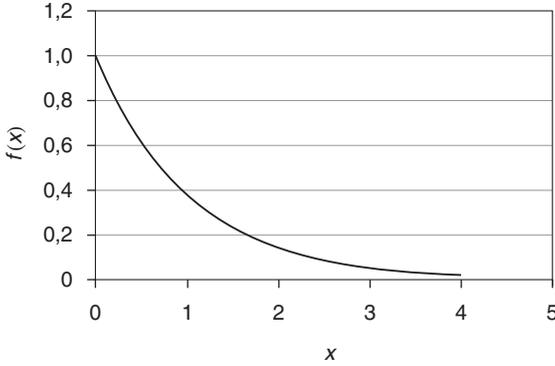
$$\text{Var}(X) = \sigma^2.$$

— $N(0, 1)$ -- $N(0, 0,5)$ — $N(1, 1)$

Figure A2.15. Loi Normale.

■ *Loi exponentielle* $\text{Exp}(x_0, \lambda)$

En hydrologie, cette loi est utilisée pour modéliser les valeurs de dépassement d'un seuil ou le temps d'attente entre deux événements de crue.



$$f(x) = \begin{cases} \frac{1}{\lambda} e^{-\frac{x-x_0}{\lambda}} & \text{si } x > x_0, \\ 0 & \text{sinon} \end{cases}$$

$$E(X) = x_0 + \lambda,$$

$$\text{Var}(X) = \lambda^2.$$

Figure A2.16. Loi exponentielle $\text{Exp}(0,1)$.

■ *Loi de Gumbel* $\text{Gu}(\mu, \lambda)$

C'est également une loi très utilisée en hydrologie puisqu'il s'agit d'une des trois distributions suivies par le maximum d'une suite de variables. Le débit maximum annuel est souvent décrit par cette loi.

$$f(x) = \frac{1}{\lambda} \exp\left(-\frac{x-\mu}{\lambda} - \exp\left(-\frac{x-\mu}{\lambda}\right)\right), \text{ et } \lambda > 0,$$

$$E(X) = \mu + \gamma\lambda, \text{ avec } \gamma = 0,5772 \text{ (constante d'Euler-Mascheroni),}$$

$$\text{Var}(X) = \frac{\pi^2}{6} \lambda^2.$$

■ *Loi généralisée des valeurs extrêmes* $\text{GEV}(\mu, \lambda, \xi)$

Cette loi généralise la loi de Gumbel.

$$f(x) = \frac{1}{\lambda} \left(1 - \frac{\xi(x-\mu)}{\lambda}\right)^{\frac{1}{\xi}-1} \exp\left(-\left(1 - \frac{\xi(x-\mu)}{\lambda}\right)^{\frac{1}{\xi}}\right),$$

$$\text{avec } \lambda > 0, \xi \neq 0 \text{ et } 1 - \frac{\xi(x-\mu)}{\lambda} > 0,$$

$$E(X) = \mu + \frac{\lambda}{\xi} (1 - \Gamma(\xi + 1)),$$

$$\text{Var}(X) = \left(\frac{\lambda}{\xi}\right)^2 (\Gamma(2\xi + 1) - \Gamma^2(\xi + 1)),$$

où $\Gamma(t) = \int_0^{+\infty} x^{t-1} e^{-x} dx$ est la fonction gamma.

Pour $\xi \rightarrow 0$, on retrouve la loi de Gumbel. Si $\xi < 0$, on obtient la loi de Fréchet, et si $\xi > 0$, on a alors la loi de Weibull.

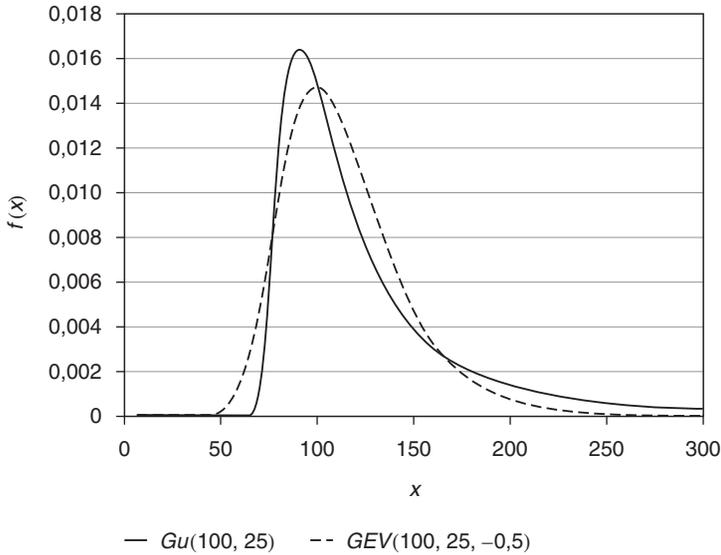


Figure A2.17. Loi généralisée des valeurs extrêmes.

Remarque : les trois lois qui suivent sont très souvent utilisées en statistiques, notamment pour effectuer des tests d'hypothèses.

■ **Loi du « Chi-2 » $\chi^2(n)$**

C'est la loi suivie par la somme des carrés de n variables gaussiennes $N(0,1)$.

$$f(x) = \frac{x^{\frac{n}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)},$$

$$E(X) = n,$$

$$\text{Var}(X) = 2n.$$

■ **Loi de Student $t(n)$**

C'est la loi suivie par la variable $\frac{U}{\sqrt{X/n}}$, où $U \sim N(0, 1)$ et $X \sim \chi^2(n)$, avec U et X indépendantes. Cette loi est principalement utilisée via la table de sa fonction de répartition (cf. les tables usuelles situées en fin d'annexe).

$$E(t(n)) = 0 \quad \text{si } n > 1,$$

$$\text{Var}(t(n)) = \frac{n}{n-2} \quad \text{si } n > 2.$$

■ **Loi de Fisher** $F(n_1, n_2)$

C'est la loi suivie par la variable $\frac{X/n_1}{Y/n_2}$, où X et Y suivent indépendamment des $\chi^2(n_1)$ et $\chi^2(n_2)$. Cette loi est également tabulée.

$$E(F(n_1, n_2)) = \frac{n_2}{n_2 - 2},$$

$$\text{Var}(F(n_1, n_2)) = 2 \frac{n_2^2}{n_1} \cdot \frac{n_1 + n_2 - 2}{n_1(n_2 - 2)^2(n_2 - 4)}.$$

Quelques mots sur le cas multivarié

Nous allons nous limiter à aborder quelques aspects concernant la loi jointe d'un couple de deux variables. Supposons donc que nous étudions une variable aléatoire $Z = (X, Y)$ à valeurs dans \mathbb{R}^2 . La **fonction de répartition jointe** du couple (X, Y) est définie par $H(x, y) = P(\{X \leq x\} \cap \{Y \leq y\})$. Il est possible de définir les **fonctions de répartition marginales** par $F(x) = P(X \leq x) = H(x, \infty)$ et $G(y) = P(Y \leq y) = H(\infty, y)$. Comme dans le cas univarié, la **densité de probabilité jointe** est égale à la dérivée de la fonction

de répartition : $h(x, y) = \frac{\partial^2 h(x, y)}{\partial x \partial y}$. Enfin, les densités marginales s'obtiennent par :

$$f(x) = \int_{\mathbb{R}} h(x, y) dy \quad \text{et} \quad g(y) = \int_{\mathbb{R}} h(x, y) dx.$$

La **covariance** entre deux vecteurs aléatoires X et Y est définie par :

$$\text{cov}(X, Y) = E[(X - E(X))(Y - E(Y))] = \int_{\mathbb{R}^2} (x - E(X))(y - E(Y))h(x, y) dx dy.$$

Exemple 6 – La loi normale de dimension 2

La densité d'un vecteur aléatoire $z = (x, y)$ suivant une loi binormale de paramètres

$$\mu = (\mu_1, \mu_2) \quad \text{et} \quad \Sigma = \begin{pmatrix} v_1 & c \\ c & v_2 \end{pmatrix} \quad \text{est :$$

$$h(x, y) = h(z) = \frac{1}{2\pi(\det \Sigma)^{1/2}} \exp\left(-\frac{1}{2}(z - u)' \Sigma^{-1}(z - u)\right).$$

La matrice Σ est la matrice des variances-covariances. On peut alors démontrer que les lois marginales sont également gaussiennes, de moyennes respectives μ_1 et μ_2 , et de variances v_1 et v_2 . c est la covariance des deux lois marginales. La *figure A2.18* permet de visualiser la loi jointe (vue de dessus) et les lois marginales de X (trait plein) et de Y (pointillés).

Une notion capitale en statistiques et probabilités est la notion d'**indépendance** entre deux variables aléatoires X et Y :

$$X \text{ et } Y \text{ sont indépendantes } \begin{cases} H(x, y) = F(x)G(y) \\ h(x, y) = f(x)g(y) \end{cases} \quad (\text{si ces densités existent}).$$

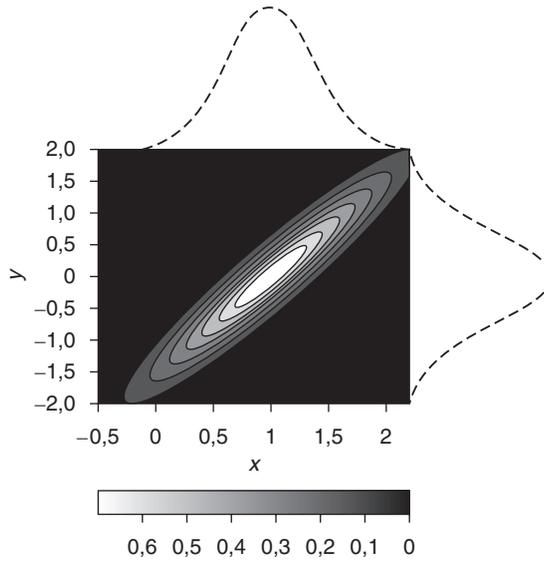


Figure A2.18. Loi normale de dimension 2.

En d'autres termes, la distribution jointe doit être le produit des distributions marginales. Si X et Y sont indépendantes, il suffit donc de connaître les distributions de X et de Y pour obtenir la loi jointe du couple. De manière plus intuitive, cela signifie que les variables X et Y n'influent pas l'une sur l'autre, et donc que la connaissance de la valeur prise par Y ne change en rien la loi de probabilité de X .

Quelques relations utiles

Soit a une constante :

- $E(a) = a$;
- $E(aX) = aE(X)$;
- $E(X + a) = E(X) + a$;
- $E(X + Y) = E(X) + E(Y)$;
- X et Y indépendantes $\Rightarrow E(XY) = E(X)E(Y)$;
- $\text{Var}(X) = E(X^2) - (E(X))^2$;
- $\text{Var}(X + a) = \text{Var}(X)$;
- $\text{Var}(aX) = a^2\text{Var}(X)$;
- $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{cov}(X, Y)$;
- $\text{cov}(X, Y) = E(XY) - E(X)E(Y)$;
- X et Y indépendantes $\Rightarrow \text{cov}(X, Y) = 0$ (réciproque fausse) ;
- X et Y indépendantes $\Rightarrow \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$ (réciproque fausse).

Statistique inférentielle

Les deux sections précédentes (*Statistiques descriptives* et *Le modèle probabiliste*) nous ont permis de nous familiariser avec les notions les plus importantes de statistiques descriptives et de probabilités, qui peuvent être abordées indépendamment. Dans cette section, nous allons nous efforcer de relier ces deux domaines en utilisant la théorie des probabilités pour quantifier plus efficacement les caractéristiques d'un échantillon, et en intégrant des notions comme le « risque », la « confiance », la « décision », etc. Dans la pratique, la statistique inférentielle est certainement l'outil le plus utilisé pour assister des décisions dans un contexte où le hasard intervient.

Retour sur la notion d'échantillon

La fluctuation d'échantillonnage

Reprenons l'exemple concernant la taille des Français, que nous avons évaluée sur un échantillon de 35 individus. Admettons que la moyenne \bar{x} de cet échantillon vaut 1,81 m. Peut-on alors affirmer que les Français mesurent en moyenne 1,81 m ? En d'autres termes, la caractéristique mesurée sur l'échantillon peut-elle être considérée comme valable pour la population entière ? La statistique inférentielle a justement pour but de quantifier, en termes de risque ou de confiance, la transmission d'une caractéristique (ici, la moyenne) entre un échantillon et sa population parente.

Pour tenter de répondre à cette question, nous décidons de répéter l'expérience, mais sur 35 autres individus. Nous obtenons alors une moyenne $\bar{x}_{(1)} = 1,70$ m. Cette valeur est différente de celle mesurée sur le premier échantillon. Nous répétons encore l'expérience un certain nombre de fois, et nous obtenons des valeurs moyennes $\bar{x}_{(2)}, \dots, \bar{x}_{(k)}$. Nous pouvons alors tracer l'histogramme de ces moyennes, et y replacer notre valeur initiale $\bar{x} = 1,81$ m (figure A2.19).

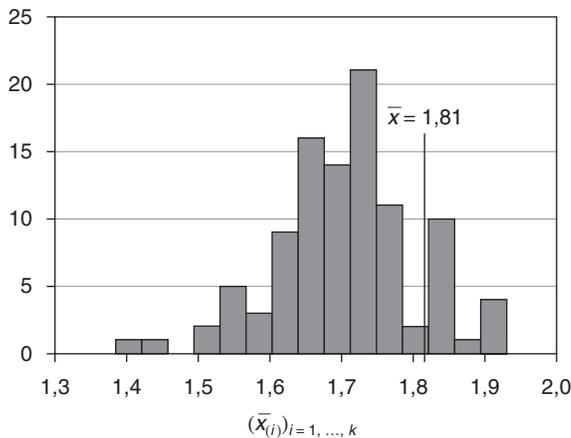


Figure A2.19. Illustration de la fluctuation d'échantillonnage.

En observant la *figure A2.19*, nous voyons apparaître une distribution liée aux moyennes empiriques calculées sur différents échantillons. Cette nouvelle source de variabilité sera appelée la **fluctuation d'échantillonnage**. Cette variabilité est due au fait que notre échantillon n'est pas notre population, donc que sa **représentativité** n'est pas absolue. Notons que le même raisonnement vaut pour la variance ou la médiane, ou pour toute autre caractéristique de l'échantillon.

Pour pallier ce problème, on peut envisager trois solutions :

- Échantillonner la population entière : c'est parfois ce qui est fait (le recensement, par exemple), mais, dans la plupart des cas, cette solution s'avère beaucoup trop coûteuse, voire impossible à mettre en œuvre (populations infinies).
- Faire comme dans l'exemple, c.-à-d. répéter l'expérience un certain nombre de fois afin d'avoir accès à cette fluctuation d'échantillonnage. Procéder de la sorte est bien souvent impossible, et ce, pour diverses raisons : coût en temps ou en argent, difficulté de répéter à l'identique une expérience en deux endroits ou deux instants différents... Une alternative consiste à utiliser une technique de rééchantillonnage que nous aborderons ultérieurement.
- Se donner un cadre théorique probabiliste qui permet de déduire le comportement de caractéristiques affectées par cette fluctuation d'échantillonnage : c'est cette approche que nous allons détailler maintenant.

L'échantillon du point de vue probabiliste

Soit (x_1, \dots, x_n) un échantillon. Nous allons considérer chaque valeur x_i de l'échantillon comme **une réalisation** d'une variable aléatoire parente X_i . Plus précisément, nous supposons que tous les X_i sont mutuellement indépendants et qu'ils ont la même loi de probabilité. Nous parlerons d'échantillons **iid** pour **indépendants et identiquement distribués**. Ainsi, l'échantillon (x_1, \dots, x_n) peut être vu comme la réalisation d'un n -uplet de variables aléatoires indépendantes et de même loi (X_1, \dots, X_n) . Par extension, ce n -uplet est aussi appelé parfois un échantillon. Malgré cette légère ambiguïté terminologique, il est capital — les notations existent également pour ça ! — de bien différencier l'échantillon (x_1, \dots, x_n) qui représente n valeurs numériques bien précises, qui n'ont donc rien d'aléatoires, du n -uplet (X_1, \dots, X_n) qui est une représentation abstraite d'un processus aléatoire ayant généré les données. Pour bien marquer cette différence, nous emploierons les termes « empirique » ou « observé » pour tout ce qui se rapporte à (x_1, \dots, x_n) , et « théorique » (on rencontre parfois « vrai ») pour ce qui concerne (X_1, \dots, X_n) .

Pour terminer avec ces définitions, nous appellerons « **statistique** » toute fonction de l'échantillon $T = f(X_1, \dots, X_n)$. Une statistique est donc une variable aléatoire dont on obtient une réalisation grâce à l'échantillon observé $T_{\text{obs}} = f(x_1, \dots, x_n)$.

Illustrons ces définitions grâce à l'exemple de la taille des Français. À partir de notre échantillon (x_1, \dots, x_{35}) , nous avons calculé la moyenne empirique $\bar{x} = \frac{1}{35} \sum_{i=1}^{35} x_i$. Cette valeur est l'observation sur l'échantillon de la statistique, que nous noterons $\bar{X} = \frac{1}{35} \sum_{i=1}^{35} X_i$.

Supposons, par exemple, que pour tout i , X_i suive une loi normale d'espérance théorique μ et d'écart-type théorique σ . Il est possible de montrer que la somme de variables gaus-

siennes demeure gaussienne. Ainsi, en utilisant les formules de la section *Le modèle probabiliste* relatives à la manipulation d'espérances et de variances, on parvient aux relations suivantes :

$$E(\bar{X}) = \frac{1}{35} \sum_{i=1}^{35} E(X_i) = \frac{1}{35} \sum_{i=1}^{35} \mu = \mu,$$

$$\text{Var}(\bar{X}) = \left(\frac{1}{35}\right)^2 \sum_{i=1}^{35} \text{Var}(X_i) = \left(\frac{1}{35}\right)^2 \sum_{i=1}^{35} \sigma^2 = \frac{\sigma^2}{35},$$

soit l'écart-type $\sigma(\bar{X}) = \frac{\sigma}{\sqrt{35}}$.

Nous avons donc caractérisé de manière théorique la fluctuation d'échantillonnage de la statistique \bar{X} , qui suit une loi normale $N\left(\mu, \frac{\sigma}{\sqrt{35}}\right)$. L'histogramme représenté en *figure A2.19* est une approximation de la densité de cette loi de probabilité. Notons que la validité de ce calcul est dépendante des trois hypothèses que nous avons posées : les données sont issues d'une même loi, qui est une loi normale, et sont indépendantes. Profitons-en pour insister sur le fait qu'indépendance et équidistribution sont des hypothèses qui sont posées a priori, et non des propriétés démontrées. Il est donc capital, lors du recueil des données, de s'assurer que le protocole garantira au mieux la validité de ces deux hypothèses. Pour finir, remarquons que dans la pratique, les valeurs de μ et σ ne nous étant pas connues, il nous faut les estimer, ce qui fera l'objet de la section *L'estimation*.

Quelques théorèmes importants

Les théorèmes que nous allons présenter *infra* sont des théorèmes de convergence de suites de variables aléatoires. La place nous manque pour aborder de manière totalement rigoureuse ces notions de convergence, aussi les théorèmes seront exposés de manière plus ou moins intuitive, le but étant ici avant tout de justifier la démarche probabiliste entreprise, et de donner des résultats dont l'importance pratique est primordiale.

Théorème de Glivenko-Cantelli

Soit (x_1, \dots, x_n) un échantillon iid issu d'une loi de probabilité de fonction de répartition F . Soit F_n la courbe empirique des fréquences cumulées obtenue par l'échantillon de taille n (cf. § *Représentations graphiques d'un échantillon*).

Notons $D_n = \sup_x |F_n(x) - F(x)|$. Alors $D_n \xrightarrow{n \rightarrow \infty} 0$.

Ce théorème affirme donc que si n est suffisamment grand, alors la différence entre les fonctions de répartition empirique et théorique est négligeable. Dit autrement, **les fréquences empiriques convergent vers les probabilités théoriques**, ce qui justifie par conséquent une analogie fréquence/probabilité que nous avons déjà relevée.

Ajoutons que la quantité D_n est à la base d'un test important en statistique, que nous verrons ultérieurement.

Loi des grands nombres

Soit (x_1, \dots, x_n) un échantillon iid issu d'une loi de probabilité d'espérance théorique m .

Notons $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ la statistique « moyenne empirique » obtenue à partir de l'échantillon théorique (X_1, \dots, X_n) . Alors $\bar{X}_n \xrightarrow[n \rightarrow \infty]{} m$.

Cette convergence mérite quelques explications. Nous avons en effet insisté sur le fait que \bar{X}_n était une variable aléatoire, alors que m est une constante. Dans ces conditions, comment l'une peut-elle converger vers l'autre ? Ce mode de convergence signifie que, justement, \bar{X}_n est de « moins en moins aléatoire » à mesure que la taille de l'échantillon augmente, en ce sens que, pour tout intervalle autour de m , la probabilité pour que \bar{X}_n soit à l'intérieur de cet intervalle peut être choisie aussi proche de 1 que l'on veut, et ce, en choisissant un rang n suffisamment grand. Notons que ce théorème reste vrai quelle que soit la loi dont est issu l'échantillon (pourvu que l'espérance existe).

Théorème central limite

Ce théorème est d'une importance pratique considérable en statistiques.

Soit (X_1, \dots, X_n) un n -uplet de variables aléatoires iid, d'espérance μ et d'écart-type σ . Alors :

$$\frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}} \xrightarrow[n \rightarrow \infty]{} N(0,1) \Leftrightarrow \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow[n \rightarrow \infty]{} N(0,1)$$

La signification du théorème est la suivante : **quelle que soit la loi dont est issu un échantillon**, du moment qu'espérance et variance existent, la statistique « moyenne empirique » suivra une loi normale si n est suffisamment grand. En pratique, une trentaine d'observations est souvent jugée suffisante pour appliquer l'approximation gaussienne. Reprenons à nouveau notre exemple sur la taille moyenne des Français : il s'avère que nos conclusions demeurent exactes, et ce, même si l'échantillon n'est pas issu d'une loi normale !

La *figure A2.20* illustre ces deux derniers théorèmes. Pour trois lois de probabilité bien distinctes, nous avons simulé des échantillons de diverses tailles, et calculé la moyenne empirique. Les histogrammes de la *figure A2.20* effectuent l'approximation de la densité de la statistique \bar{X}_n . Nous observons alors les deux phénomènes précédemment décrits : d'une part, quand n devient grand, \bar{X}_n est de « moins en moins aléatoire » et ses valeurs se concentrent de plus en plus autour de la valeur théorique de l'espérance. D'autre part, quelle que soit la forme de la distribution parente (première colonne), la distribution d'échantillonnage de \bar{X}_n prend peu à peu une forme gaussienne, comme le prédit le théorème central limite.

Théorème des valeurs extrêmes

C'est, en quelque sorte, l'équivalent du théorème central limite pour la statistique $M_n = \max(X_1, \dots, X_n)$. La formulation étant un peu délicate, nous le retiendrons sous la forme suivante : si la distribution de M_n converge, elle le fait forcément vers une loi généralisée des valeurs extrêmes $GEV(\mu, \lambda, \xi)$. La *figure A2.21* permet d'illustrer cette convergence pour les lois $N(0,1)$, $U(0,1)$ et de Cauchy, dont les maximums convergent vers une GEV où le paramètre ξ est respectivement nul (loi de Gumbel), positif (loi de Weibull) et négatif (loi de Fréchet). Ce théorème justifie l'utilisation (en hydrologie) de ces lois

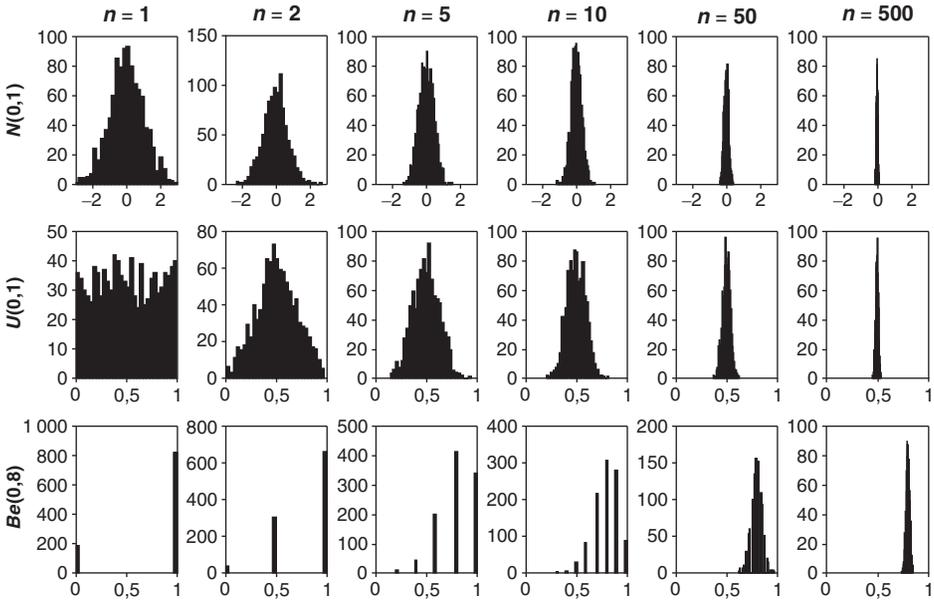


Figure A2.20. Illustration du théorème central limite.

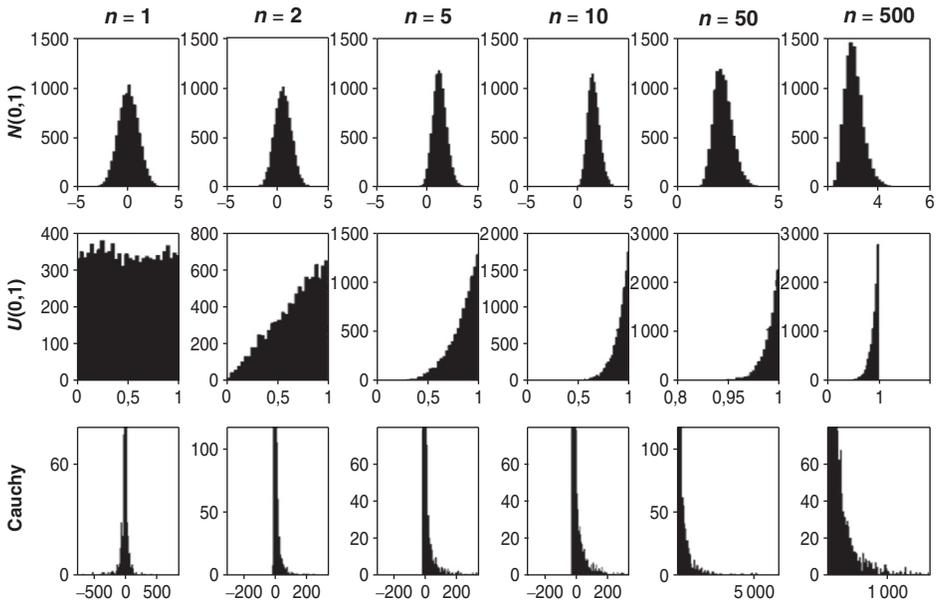


Figure A2.21. Illustration du théorème des valeurs extrêmes.

des valeurs extrêmes pour décrire le comportement probabiliste des crues. Notons une nouvelle fois que la loi limite du maximum ne dépend pas de la loi parente dont est issu l'échantillon.

Techniques de rééchantillonnage

Une autre approche pour accéder à la distribution d'échantillonnage d'une statistique $T = f(X_1, \dots, X_n)$ consiste à utiliser des techniques comme le *bootstrap* qui ont gagné en popularité ces dernières années du fait de leur simplicité et de la puissance accrue des moyens informatiques. Voici l'algorithme du *bootstrap* à partir d'un échantillon observé (x_1, \dots, x_n) :

- faire pour $i = 1, \dots, Nb$;
 - tirer au sort avec remise n individus (certains individus apparaîtront plusieurs fois, d'autres aucune),
 - calculer $T_{\text{obs}}^{(i)}$ sur cet échantillon ;
- fin.

Nous obtenons ainsi un nouvel échantillon $(T_{\text{obs}}^{(1)}, \dots, T_{\text{obs}}^{(Nb)})$ dont on montre qu'il est issu d'une bonne approximation de la distribution d'échantillonnage. La simplicité de cette méthode tient au fait qu'elle ne fait plus appel au cadre probabiliste (même si celui-ci est nécessaire pour démontrer que la technique fonctionne...), et qu'elle permet donc de prendre en compte cette fluctuation dans un contexte de statistiques descriptives.

L'estimation

La théorie de l'estimation est un domaine important et vaste des statistiques. Nous n'aborderons que quelques aspects en expliquant comment mesurer la qualité d'un estimateur, et en présentant deux méthodes d'estimation très utilisées en pratique.

Quelques définitions

Considérons un échantillon (x_1, \dots, x_n) iid, réalisation du n -uplet de variables aléatoires (X_1, \dots, X_n) . Il est fréquent d'avoir une idée a priori sur la loi dont devrait être issu cet échantillon, par expérience ou en considérant un des théorèmes limites *supra* (exemple : la crue maximale annuelle suit une loi de Gumbel). Nous allons donc supposer que l'échantillon est issu d'une certaine loi de densité f paramétrée par un certain nombre de paramètres, et que nous noterons sous forme vectorielle θ . Soit $f(x|\theta)$ cette densité. Par exemple, pour une loi normale, $\theta = (\mu, \sigma)$. La question est donc à présent d'**estimer** la valeur des paramètres à partir des données observées. Pour cela, nous allons faire appel à des statistiques $T = g(X_1, \dots, X_n)$ particulières, appelées **estimateurs**. Nous réclamerons que ces estimateurs satisfassent à un certain nombre de critères garantissant leur **qualité**. Parmi ces critères, citons :

- la **convergence** : si T_n est un estimateur d'un paramètre α , obtenu à partir d'un échantillon de taille n , alors il est souhaitable que $T_n \xrightarrow[n \rightarrow \infty]{} \alpha$ (intuitivement, quand l'échantillon tend vers la population, on retrouve la vraie valeur) ;
- l'absence de biais : $E(T_n) = \alpha$;
- la **précision** : $E((T_n - \alpha)^2)$, l'**erreur quadratique moyenne**, est minimale.

Notons qu'il n'est pas forcément possible de trouver un estimateur qui satisfasse ces trois critères simultanément.

Exemple 7 – Estimateurs d'une espérance et d'une variance

Soit μ l'espérance de la loi dont est issu un échantillon. Il est alors aisé de montrer que \bar{X}_n est un estimateur de μ convergent et sans biais :

– convergent : c'est la loi des grands nombres (cf. § *Retour sur la notion d'échantillon, Quelques théorèmes importants*) ;

– sans biais :
$$E(\bar{X}_n) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \mu .$$

Supposons à présent que la loi dont est issu l'échantillon est de variance V que nous souhaitons estimer. Les estimateurs naturels sont (cf. § *Comment résumer un échantillon ?*,

Les indicateurs de dispersion) :
$$T_1 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \quad \text{et} \quad T_2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 .$$

Calculez l'espérance de chacun de ces estimateurs (conseil : utiliser l'expression

$$T_1 = \left(\frac{1}{n} \sum_{i=1}^n X_i^2 \right) - (\bar{X}^2) .$$
 Lequel d'entre eux est-il sans biais ?

Estimation par la méthode des moments

L'idée de cette méthode est la suivante : si les paramètres sont bien estimés, il devrait alors y avoir adéquation entre les caractéristiques observées (ou empiriques) et les caractéristiques théoriques. Nous allons rechercher cette adéquation sur les moments, en général le premier moment non centré (c.-à-d. la moyenne) et le second moment centré (c.-à-d. la variance). Plus précisément, supposons que nous ayons p paramètres $\theta_1, \dots, \theta_p$ à estimer ; leurs estimateurs $\hat{\theta}_1, \dots, \hat{\theta}_p = \hat{\theta}$ seront alors solutions du système :

$$\begin{cases} E_{\hat{\theta}}(X) = m_1 \\ \text{Var}_{\hat{\theta}}(X) = m'_2 \\ \dots \\ E_{\hat{\theta}}((X - E_{\hat{\theta}}(X))^p) = m'_p \end{cases} .$$

La notation $E_{\hat{\theta}}$ a pour but de bien montrer que ces moments sont des fonctions des paramètres à estimer.

Exemple 8 – Estimateurs des paramètres d'une loi normale

Soit (x_1, \dots, x_n) iid, un échantillon issu d'une loi normale $N(\mu, \sigma)$. L'espérance vaut donc μ et la variance σ^2 . Les estimateurs des moments sont donc définis par :

$$\hat{\mu} = m_1 = \bar{x} ,$$

$$\hat{\sigma}^2 = m_2 \Rightarrow \hat{\sigma} = \sqrt{m_2} .$$

En d'autres termes, les estimateurs des moments sont égaux à la moyenne et à l'écart-type empiriques.

Exercice

Exprimez les estimateurs des moments pour une loi exponentielle $\text{Exp}(0, \lambda)$ et uniforme $U(a, b)$.

Estimation par maximum de vraisemblance

Cette méthode consiste, étant donné un échantillon (x_1, \dots, x_n) iid, de choisir comme estimateur de $\theta_1, \dots, \theta_p$ les valeurs qui rendent l'échantillon le plus « probable » possible. Plus précisément, comme l'échantillon est indépendant, la densité jointe du vecteur aléatoire (X_1, \dots, X_n) est égale au produit des lois marginales. Nous appellerons **vraisemblance** l'expression de cette densité prise pour les valeurs (x_1, \dots, x_n) , soit :

$$L(x_1, \dots, x_n | \theta) = \prod_{i=1}^n f(x_i | \theta).$$

Il faut donc trouver les valeurs $\hat{\theta}_1, \dots, \hat{\theta}_p$ qui maximisent cette vraisemblance.

Dans la pratique, on aura souvent intérêt à maximiser le **log-vraisemblance** :

$$\text{Log}L(x_1, \dots, x_n | \theta) = \sum_{i=1}^n \log(x_i | \theta).$$

Pour ce faire, il faut résoudre une équation aux dérivées partielles :

$$\frac{\partial \log L}{\partial \theta_i} = 0, \forall i = 1, \dots, p.$$

Exemple 9 – Estimation du paramètre d'une loi exponentielle $\text{Exp}(0, \lambda)$

Soient (x_1, \dots, x_n) les observations issues de cette loi exponentielle dont la densité s'écrit :

$$f(x | \lambda) = \frac{1}{\lambda} e^{-\frac{x}{\lambda}}.$$

La log-vraisemblance vaut donc :

$$\begin{aligned} \text{Log}L(x_1, \dots, x_n | \lambda) &= \sum_{i=1}^n \log(f(x_i | \lambda)) \\ &= \sum_{i=1}^n \log\left(\frac{1}{\lambda} e^{-\frac{x_i}{\lambda}}\right) \\ &= n \log\left(\frac{1}{\lambda}\right) - \sum_{i=1}^n \frac{x_i}{\lambda} \\ &= -n \log(\lambda) - \sum_{i=1}^n \frac{x_i}{\lambda} \end{aligned}$$

D'où :

$$\begin{aligned} \frac{\partial \log L}{\partial \lambda} &= -\frac{n}{\lambda} + \sum_{i=1}^n \frac{x_i}{\lambda^2} = 0 \Leftrightarrow -n + \sum_{i=1}^n \frac{x_i}{\lambda} = 0 \Leftrightarrow \sum_{i=1}^n \frac{x_i}{\lambda} = n \\ &\Leftrightarrow \lambda = \frac{1}{n} \sum_{i=1}^n x_i \end{aligned}$$

L'estimateur du maximum de vraisemblance correspond ici à l'estimateur des moments.

Exercice

Exprimer l'estimateur du maximum de vraisemblance pour les lois $N(\mu, \sigma)$ et $U(a, b)$.

Quel estimateur choisir ?

Il n'existe malheureusement pas de réponse à cette question... Dans certains cas, les deux estimateurs présentés *supra* coïncident et sont « optimaux ». Dans d'autres cas, l'un aura des propriétés que l'autre n'aura pas, et vice versa. En général, l'estimateur du maximum de vraisemblance est utilisé pour les grands échantillons, car la théorie permet d'en déterminer les propriétés asymptotiques (loi, variance). En contrepartie, les calculs détaillés *supra* étant souvent inextricables à la main dans des cas plus compliqués, il faut recourir à des méthodes d'optimisation numérique qui ne garantissent d'ailleurs pas la convergence vers un maximum absolu.

Notons que la méthode qui consiste à réaliser un ajustement des moindres carrés sur un graphique de fréquence linéarisé se révèle en fait inférieure à tous les autres estimateurs ! Ce point a été montré dans le cas d'une loi de Gumbel sur les valeurs maximales annuelles (Lowery et Nash, 1970), et dans le cas d'une loi exponentielle sur les valeurs supérieures à un seuil (Sambou, 2004). Sur ces deux exemples, on a le classement suivant :

$$EQM \text{ (maxi vrais)} < EQM \text{ (moments)} < EQM \text{ (moindres carrés)},$$

la dernière méthode présentant à la fois le plus fort biais et la plus forte dispersion !

Intervalles de confiance

Les estimateurs étant en particulier des statistiques, ils sont soumis à la fluctuation d'échantillonnage dont nous avons longuement parlé précédemment : si l'on prend un autre échantillon, la valeur des estimations changera. Il est donc important de prendre en compte cette fluctuation afin de rendre plus interprétables les estimations ponctuelles.

Supposons que $T = g(X_1, \dots, X_n)$ est un estimateur dont on connaît la loi de probabilité F (qui dépend de θ). Un intervalle de probabilité au risque α (ou au niveau de confiance $1 - \alpha$), noté $I_\alpha(\theta)$, est un intervalle vérifiant la propriété :

$$P(T \in I_\alpha) = 1 - \alpha.$$

Dans la pratique, on choisira souvent un intervalle $[t^- ; t^+]$, tel que :

$$P(T < t^-) = \frac{\alpha}{2} \quad \text{et} \quad P(T > t^+) = \frac{\alpha}{2} \quad \Leftrightarrow \quad P(T \leq t^+) = 1 - \frac{\alpha}{2}.$$

Les limites de l'intervalle de probabilité sont donc les quantiles d'ordre $\frac{\alpha}{2}$ et $1 - \frac{\alpha}{2}$ de la distribution d'échantillonnage de notre estimateur, ils dépendent par conséquent toujours de θ . Or, les observations nous conduisent à une valeur observée de l'estimateur, notée $\hat{\theta}$. Une valeur « acceptable » de θ devrait donc conduire à une valeur observée $\hat{\theta} \in I_\alpha(\theta)$. C'est l'ensemble de ces valeurs acceptables que nous appellerons intervalle de confiance au niveau $1 - \alpha$.

Exemple 10 – Intervalle de confiance de l'estimateur d'une moyenne

D'après le théorème central limite, $\frac{\bar{X}_n - \mu}{\sigma / \sqrt{n}}$ converge vers une $N(0,1)$. Notons u_p le quantile d'ordre p de cette loi. Nous avons donc, par définition du quantile :

$$P\left(u_{\frac{\alpha}{2}} < \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \leq u_{1-\frac{\alpha}{2}}\right) = 1 - \alpha.$$

Un intervalle de probabilité au niveau $1 - \alpha$ pour $N(0,1)$ est donc :

$$\mu + \frac{\sigma}{\sqrt{n}} u_{\frac{\alpha}{2}} < \bar{X}_n \leq \mu + \frac{\sigma}{\sqrt{n}} u_{1-\frac{\alpha}{2}}.$$

La valeur \bar{x} ayant été observée, nous recherchons l'ensemble des μ vérifiant :

$$\mu + \frac{\sigma}{\sqrt{n}} u_{\frac{\alpha}{2}} < \bar{x} \leq \mu + \frac{\sigma}{\sqrt{n}} u_{1-\frac{\alpha}{2}} \Leftrightarrow \bar{x} - \frac{\sigma}{\sqrt{n}} u_{1-\frac{\alpha}{2}} \leq \mu < \bar{x} - \frac{\sigma}{\sqrt{n}} u_{\frac{\alpha}{2}}.$$

Si la valeur de σ est connue, alors nous disposons d'un intervalle de confiance au niveau $1 - \alpha$. Dans le cas contraire, il faut recommencer en remplaçant σ par son estimateur

naturel, mais la loi de $\frac{\bar{X}_n - \mu}{\frac{\text{Estim}(\sigma)}{\sqrt{n}}}$ n'est alors plus gaussienne. C'est en cela que réside la

principale difficulté de la détermination des intervalles de confiance : il n'est pas toujours aisé d'accéder à la distribution d'échantillonnage de l'estimateur. Fort heureusement, de nombreux résultats asymptotiques existent qui permettent d'effectuer ce type de calculs, ces derniers étant absolument nécessaires afin d'avoir une vision un tant soit peu objective de la confiance que l'on peut accorder à des résultats statistiques.

Les tests statistiques

Le principe des tests statistiques

D'après Saporta G. (1996), « un test est un mécanisme qui permet de trancher entre deux hypothèses, dont une et une seule est vraie, au vu des résultats d'un échantillon. » Soit H_0 et H_1 ces deux hypothèses. Il existe alors quatre possibilités de probabilités différentes, résumées dans le *tableau A2.6*.

Tableau A2.6. Risques d'erreur.

Vérité Décision	H_0	H_1
H_0	$1 - \alpha$	β
H_1	α	$1 - \beta$

H_0 est appelée l'**hypothèse nulle**, et H_1 l'**alternative**. Nous allons voir que ces hypothèses ne jouent pas un rôle symétrique, contrairement à ce que le *tableau A2.6* pourrait laisser croire. α et β sont appelés les **risques de première** et de **seconde espèce**. $1 - \beta$ est la **puissance** du test.

Ces définitions étant posées, comment effectue-t-on la prise de décision ? Nous allons tout d'abord tenter de comprendre le canevas général via l'exemple 11.

Exemple 11 – La température globale moyenne sur l'année de la planète Terre peut être considérée comme une variable aléatoire gaussienne de moyenne 14 °C et d'écart-type 0,25. Au début des années quatre-vingt, les valeurs contenues dans le *tableau A2.7* ont été calculées.

Tableau A2.7. Tableau de données.

Année	1980	1981	1982	1983	1984	1985
T (en °C)	14,27	14,4	14,1	14,34	14,16	14,13

Nous souhaitons tester l'hypothèse du réchauffement de la Terre. Nous allons donc choisir comme hypothèses H_0 « la température moyenne est de 14 °C », contre H_1 « la température moyenne est supérieure à 14 °C ». Pour cela, nous allons commencer par choisir un risque de première espèce α , qui représente le risque d'affirmer à tort que la planète est plus chaude qu'autrefois. Étant donné la portée d'une telle affirmation, nous allons rester prudents et choisir un risque faible, disons de 1 %. À partir de notre échantillon de 6 années, il est possible de calculer la moyenne empirique \bar{x} , réalisation de la statistique \bar{X} que l'on appelle la **statistique de test**. Si H_0 est vraie, alors l'échantillon suit une loi normale $N(14, 0, 25)$. Comme nous l'avons vu précédemment, la statistique \bar{X} suit alors une loi normale $N\left(14, \frac{0,25}{\sqrt{6}}\right)$. L'hypothèse H_0 devrait donc être rejetée (c.-à-d. H_1 devrait être acceptée) si la réalisation de cette statistique conduit à une valeur trop grande pour être vraisemblable, c.-à-d. si $\bar{x} > k$, où k est à déterminer. La région $[k ; +\infty]$ est appelée la **zone de rejet** de l'hypothèse H_0 . Pour calculer k , nous allons nous appuyer sur la définition de α :

$$\begin{aligned} \alpha &= P(\text{rejeter } H_0 | H_0) \\ &= P(\bar{X} > k | H_0) \\ &= 1 - P(\bar{X} \leq k | H_0) \end{aligned}$$

Or, si H_0 est vraie, alors nous connaissons la loi de \bar{X} qui est une $N\left(14, \frac{0,25}{\sqrt{6}}\right)$.

De l'équation précédente, il vient que k est le $1 - \alpha$ quantile de cette distribution, qui vaut 14,237 (valeur tabulée ou fonction prédéfinie de Microsoft Excel, par exemple). Nous allons donc rejeter H_0 si $\bar{x} > k$. Avec les données ci-dessus, on trouve $\bar{x} = 14,233$, ce qui nous conduit à ne pas rejeter H_0 : on dit que le test est **non significatif**, ou encore que la température n'a pas significativement augmenté (au risque 1 %).

Il subsiste cependant un risque d'erreur, à savoir celui d'avoir conservé à tort H_0 :

$$\beta = P(\text{conserver } H_0 | H_1) = P(\bar{X} \leq k | H_1).$$

Or, il nous est impossible ici de calculer ce risque, car, sous l'hypothèse H_1 , nous ne connaissons pas la loi dont sont issues les données. Nous voyons donc le rôle asymétrique joué par les deux hypothèses : l'hypothèse H_0 est dite **simple** (le paramètre vaut une valeur précise), tandis que H_1 est dite **composite** (le paramètre appartient à un sous-ensemble de \mathbb{R}). Le risque de seconde espèce n'est ici calculable que si H_1 est également simple.

Exercice

Refaites le test en intégrant les données des années 1986 à 2004 (tableau A2.8).

Tableau A2.8. Tableau de données.

Année	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995
T (en °C)	14,19	14,35	14,42	14,28	14,49	14,44	14,16	14,18	14,31	14,47
Année	1996	1997	1998	1999	2000	2001	2002	2003	2004	
T (en °C)	14,36	14,4	14,71	14,44	14,41	14,56	14,7	14,64	14,61	

La quasi-totalité des tests d'hypothèses suivent le même schéma qui peut se résumer comme suit :

1. choix des hypothèses H_0 et H_1 ;
2. choix du risque de première espèce α ;
3. choix de la statistique de test ;
4. détermination de la loi de cette statistique sous H_0 ;
5. calcul de la zone de rejet ;
6. calcul de la valeur expérimentale de la variable de test et décision ;
7. si possible, calcul du risque de seconde espèce et de la puissance.

Par rapport au test relativement simple donné en exemple, l'étape la plus difficile, en général, est le calcul de la loi de la statistique de test sous H_0 . Bien souvent, des résultats asymptotiques (c.-à-d. valables pour un échantillon suffisamment grand) sont utilisés.

Quelques tests classiques

Tests sur les paramètres d'une loi de Gauss

Condition d'application : l'échantillon *iid* est issu d'une loi normale $N(\mu, \sigma)$.

Test 1 – $H_0 : \mu = m_0$ contre H_1 : hypothèse alternative ($\mu > m_0$, ou $\mu = m_1$, par exemple) ; σ connu.

Statistique de test : $\bar{X} \sim N\left(m_0, \frac{\sigma}{\sqrt{n}}\right)$ sous H_0 .

Test 2 – $H_0 : \mu = m_0$ contre H_1 : hypothèse alternative ($\mu > m_0$, ou $\mu = m_1$, par exemple) ; σ inconnu.

Statistique de test : $T = \frac{\bar{X} - m_0}{S} / \sqrt{n - 1}$ Student ($n - 1$) sous H_0 ,

où $S = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2}$ est l'estimateur de l'écart-type.

Note : ces deux tests restent applicables si les données ne sont pas gaussiennes avec un échantillon d'effectif au moins 30.

Test 3 – $H_0 : \sigma = \sigma_0$ contre H_1 : hypothèse alternative ; μ connu.

$$\text{Statistique de test : } \frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma_0^2} \sim \chi^2(n) \text{ sous } H_0.$$

Test 4 – $H_0 : \sigma = \sigma_0$ contre H_1 : hypothèse alternative ; μ inconnu.

$$\text{Statistique de test : } \frac{nS^2}{\sigma_0^2} \approx \chi^2(n-1) \text{ sous } H_0.$$

Tests d'ajustement à une distribution

Les deux tests présentés ci-après ont pour but de vérifier que les données sont issues d'une distribution $F_0(x)$.

■ Test de Kolmogorov

Conditions d'applications : $F_0(x)$ entièrement spécifiée (*i.e.* pas de paramètres) et continue.

H_0 : données issues de $F_0(x)$ contre H_1 : données issues d'une autre distribution.

$$\text{Statistique de test : } D_n = \sup_x |F_n(x) - F_0(x)|,$$

où $F_n(x)$ est la fonction de répartition empirique.

Sous H_0 , la loi de D_n est tabulée.

■ Test du χ^2

Conditions d'applications : $F_0(x)$ doit être discrète ou discrétisée. p_1, \dots, p_k les probabilités théoriques de chaque classe, et N_1, \dots, N_k les effectifs observés pour chaque classe.

H_0 : données issues de $F_0(x)$ contre H_1 : données issues d'une autre distribution.

$$\text{Statistique de test : } D^2 = \sum_{i=1}^k \frac{(N_i - np_i)^2}{np_i}.$$

Loi sous H_0 :

- si $F_0(x)$ est entièrement spécifiée, $D^2 \sim \chi^2(k-1)$ asymptotiquement ;
- si $F_0(x)$ dépend de ℓ paramètres, alors il faut estimer ces paramètres par maximum de vraisemblance à partir de la loi discrétisée, et $D^2 \sim \chi^2(k-1-\ell)$ asymptotiquement.

Note : on admettra l'approximation asymptotique si $np_i > 5$ pour toutes les classes (procéder éventuellement à des regroupements).

Tests de comparaison d'échantillons

■ Test des variances de Fischer

Conditions d'application : deux échantillons indépendants, de tailles n_1 et n_2 , et de lois normales $N(\mu_1, \sigma_1)$ et $N(\mu_2, \sigma_2)$.

$H_0 : \sigma_1 = \sigma_2$ contre $H_1 : \sigma_1 \neq \sigma_2$.

$$\text{Statistique de test : } F = \frac{(n_1 S_1^2)/(n_1 - 1)}{(n_2 S_2^2)/(n_2 - 1)} = \frac{S_1^{*2}}{S_2^{*2}},$$

où S^{*2} est l'estimateur sans biais de la variance. On mettra au numérateur l'échantillon conduisant à la variance estimée la plus grande.

Loi sous H_0 : $F \sim \text{Fisher}(n_1 - 1, n_2 - 1)$.

■ *Test des moyennes de Student*

Conditions d'application : deux échantillons indépendants, de tailles n_1 et n_2 , et de lois normales $N(\mu_1, \sigma)$ et $N(\mu_2, \sigma)$ (écarts-types inconnus mais égaux).

$H_0 : \mu_1 = \mu_2$ contre H_1 : hypothèse alternative.

$$\text{Statistique de test : } T = \frac{(\bar{X}_1 - \bar{X}_2)\sqrt{n_1 + n_2 - 2}}{\sqrt{(n_1 S_1^2 + n_2 S_2^2)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Note : si les deux échantillons sont suffisamment grands (quelques dizaines d'individus), le test de Student peut être appliqué même si les données ne sont pas gaussiennes, ou ont des variances inégales. On dit que ce test est « **robuste** » (*i.e.* peu sensible au non-respect des conditions d'application).

■ *Test des rangs de Wilcoxon*

Conditions d'application : deux échantillons (x_1, \dots, x_{n_1}) et (y_1, \dots, y_{n_2}) indépendants, de tailles $n_1 < n_2$. On mélange les échantillons, et on note $\text{rg}(x_i)$ le rang d'une observation x_i dans cet échantillon mélangé.

H_0 : les échantillons sont issus d'une même population contre H_1 : hypothèse alternative.

$$\text{Statistique de test : } S = \frac{\sum_{i=1}^n \text{rg}(x_i) - \mu}{\sigma},$$

où $\mu = \frac{n_1(n_1 + n_2 + 1)}{2}$ et $\sigma = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$.

Loi sous H_0 : $S \sim N(0,1)$ dès que les deux échantillons sont d'effectifs supérieurs à 8.

■ *Test de corrélation*

Conditions d'application : deux variables X et Y gaussiennes, dont on cherche à savoir si elles sont corrélées. Soit $\rho = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$, le coefficient de corrélation.

$H_0 : \rho = 0$ contre H_1 : hypothèse alternative.

$$\text{Statistique de test : } R = \rho \sqrt{\frac{n-2}{1-\rho^2}} \sim \text{Student}(n-2) \text{ sous } H_0.$$

Note : ce test permet également de tester la nullité du paramètre a d'une régression

$$Y = aX + b, \text{ puisque } \hat{a} = \frac{\text{cov}(X, Y)}{\text{cov}(X)} = \rho \frac{\sigma_Y}{\sigma_X}.$$

Test de stationnarité

■ *Test de Pettitt*

Conditions d'application : un échantillon (x_1, \dots, x_n) , dont on cherche à savoir s'il présente une rupture.

H_0 : échantillon stationnaire contre H_1 : hypothèse alternative.

$$\text{Statistique de test : } S = \max_k (|U(k)|),$$

où
$$U(k) = \sum_{i=1}^k \sum_{j=k+1}^n \text{sgn}(x_i - x_j).$$

Loi sous H_0 :
$$P(S \geq s_0) = 2 \exp\left(\frac{-6s_0^2}{n^3 + n^2}\right).$$

■ **Test de Spearman**

Conditions d'application : un échantillon (x_1, \dots, x_n) , pour lequel on cherche à savoir s'il existe une corrélation significative entre le rang des observations et le temps. Soit

$$\rho = \frac{\text{cov}(Y, Z)}{(\sigma_Y \sigma_Z)},$$
 le coefficient de corrélation entre $Y = \text{rg}(X)$ et $Z = t(X)$.

H_0 : échantillon stationnaire ($\rho = 0$) contre H_1 : hypothèse alternative.

Statistique de test :
$$R = \rho \sqrt{\frac{n-2}{1-\rho^2}}$$
 suit la loi Student $(n-2)$ sous H_0 .

■ **Test de Mann-Kendall**

Conditions d'application : recherche d'une tendance dans l'échantillon (x_1, \dots, x_n) .

H_0 : échantillon stationnaire contre H_1 : hypothèse alternative.

Statistique de test :
$$Z = \begin{cases} \frac{S-1}{\sqrt{\text{var}(S)}} & \text{si } S > 0 \\ 0 & \text{si } S = 0 \\ \frac{S+1}{\sqrt{\text{var}(S)}} & \text{si } S < 0 \end{cases},$$

avec
$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sgn}(x_j - x_i)$$
 et
$$\text{var}(S) = \frac{n(n-1)(2n+5)}{18}.$$

En cas d'*ex aequo* dans la série, la variance est calculée par la formule :

$$\text{var}(S) = \frac{\left(n(n-1)(2n+5) - \sum_{k=1}^n t_k k(k-1)(2k+5) \right)}{18},$$

où t_k désigne le nombre d'égalités impliquant k valeurs.

Loi sous H_0 : $Z \sim N(0 ; 1)$ si l'échantillon comporte au moins une dizaine d'observations.

■ **Test de Buishand**

Conditions d'application : un échantillon gaussien (x_1, \dots, x_n) de valeurs indépendantes, dont on cherche à savoir s'il présente une rupture.

H_0 : échantillon stationnaire contre H_1 : hypothèse alternative.

Statistique de test :
$$Z = \max_{k=1, \dots, n} \left(\frac{|U(k)|}{\sqrt{n} \sigma_X} \right),$$
 avec
$$U(k) = \sum_{i=1}^k (x_i - \bar{X}).$$

$$\text{Instant de rupture : } \hat{k} = \arg \max_{k=1, \dots, n} \left(\frac{|U(k)|}{\sqrt{n}\sigma_X} \right).$$

Loi sous H_0 : tabulée par Buishand (1982) (cf. *tableau A2.9*).

Tableau A2.9. Valeurs critiques de la statistique de Buishand (1982).

n	α		
	0,10	0,05	0,01
10	1,05	1,14	1,29
20	1,10	1,22	1,42
30	1,12	1,24	1,46
40	1,13	1,26	1,50
50	1,14	1,27	1,52
100	1,17	1,29	1,55
∞	1,22	1,36	1,63

■ *Test sur les dates d'occurrence d'événements supérieurs à un seuil*

Conditions d'application : échantillon (t_1, \dots, t_n) des dates d'occurrence des valeurs supérieures à un seuil, dont on cherche à savoir si elles sont distribuées suivant un processus de Poisson stationnaire.

H_0 : processus d'occurrence de Poisson stationnaire contre H_1 : hypothèse alternative.

Statistique de test : Pour une date donnée t , nombre d'événements m_t observés dans l'intervalle $[0; t]$, sachant que N valeurs ont été observées sur la période complète $[0; t_{\text{fin}}]$ ($t < t_{\text{fin}}$).

Loi sous H_0 : $\text{Prob}[m_t = k | m_{\text{fin}} = N] = \binom{N}{k} \left(\frac{t}{t_{\text{fin}}} \right) \left(1 - \frac{t}{t_{\text{fin}}} \right)^{N-k}$ qui définit un intervalle

de tolérance $I_{1-\alpha}$, avec un risque d'erreur α : $\text{Prob}[m \in I_{1-\alpha}] = 1 - \alpha$.

Dans le cas général où l'on ne suspecte pas de changement à une date donnée, on comptabilise le nombre de points $N(H_1)$ de la courbe expérimentale (m_t, t) sortant de l'intervalle de tolérance $I_{1-\alpha}$. Le *tableau A2.10* (Lang *et al.*, 2004) donne le pourcentage minimum de points pour rejeter l'hypothèse H_0 avec le risque α . Si la taille de l'échantillon est supérieure à 25, une règle simplifiée donne un pourcentage limite de points hors intervalle de respectivement 29 %, 26 % et 21 %, pour les risques $\alpha = 10$ %, 5 % et 1 %.

Tableau A2.10. Pourcentage critique de points associé au rejet de l'hypothèse H_0 , quand le test est appliqué sur l'ensemble d'une période d'observations, avec un risque d'erreur $\alpha = 10 \%$, 5% et 1% ^a.

m N	1/10	1/5	1/4	1/3	1/2	1	2	3	4	5
10	100	50			40	30	30	30	27,5	28
	100	50			40	30	25	23,33	25	24
	0	50			20	20	20	20	20	20
20	50	25	40		30	30	27,5	28,33	27,5	27
	50	25	40		30	25	25	23,33	23,75	23
	50	25	20		20	20	20	18,33	18,75	17
30	33,33	33,33		30	26,67	30	26,67	26,67	27,5	26
	33,33	33,33		30	26,67	26,67	23,33	23,33	24,17	24,67
	33,33	33,33		20	20	23,33	18,33	17,78	19,17	18,67
40	25	37,5	30		30	30	28,75	28,33	28,13	26,5
	25	25	30		25	25	26,25	25	23,75	23,5
	25	25	20		20	20	21,25	19,17	18,13	17,5
50	40	30			28	30	29	28	27,5	27,2
	40	30			24	26	26	24	23,5	23,2
	20	20			20	22	21	21,33	17,5	19,2
60	33,33	33,33	26,67	30	30	28,33	30	29,44	29,17	28,33
	33,33	25	26,67	25	26,67	25	25,83	25,56	27,5	26
	33,33	25	20	20	20	20	20	21,67	22,92	20,33
70	28,57	28,57			28,57	27,14	27,86	29,05	29,29	27,71
	28,57	28,57			25,71	24,29	24,29	25,71	25,71	24,57
	28,57	21,43			20	20	18,57	20,95	19,64	17,14
80	25	31,25	30		30	27,5	28,13	29,58	26,88	26,25
	25	25	25		25	23,75	23,75	26,67	23,13	22,5
	25	18,75	20		20	20	19,38	20,83	19,06	18,75
90	33,33	27,78		26,67	28,89	27,78	27,22	27,78	26,94	27,33
	22,22	27,78		23,33	24,44	24,44	25	25,19	23,89	24,22
	22,22	16,67		20	20	18,89	23,33	18,52	19,17	17,11
100	30	30	28		30	27	29	27,67	28,5	27
	30	25	28		26	24	25,5	24,67	24,75	24,2
	20	20	20		20	21	19	16	18	18,8

a. Lang *et al.* (2004).

Applications à l'étude des crues

Nous allons à présent appliquer les méthodes décrites précédemment à l'étude des crues en hydrologie. Il existe quelques particularités liées aux types de données utilisées, ou aux « habitudes » des hydrologues, mais le cadre théorique est celui de la statistique inférentielle.

Les données

Il n'existe pas de manière unique de décrire une crue ; on peut s'intéresser à une hauteur d'eau, un temps de submersion, une vitesse d'écoulement... La variable la plus utilisée par les hydrologues est le débit, car c'est la variable la moins influencée par les conditions locales aux sites de mesure, et c'est aussi celle qui est physiquement la plus maniable (lien avec la pluie). Malheureusement, un débit ne peut se mesurer instantanément, contrairement à une hauteur ou une température. Pour pouvoir disposer de chroniques de débits les plus continues possibles, on utilise une transformation hauteur/débit matérialisée par une **courbe de tarage**. Pour tracer cette courbe, on effectue quelques mesures conjointes de débit et de hauteur (*cf.* cours de Leblais E. pour la métrologie des débits), et on ajuste à ce nuage de points une courbe par régression (*figure A2.22*).

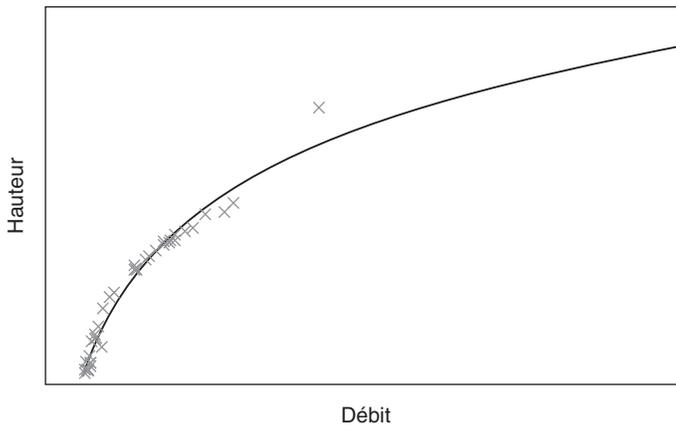


Figure A2.22. Courbe de tarage.

Cette transformation permet de passer d'une chronique de hauteurs, mesurées quasiment en continu, à une chronique de débits (*figure A2.23*).

Dans cette chaîne de traitements, plusieurs sources d'**incertitudes** apparaissent :

- Incertitude métrologique liée à la mesure de la hauteur.
- Incertitude métrologique et opératoire liée aux mesures de débits. Ce type d'erreur peut être assez important pour les débits de crues et d'étiages.

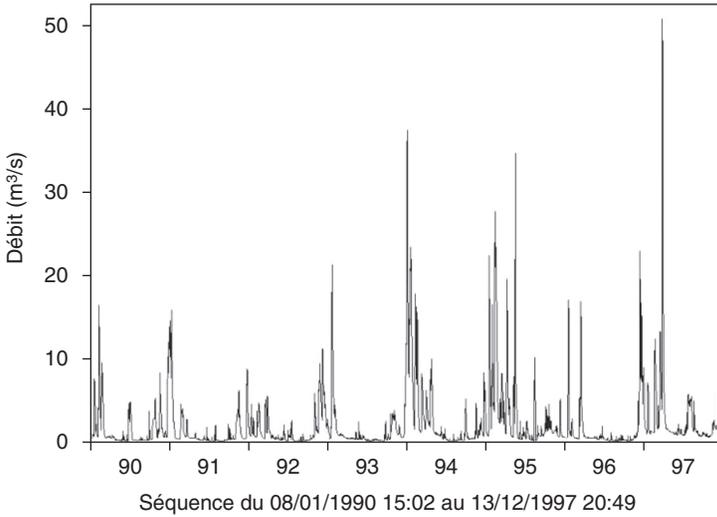


Figure A2.23. Chronique de débits.

- Incertitudes liées à l’ajustement d’une courbe de tarage : plusieurs types de courbes sont ajustables, et plusieurs méthodes d’ajustement peuvent être utilisées. Les lois de l’hydraulique permettent de guider cette étape.
- Incertitude liée à l’extrapolation : par nature, les événements extrêmes sont rares, donc difficiles à mesurer. De plus, effectuer un jaugeage en crue est une opération délicate, voire périlleuse (pour le matériel et les hommes). Ainsi, les débits très faibles ou très forts peuvent ne jamais avoir été jaugés, ce qui conduit, lors de la survenue d’un tel événement, à calculer le débit par une extrapolation éloignée du domaine d’observation, d’où un important risque d’erreur.

En général, toutes ces incertitudes ne sont pas prises en compte dans l’analyse probabiliste des crues, qui s’attache à quantifier la fluctuation d’échantillonnage. Il est cependant important de garder à l’esprit qu’un débit est, à la base, une grandeur entachée d’incertitudes. Il sera en particulier totalement inutile de travailler avec de trop nombreux chiffres significatifs.

L’échantillonnage

À partir d’une chronique de débit, il s’agit d’extraire des variables caractéristiques des crues. Nous nous intéresserons ici aux pics de débits, mais d’autres variables peuvent être étudiées : volume de crue, durée, temps de montée ou de descente, etc.

Techniques d’échantillonnage

Nous allons nous intéresser aux deux principales techniques utilisées par les hydrologues : l’échantillonnage par valeurs maximales annuelles (MAXAN) qui consiste à sélectionner chaque année le débit observé le plus fort (*figure A2.24*), et l’échantillonnage par valeurs supérieures à un seuil (SUPSEUIL) où, après avoir choisi un seuil, on sélectionne toutes les pointes des événements dépassant ce seuil (*figure A2.25*). Dans la pratique,

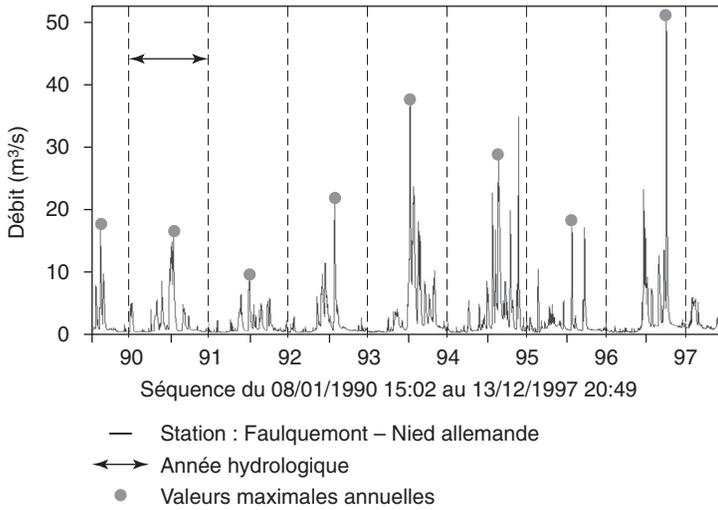


Figure A2.24. Échantillonnage de type MAXAN.

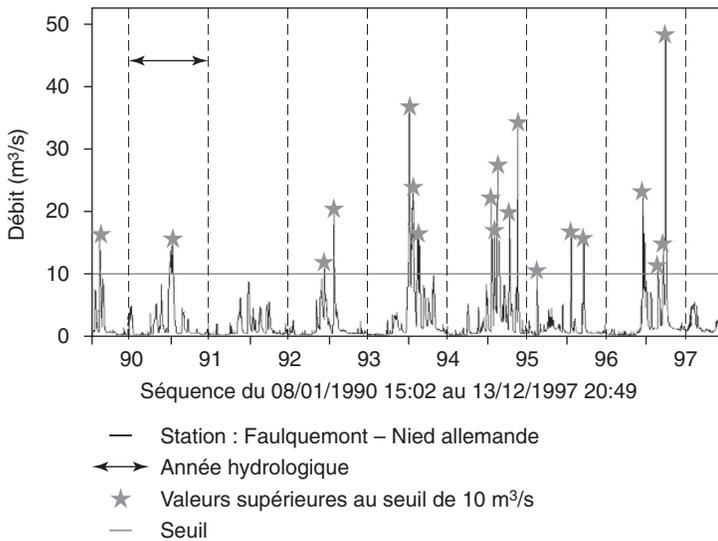


Figure A2.25. Échantillonnage de type SUPSEUIL.

plutôt qu'un seuil, on se donne en général un nombre d'événements à sélectionner par an (en moyenne), puis, par itérations successives, on calcule le seuil conduisant à cet objectif.

Chacune de ces méthodes présente des avantages et des inconvénients, notamment dans l'optique d'une analyse statistique où nous souhaiterions disposer d'un échantillon iid :

– MAXAN : la mise en œuvre de cette méthode est très simple. En ne sélectionnant qu'un unique événement par an, on s'assure de l'indépendance entre les valeurs, sauf éventuel-

lement si une crue importante se produit à cheval sur deux années successives, conduisant, par exemple, à échantillonner un débit le 31 décembre 2000 et un autre le 1^{er} janvier 2001. Pour cette raison, on échantillonne souvent par année hydrologique, chaque année étant séparée par une période d'étiage. L'inconvénient de cette méthode est d'ignorer un certain nombre d'événements lors des années où beaucoup de crues se sont produites, et inversement de prendre en compte des événements peu importants lors des années peu actives. L'homogénéité de l'échantillon n'est donc pas optimale.

– SUPSEUIL : cette technique est plus difficile à mettre en œuvre. En effet, il faut ajouter des contraintes d'indépendance afin de ne pas échantillonner plusieurs fois le même événement hydrologique, comme dans l'exemple de la *figure A2.26*.

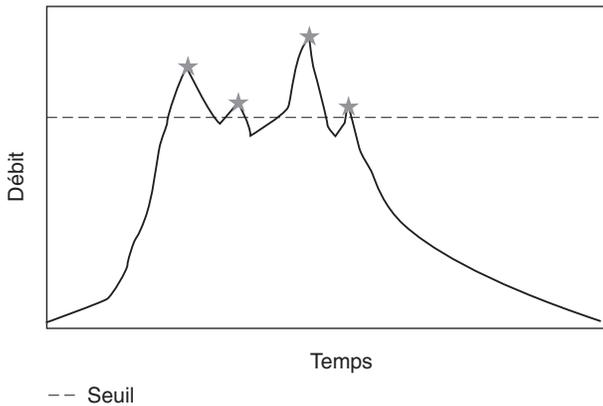


Figure A2.26. Illustration de la nécessité des contraintes d'indépendance.

On impose en général une contrainte d'espacement temporel minimal entre deux points sélectionnées, ainsi qu'une contrainte de redescente vers un débit de base. Bien choisies, ces contraintes permettent de garantir l'indépendance de l'échantillon. Cet échantillon sera également plus homogène que celui fourni par la méthode MAXAN. Autre avantage, il est possible d'étoffer l'échantillon en choisissant, en moyenne, plus d'un événement par an.

D'autres subtilités peuvent être employées pour améliorer l'échantillonnage. Citons notamment l'échantillonnage saisonnalisé qui permet, par exemple, d'obtenir deux échantillons pour les cours d'eau à régime mixte (pluvio-nival).

Propriétés probabilistes des échantillons

Nous noterons X^* la variable aléatoire génératrice des données fournies par la méthode MAXAN, et X_s la variable aléatoire génératrice des données fournies par la méthode SUPSEUIL.

L'expérience (et le théorème des valeurs extrêmes) montre que les lois suivantes sont bien adaptées pour décrire ces variables :

– X^* suit une loi généralisée des valeurs extrêmes (GEV) à trois paramètres. On suppose parfois directement que le paramètre de forme est nul, c.-à-d. que l'on choisit plutôt de modéliser X^* par une loi de Gumbel à deux paramètres :

- loi de Gumbel $Gu(\mu, \lambda)$:

$$f(x) = \frac{1}{\lambda} \exp\left(-\frac{x-\mu}{\lambda} - \exp\left(-\frac{x-\mu}{\lambda}\right)\right),$$

$$F(x) = \exp\left[-\exp\left(-\frac{x-\mu}{\lambda}\right)\right],$$

$$\mu > 0, \lambda > 0,$$

$$E(X) = \mu + \gamma\lambda, \text{ avec } \gamma = 0,5772 \text{ (constante d'Euler-Mascheroni),}$$

$$Var(X) = \frac{\pi^2}{6}\lambda^2 ;$$

- loi généralisée des valeurs extrêmes $GEV(\mu, \lambda, \xi)$:

$$f(x) = \frac{1}{\lambda} \left(1 - \frac{\xi(x-\mu)}{\lambda}\right)^{\frac{1}{\xi}-1} \exp\left(-\left(1 - \frac{\xi(x-\mu)}{\lambda}\right)^{\frac{1}{\xi}}\right),$$

$$F(x) = \exp\left(-\left(1 - \frac{\xi(x-\mu)}{\lambda}\right)^{\frac{1}{\xi}}\right),$$

$$\mu > 0, \lambda > 0, \xi \neq 0, 1 - \frac{\xi(x-\mu)}{\lambda} > 0,$$

$$E(X) = \mu + \frac{\lambda}{\xi}(1 - \Gamma(\xi + 1)),$$

$$Var(X) = \left(\frac{\lambda}{\xi}\right)^2 [\Gamma(2\xi + 1) - \Gamma^2(\xi + 1)].$$

– X_s suit une loi de Pareto généralisée à trois paramètres. De même que pour les échantillons MAXAN, on se contente souvent d'ajuster une loi exponentielle à deux paramètres, ce qui revient à supposer que le paramètre de forme est nul :

- Loi exponentielle $\exp(x_0, \lambda)$:

$$f(x) = \begin{cases} \frac{1}{\lambda} e^{-\frac{x-x_0}{\lambda}} & \text{si } x > x_0, \\ 0 & \text{sinon} \end{cases}$$

$$F(x) = \begin{cases} 1 - e^{-\frac{x-x_0}{\lambda}} & \text{si } x > x_0, \\ 0 & \text{sinon} \end{cases}$$

$$E(X) = x_0 + \lambda,$$

$$Var(X) = \lambda^2 ;$$

- Loi de Pareto généralisée $GP(x_0, \lambda, \xi)$:

$$f(x) = \frac{1}{\lambda} \left(1 - \frac{\xi(x - x_0)}{\lambda} \right)^{\frac{1}{\xi} - 1},$$

$$F(x) = 1 - \left(1 - \frac{\xi(x - x_0)}{\lambda} \right)^{\frac{1}{\xi}},$$

$$\lambda > 0, \xi \neq 0, 1 - \frac{\xi(x - x_0)}{\lambda} > 0,$$

$$E(X) = \frac{\lambda}{1 + \xi} + x_0,$$

$$Var(X) = \frac{\lambda^2}{(1 + \xi)^2(1 + 2\xi)}.$$

L'échantillonnage SUPSEUIL conduit à s'intéresser à une autre variable décrivant le processus d'occurrence des crues. Il est en effet possible de considérer une des variables suivantes :

– N_t le nombre de crues dans un intervalle de temps $[0, t]$. On suppose généralement que cette variable suit une loi de Poisson :

$$P(N_t = k) = e^{-\mu t} \cdot \frac{(\mu t)^k}{k!}.$$

D'autres lois sont parfois utilisées, comme la loi binomiale ou la loi binomiale négative.

– θ le temps séparant deux occurrences de crue, modélisé par une loi exponentielle simple :

$$P(0 \leq t) = \begin{cases} 1 - e^{-\mu t} & \text{si } t > 0 \\ 0 & \text{sinon} \end{cases}.$$

Exercice

Démontrez que si N_t suit une loi de poisson de paramètre μ , alors θ suit une loi exponentielle simple de même paramètre μ .

Il est possible de démontrer que si le processus d'occurrence suit une loi de Poisson et le processus de dépassement du seuil une loi exponentielle (échantillonnage SUPSEUIL), ces deux variables étant supposées indépendantes, alors le maximum annuel suit une loi de Gumbel (MAXAN). De même, il existe une relation reliant les lois de Poisson et de Pareto généralisée à la loi GEV.

Exercice

Démontrez que si X_s suit une loi exponentielle et N_t suit une loi de Poisson, avec X_s et N_t indépendantes, alors X^* suit une loi de Gumbel.

Estimation des paramètres des distributions

Une distribution ayant été choisie pour les variables X^* , X_s ou N_t , l'estimation des paramètres peut se faire par une des deux méthodes décrites à la section *Statistique inférentielle* : la méthode des moments ou la méthode du maximum de vraisemblance. Pour information, les hydrologues utilisent parfois une variante de la méthode des moments,

appelée méthode des L-moments, qui a pour but de rendre l'estimation plus robuste vis-à-vis des observations extrêmes. Nous nous contentons, *infra*, de donner les formules des estimateurs classiques.

Formulaire des estimateurs

Cf. tableau ci-contre.

Représentation graphique

Pour juger de la qualité de l'ajustement, les hydrologues représentent en général sur un même graphique les fonctions de répartition empirique et estimée. Cependant, par rapport aux représentations classiques que nous avons vues précédemment, il existe quelques originalités.

Nous allons illustrer ces différences à partir du jeu de données du *tableau A2.11*, qui représente les crues supérieures à $72 \text{ m}^3/\text{s}$ entre 1960 et 2002 pour l'Ubaye à Barcelonnette.

Tableau A2.11. Tableau de données.

<i>QJX</i> classé	Année	Jour	<i>QJX</i> classé	Année	Jour
72,6	1985	158	84,1	1983	160
73,5	1973	125	84,7	1994	267
74	1979	152	91	1983	136
75,6	1994	310	98,1	2000	165
75,8	1984	174	100	2001	151
77,7	1986	141	111	2000	289
79,9	1977	165	112	1978	162
83	1972	158	120	1963	320

Nous allons ajuster une loi exponentielle à cet échantillon. Les formules du formulaire des estimateurs nous donnent $\lambda = X_s - x_0 = 88,3 - 72 = 16,3$. On peut donc à présent tracer sur un même graphique la fonction de répartition estimée et la fonction de distribution empirique (c.-à-d. la courbe des fréquences cumulées dont nous ne représentons que les points observés en *figure A2.27*).

En hydrologie, il est fréquent que les axes soient inversés, c.-à-d. que l'on portera en abscisse la fréquence cumulée, ou **probabilité de non-dépassement**, et en ordonnée le débit correspondant (*figure A2.28*).

Distribution	Variabile	Méthode des moments	Maximum de vraisemblance
Exponentielle simple $Exp(\mu)$	θ , intervalle inter-occurrences	$\hat{\mu} = \frac{1}{\theta}$	Idem moments
Poisson $P(\mu)$	N_t , nombre d'événements dans $[0 ; t]$	$\hat{\mu} = \bar{N}_t = \frac{\text{Nombre de crues}}{\text{Durée d'observation}}$	Idem moments
Gumbel $Giu(\mu, \lambda)$	X^* , débit MAXXAN	$\begin{cases} \hat{\lambda} = \frac{\sqrt{6}}{\pi} S_{X^*} \\ \hat{\mu} = \bar{X}^* - \gamma \hat{\lambda} \end{cases}$	Méthode numérique
$GEV(\mu, \lambda, \xi)$	X^* , débit MAXXAN	$\begin{cases} -\frac{\hat{\xi}}{ \hat{\xi} } \left[\frac{\Gamma(3\hat{\xi} + 1) - 3\Gamma(\hat{\xi} + 1)\Gamma(2\hat{\xi} + 1) + 2\Gamma^2(\hat{\xi} + 1)}{ \hat{\xi} } \right] = \beta_1^{X^*} \\ \hat{\lambda} = \frac{ \hat{\xi} }{S_{X^*}} [\Gamma(2\hat{\xi} + 1) - \Gamma^2(\hat{\xi} + 1)]^{3/2} \\ \hat{\mu} = \bar{X}^* - \frac{\hat{\lambda}}{\hat{\xi}} [1 - \Gamma(\hat{\xi} + 1)] \end{cases}$ <p>La première équation est résolue numériquement. $\beta_1^{X^*}$ est le coefficient d'asymétrie de l'échantillon.</p>	Méthode numérique. Préférable, car l'existence des moments n'est pas assurée.
Exponentielle $Exp(x_0, \lambda)$	X_s , débit SUPSEUIL	$\hat{\lambda} = \bar{X}_s - x_0$ x_0 supposé connu (valeur du seuil).	Idem moments
Pareto généralisée $GP(x_0, \lambda, \xi)$	X_s , débit SUPSEUIL	$\hat{\lambda} = \frac{1}{2} (\bar{X}_s - x_0) \left[\frac{(\bar{X}_s - x_0)^2}{S_{X_s}^2} + 1 \right]$ $\hat{\xi} = \frac{1}{2} \left[\frac{(\bar{X}_s - x_0)^2}{S_{X_s}^2} - 1 \right]$ x_0 supposé connu (valeur du seuil).	Méthode numérique. Préférable, car l'existence des moments n'est pas assurée.

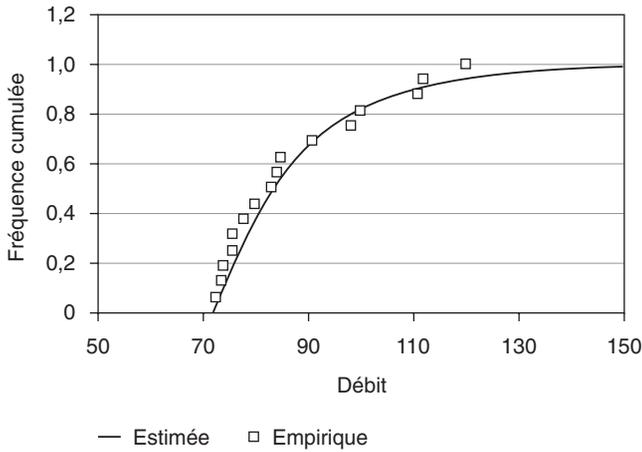


Figure A2.27. Fonction de répartition.

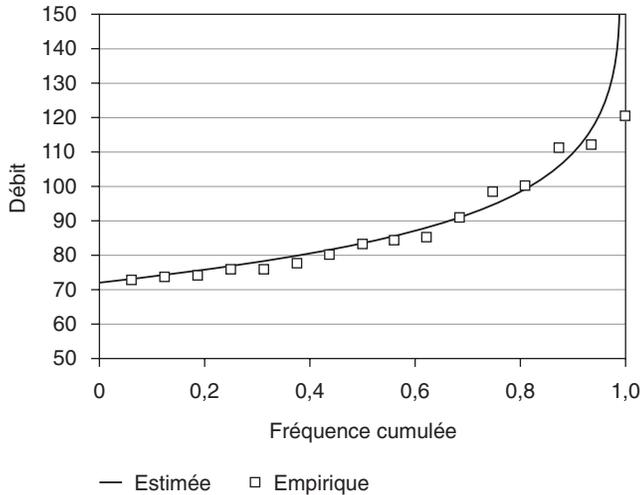


Figure A2.28. Fonction de répartition.

La seconde originalité est liée au calcul des fréquences cumulées empiriques : jusqu'ici, nous utilisons la formule $\frac{i}{N}$. Cependant, la fréquence cumulée étant une fonction des observations, c'est donc une statistique soumise à la fluctuation d'échantillonnage. Rien ne prouve que l'estimateur $\frac{i}{N}$ soit optimal dans ce cadre. En fait, des études ont montré que cet estimateur est biaisé pour les distributions que nous utilisons ; on lui préférera donc une formule du type : $\frac{i-a}{N+1-2a}$, avec $a = 0,3$. Notons que ceci ne change en rien nos estimations qui ne dépendent pas des fréquences empiriques, mais impacte seulement l'appréciation graphique.

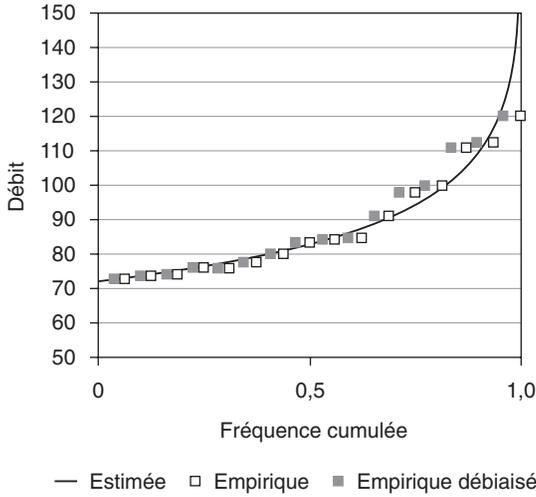


Figure A2.29. Fonction de répartition.

Enfin, il est d'usage de linéariser la figure A2.29, en effectuant une transformation sur l'axe des abscisses. Pour un échantillonnage SUPSEUIL, étant donné la définition de la

loi exponentielle, $F(x) = \begin{cases} 1 - e^{-\frac{x-x_0}{\lambda}} & \text{si } x > x_0, \\ 0 & \text{sinon} \end{cases}$, on tracera les débits en fonction de la

variable $-\log(1 - p)$ (figure A2.30). Pour un échantillonnage MAXAN, le changement de variable consistera à reporter en abscisses la variable $-\log(-\log(p))$.

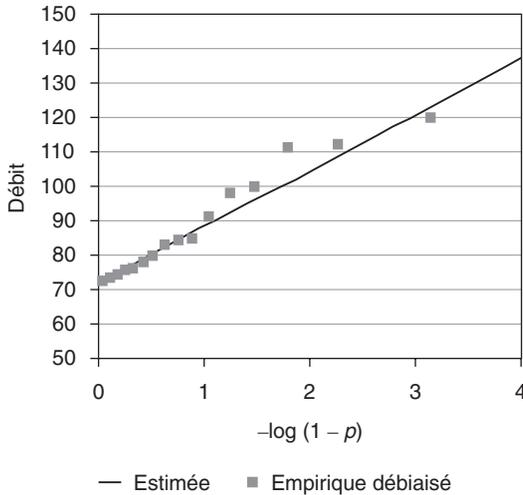


Figure A2.30. Fonction de répartition.

Détails sur le mode d'obtention de la fréquence empirique

Si l'on dispose d'un échantillon classé de taille N :

$$x(1) < \dots < x(i) < \dots < x(N),$$

on définit la probabilité au non-dépassement de la i -ème valeur par :

$$F(i) = \text{Prob}[X < x(i)].$$

La formule générale :

$$F(i) = \frac{i - a}{N + 1 - 2a}$$

donne de bons résultats pour les lois à deux paramètres (tableau A2.12).

Tableau A2.12. Valeur du coefficient a dans la formule de fréquence empirique.

Loi	Gauss	Gumbel	Exponentielle	Uniforme
a	3/8	0,44	0,44	0

La formule de Cunnane, avec $a = 0,4$, apparaît comme un bon compromis en termes d'indépendance vis-à-vis de la distribution parentale.

Pour les lois à trois paramètres, le choix optimum d'une fréquence empirique dépend du paramètre de forme. Cunnane (1978) a exposé les principes de choix d'une fréquence empirique. Comme le positionnement de la valeur $X(i)$ de rang i est sujet à la distribution d'échantillonnage (figure A2.31), on choisit la fréquence empirique F_i de sorte que le point expérimental se situe en moyenne sur la courbe théorique.

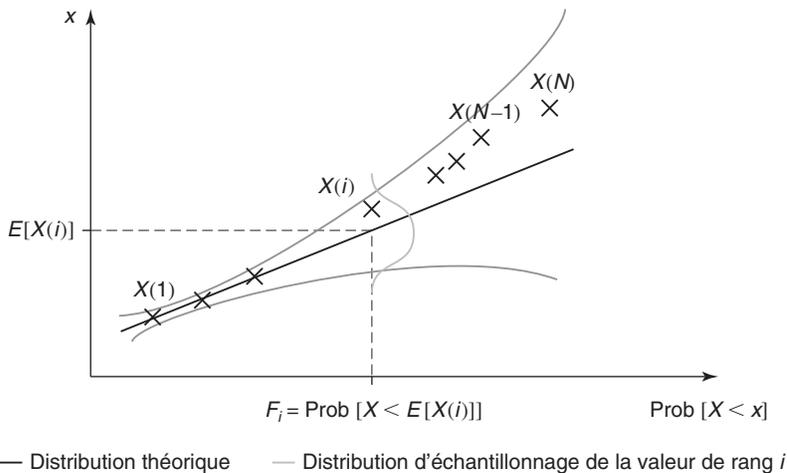


Figure A2.31. Principe du choix d'une formule de fréquence empirique.

Soit :

$$F_i = F\{E[X(i)]\}.$$

Le calcul de la valeur moyenne $E[X(i)]$ de la valeur $X(i)$ de rang i se développe à partir de l'expression de la densité :

$$\begin{aligned} f_{X(i)}(x) &= \text{Prob}[x < X(i) < x + dx] \\ &= \text{Prob}[(i-1) \text{ valeurs parmi } N < x] \text{Prob}[x < X < x + dx] \\ &\quad \text{Prob}[(N-i) \text{ valeurs parmi } N - (i-1) < x] \\ &= C_N^{i-1} F(x)^{i-1} f(x) C_{N-i+1}^{N-i} [1 - F(x)]^{N-i} \end{aligned}$$

soit :

$$f_{X(i)}(x) = \frac{N!}{(i-1)!(N-i)!} f(x) F(x)^{i-1} [1 - F(x)]^{N-i},$$

puis :

$$E[X(i)] = \int_0^{+\infty} x f_{X(i)}(x) dx.$$

Dans le cas d'une loi à deux paramètres x_0 et a , le calcul de $F\{E[X(i)]\}$ se trouve simplifié en raisonnant sur la variable transformée $U = \frac{X - x_0}{a}$, car on ne fait plus intervenir la valeur de ces deux paramètres.

$$\begin{aligned} F\{E[X(i)]\} &= \text{Prob}\{X < E[X(i)]\} \\ &= \text{Prob}\left\{\frac{X - x_0}{a} < \frac{E[X(i)] - x_0}{a}\right\} \\ &= \text{Prob}\{U < E[U(i)]\} \end{aligned}$$

À titre d'exemple, pour une loi exponentielle :

$$\begin{aligned} F_X(x) &= 1 - e^{-\frac{x-x_0}{a}}, \\ F_U(u) &= 1 - e^{-u}. \end{aligned}$$

La valeur de U de rang i a pour espérance¹ :

$$E[U(i)] = \sum_{j=1}^i \frac{1}{N+1-j},$$

dont la probabilité au non-dépassement :

$$F\{E[U(i)]\} = 1 - e^{E[U(i)]}$$

est correctement approchée par la formule de Gringorten :

$$F_i = \frac{i - 0,44}{N + 0,12}.$$

1. Sukhatme, 1938.

Commentaire sur le choix d'une formule de fréquence empirique

La formule optimale de fréquence empirique est donc fonction de la distribution étudiée. Dans la pratique, le choix d'une formulation n'est pas déterminant. En effet, la plupart des formules donnent des résultats voisins pour le corps de la distribution, et ne sont significativement différentes que pour les premiers points de la distribution. Or, le diagnostic sur l'adéquation entre la distribution expérimentale et la distribution théorique doit surtout porter sur le corps de la distribution, sachant que les points les plus forts ne peuvent être positionnés précisément (forte incertitude d'échantillonnage).

La formule de Chegodayev ($a = 0,3$), souvent utilisée au Cemagref, est une variante basée sur un positionnement médian de la courbe des fréquences empiriques ($F_i = F\{\text{Médiane}[X(i)]\}$) :

$$F_i = \frac{i - 0,3}{N + 0,4}.$$

Cas particulier d'un échantillon à valeurs tronquées

Dans le cas d'un échantillon mixte $x(1) \geq \dots \geq x(i) \geq \dots \geq x(N)$ composé d'une part de N_1 valeurs maximales annuelles (période récente), et d'autre part de k valeurs supérieures à un seuil S sur une période plus large de N années (incluant la période récente), avec N_2 crues appartenant à la période récente, Hirsch et Stedinger¹ proposent la formule suivante :

- pour les k crues les plus fortes : $F_i = 1 - \left(\frac{i}{k+1}\right) \cdot \left(\frac{k}{N}\right)$, pour ($i \leq k$) ;
- pour les autres crues : $F_i = \left(1 - \frac{k}{N}\right) \cdot \left(1 - \frac{i-k}{N_1 - N_2 + 1}\right)$, pour ($i > k$).

Quantiles, périodes de retour et intervalles de confiance

Les graphiques *supra* permettent, étant donné une probabilité p , de calculer le p -quantile de la distribution estimée des débits de crue. Ces quantiles sont des descripteurs du comportement des débits extrêmes sur un site bien précis. Ils ont également une importance réglementaire, en termes d'aménagement du territoire notamment, puisque les ouvrages à risque (centrales nucléaires, réseaux routiers, zones résidentielles ou industrielles) doivent être protégés d'un événement de probabilité donnée, dépendant de la vulnérabilité de l'aménagement. Dans le domaine de l'hydrologie, et plus généralement dans celui de la gestion des risques, la notion de probabilité de non-dépassement est généralement remplacée par la notion de **période de retour**. La période de retour T d'un événement est la durée *moyenne* qui sépare deux crues supérieures ou égales à cet événement. Inversement, une crue de période de retour T est une crue qui, *en moyenne*, est égalée ou dépassée toutes les T unités de temps. En général, l'unité de temps est l'année, et les relations suivantes lient les probabilités de non-dépassement et les périodes de retour :

- échantillonnage MAXAN : $1 - p = \frac{1}{T}$;
- échantillonnage SUPSEUIL avec μ valeurs par an : $1 - p = \frac{1}{\mu T}$.

1. Hirsch., Stedinger, 1987.

Il est bien important d'avoir conscience que malgré l'utilisation du terme « période », il n'y a pas de périodicité déterministe dans l'occurrence des crues ; il est donc tout à fait possible d'observer plusieurs crues de période de retour 10 ans au cours de la même année. Il faut dès lors bien garder à l'esprit que la période de retour n'est rien d'autre qu'une probabilité transformée.

Une crue de période de retour 10 ans ($p = 0,9$) est dite **décennale**, **centennale** si $T = 100$ ans ($p = 0,99$), **millénaire** si $T = 1\,000$ ans ($p = 0,999$).

Il est également possible de calculer analytiquement ces quantiles en inversant les fonctions de répartition. En effet, par définition, un quantile q_p vérifie $F(q_p) = p$, soit $q_p = F^{-1}(p)$.

Les formules suivantes permettent de calculer les quantiles :

- loi de Gumbel : $q_p = \mu - \lambda \log(-\log p)$;
- GEV : $q_p = \mu + \frac{\lambda}{\xi} [1 - (-\log p)^\xi]$;
- loi exponentielle : $q_p = x_0 - \lambda \log(1 - p)$;
- loi de Pareto généralisée : $q_p = x_0 + \frac{\lambda}{\xi} [1 - (1 - p)^\xi]$.

Dans la pratique, on calcule ces quantiles en remplaçant les paramètres par leur estimation. Cela implique que les quantiles, comme toute statistique, sont soumis aux fluctuations d'échantillonnage. Il est donc important de quantifier cette incertitude, de manière à bien cerner les limites de notre approche statistique. Il est en effet analytiquement possible de donner une valeur pour un quantile de période de retour 10 000 ans, mais, intuitivement, on sent bien que cette valeur ne signifierait pas grand-chose avec 20 ans de données disponibles pour l'analyse...

Le calcul des intervalles de confiance est assez fastidieux. Nous nous bornerons ici à donner quelques formules utiles. La formule suivante donne la variance des quantiles pour p paramètres estimés par la méthode des moments :

$$Var(\hat{q}_p) \approx \sum_{i=1}^p \left(\frac{\partial q_p}{\partial m_i} \right)^2 \cdot Var(m_i) + 2 \sum_{i=1}^p \sum_{j \neq i} \frac{\partial q_p}{\partial m_i} \cdot \frac{\partial q_p}{\partial m_j} \cdot cov(m_i, m_j).$$

En appliquant cette équation à l'exemple de la loi exponentielle *supra*, on obtient :

$$Var(\hat{q}_p) \approx [\log(1 - p)]^2 \cdot \frac{\sigma_{X_s}^2}{n}.$$

Si l'on suppose que le quantile est asymptotiquement gaussien et non biaisé, alors l'intervalle de confiance au niveau α est de la forme $\hat{q}_p \pm u_{1 - \frac{1 - \alpha}{2}} \sqrt{Var(\hat{q}_p)}$, où $u_{1 - \frac{1 - \alpha}{2}}$ est le quantile de la loi normale $N(0, 1)$. La *figure A2.32* présente l'intervalle de confiance à 90 % :

Cette hypothèse de normalité asymptotique des quantiles est à considérer avec précaution, surtout si l'effectif de l'échantillon n'est pas très important. Il existe des théorèmes proches du théorème central limite qui, théoriquement, garantissent cette normalité ; néanmoins, certaines lois que nous utilisons peuvent invalider les hypothèses de ces

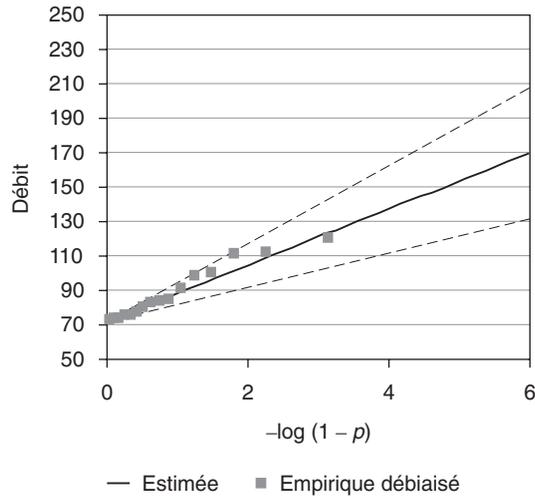


Figure A2.32. Fonction de répartition et intervalle de confiance à 90 %.

théorèmes. C'est notamment le cas des lois GEV et de Pareto généralisée qui peuvent avoir des moments infinis.

Plusieurs formules plus ou moins empiriques ont donc été établies pour calculer des intervalles de confiance plus réalistes. Pour une loi de Gumbel, on utilise en général l'approximation suivante :

– borne inférieure : $\hat{q}_p - h_1 \hat{\sigma}$;

– borne supérieure : $\hat{q}_p + h_2 \hat{\sigma}$;

avec :

$$h_1 = \frac{A - B}{C} \quad \text{et} \quad h_2 = \frac{A + B}{C} ;$$

$$A = u_{1 - \frac{1-\alpha}{2}} \cdot \sqrt{\frac{1 + 1,13t_p + 1,1t_p^2}{n}} ;$$

$$B = \left(u_{1 - \frac{1-\alpha}{2}}\right)^2 \cdot \frac{1,1t_p + 0,57}{n} ;$$

$$C = 1 - \frac{1,1}{n} \left(u_{1 - \frac{1-\alpha}{2}}\right) ;$$

$$t_p = \frac{-\log(-\log p) - 0,577}{1,28} .$$

Notons que si n est grand, alors B devient négligeable devant A , et l'intervalle de confiance devient symétrique, signe de convergence vers la normalité.

En *figure A2.33*, on dispose d'un exemple d'intervalle de confiance à 90 % à partir d'un échantillon MAXAN d'effectif 21.

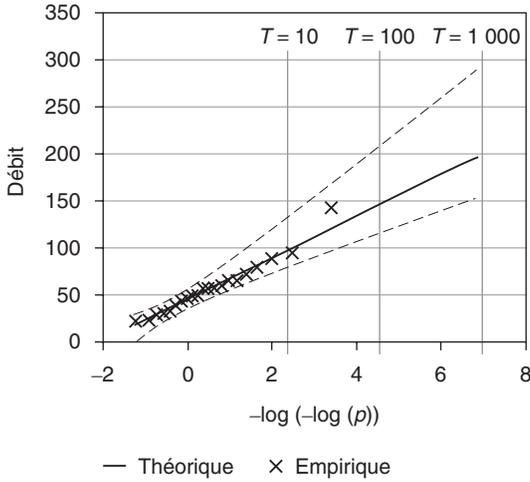


Figure A2.33. Fonction de répartition et intervalle de confiance à 90 %.

On voit clairement que cet intervalle n'est pas symétrique et que l'estimation de débits de périodes de retour 100 ou 1 000 ans avec seulement 20 années de données est illusoire, et ce, du fait que les valeurs varient du simple au double dans l'intervalle de confiance.

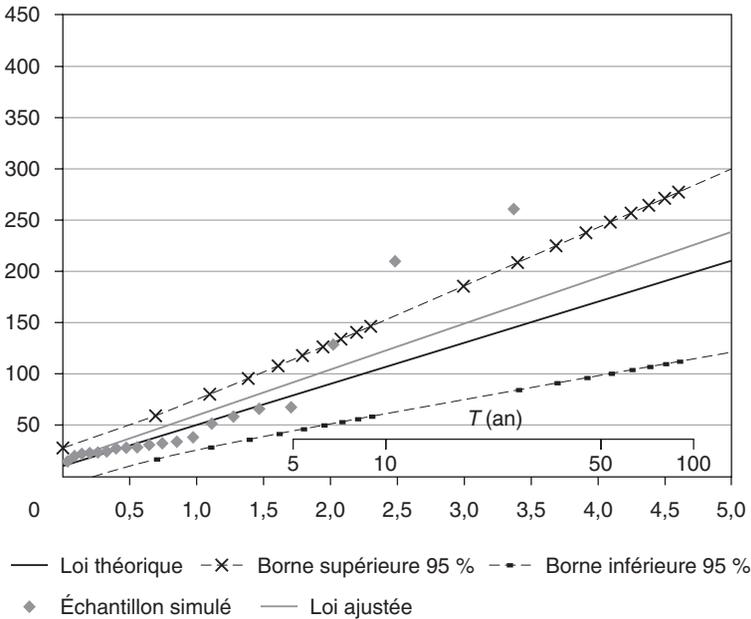


Figure A2.34. Intervalle de confiance à 90 % d'un quantile de crue issu d'une loi exponentielle.

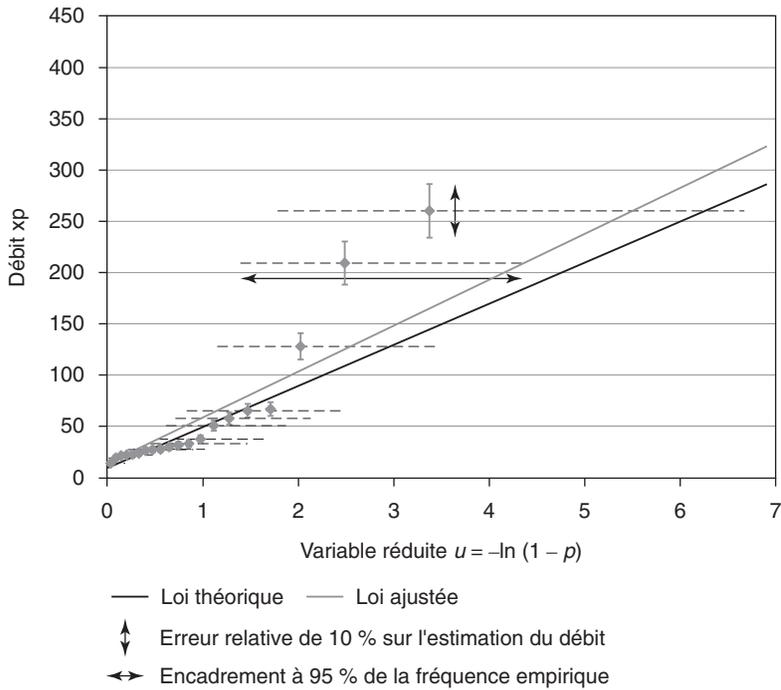


Figure A2.35. Incertitudes sur la fréquence empirique de valeurs issues d'une loi exponentielle.

Une alternative intéressante et simple à mettre en œuvre pour déterminer ces intervalles est l'utilisation des méthodes de rééchantillonnage, le *bootstrap* notamment, qui permettent de ne pas poser d'hypothèses a priori sur la distribution des quantiles.

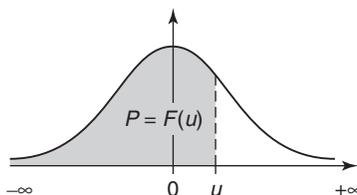
Enfin, il est à noter que la distribution empirique est elle-même entachée d'erreur d'échantillonnage puisque la valeur classée de rang i , la statistique d'ordre $X(i)$, varie suivant l'échantillon utilisé (à taille donnée). On peut ainsi associer un intervalle de confiance à la fréquence empirique $F_i = \text{Prob}[X < X(i)]$. La *figure A2.34* donne une représentation des incertitudes liées à l'estimation de la distribution des crues (incertitude à p fixé, sur la verticale), alors que la *figure A2.35* donne une indication de l'erreur sur la distribution empirique (incertitude à la fois sur les débits — trait vertical — et les fréquences — trait horizontal).

Tables usuelles

Cf. tableaux pages suivantes.

Fonction de répartition de la loi normale réduite

Probabilité de trouver une valeur inférieure à u .



P	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359	0,0
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753	0,1
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141	0,2
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517	0,3
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879	0,4
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224	0,5
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549	0,6
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852	0,7
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133	0,8
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389	0,9
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621	1,0
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830	1,1
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015	1,2
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177	1,3
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319	1,4
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441	1,5
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545	1,6
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633	1,7
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706	1,8
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767	1,9
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817	2,0
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857	2,1
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890	2,2
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916	2,3
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936	2,4
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952	2,5
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964	2,6
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974	2,7
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981	2,8
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986	2,9
	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	P

u	3	3,1	3,2	3,3	3,4	3,5	3,6	3,8	4	4,5
$F(u)$	0,998650	0,999032	0,999313	0,999517	0,999663	0,999767	0,999841	0,999928	0,999968	0,999997

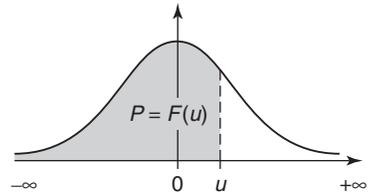
Lorsque u est négatif, il faut prendre le complément à l'unité de la valeur lue dans la table :

$$F(u) = 1 - F(-u).$$

Ex. : pour $u = 1,21$, $F(u) = 0,8869$; pour $u = -1,21$, $F(u) = 0,1131$.

Fractiles de la loi normale centrée réduite

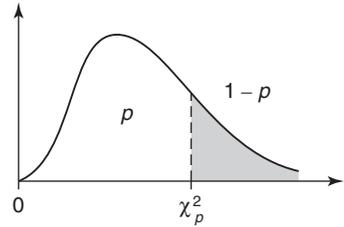
Valeur de u telle que $\text{Prob}(X < u) = 1 - P$
avec P donnée.



P	0,000	0,001	0,002	0,003	0,004	0,005	0,006	0,007	0,008	0,009	0,010	
0,00		3,0902	2,8782	2,7478	2,6521	2,5758	2,5121	2,4573	2,4089	2,3656	2,3263	0,00
0,01	2,3263	2,2904	2,2571	2,2262	2,1973	2,1701	2,1444	2,1201	2,0969	2,0748	2,0537	0,01
0,02	2,0537	2,0335	2,0141	1,9954	1,9774	1,9600	1,9431	1,9268	1,9110	1,8957	1,8808	0,02
0,03	1,8808	1,8663	1,8522	1,8384	1,8250	1,8119	1,7991	1,7866	1,7744	1,7624	1,7507	0,03
0,04	1,7507	1,7392	1,7279	1,7169	1,7060	1,6954	1,6849	1,6747	1,6646	1,6546	1,6449	0,04
0,05	1,6449	1,6352	1,6258	1,6164	1,6072	1,5982	1,5893	1,5805	1,5718	1,5632	1,5548	0,05
0,06	1,5548	1,5464	1,5382	1,5301	1,5220	1,5141	1,5063	1,4985	1,4909	1,4833	1,4758	0,06
0,07	1,4758	1,4684	1,4611	1,4538	1,4466	1,4395	1,4325	1,4255	1,4187	1,4118	1,4051	0,07
0,08	1,4051	1,3984	1,3917	1,3852	1,3787	1,3722	1,3658	1,3595	1,3532	1,3469	1,3408	0,08
0,09	1,3408	1,3346	1,3285	1,3225	1,3165	1,3106	1,3047	1,2988	1,2930	1,2873	1,2816	0,09
0,10	1,2816	1,2759	1,2702	1,2646	1,2591	1,2536	1,2481	1,2426	1,2372	1,2319	1,2265	0,10
0,11	1,2265	1,2212	1,2160	1,2107	1,2055	1,2004	1,1952	1,1901	1,1850	1,1800	1,1750	0,11
0,12	1,1750	1,1700	1,1650	1,1601	1,1552	1,1503	1,1455	1,1407	1,1359	1,1311	1,1264	0,12
0,13	1,1264	1,1217	1,1170	1,1123	1,1077	1,1031	1,0985	1,0939	1,0893	1,0848	1,0803	0,13
0,14	1,0803	1,0758	1,0714	1,0669	1,0625	1,0581	1,0537	1,0494	1,0451	1,0407	1,0364	0,14
0,15	1,0364	1,0322	1,0279	1,0237	1,0194	1,0152	1,0110	1,0069	1,0027	0,9986	0,9945	0,15
0,16	0,9945	0,9904	0,9863	0,9822	0,9782	0,9741	0,9701	0,9661	0,9621	0,9581	0,9542	0,16
0,17	0,9542	0,9502	0,9463	0,9424	0,9385	0,9346	0,9307	0,9269	0,9230	0,9192	0,9154	0,17
0,18	0,9154	0,9116	0,9078	0,9040	0,9002	0,8965	0,8927	0,8890	0,8853	0,8816	0,8779	0,18
0,19	0,8779	0,8742	0,8706	0,8669	0,8632	0,8596	0,8560	0,8524	0,8488	0,8452	0,8416	0,19
0,20	0,8416	0,8381	0,8345	0,8310	0,8274	0,8239	0,8204	0,8169	0,8134	0,8099	0,8064	0,20
0,21	0,8064	0,8030	0,7995	0,7961	0,7926	0,7892	0,7858	0,7824	0,7790	0,7756	0,7722	0,21
0,22	0,7722	0,7688	0,7655	0,7621	0,7588	0,7554	0,7521	0,7488	0,7454	0,7421	0,7388	0,22
0,23	0,7388	0,7356	0,7323	0,7290	0,7257	0,7225	0,7192	0,7160	0,7128	0,7095	0,7063	0,23
0,24	0,7063	0,7031	0,6999	0,6967	0,6935	0,6903	0,6871	0,6840	0,6808	0,6776	0,6745	0,24
0,25	0,6745	0,6713	0,6682	0,6651	0,6620	0,6588	0,6557	0,6526	0,6495	0,6464	0,6433	0,25
0,26	0,6433	0,6403	0,6372	0,6341	0,6311	0,6280	0,6250	0,6219	0,6189	0,6158	0,6128	0,26
0,27	0,6128	0,6098	0,6068	0,6038	0,6008	0,5978	0,5948	0,5918	0,5888	0,5858	0,5828	0,27
0,28	0,5828	0,5799	0,5769	0,5740	0,5710	0,5681	0,5651	0,5622	0,5592	0,5563	0,5534	0,28
0,29	0,5534	0,5505	0,5476	0,5446	0,5417	0,5388	0,5359	0,5330	0,5302	0,5273	0,5244	0,29
0,30	0,5244	0,5215	0,5187	0,5158	0,5129	0,5101	0,5072	0,5044	0,5015	0,4987	0,4958	0,30
0,31	0,4958	0,4930	0,4902	0,4874	0,4845	0,4817	0,4789	0,4761	0,4733	0,4705	0,4677	0,31
0,32	0,4677	0,4649	0,4621	0,4593	0,4565	0,4538	0,4510	0,4482	0,4454	0,4427	0,4399	0,32
0,33	0,4399	0,4372	0,4344	0,4316	0,4289	0,4261	0,4234	0,4207	0,4179	0,4152	0,4125	0,33
0,34	0,4125	0,4097	0,4070	0,4043	0,4016	0,3989	0,3961	0,3934	0,3907	0,3880	0,3853	0,34
0,35	0,3853	0,3826	0,3799	0,3772	0,3745	0,3719	0,3692	0,3665	0,3638	0,3611	0,3585	0,35
0,36	0,3585	0,3558	0,3531	0,3505	0,3478	0,3451	0,3425	0,3398	0,3372	0,3345	0,3319	0,36
0,37	0,3319	0,3292	0,3266	0,3239	0,3213	0,3186	0,3160	0,3134	0,3107	0,3081	0,3055	0,37
0,38	0,3055	0,3029	0,3002	0,2976	0,2950	0,2924	0,2898	0,2871	0,2845	0,2819	0,2793	0,38
0,39	0,2793	0,2767	0,2741	0,2715	0,2689	0,2663	0,2637	0,2611	0,2585	0,2559	0,2533	0,39
0,40	0,2533	0,2508	0,2482	0,2456	0,2430	0,2404	0,2378	0,2353	0,2327	0,2301	0,2275	0,40
0,41	0,2275	0,2250	0,2224	0,2198	0,2173	0,2147	0,2121	0,2096	0,2070	0,2045	0,2019	0,41
0,42	0,2019	0,1993	0,1968	0,1942	0,1917	0,1891	0,1866	0,1840	0,1815	0,1789	0,1764	0,42
0,43	0,1764	0,1738	0,1713	0,1687	0,1662	0,1637	0,1611	0,1586	0,1560	0,1535	0,1510	0,43
0,44	0,1510	0,1484	0,1459	0,1434	0,1408	0,1383	0,1358	0,1332	0,1307	0,1282	0,1257	0,44
0,45	0,1257	0,1231	0,1206	0,1181	0,1156	0,1130	0,1105	0,1080	0,1055	0,1030	0,1004	0,45
0,46	0,1004	0,0979	0,0954	0,0929	0,0904	0,0878	0,0853	0,0828	0,0803	0,0778	0,0753	0,46
0,47	0,0753	0,0728	0,0702	0,0677	0,0652	0,0627	0,0602	0,0577	0,0552	0,0527	0,0502	0,47
0,48	0,0502	0,0476	0,0451	0,0426	0,0401	0,0376	0,0351	0,0326	0,0301	0,0276	0,0251	0,48
0,49	0,0251	0,0226	0,0201	0,0175	0,0150	0,0125	0,0100	0,0075	0,0050	0,0025	0,0000	0,49
	0,000	0,001	0,002	0,003	0,004	0,005	0,006	0,007	0,008	0,009	0,010	P

Fractiles pour la loi du Chi-2 à ndl degrés de liberté

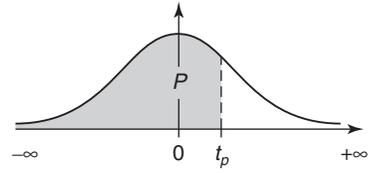
Valeur χ_p^2 telle que $\text{Prob}(X < \chi_p^2) = P$
avec P donnée.



ndl	$\chi^2_{0,005}$	$\chi^2_{0,01}$	$\chi^2_{0,025}$	$\chi^2_{0,05}$	$\chi^2_{0,10}$	$\chi^2_{0,25}$	$\chi^2_{0,50}$	$\chi^2_{0,75}$	$\chi^2_{0,90}$	$\chi^2_{0,95}$	$\chi^2_{0,975}$	$\chi^2_{0,99}$	$\chi^2_{0,995}$	$\chi^2_{0,999}$
1	0,0000	0,0002	0,0010	0,0039	0,0158	0,102	0,455	1,32	2,71	3,84	5,02	6,63	7,88	10-83
2	0,0100	0,0201	0,0506	0,103	0,211	0,575	1,39	2,77	4,61	5,99	7,38	9,21	10,60	13,82
3	0,0717	0,115	0,216	0,352	0,584	1,21	2,37	4,11	6,25	7,81	9,35	11,34	12,84	16,27
4	0,207	0,297	0,484	0,711	1,06	1,92	3,36	5,39	7,78	9,49	11,14	13,28	14,86	18,47
5	0,412	0,554	0,831	1,15	1,61	2,67	4,35	6,63	9,24	11,07	12,83	15,09	16,75	20,51
6	0,676	0,872	1,24	1,64	2,20	3,45	5,35	7,84	10,64	12,59	14,45	16,81	18,55	22,46
7	0,989	1,24	1,69	2,17	2,83	4,25	6,35	9,04	12,02	14,07	16,01	18,48	20,28	24,32
8	1,34	1,65	2,18	2,73	3,49	5,07	7,34	10,22	13,36	15,51	17,53	20,09	21,95	26,12
9	1,73	2,09	2,70	3,33	4,17	5,90	8,34	11,39	14,68	16,92	19,02	21,67	23,59	27,88
10	2,16	2,56	3,25	3,94	4,87	6,74	9,34	12,55	15,99	18,31	20,48	23,21	25,19	29,59
11	2,60	3,05	3,82	4,57	5,58	7,58	10,34	13,70	17,28	19,68	21,92	24,73	26,76	31,26
12	3,07	3,57	4,40	5,23	6,30	8,44	11,34	14,85	18,55	21,03	23,34	26,22	28,30	32,91
13	3,57	4,11	5,01	5,89	7,04	9,30	12,34	15,98	19,81	22,36	24,74	27,69	29,82	34,53
14	4,07	4,66	5,63	6,57	7,79	10,17	13,34	17,12	21,06	23,68	26,12	29,14	31,32	36,12
15	4,60	5,23	6,26	7,26	8,55	11,04	14,34	18,25	22,31	25,00	27,49	30,58	32,80	37,70
16	5,14	5,81	6,91	7,96	9,31	11,91	15,34	19,37	23,54	26,30	28,85	32,00	34,27	39,25
17	5,70	6,41	7,56	8,67	10,09	12,79	16,34	20,49	24,77	27,59	30,19	33,41	35,72	40,79
18	6,26	7,01	8,23	9,39	10,86	13,68	17,34	21,60	25,99	28,87	31,53	34,81	37,16	42,31
19	6,84	7,63	8,91	10,12	11,65	14,56	18,34	22,72	27,20	30,14	32,85	36,19	38,58	43,82
20	7,43	8,26	9,59	10,85	12,44	15,45	19,34	23,83	28,41	31,41	34,17	37,57	40,00	45,31
21	8,03	8,90	10,28	11,59	13,24	16,34	20,34	24,93	29,62	32,67	35,48	38,93	41,40	46,80
22	8,64	9,54	10,98	12,34	14,04	17,24	21,34	26,04	30,81	33,92	36,78	40,29	42,80	48,27
23	9,26	10,20	11,69	13,09	14,85	18,14	22,34	27,14	32,01	35,17	38,08	41,64	44,18	49,73
24	9,89	10,86	12,40	13,85	15,66	19,04	23,34	28,24	33,20	36,42	39,36	42,98	45,56	51,18
25	10,52	11,52	13,12	14,61	16,47	19,94	24,34	29,34	34,38	37,65	40,65	44,31	46,93	52,62
26	11,16	12,20	13,84	15,38	17,29	20,84	25,34	30,43	35,56	38,89	41,92	45,64	48,29	54,05
27	11,81	12,88	14,57	16,15	18,11	21,75	26,34	31,53	36,74	40,11	43,19	46,96	49,65	55,48
28	12,46	13,56	15,31	16,93	18,94	22,66	27,34	32,62	37,92	41,34	44,46	48,28	50,99	56,89
29	13,12	14,26	16,05	17,71	19,77	23,57	28,34	33,71	39,09	42,56	45,72	49,59	52,34	58,30
30	13,79	14,95	16,79	18,49	20,60	24,48	29,34	34,80	40,26	43,77	46,98	50,89	53,67	59,70
40	20,71	22,16	24,43	26,51	29,05	33,66	39,34	45,62	51,81	55,76	59,34	63,69	66,77	73,40
50	27,99	29,71	32,36	34,76	37,69	42,94	49,33	56,33	63,17	67,50	71,42	76,15	79,49	86,66
60	35,53	37,48	40,48	43,19	46,46	52,29	59,33	66,98	74,40	79,08	83,30	88,38	91,95	99,61
70	43,28	45,44	48,76	51,74	55,33	61,70	69,33	77,58	85,53	90,53	95,02	100,43	104,21	112,32
80	51,17	53,54	57,15	60,39	64,28	71,14	79,33	88,13	96,58	101,88	106,63	112,33	116,32	124,84
90	59,20	61,75	65,65	69,13	73,29	80,62	89,33	98,65	107,57	113,15	118,14	124,12	128,30	137,21
100	67,33	70,06	74,22	77,93	82,36	90,13	99,33	109,14	118,50	124,34	129,56	135,81	140,17	149,45

Centiles t_p pour la distribution t de Student à ndl degrés de liberté

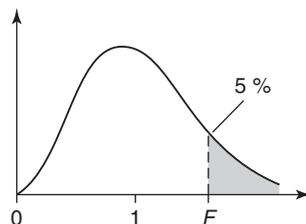
Valeur de t_p telle que $\text{Prob}(X < t_p) = P$
avec P donnée.



ndl	$t_{0,55}$	$t_{0,60}$	$t_{0,70}$	$t_{0,75}$	$t_{0,80}$	$t_{0,90}$	$t_{0,95}$	$t_{0,975}$	$t_{0,99}$	$t_{0,995}$
1	0,158	0,325	0,727	1,000	1,376	3,08	6,31	12,71	31,82	63,66
2	0,142	0,289	0,617	0,816	1,061	1,89	2,92	4,30	6,96	9,92
3	0,137	0,277	0,584	0,765	0,978	1,64	2,35	3,18	4,54	5,84
4	0,134	0,271	0,569	0,741	0,941	1,53	2,13	2,78	3,75	4,60
5	0,132	0,267	0,559	0,727	0,920	1,48	2,02	2,57	3,36	4,03
6	0,131	0,265	0,553	0,718	0,906	1,44	1,94	2,45	3,14	3,71
7	0,130	0,263	0,549	0,711	0,896	1,41	1,89	2,36	3,00	3,50
8	0,130	0,262	0,546	0,706	0,889	1,40	1,86	2,31	2,90	3,36
9	0,129	0,261	0,543	0,703	0,883	1,38	1,83	2,26	2,82	3,25
10	0,129	0,260	0,542	0,700	0,879	1,37	1,81	2,23	2,76	3,17
11	0,129	0,260	0,540	0,697	0,876	1,36	1,80	2,20	2,72	3,11
12	0,128	0,259	0,539	0,695	0,873	1,36	1,78	2,18	2,68	3,05
13	0,128	0,259	0,538	0,694	0,870	1,35	1,77	2,16	2,65	3,01
14	0,128	0,258	0,537	0,692	0,868	1,35	1,76	2,14	2,62	2,98
15	0,128	0,258	0,536	0,691	0,866	1,34	1,75	2,13	2,60	2,95
16	0,128	0,258	0,535	0,690	0,865	1,34	1,75	2,12	2,58	2,92
17	0,128	0,257	0,534	0,689	0,863	1,33	1,74	2,11	2,57	2,90
18	0,127	0,257	0,534	0,688	0,862	1,33	1,73	2,10	2,55	2,88
19	0,127	0,257	0,533	0,688	0,861	1,33	1,73	2,09	2,54	2,86
20	0,127	0,257	0,533	0,687	0,860	1,33	1,72	2,09	2,53	2,85
21	0,127	0,257	0,532	0,686	0,859	1,32	1,72	2,08	2,52	2,83
22	0,127	0,256	0,532	0,686	0,858	1,32	1,72	2,07	2,51	2,82
23	0,127	0,256	0,532	0,685	0,858	1,32	1,71	2,07	2,50	2,81
24	0,127	0,256	0,531	0,685	0,857	1,32	1,71	2,06	2,49	2,80
25	0,127	0,256	0,531	0,684	0,856	1,32	1,71	2,06	2,49	2,79
26	0,127	0,256	0,531	0,684	0,856	1,31	1,71	2,06	2,48	2,78
27	0,127	0,256	0,531	0,684	0,855	1,31	1,70	2,05	2,47	2,77
28	0,127	0,256	0,530	0,683	0,855	1,31	1,70	2,05	2,47	2,76
29	0,127	0,256	0,530	0,683	0,854	1,31	1,70	2,05	2,46	2,76
30	0,127	0,256	0,530	0,683	0,854	1,31	1,70	2,04	2,46	2,75
40	0,126	0,255	0,529	0,681	0,851	1,30	1,68	2,02	2,42	2,70
60	0,126	0,254	0,527	0,679	0,848	1,30	1,67	2,00	2,39	2,66
120	0,126	0,254	0,526	0,677	0,845	1,29	1,66	1,98	2,36	2,62
99999999	0,126	0,253	0,524	0,674	0,842	1,28	1,64	1,96	2,33	2,58

Distribution F (percentile 95 %)

La table donne le percentile 95 % (valeur ayant 5 chances sur 100 d'être égalée ou dépassée) de la fonction F en fonction des nombres de degrés de liberté du numérateur dln et du dénominateur dld avec $dln > dld$.



dld \ dln	1	2	3	4	5	6	7	8	9	10	15	20	40	60
1	161,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9	240,5	241,9	245,9	248,0	251,1	252,2
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,43	19,45	19,47	19,48
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,70	8,66	8,59	8,57
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,86	5,80	5,72	5,69
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,62	4,56	4,46	4,43
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	3,94	3,87	3,77	3,74
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,51	3,44	3,34	3,30
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,22	3,15	3,04	3,01
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,01	2,94	2,83	2,79
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,85	2,77	2,66	2,62
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,72	2,65	2,53	2,49
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,62	2,54	2,43	2,38
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,53	2,46	2,34	2,30
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,46	2,39	2,27	2,22
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,40	2,33	2,20	2,16
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,35	2,28	2,15	2,11
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,31	2,23	2,10	2,06
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,27	2,19	2,06	2,02
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,23	2,16	2,03	1,98
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,20	2,12	1,99	1,95
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,18	2,10	1,96	1,92
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,15	2,07	1,94	1,89
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,13	2,05	1,91	1,86
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,11	2,03	1,89	1,84
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,09	2,01	1,87	1,82
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,07	1,99	1,85	1,80
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20	2,06	1,97	1,84	1,79
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,04	1,96	1,82	1,77
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18	2,03	1,94	1,81	1,75
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,01	1,93	1,79	1,74
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	1,92	1,84	1,69	1,64
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,84	1,75	1,59	1,53
120	3,92	3,07	2,68	2,45	2,29	2,18	2,09	2,02	1,96	1,91	1,75	1,66	1,50	1,43
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,67	1,57	1,39	1,32

dln = nombre de degrés de liberté pour la plus grande estimation de fluctuation.

dld = nombre de degrés de liberté pour la plus petite estimation de fluctuation.

Tables du test de Kolmogorov-Smirnov

Valeur de dn telle que $\text{Prob}(Dn < dn) = P$.

n	$P = ,80$	$P = ,90$	$P = ,95$	$P = ,98$	$P = ,99$
1	,90000	,95000	,97500	,99000	,99500
2	,68377	,77369	,84189	,90000	,92929
3	,56481	,63604	,70760	,78456	,82900
4	,49265	,56522	,62394	,68887	,73424
5	,44698	,50945	,56328	,62718	,66853
6	,41037	,46799	,51926	,57741	,61661
7	,38148	,43607	,48342	,53844	,57581
8	,35831	,40962	,45427	,50654	,54179
9	,33910	,38746	,43001	,47960	,51332
10	,32260	,36866	,40925	,45662	,48893
11	,30829	,35242	,39122	,43670	,46770
12	,29577	,33815	,37543	,41918	,44905
13	,28470	,32549	,36143	,40362	,43247
14	,27481	,31417	,34890	,38970	,41762
15	,26588	,30397	,33760	,37713	,40420
16	,25778	,29472	,32733	,36571	,39201
17	,25039	,28627	,31796	,35528	,38086
18	,24360	,27851	,30936	,34569	,37062
19	,23735	,27136	,30143	,33685	,36117
20	,23156	,26473	,29408	,32866	,35241
21	,22617	,25858	,28724	,32104	,34427
22	,22115	,25283	,28087	,31394	,33666
23	,21645	,24746	,27490	,30728	,32954
24	,21205	,24242	,26931	,30104	,32286
25	,20790	,23768	,26404	,29516	,31657
26	,20399	,23320	,25907	,28962	,31064
27	,20030	,22898	,25438	,28438	,30502
28	,19680	,22497	,24993	,27942	,29971
29	,19348	,22117	,24571	,27471	,29466
30	,19032	,21756	,24170	,27023	,28987
31	,18732	,21412	,23788	,26596	,28530
32	,18445	,21085	,23424	,26189	,28094
33	,18171	,20771	,23076	,25801	,27677
34	,17909	,20472	,22743	,25429	,27279
35	,17659	,20185	,22425	,25073	,26897
36	,17418	,19910	,22119	,24732	,26532
37	,17188	,19646	,21826	,24404	,26180
38	,16966	,19392	,21544	,24089	,25843
39	,16753	,19148	,21273	,23786	,25518
40	,16547	,18913	,21012	,23494	,25205
41	,16349	,18687	,20760	,23213	,24904
42	,16158	,18468	,20517	,22941	,24613
43	,15974	,18257	,20283	,22679	,24332
44	,15796	,18053	,20056	,22426	,24060
45	,15623	,17856	,19837	,22181	,23798
46	,15457	,17665	,19625	,21944	,23544
47	,15295	,17481	,19420	,21715	,23298
48	,15139	,17302	,19221	,21493	,23059
49	,14987	,17128	,19028	,21277	,22828
50	,14840	,16959	,18841	,21068	,22604

n	$P = ,80$	$P = ,90$	$P = ,95$	$P = ,98$	$P = ,99$
51	,14697	,16796	,18659	,20864	,22386
52	,14558	,16637	,18482	,20667	,22174
53	,14423	,16483	,18311	,20475	,21968
54	,14292	,16332	,18144	,20289	,21768
55	,14164	,16186	,17981	,20107	,21574
56	,14040	,16044	,17823	,19930	,21384
57	,13919	,15906	,17669	,19758	,21199
58	,13801	,15771	,17519	,19590	,21019
59	,13686	,15639	,17373	,19427	,20844
60	,13573	,15511	,17231	,19267	,20673
61	,13464	,15385	,17091	,19112	,20506
62	,13357	,15263	,16956	,18960	,20343
63	,13253	,15144	,16823	,18812	,20184
64	,13151	,15027	,16693	,18667	,20029
65	,13052	,14913	,16567	,18525	,19877
66	,12954	,14802	,16443	,18387	,19729
67	,12859	,14693	,16322	,18252	,19584
68	,12766	,14587	,16204	,18119	,19442
69	,12675	,14483	,16088	,17990	,19303
70	,12586	,14381	,15975	,17863	,19167
71	,12499	,14281	,15864	,17739	,19034
72	,12413	,14183	,15755	,17618	,18903
73	,12329	,14087	,15649	,17498	,18776
74	,12247	,13993	,15544	,17382	,18650
75	,12167	,13901	,15442	,17268	,18528
76	,12088	,13811	,15342	,17155	,18408
77	,12011	,13723	,15244	,17045	,18290
78	,11935	,13636	,15147	,16938	,18174
79	,11860	,13551	,15052	,16832	,18060
80	,11787	,13467	,14960	,16728	,17949
81	,11716	,13385	,14868	,16656	,17840
82	,11645	,13305	,14779	,16526	,17732
83	,11576	,13226	,14691	,16428	,17627
84	,11508	,13148	,14605	,16331	,17523
85	,11442	,13072	,14520	,16236	,17421
86	,11376	,12997	,14437	,16143	,17321
87	,11311	,12923	,14355	,16051	,17223
88	,11248	,12850	,14274	,15961	,17126
89	,11186	,12779	,14195	,15873	,17031
90	,11125	,12709	,14117	,15786	,16938
91	,11064	,12640	,14040	,15700	,16846
92	,11005	,12572	,13965	,15616	,16755
93	,10947	,12506	,13891	,15533	,16666
94	,10889	,12440	,13818	,15451	,16579
95	,10833	,12375	,13746	,15371	,16493
96	,10777	,12312	,13675	,15291	,16408
97	,10722	,12249	,13606	,15214	,16324
98	,10668	,12187	,13537	,15137	,16242
99	,10615	,12126	,13469	,15061	,16161
100	,10563	,12067	,13403	,14987	,16081
$n > 100$	$1,073/\sqrt{n}$	$1,223/\sqrt{n}$	$1,358/\sqrt{n}$	$1,518/\sqrt{n}$	$1,629/\sqrt{n}$

Annexe 3

Références bibliographiques

Références citées dans le texte

- Antoine J.-M., 1992. *Les avatars de l'inondation, du risque et de l'aménagement dans la vallée de l'Ariège (XVII^e-XX^e siècles)*. Thèse de doctorat de géographie, université de Toulouse-Le Mirail, 495 p.
- Armand M., 2002. *Instruction technique. Études hydrologiques. Critique/reconstitution des données*. Grenoble, EDF/DTG, 6 p.
- Arnaud P., Lavabre J., 2000. La modélisation stochastique des pluies horaires et leur transformation en débits pour la prédétermination des crues, *Revue des Sciences de l'Eau*, (13/4) : 441-462.
- Arnaud P., Lavabre J., 2002. Coupled rainfall model and discharge model for flood frequency estimation, *Water Resources Research*, 38 (6), 11.1-11.11.
- Astier J., Desbordes M., Lefort P., Limandat A., 1993. *Projet de ligne TGV Méditerranée. Méthodologie et principes de calcul pour le dimensionnement hydraulique des ouvrages de franchissement des petits bassins versants*. 47 p.
- Benichou P., Le Breton O., 1986. Prise en compte de la topographie pour la cartographie des champs pluviométriques statistiques, *La Météorologie*, 7^e série : 23-34.
- Benito G., Lang M., Barriendos M., Llasat M. C., Frances F., Ouarda T., Thorndycraft V., Enzel Y., Bardossy A., Cœur D., Bobee B., 2004. Use of Systematic, Palaeoflood and Historical data for the improvement of flood risk estimation, *Review of scientific methods. Natural Hazard, Strategies and Applications in Natural Hazard Research using Historical Data*, Glade T., Lang M. (Eds), Kluwer Academic Publishers, 31 (3) : 623-643.
- Benoist J.-C., Birgand F., 2002. Les dispositifs de mesure des débits dans les bassins versants agricoles, *Ingénieries EAT*, (32) : 51-63.
- Bertrand-Krajewski J.-L., Laplace D., Joannis C., Chebbo G., 2000. *Mesures en hydrologie urbaine et assainissement*. Tec&Doc, 794 p.
- Bressand F., Golossov G., 1996. *Méthode de calcul des débits rares et exceptionnels d'eaux pluviales sur les petits bassins versants naturels situés dans l'arc méditerranéen*. DDE du Gard, service Eau et Environnement, 42 p.

- Brochot S., Marchi L., Lang M., 2002. L'estimation des volumes des laves torrentielles : méthodes disponibles et application au torrent du Poucet (Savoie), *Bull. Eng. Geol. Env.*, 61 : 389-402.
- Buishand T.A., 1982. Some Methods for Testing the Homogeneity of Rainfall Records, *J. of Hydrology*, (58) :11-27.
- Cayla O., 1993. *Probabilistic calculation of design floods: SPEED*. International Symposium on Engineering Hydrology, American Society of Civil Engineers, San Francisco, July.
- Cemagref, 1993. *Étude d'inondabilité du Rival*. Rapport Groupement de Lyon, Division Hydrologie-Hydraulique, 114 p.
- Cemagref, Météo-France, CETE Méditerranée, 2002. *Synthèse des débits de crue sur les régions PACA et Languedoc-Roussillon : présentation des résultats*.
- Cernesson F., Lavabre J., Masson J.-M., 1996. Stochastic model for generating hourly hyetographs, *Atmospheric Research*, 42 (1-4).
- CFGB, 1994. Les crues de projet des barrages : méthode du Gradex. Design Flood Determination by the Gradex Method. 18^e congrès CIGB-ICOLD, *Bulletin du Comité Français des Grands Barrages*, (2), 96 p.
- Champion M., 1858-1864. *Les inondations en France du VI^e siècle à nos jours*. Paris, Cemagref, 2000, 6 tomes (reprint).
- Cœur D., Lang M., 2000. L'information historique des inondations : l'histoire ne donne-t-elle que des leçons ?, *La Houille Blanche*, (2) : 79-84.
- CTGREF, SRAE, Diame S.-H., 1980-1982. *Synthèse nationale sur les crues des petits bassins versants*. Fascicule 2 : la méthode Socose ; Information Technique n° 8-2 (juin 1980) ; Fascicule 3 : la méthode Crupedix.
- Cunnane, C., 1973. A particular comparison of annual maxima and partial duration series methods of flood frequency prediction, *Journal of Hydrology*, (18) : 257-271.
- Cunnane C., 1978. Unbiased plotting positions, *Journal of Hydrology*, (37) : 205-222.
- Cunnane C., 1979. A note on the Poisson assumption in partial duration series models, *Water Resources Research*, 15 (2) : 489-494.
- Dalrymple T., 1960. Flood frequency analysis, *U.S. Geol. Surv. Water Supply Pap.*, 1543A.
- Davison A.C., Smith R.L., 1990. Models for Exceedances over High Thresholds, *J.R. Statist. Soc.*, B, 52 (3) : 393-442.
- Duband D., Michel C., Garros H, Astier J., 1988. *Évaluation des crues extrêmes et de la crue de projet par la méthode du Gradex*. San Francisco, CIGB, 16^e congrès.
- Edijatno, Michel C., 1989. Un modèle pluie-débit journalier à trois paramètres, *La Houille Blanche*, (2) : 113-121.
- Forray N., Auer J.-C., Brochot J.-F., Eraud Y., Favriau A., Ghio M., Lallement C., Lang M., Odier M., Scherer C., 1998. *Charte de qualité de l'hydrométrie ; code de bonnes pratiques*. Ministère de l'Environnement, Groupe qualité des données, banque Hydro, 50 p.
- Francou J., Rodier J.-A., 1967. Essai de classification des crues maximales observées dans le monde, *Cahiers de l'Orstom*, série Hydrologie, IV (3).
- Galéa G., Prudhomme C., 1994. Modèles Débit-durée-Fréquence et conceptualisation d'un hydrogramme de crue synthétique : validation sur le BVRE de Draix, *Hydrologie Continentale*, 9 (2) : 139-151.
- Galéa G., Prudhomme C., 1997. Notions de base et concepts utiles pour la compréhension de la modélisation synthétique des régimes de crue des bassins-versants au sens des modèles QdF, *Revue des Sciences de l'Eau*, 1 : 83-101.
- Galéa G., Ramez P., 1995. *Maîtrise du ruissellement et de l'érosion d'un coteau de vignoble : guide à l'usage des aménageurs*. Cemagref Éditions, coll. « Étude », n° 10, 128 p.

- Garçon R., Bossard E., Lang M., Le Clerc S., 2002. *Revisiter la notion de scénario hydrologique de référence pour la caractérisation du risque d'inondation*. Rapport commun EDF/DTG-Cemagref pour le compte du ministère de l'Écologie et du Développement Durable, Programme risque inondation RIO 1, 76 p.
- Garry G., Masson M., Ballais J.-L., 1996. *Cartographie des zones inondables : approche hydro-géomorphologique*. Paris La Défense, ministère de l'Équipement, ministère de l'Environnement, Les Éditions Ville et Territoires, 100 p.
- Greenwood J.A., Landwehr J.M., Matalas N.C., Wallis J.R., 1979. Probability weighted moments: definition and relation to parameters of several distributions expressible in inverse form, *Water Resources Research*, 15 (5) : 1049-1054.
- Guillot P., Duband D., 1967. *La méthode du Gradex pour le calcul de la probabilité des crues à partir des pluies*. Colloque international sur les crues et leur évaluation, Leningrad, 15 au 18 août, IASH.
- Gumbel E.J., 1958. *Statistics of extremes*. Columbia University Press.
- Hirsch R.M., Stedinger J.R., 1987. Plotting positions for historical floods and their precision, *Water Resources Research*, 23 (4) : 715-727.
- Hosking J.R., Wallis J.R., Wood E.F., 1985. An appraisal of the regional flood frequency procedure in the UK flood studies report, *Hydrology Science Journal*, (30) : 85-109.
- House K.P., Webb R.H., Baker V.R., Levish D. R., 2002. Ancient Floods, Modern Hazards: Principles and Applications of Paleoflood Hydrology, *Water science and Application*, 5 AGU.
- Jaccon G., 1987. *Manuel d'hydrométrie. Tome V : tracé de la courbe de tarage et calcul des débits*. Orstom, 225 p.
- Javelle P., 2001. *Caractérisation du régime des crues : le modèle débit-durée-fréquence convergent. Approche locale et régionale*. Thèse de doctorat, Grenoble, Institut national polytechnique, 268 p.
- Javelle P., Galéa G., Grésillon J.-M., 2000. L'approche débit-durée-fréquence : historique et avancées, *Revue des Sciences de l'Eau*, 13 (3) : 305-323.
- Javelle P., Ouarda T., Lang M., Bobée B., Galéa G., Grésillon J.-M., 2002. Development of regional flood-duration-frequency curves based on the index-flood method. *Journal of Hydrology*, 258 (1-4) : 249-259.
- Laborde J.-P., Mahous N., 1998. *Notice d'installation et d'utilisation du logiciel HYDROLAB*. UMR 5651 « Espace » du CNRS, 42 p.
- Lang M., 1995. *Les chroniques en hydrologie : Modélisation comparée par un système de gestion de bases de données relationnel et orienté-objet ; Traitements de base et intervalles de confiance des quantiles de crues ; Techniques d'échantillonnage par la méthode du renouvellement*. Thèse de doctorat, université J. Fourier Grenoble, Cemagref Lyon, 296 p.
- Lang M., 1997. New developments with Agregée, a statistical model using hydrometeorological information. In Oberlin G., Desbos E. (dir.), *FRIEND projects H-5-5 et H-1-1. Third report 1994-1997* », Cemagref Éditions, 181-191.
- Lang M., Claudet R., 2005. Les échelles de gravité sur les inondations : réflexion nationale et exemple dans l'Hérault, *La Houille Blanche*, (1) : 52-59.
- Lang M., Ouarda T., Bobée B., 1999. Towards operational guidelines for over-threshold modelling, *Journal of Hydrology*, (225) : 103-117.
- Lang M., Cœur D., Brochot S., Naulet R., 2003. Information historique et ingénierie des risques naturels : l'Isère et le torrent du Manival. Cemagref Éditions, 178 p.

- Lang M., Cœur D., Lallement C., Naulet R., 1998. Valorisation de l'information historique pour la prédétermination du risque d'inondation : application sur le bassin du Guiers, *Ingénieries EAT*, (16) : 3-13.
- Lang M., Bois P., Mestre O., Lubès-Niel H., Sauquet E., 2003. *Détection de changements éventuels dans le régime des crues*. Rapport PNRH, Convention INSU n° 02CV036, Cemagref Lyon, Lthe Grenoble, Météo-France Toulouse, UMR HydroSciences Montpellier, 75 p.
- Lang M., Naulet R., Recking A., Cœur D., Gigon C., 2002. Études de cas : l'analyse des pluies et crues extrêmes observées depuis 200 ans dans un bassin cévenol, l'Ardèche, *La Houille Blanche*, (6/7) : 131-138.
- Lang M., Perret C., Renouf E., Sauquet E., Paquier A., 2006. Incertitudes sur les débits de crue, *181^e session du Comité scientifique et technique « Valeurs rares et extrêmes de précipitations et de débits. Pour une meilleure maîtrise des risques »*, congrès de la SHF, Lyon, 15-16 mars, 103-112.
- Lang M., Renard B., Dindar L., Lemaitre F., Bois P., 2004. Use of a statistical test based on Poisson process for the detection of changes in peak-over-threshold series, *Hydrology: Science & Practice for the 21st Century, British Hydrological Society, Proceedings of the London Conference, July 2004*, 1 : 158-164.
- Lavabre J., Arnaud P., Masson J.-M., Folton N., 1998. *Apport de la modélisation de la pluie en débit pour la connaissance de la ressource en eau et la prédétermination des crues*. Congrès de la Société hydrotechnique de France, 25^e journées de l'hydraulique, Chambéry, 15-18/09/98, Publications SHF, tome 2 : 349-358.
- Le Clerc S., 2004. *Revisiter la notion de scénario hydrologique de référence pour la caractérisation des conséquences des crues*. Thèse de doctorat, université J. Fourier Grenoble, Cemagref Lyon, 373 p.
- Lowery M.D., Nash J.E., 1970. A comparison of methods of fitting the double exponential distribution, *Journal of Hydrology*, (10) : 259-275.
- Lubes H., Masson J.-M., Raous P., Tapiiau M., 1994. *Safarhy : logiciel de calculs statistiques et d'analyse fréquentielle adapté à l'évaluation du risque en hydrologie. Manuel de référence*. Orstom, université Montpellier II, version 2.111, mai.
- Madsen H., Rasmussen P.F., Rosbjerg D., 1997. Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events. 1. At-site modeling, *Water Resour. Res.*, 33 (4) : 747-757.
- Margoum M., 1992. *Estimation des crues rares et extrêmes : le modèle Agregee. Conceptions et premières validations* Thèse de doctorat, École des mines de Paris, Cemagref Lyon, GIS Hydrologie Friend-Amhy, 252 p.
- Margoum M., Oberlin G., Lang M., Weingartner R., 1994. Estimation des crues rares et extrêmes : principes du modèle Agregee, *Hydrologie Continentale*, 9 (1) : 85-100.
- Meunier M., 2001. Regional flow-duration-frequency model for the tropical island of Martinique, *Journal of Hydrology*, 247 (1-2) : 31-53.
- Michel C., 1982. *Extrapolation par la méthode du Gradex*. Note du 03-05, Cemagref Antony.
- Ministère de l'Écologie, du Développement et de l'Aménagement durables, 2007. *Guide de pilotage des études hydrauliques*.
- Miquel J., 1984. *Guide pratique d'estimation des probabilités de crue*. Paris, Eyrolles, 160 p.
- Muller A., 2006. *Comportement asymptotique de la distribution des pluies extrêmes en France*. Thèse de doctorat, université Montpellier II, Cemagref Lyon et Aix, 246 p.
- Naden P.S., Bayliss A.C., 1993. *Flood estimation: peak-over-threshold techniques*. MAFF Conf. of River and Coastal Engineers, Univ. of Loughborough, UK, 5-7 July, 9.1.1-9.1.18.

- 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-Martin-d'Ardèche*. Ph.D. dissertation, université J. Fourier Grenoble, INRS-Eau Québec, Cemagref Lyon, 322 p.
- Naulet R., Lang M., Ouarda T., Cœur D., Bobée B., Recking A., Moussay D., 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*, Benito G., Ouarda T. Bárdossy A., (guest Eds), (312) : 58-78.
- Neppel L., Bouvier C., Lavabre J., 2003. Abattement spatial des précipitations en Languedoc-Roussillon, *Hydrology of mediterranean and semi-arid regions*, IAHS, (278) : 276-285.
- Ouarda T., Lang M., Bobée B., Bernier J., Bois P., 1999. Synthèse de modèles régionaux d'estimation de crue utilisés en France et au Québec, *Revue des Sciences de l'Eau*, 12 (1) : 155-182.
- Paquet E., Gailhard J., Garçon R., 2006. Évolution des extrêmes hydrométriques en France à partir de séries observées, *181^e session du Comité scientifique et technique « Valeurs rares et extrêmes de précipitations et de débits. Pour une meilleure maîtrise des risques »*, congrès de la SHF, Lyon, 15-16 mars, 37-50.
- Payrastra O., 2005. *Faisabilité et utilité du recueil de données historiques pour l'étude des crues extrêmes de petits cours d'eau. Étude de cas de quatre bassins-versants affluents de l'Aude*. Thèse Enpc, 390 p.
- Pickands J., 1975. Statistical inference using extreme order statistics, *The Annals of Statistics*, 3 (1) : 119-131.
- Prudhomme C., 1995. *Modèles synthétiques des connaissances en hydrologie : application à la régionalisation des crues en Europe alpine et méditerranéenne*. Thèse de doctorat, université Montpellier II, Cemagref Lyon, 397 p.
- Reed D. et al., 1999. *Flood Estimation Handbook*. Wallingford (Royaume-Uni), Institute of Hydrology, 5 vol.
- Safege, 2000. *Étude globale pour une stratégie de réduction des risques dus aux crues du Rhône. Analyse hydrologique*. Territoire Rhône, 200 p.
- Sambou M.S., 2004. Comparaison par simulation de Monte-Carlo des propriétés de deux estimateurs du paramètre de la loi exponentielle : méthode du maximum de vraisemblance (MMV) et méthode des moindres carrés, *Revue des Sciences de l'Eau*, 17 (1) : 23-47.
- Saporta G., 1990. *Probabilités, Analyse de données et Statistiques*. Paris, Éditions Technip.
- Sauquet É., Javelle P., Le Clerc S., 2003. Description des régimes hydrologiques des hautes eaux : nouvelle formulation pour l'analyse en débit-durée-fréquence et applications en ingénierie, *Ingénieries EAT*, (34) : 3-15.
- Sheffer N.A., Enzel Y., Benito G., Grodek T., Poart N., Lang M., Naulet R., Cœur D., 2003. Paleofloods and historical floods of the Ardèche river, France, *Water Resources Research*, 39 (12) : 1376.
- Soil Conservation Service, 2003. *Curve number (SCS-CN) methodology*. Kluwer academic publishers, 513 p.
- Stanescu V., 1997. Les grandes crues en Europe. In Desbos E., Oberlin G. (dir.), *FRIEND projects H-5-5 et H-1-1. Third report 1994-1997*, Éditions de l'Unesco, Cemagref Éditions, 229-236.
- Stifel W.J., Mc Pherson M.B., 1964. In *Handbook of applied hydrology*, McGraw-Hill book Company, 20.8-20.9.

- Sukhatme P.V., 1938. Test of significance for samples of the χ^2 population with two degrees of freedom, *Ann. Eugen.*, (8) : 52-56.
- United States Department of Agriculture, 1985. *National Engineering Handbook*. United States Department of Agriculture, Soil Conservation Service, Section 4 : Hydrology.
- Van Te Chow, Maidment D.R., Mays L.W., 1998. *Applied of hydrology*. McGraw-Hill, 572 p.
- Water Resources Council, 1981. *Guidelines for determining flood flow frequency*. Bulletin 17B, available from Office of Water Data Coordination, US Geological Survey, Reston.
- Wonnacott T.H., Wonnacott R.J., 1991. *Statistiques*. Économica.

Quelques ouvrages francophones de référence en hydrologie et en statistiques

- Ambroise B., 1999. *La dynamique du cycle de l'eau dans un bassin versant – Processus, facteurs, modèles*. Bucarest, HGA, 200 p.
- Ceresta, 1986. *Aide-mémoire pratique des techniques statistiques*. Revue de statistique appliquée, vol. XXXIV, numéro spécial, 274 p.
- Dubreuil P., 1974. *Introduction à l'analyse hydrologique*. Paris, Masson, Orstom, 216 p.
- Llamas J., 1985. *Hydrologie générale : principes et applications*. Chicoutimi, Boucherville (Québec), Gaëtan Morin, 487 p.
- Miquel J., 1984. *Guide pratique d'estimation des probabilités de crue*. Paris, Eyrolles, 160 p.
- Remenieras G., 1986. *L'hydrologie de l'ingénieur*, Paris, Eyrolles, 2^e éd., 456 p.
- Roche M., 1963. *Hydrologie de surface*. Paris, Gauthier-Villars, Orstom, 430 p.
- Saporta G., 1990. *Probabilités, analyse de données et statistiques*. Paris, Éditions Technip, 493 p.
- Wonnacott T.H., Wonnacott R.J., 1991. *Statistiques*. Économica, 919 p.

Sites Web

- Cours « Hydrologie générale » du professeur André Musy. Laboratoire d'hydrologie et aménagements (Hydram), Institut des sciences et technologies de l'environnement (Iste), École Polytechnique Fédérale (EPFL). <<http://hydram.epfl.ch/e-drologie/>>.
- Cours « Les techniques de la statistique ». Collectif Cnam/CRA Languedoc-Roussillon/Agro de Montpellier/Université Montpellier II.
<<http://www.agro-montpellier.fr/cnam-lr/statnet>>
- Cours « Statistique descriptive » d'Alain Baccini. Université Toulouse III.
<<http://www.lsp.ups-tlse.fr/Besse/pub/sdm1.pdf>>
- Documents pédagogiques en statistique. Université de technologie de Compiègne.
<<http://www.hds.utc.fr/sy02/>>
- Projet SMEL (Statistique médicale en ligne) : initiation interactive à la statistique. Inria Rhône-Alpes, Université Paris V.
<<http://www.math-info.univ-paris5.fr/smel>>
- Site « Hydrologie.org ». Comité national français des sciences hydrologiques.
<<http://www.cig.ensmp.fr/~hydro/>>

Ce guide méthodologique, réalisé pour le compte du ministère chargé de l'Écologie, a pour objectif d'améliorer la qualité des études hydrologiques, d'assurer une certaine homogénéité dans le mode de détermination des crues de référence à l'échelle du territoire national, et d'exploiter au mieux les informations hydrologiques disponibles sur les événements passés. Il dresse un inventaire des différentes méthodes probabilistes de prédétermination des crues avec une série d'exemples d'application et une partie bibliographique (logiciels, statistique, hydrologie).

Il est destiné aux services de l'État chargés de l'instruction des PPR (plans de prévention des risques), et aux bureaux d'études et étudiants concernés par le risque inondation.

Michel Lang, hydrologue, est responsable de l'unité de recherche Hydrologie-Hydraulique au Cemagref de Lyon (HHLY). Il travaille sur les méthodes d'estimation des crues extrêmes..

Jacques Lavabre, ingénieur de recherche, animateur de l'équipe Hydrologie au Cemagref d'Aix-en Provence, est spécialisé en hydrologie rurale et forestière.

Prix TTC : 49 €

ISBN : 978-2-7592-0067-2

éditions
Quæ

 **Cemagref**



Éditions Cemagref, Cirad, Ifremer, Inra
www.quae.com

ISSN : 1773-7923
Réf. : 02054