



HAL
open science

Caractérisation du régime des crues : le modèle débit-durée-fréquence convergent. Approche locale et régionale

P. Javelle

► **To cite this version:**

P. Javelle. Caractérisation du régime des crues : le modèle débit-durée-fréquence convergent. Approche locale et régionale. Sciences de l'environnement. Doctorat Mécanique de milieux géophysiques et environnement, Institut national polytechnique de Grenoble, 2001. Français. NNT : . tel-02579661

HAL Id: tel-02579661

<https://hal.inrae.fr/tel-02579661>

Submitted on 14 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

INSTITUT NATIONAL POLYTECHNIQUE DE GRENOBLE

\\\nN° Bibliothèque

THESE

pour obtenir le grade de

DOCTEUR DE L'INPG

Spécialité : Mécanique des Milieux Géophysiques et Environnement
dans le cadre de l'Ecole Doctorale « Terre, Univers, Environnement »

préparée dans l'Unité de Recherche Hydrologie-Hydraulique, Cemagref (Lyon)

présentée et soutenue publiquement par

Pierre JAVELLE

le 1^{er} mars 2001

Caractérisation du régime des crues : le modèle débit-durée-fréquence convergent Approche locale et régionale

Directeur de thèse Jean-Michel GRESILLON

Composition du jury

G. DE MARSILY	Université Paris VI	Président
L. GOTTSCHALK	Université d'Oslo	Rapporteur
E. SERVAT	IRD	Rapporteur
T. B. M. J. OUARDA	INRS-Eau	Examineur
G. GALEA	Cemagref	Examineur (co-encadrant)

Remerciements

Je remercie toutes les personnes qui m'ont permis d'aller au bout de ce travail. Sans elles rien n'aurait été possible.

Pierrick Givone, chef de l'Unité de Recherche en Hydrologie Hydraulique au Cemagref, puis Nicolas Gendreau qui lui a succédé, m'ont accueilli au sein de leur équipe. Grâce à eux, j'ai bénéficié de conditions matérielles idéales.

Un immense merci à Gilles Galéa et Michel Lang qui ont encadré mon travail au Cemagref. L'un a su me faire partager toute son expérience « d'expert », l'autre m'a initié aux lois de Poisson et autres Binomiales Négatives. Tous deux ont su se montrer disponibles, à chaque fois que j'avais besoin d'aide.

Jean-Michel Grésillon m'a fait confiance depuis mes premiers pas en « hydrologie » bien avant qu'il ne soit mon Directeur de Thèse. Son soutien a été sans faille, ne ménageant ni son temps, ni sa patience. Il m'a fait profiter de sa rigueur et de son esprit synthèse ; il a su me redonner confiance dans les moments difficiles. Je n'oublierai jamais cette réunion de travail en plein air, dans le cadre magnifique de la Vanoise. Je m'excuse auprès de lui et de sa famille pour les soirées, les week-ends et les jours de vacances qu'il a dû sacrifier pour pouvoir relire mon manuscrit.

Il ne peut y avoir de thèse sans jury. Ghislain de Marsily a accepté d'en être le président, Lars Gottschalk et Eric Servat, les rapporteurs. Je les remercie pour le temps qu'ils m'ont consacré. Grâce à leur intérêt et leurs critiques constructives, ma soutenance de thèse a donné lieu à un réel échange, qui fut pour moi très enrichissant.

Bernard Bobée et Taha Ouarda m'ont accueilli lors de mon séjour à Québec, à l'INRS-Eau. J'ai ainsi pu me familiariser avec d'autres approches en hydrologie. Taha Ouarda s'est personnellement impliqué dans mon travail. J'ai grâce à lui pu bénéficier de toute une base de données relative au Québec et à l'Ontario.

Joël Noilhan a été mon « parrain » auprès du CNES, c'est-à-dire qu'il était chargé de donner un avis extérieur sur le déroulement de mes travaux et la reconduction ou non de ma bourse. Grâce à lui, le CNES ne m'a pas coupé les vivres !

Une collaboration étroite s'est instaurée avec l'UMR 3S Cemagref Engref de Montpellier (anciennement Laboratoire Commun de Télédétection), dans le cadre du projet Noah. J'ai été accueilli à plusieurs reprises au « LCT » par Sylvain Labbé, Christian Puech, et Julie Charleux-Demargne. J'ai pu profiter de leurs compétences en Système d'Information Géographique, mais aussi du soleil montpelliérain ! Je souhaite à Julie bonne chance pour sa toute proche soutenance de thèse.

Mes nombreux passages sur le campus universitaire de Grenoble m'ont été très utiles, en particulier grâce à Philippe Bois et Charles Obled pour leurs conseils, Daniel Duband qui m'a fait profiter de son expérience, notamment concernant la *ridge regression* ; Maurice Meunier qui s'est intéressé à mon travail et y a apporté sa contribution.

Cette thèse est illustrée par quelques gravures représentant la crue de l'Isère à Grenoble en 1859. Celles-ci ont été fournies par Denis Cœur qui termine une thèse d'histoire à l'Université Pierre Mendès France de Grenoble : « *La maîtrise des inondations dans la plaine de Grenoble (XVIIe -XXe s.) : enjeux techniques, politiques et urbains* ».

Il n'est pas possible de citer toutes les personnes qui ont fait que ces trois années de dur labeur au Cemagref sont passées en un éclair. En voici quelques unes, que les oubliés me pardonnent : Frédéric, Sandrine et Nathalie, mes différents compagnons de bureau, unis dans la même « galère », Etienne, Christine, Cécile qui ont relu ma thèse en un temps record, Eric qui même lorsqu'il est outre-manche déverse par mail sa gentillesse corrosive, Pierre F., David, Heckel et Jeckel pour leur Equipe Super Détonante et Inventive, Pascal R. José, Céline, Marie-B pour les sorties terrain ; Sébastien P., Pierre B. pour leur philosophie de la vie ; Sébastien G. pour nos vidéos conférences entre les bâtiments Rhône et Saône ; Robin pour sa télé ; Karim pour sa Chekchouka ; Hélène F. et Anne pour avoir résolu avec le sourire toutes mes tracasseries administratives ; Erwan pour la Bretagne ; Hélène P. et Christine F. pour leurs pauses thé ; Pascal B. pour son Beaujolais, Sébastien S. pour son efficacité et tous les autres pour m'avoir pardonné de les avoir oubliés...

Grenoble et les amis que j'y ai m'ont permis de m'aérer l'esprit. Je pense à Mathieu pour nos pause midi ; Greg pour ses rollers ; Aubin pour l'aviron ; Alain pour s'être penché avec moi sur ce problème crucial : « Comment sauver la planète ? ».

Je remercie aussi bien sûr mes parents sans qui je ne serais pas en vie.

Et pour terminer, une mention spéciale pour Daniela, qui a supporté mes sautes d'humeur, a corrigé mes phautes d'ortagrafes, et organisé le pot de thèse.

Résumé

Caractérisation du régime des crues : le modèle débit-durée-fréquence convergent. Approche locale et régionale

L'objectif de ce travail est de développer un modèle fréquentiel capable de rendre compte du régime des crues d'un bassin versant. L'approche retenue est l'analyse débit-durée-fréquence (QdF) car elle intègre indirectement l'évolution des débits avec le temps, de façon analogue à l'approche intensité-durée-fréquence (IdF) pour les pluies. Le modèle proposé permet pour un bassin donné, de calculer ses courbes QdF à l'aide d'un minimum de paramètres. Ce modèle est appelé modèle "convergent", en raison de la convergence observée des différentes distributions vers les faibles périodes de retour. Une approche locale, puis régionale est développée. Dans ce dernier cas, le modèle s'inspire de méthodes classiquement utilisées en hydrologie régionale (méthode de l'indice de crue), et peut s'appliquer à des bassins ne disposant d'aucune mesure sur les débits. Les résultats, obtenus sur environ deux cents bassins, en France et au Canada, sont encourageants et ouvrent de réelles perspectives d'application.

Mots clefs

Régime hydrologique, crue, analyse fréquentielle, approche débit-durée-fréquence, modèle convergent, régionalisation

Abstract

Flood regime characterisation : the converging flow-duration-frequency model. Local and regional approach

The aim of this work is to develop a statistical model describing the floods regime of a catchment. We adopted an approach based on the flow-duration-frequency (QdF) analysis, which takes into account the temporal variability of floods. This approach is analogous to the intensity-duration-frequency (IdF) applied to the rainfall. The proposed model makes it possible to calculate QdF curves of a basin, using a minimum of parameters. This model is called "converging" model, because of the observed convergence of distributions, towards small return periods. A local and regional approach is developed. In this last case, the model uses classical concepts of regional hydrology (index flood method) and can be applied to basins where no flow data is available. Results obtained from about two hundred catchments, in France and Canada, are fairly good and open new perspectives of application.

Key words

Flow regime, flood frequencial analysis, flow-duration-frequency approach, converging model, regionalization

Sommaire

Introduction

Objectif de la thèse et présentation des données utilisées 1

PARTIE I : ANALYSE LOCALE

Chapitre 1

Description du régime hydrologique : introduction à l'approche débit-durée-fréquence 13

1. La notion de régime hydrologique 14
 - 1.1. Définition du régime hydrologique 14
 - 1.2. Régimes et échelles temporelles 14
 - 1.3. L'information de base : la chronique de débit 15
2. Variations saisonnières des débits interannuels : le choix du pas de temps 19
 - 2.1. Le cas du pas de temps mensuel : les régimes de Pardé 19
 - 2.2. Variations interannuelles au pas de temps journalier 20
 - 2.3. Exemple montrant l'importance du choix du pas de temps 22
 - 2.4. Commentaire sur les différentes « durées » de l'analyse 24
3. Analyse du régime des crues : introduction à l'approche débit-durée-fréquence 26
 - 3.1. Echantillonnage des valeurs maximales 27
 - 3.2. La prise en compte de la date de crues 28
 - 3.3. Les limites de l'analyse fréquentielle des pointes de crues 31
 - 3.4. Analyse des durées de crue : définition de la durée de SOCOSE 32
 - 3.5. L'analyse combinée pointe-volume des crues 35
 - 3.6. L'approche débit-durée-fréquence 36

Chapitre 2

Echantillonnage par valeurs supérieures à un seuil 41

1. Eléments théoriques sur la modélisation des valeurs supérieures à un seuil 42
 - 1.1. Etude du processus d'occurrence : le modèle du renouvellement 42
 - 1.2. Etude de la distribution fréquentielle des débits maximaux 47
2. Outils et tests aidant au choix des paramètres d'échantillonnage 49
 - 2.1. Les critères d'indépendances $COEF1$ et $COEF2$ 49
 - 2.2. Etude du nombre moyen d'événements sélectionnés par an en fonction du seuil S 50
 - 2.3. Test vérifiant si l'indice de dispersion I_1 est égal à 1 51
 - 2.4. Test vérifiant la stationnarité du processus de Poisson 51
 - 2.5. Outils développés 52
3. Résultats obtenus 54
 - 3.1. Bassins disponibles et critique des données 54
 - 3.2. Influence du nombre d'événements sélectionnés $\mu_{objectif}$ 56
 - 3.3. Echantillonnage utilisant les critères d'indépendance $COEF1$ et $COEF2$ 61
 - 3.4. Comparaison des quantiles obtenus avec et sans critères d'espacement 64
 - 3.5. Discussion sur le respect des hypothèses de la loi de Poisson 66

Chapitre 3

Modélisation débit-durée-fréquence : le modèle local convergent 71

1. Les différentes étapes de l'approche débit-durée-fréquence	72
1.1. Echantillonnage « multi-durées »	72
1.2. Objectif : une modélisation « continue »	72
1.3. Quelques modèles possibles	75
2. Présentation du modèle <i>QdF</i> convergent	76
2.1. Hypothèses et formulation du modèle	76
2.2. Ajustement des paramètres du modèle	80
2.3. Validation du modèle sur les trois exemples étudiés	87
2.4. Signification et propriété du paramètre Δ	89
3. Validation du modèle <i>QdF</i> convergent sur 105 bassins français	92
3.1. Automatisation de l'échantillonnage	92
3.2. Validation du modèle <i>QdF</i> convergent	97
3.3. Corrélation entre Δ et la durée de Socose	101
3.4. Application : désagrégation du pas de temps	102
4. Tentative d'interprétation des propriétés	103
4.1. Hypothèse explicative : l'affinité des hydrogrammes de crue	103
4.2. Vérification de l'hypothèse sur l'exemple de l'Orgeval au Theil	105

Chapitre 4

Autres développements du modèle *QdF* local convergent 109

1. Utilisation d'une loi statistique à trois paramètres sur de longues séries	110
1.1. Présentation de la loi <i>GEV</i> (Generalised Extreme Value)	110
1.2. Généralisation des hypothèses du modèle convergent	112
1.3. Application au modèle local convergent	113
1.4. Conclusion partielle	115
2. Etudes des « débits seuils » maximaux	116
2.1. Définition des débits seuils	116
2.2. Propriété expérimentale observée entre les courbes <i>QdF</i> en débits seuils et en débits moyens	118
2.3. Application au modèle <i>QdF</i> local convergent	122
2.4. Interprétation des propriétés observées	125
3. Etudes des étiages	128
3.1. Variables étudiées et échantillonnage	128
3.2. Illustration de la propriété d'affinité des hydrogrammes d'étiages	130
3.3. Interprétation des propriétés en terme de forme des hydrogrammes	132
3.4. Conclusion	134

PARTIE II : ANALYSE REGIONALE

Chapitre 5

Présentation du modèle *QdF* régional convergent 139

1. Quelques rappels sur l'analyse régionale	140
1.1. Qu'est-ce que l'analyse régionale ?	140
1.2. Constitution d'une région hydrologique homogène	141
1.3. Estimation des quantiles de crue par régressions multiples	143
1.4. La méthode de l'indice de crue	143
1.5. Remarques sur la notion de région hydrologique homogène	145
2. Généralisation de la méthode de l'indice de crue : le modèle « régional » convergent	147
2.1. Présentation du modèle régional convergent	147
2.2. Ajustement de la distribution $V_R^*(0,T)$ régionale : application à la loi <i>GEV</i>	148
3. Présentation des données étudiées	150
3.1. Les données disponibles	150
3.2. Etude préliminaire sur la saisonnalité et la dynamique des crues	151
4. Résultats du modèle convergent (local et régional)	153
4.1. Résultats du modèle <i>QdF</i> local	153
4.2. Résultats du modèle <i>QdF</i> régional	154

Chapitre 6

Modélisation *QdF* « locale-régionale » : comparaison avec l'approche « bassin de référence » 161

1. Rappels sur l'approche <i>QdF</i> « bassin de référence »	162
1.1. Analyse locale	162
1.2. Analyse régionale : l'approche « bassin de référence »	163
1.3. Remarques sur l'approche « bassin de référence »	165
2. Données étudiées et méthodologie de comparaison	167
2.1. Données	167
2.2. Méthodologie	168
3. Résultats de l'approche « bassin de référence » : incidence du choix du bassin de référence	169
3.1. Comparaison des performances des trois bassins de références	169
3.2. Comparaison avec les résultats de Prudhomme	172
4. Comparaison avec les résultats du modèle convergent	175
4.1. Comparaison avec le modèle local convergent	175
4.2. Comparaison avec le modèle régional convergent	177
5. Synthèse des résultats	183

Chapitre 7	
Modélisation <i>QdF</i> « purement régionale »	187
1. Présentation du contexte de l'étude et des données	188
1.1. Objectif	188
1.2. Le projet NOAH et sa zone d'étude	189
1.3. Les données disponibles	189
1.4. Les caractéristiques calculées à partir des données disponibles	192
2. Estimation des caractéristiques hydrologiques par régressions multiples	195
2.1. Présentation des outils statistiques utilisés	195
2.2. Préparation des données	198
2.3. Sélection des variables explicatives pertinentes par la méthode Stepwise	199
2.4. Premiers commentaires	204
3. Test à l'aide de la "ridge regression" de la robustesse des régressions obtenues	205
3.1. Présentation de la ridge regression (régression de crête)	205
3.2. Application de la ridge regression	207
3.3. Résultats finaux et interprétation	209
3.4. Conclusion sur les régressions	216
4. Application « purement régionale » de l'approche débit-durée-fréquence	216
4.1. Objectif et intérêt de la démarche	216
4.2. Rappel de la méthodologie	217
4.3. Résultats et discussion	219
Conclusion générale et perspectives	226
Références bibliographiques	231
Annexes	239
Annexe 1 : Rappels sur les lois de distributions et comparaison entre valeurs supérieures à un seuil et maximales annuelles	241
Annexe 2 : Liste des bassins étudiés au chapitre 3	247
Annexe 3 : Paramètres du modèle local convergent ajusté avec $P=0$	249
Annexe 4 : Paramètres du modèle local convergent ajusté avec $P=\overline{QA}$	251
Annexe 5 : Application du modèle <i>QdF</i> convergent aux étiages	253
Annexe 6 : Liste des bassins étudiés au chapitre 5	261
Annexe 7 : Liste des bassins étudiés au chapitre 6	265
Annexe 8 : Liste des bassins étudiés au chapitre 7	267

Introduction

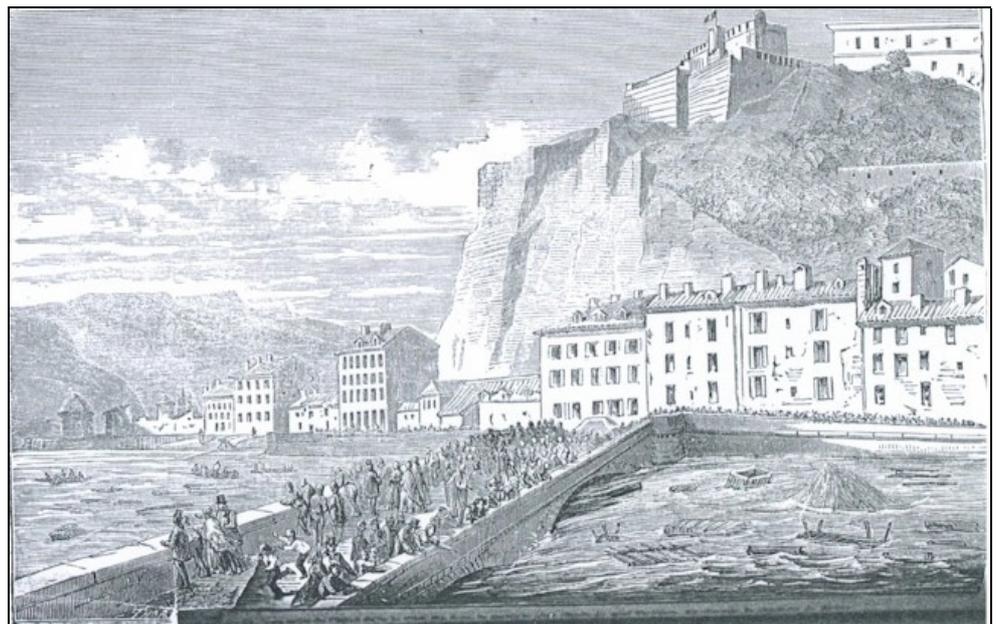


Illustration 1 : Crue de l'Isère du 3 novembre 1859 au Pont de Pierre, Grenoble
Source : Pilot J.-J.-A., Grenoble inondé, Maisonville & Jourdain, 1859. Dessin de D. Raoult

Introduction

Objectif de la thèse et présentation des données utilisées

Les caprices de la nature

Comme l'illustre la gravure précédente représentant l'inondation de l'Isère en 1859 à Grenoble, depuis que l'homme s'est installé aux abords des rivières pour profiter de leurs avantages : transport des marchandises, pêche, alimentation en eau, source d'énergie hydraulique..., il doit, en contre partie, en subir les caprices.

Les plus spectaculaires sont bien entendu liés aux crues. Celles-ci peuvent entraîner de lourdes pertes humaines, matérielles, ainsi que la paralysie de la région touchée : coupure du réseau routier, électrique, etc. A titre d'exemple, durant les inondations qui ont frappé les départements de l'Aude, de l'Hérault, des Pyrénées-Orientales et du Tarn les 12 et 13 novembre 1999, 35 morts et un disparu ont été déplorés, ainsi que des dégâts matériels estimés à 3,5 milliards de francs (*cf.* retour d'expérience commandé par le Ministère de l'Aménagement du Territoire et de l'Environnement, Lefrou *et al.*, 2000). Ce dernier chiffre représente plus du double du coût moyen annuel des dommages causés par les inondations en France, soit 1,5 milliards de francs d'après l'Office International de l'Eau (<http://www.oieau.fr/>).

Si, comme nous l'avons noté, l'excès de ressource en eau s'avère préjudiciable aux activités humaines, son manque peut aussi générer des situations de crise. Ainsi durant les forts étiages, des problèmes liés au manque de la ressource (pour l'irrigation par exemple), ou bien à la dégradation de la qualité de l'eau peuvent survenir.

Même si certaines approches, notamment statistiques, sont communes pour l'étude de ces deux cas extrêmes (crues et étiages), nous nous concentrerons dans la suite de ce travail en priorité sur l'étude des crues.

Mieux vaut prévenir que guérir

Pour pouvoir faire face au risque d'inondation, les décideurs ont besoin d'outils leur permettant de mieux gérer ces situations de crise. Certes, « on n'empêchera pas la pluie de tomber », cependant les efforts pour réduire les risques peuvent porter dans deux directions complémentaires :

- la prévision des événements exceptionnels qui consiste à anticiper ces événements en temps réel ;
- la prévention de ces mêmes événements, encore appelée **prédétermination**, où il s'agit de déterminer quels événements sont susceptibles d'intervenir « un jour ou un autre ».

Les travaux de cette thèse se situent clairement dans le deuxième axe énoncé ci-dessus. Il s'agit pour nous de quantifier l'**aléa hydrologique**, c'est-à-dire - pour simplifier - de définir le plus précisément

possible la probabilité de dépassement d'un débit donné. Nous verrons qu'en réalité l'aléa hydrologique peut être caractérisé de manière plus complète en associant probabilités et débits pour différentes durées.

En confrontant cet aléa à la **vulnérabilité**, c'est à dire aux dégâts engendrés si la zone étudiée était inondée, il est possible de définir des **cartes de risques**. Le rapport du député Yves Dauge concernant le risque d'inondation en France (Dauge, 1998) et celui sur la crue de l'Aude en novembre 1999 (Lefrou *et al.*, 2000) recommandent la généralisation de ces cartes de risques aux communes les plus exposées (soit en France, un tiers des communes ou encore 10 % du territoire). Au 14 décembre 2000, 2500 communes étaient dotées d'un Plan de Prévention des Risques (PPR). On estime qu'elles seront 5000 d'ici 2005 (*cf.* MATE, <http://www.environnement.gouv.fr>). A titre comparatif, 7000 communes ont déjà fait l'objet d'un arrêté pour catastrophe naturelle, uniquement entre 1993 et 1995, et 8 fois sur 10 pour des inondations (*cf.* Office International de l'Eau, <http://www.oieau.fr>). On constate donc qu'un grand nombre de communes « exposées » restent dépourvues de PPR. Or, l'absence de PPR entraîne bien souvent une sous-estimation des risques encourus, comme souligne le rapport sur la crue de l'Aude (Lefrou *et al.*, 2000) :

« La vulnérabilité des constructions a beaucoup augmenté dans certaines communes en un demi-siècle ; les causes sont notamment les suivantes :

La réalisation de constructions :

- dans des zones où l'aléa était inconnu ou oublié [...] ou encore sous-estimé,
- dans des secteurs manifestement inondables où la pression foncière [...] a fait passer au second plan les risques d'inondation, même forts [...],
- dans des secteurs où des travaux de protection de grande envergure ont pu donner l'illusion « du risque nul », illusion entretenue par l'absence d'inondations majeures [...].

Le changement d'usage du bâti traditionnel où le rez-de-chaussée inondable était toujours occupé par le cheptel et le matériel agricole ou de peu de valeur, et les étages par le logement.

La transformation en habitations de bâtiments anciens (moulins, bâtiments agricoles) localisés en secteurs inondables, pour lesquels le risque encouru était considéré comme « acceptable ».

Le manque d'information des occupants récents d'un bien situé en zone inondable, en cas de ventes successives ou de location. »

Parmi les mesures prises par les PPR pour réduire les risques, se trouve l'interdiction de construire dans les zones les plus exposées. Cette mesure réduit la vulnérabilité, mais en contre-partie, peut entraver, à court et moyen terme, le développement de la commune concernée : les zones classées « à risque » ne pouvant plus accueillir, par exemple, entreprises ou lotissements. Les PPR se doivent donc d'être le plus pertinent possible : une erreur d'appréciation, dans un sens comme dans l'autre (sur-estimation ou au contraire sous-estimation des risques) s'avèrerait préjudiciable pour la commune « protégée ». Les efforts de recherche doivent donc porter sur une meilleure caractérisation des deux composantes servant à définir le risque : aussi bien l'aléa que la vulnérabilité.

Le rapport sur la crue de l'Aude recommande « l'adhésion des élus et habitants sur la définition du risque acceptable » (Lefrou *et al.*, 2000). Ce type de démarche est mis en pratique par exemple par la

méthode Inondabilité (Gilard, 1998). Ses récents développements (Grelot, 1999) portent sur la méthode d'évaluation contingente, qui permet à l'aide d'interviews de connaître le consentement à payer et à recevoir des riverains interrogés. Il s'agit là d'une approche dominée par l'évaluation de la vulnérabilité des zones exposées.

Le travail de cette thèse concerne quant à lui la caractérisation de l'autre composante du risque : l'aléa.

Les différentes approches de caractérisation de l'aléa

L'approche la plus appropriée à la prédétermination des crues, est l'analyse fréquentielle. En attribuant une probabilité à une grandeur caractéristique telle qu'un débit, elle fournit un outil remarquable d'aide à la décision. Pour une synthèse des différentes approches existantes, on peut par exemple se référer à Lang (2000).

En étudiant la chronique des débits, l'analyse fréquentielle permet de décrire en terme de probabilité d'une part les événements observés, et d'autre part les événements plus rares que ceux observés, mais susceptibles de se produire un jour. En donnant une information en terme probabiliste, l'analyse fréquentielle est un outil bien adapté en ingénierie pour le dimensionnement des ouvrages ou en aménagement pour l'établissement de cartes d'aléa précédemment évoquées. L'approche fréquentielle présente néanmoins deux inconvénients qui lui ont valu quelques critiques plus ou moins justifiées : d'une part il n'est possible d'attribuer une probabilité qu'à une variable ou un jeu de variables. Or l'hydrogramme d'une crue est une fonction continue $Q(t)$. L'analyse fréquentielle choisit dans cette chronique une valeur (débit maximum par exemple, ou volume global...) et lui affecte une probabilité : on ne sait donc pas attribuer une fréquence à une crue mais à une variable choisie pour caractériser cette crue. D'autre part, une hypothèse classique des analyses fréquentielles est la stationnarité des séries. La population décrite est supposée issue d'une loi de probabilité à laquelle les échantillons observés nous permettent d'accéder. La non-stationnarité des séries, pour des raisons de changement d'occupation du sol ou pour des raisons climatiques, implique une évolution temporelle de la loi de probabilité d'où chaque valeur de l'échantillon est issue. A l'heure actuelle on ne sait guère tenir compte de ces évolutions. Enfin l'approche fréquentielle ne tient absolument pas compte des processus physiques internes au bassin, contrairement aux modèles pluie/débit qui, directement ou indirectement, décrivent le comportement des bassins versants.

Le but des modèles pluie/débit est d'estimer une chronique de débit en fonction du temps. Il existe un grand nombre de modèles, des simples boîtes noires avec en entrée une chronique de pluie, aux plus sophistiqués, décrivant les processus physiques de façon distribuée sur l'ensemble du bassin. Nous ne nous étendrons pas plus sur ce vaste domaine. Pour plus d'information, on peut par exemple se référer aux articles de synthèse d'Ambroise (1998, 1999) et de Marsily (1994).

Pour l'analyse des aléas, on voit maintenant apparaître des approches connues sous le nom de « continuous simulation » consistant à faire l'analyse fréquentielle non plus sur les débits observés

mais sur les débits simulés par un modèle pluie/débit (Lamb, 1999). Ces méthodes sont censées permettre par exemple la caractérisation fréquentielle, pour un même bassin, des modifications de son occupation du sol, ou bien de sa pluviométrie dues à un supposé changement climatique. Différents scénarios peuvent ainsi être testés (Calver *et al.*, 2000). Ces approches sont aussi utilisées pour agrandir la chronique de débit disponible et donc avoir une estimation des quantiles rares, en entrant dans le modèle une chronique de pluie simulée sur un grand nombre d'années (par exemple : Blazkova et Beven, 1997 ; Arnaud et Lavabre, 1999 ; Cameron *et al.*, 1999). Ce type d'approche nécessite cependant de rester prudent quant aux résultats obtenus. La grande taille des chroniques simulées ne rime pas forcément avec diminution des incertitudes, puisque bien souvent on sort du domaine de validité pour lequel le modèle a été calibré.

La suite de ce travail se concentre uniquement sur l'analyse fréquentielle, les quelques références aux modèles pluie/débit ci-dessus servent simplement à resituer notre sujet.

Objectif et plan de la thèse

L'objectif principal de la thèse est de développer un modèle fréquentiel capable de rendre compte du régime hydrologique des crues d'un bassin. Cette information est par exemple requise pour caractériser en terme probabiliste l'aléa hydrologique précédemment évoqué, ou de façon plus générale pour décrire le régime hydrologique.

Le travail présenté est divisé en deux parties : l'analyse locale et l'analyse régionale. Dans la première, l'analyse concerne uniquement les données de débits locales. Dans la seconde, on utilise l'ensemble de l'information disponible pour toute une région donnée. Cette dernière approche permet « d'extrapoler » des résultats observés sur des bassins où l'on dispose de données hydrologiques à des bassins où l'on n'en dispose pas.

Analyse locale

Lorsque l'on cherche à caractériser un régime hydrologique, et en particulier celui des crues, il faut dans un premier temps définir les variables pertinentes : débit de pointe, forme et durée. Le chapitre 1 présente l'approche adoptée : l'analyse débit-durée-fréquence. Comme nous le verrons plus en détail, cette approche fréquentielle permet de prendre en compte une notion essentielle dans le régime : **la durée**, contrairement à une analyse fréquentielle classique effectuée uniquement sur les débits de pointe de crue.

Le chapitre 2 revient sur un outil de base de l'analyse fréquentielle : l'échantillonnage. La méthode par sélection de valeurs supérieures à un seuil est adoptée. Des outils de traitements sont développés, intégrant notamment des tests statistiques utiles pour la suite du travail.

Les chapitres 3 et 4 sont le cœur de cette thèse. Ils présentent le modèle débit-durée-fréquence développé au cours de ces trois années. Partant des approches déjà développées au Cemagref (Galéa

et Prudhomme, 1997), nous avons cherché à élaborer un modèle économe en paramètres, ajustable **localement** et sans introduire d'hypothèses sur les extrapolations. Ce modèle a été appelé modèle « convergent », suite à des propriétés observées sur les distributions de débits.

Analyse régionale

Les chapitres 5 et 6 et 7 abordent le problème d'un point de vue **régional**. L'idée est de voir comment l'information disponible en différents sites peut être intégrée dans le modèle que nous avons développé, afin de pouvoir l'utiliser sur des bassins pour lesquels il n'y a pas ou peu d'informations sur les débits.

Au chapitre 5, la méthode de l'indice de crue, couramment utilisée en analyse fréquentielle régionale, est généralisée afin d'être applicable dans le cadre de l'analyse débit-durée-fréquence présentée dans la première partie. Le modèle obtenu est appelé modèle régional convergent.

Ce nouveau modèle remplit les mêmes fonctions que l'approche dite par « bassin de référence » anciennement développée au Cemagref (Galéa et Prudhomme, 1997). Les chapitres 6 et 7 comparent ces deux approches, afin de montrer leurs points communs et leurs différences. Deux niveaux d'analyse sont distingués : les estimations « locales-régionales » et les estimations « purement régionales ». Cunnane (1988) regroupe sous la première dénomination les estimations utilisant à la fois l'information hydrologique locale du site étudié et l'information régionale. La seconde signifie que seule l'information régionale est étudiée, ce qui autorise des estimations sur des bassins ne disposant d'aucune donnée de débit.

Le chapitre 6 compare les estimations « locales-régionales » du modèle régional convergent et celles de l'approche « bassin de référence ». Le chapitre 7 s'intéresse aux estimations « purement régionales ». Pour cela, nous faisons appel à l'analyse des régressions multiples, appliquée sur les caractéristiques hydrologiques, les caractéristiques pluviométriques et les caractéristiques « physiques » des bassins étudiés (géomorphologie, occupation du sol).

Précisons que le travail développé dans le cadre de cette thèse concerne principalement les fréquences observables, et non les extrapolations des distributions. En effet les extrapolations dépendent de la méthode utilisée. Il ne s'agit pas pour nous de discuter du bien fondé de chacune de ces méthodes, mais d'établir les bases rigoureuses d'une analyse débit-durée-fréquence **locale** et **régionale**. Son domaine d'application reste donc a priori celui des fréquences observables, avec une durée de retour de 20 ans à 50 ans maximum, selon la taille de l'échantillon disponible.

Données étudiées

Le travail énoncé aux pages précédentes a été rendu possible grâce à la disponibilité de nombreuses données, pour développer puis valider les différents modèles étudiés. On distingue trois principaux jeux de données utilisés au cours de cette thèse.

Le premier jeu est constitué d'une centaine de bassins localisés dans toute la France (Figure 0-1). Leurs surfaces sont comprises entre 7 et 10 000 km². Ces données sont utilisées (en totalité ou en partie selon les cas) dans les chapitres 1 à 4, principalement pour valider sur un grand nombre de sites le modèle débit-durée-fréquence développé. Ces données ont été achetées par le Cemagref-Lyon à la banque HYDRO au cours des différentes études et appuis techniques qu'il a réalisés ces dernières années. La liste des bassins figure en annexe 2, page 247.

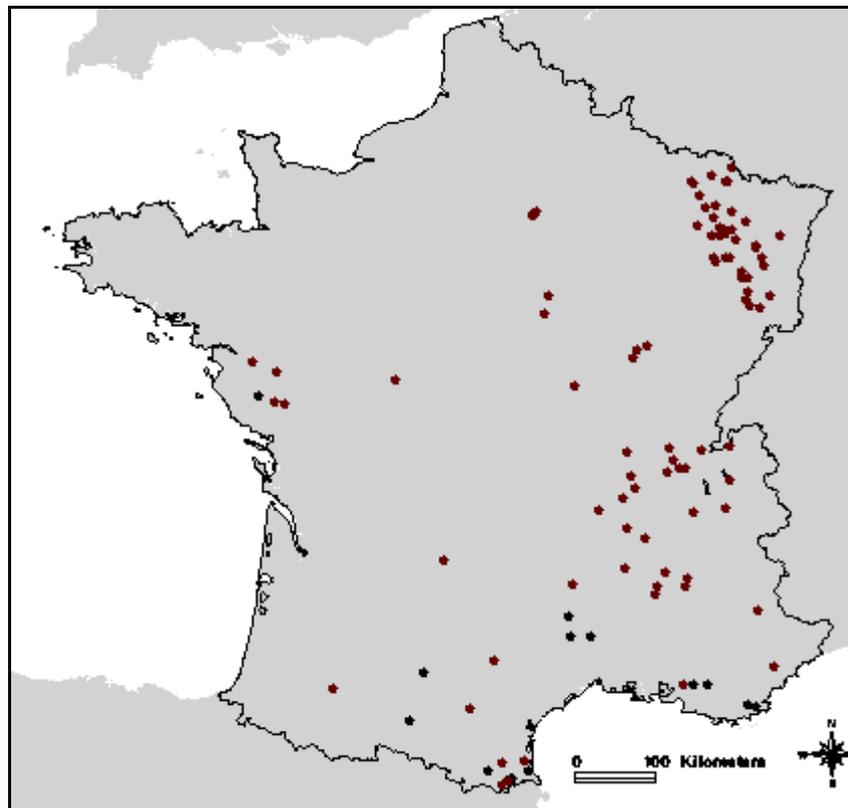


Figure 0-1 : Localisation des bassins français étudiés

Le second jeu de données est situé au Québec et en Ontario (Figure 0-2). Ces données ont été mises à disposition dans le cadre d'un projet franco-québécois avec l'INRS-Eau. Environ 150 chroniques de débits sont disponibles, ainsi que pour chaque bassin correspondant, des caractéristiques « physiques » (géomorphologie et occupation du sol). Ces données sont utilisées dans le chapitre 5 pour appliquer nos modèles dans des conditions climatiques différentes de celles du premier jeu de données, et ainsi s'assurer de leur robustesse. La liste des bassins se trouve en annexe 5, page 261.

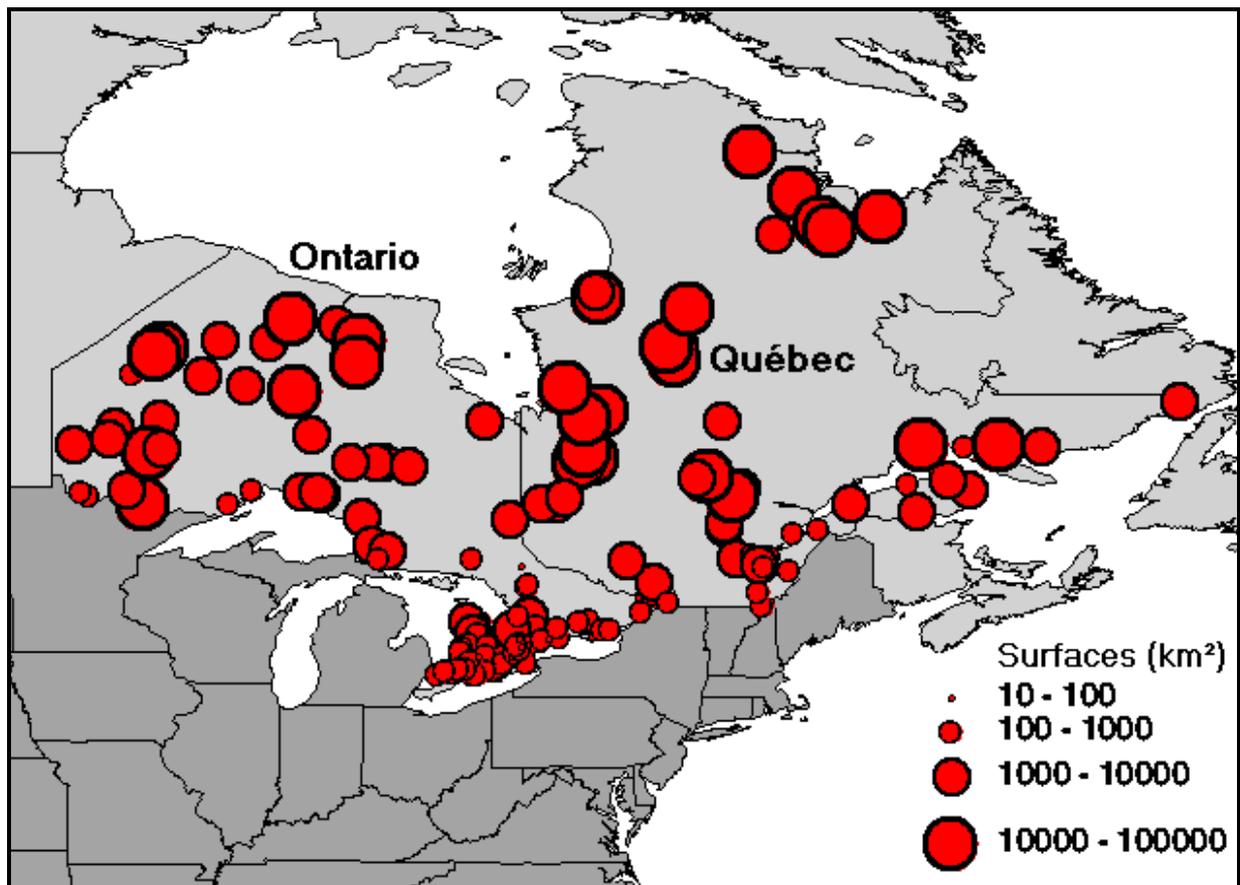


Figure 0-2 : Bassins étudiés au Québec et en Ontario

Le troisième jeu de données concerne l'ensemble du bassin hydrographique de la Moselle (Figure 0-3). Les chroniques de débits d'une cinquantaine de bassins de surfaces comprises entre 16 et 1250 km² sont disponibles, ainsi que des chroniques de pluie sur une centaine de postes. Ces bassins sont situés en Allemagne et en France, ces derniers faisant partie du premier jeu de données présenté. De plus, une carte d'occupation du sol et un modèle numérique de terrain obtenus à partir d'images du satellite SPOT permettent pour chaque bassin étudié de calculer différentes caractéristiques physiques. Toutes ces données ont été obtenues dans le cadre d'un projet européen et sont utilisées au chapitre 7 pour établir des régressions entre caractéristiques hydrologiques, pluviométriques et physiques de bassin. La liste des stations se trouve en annexe 8, page 267.

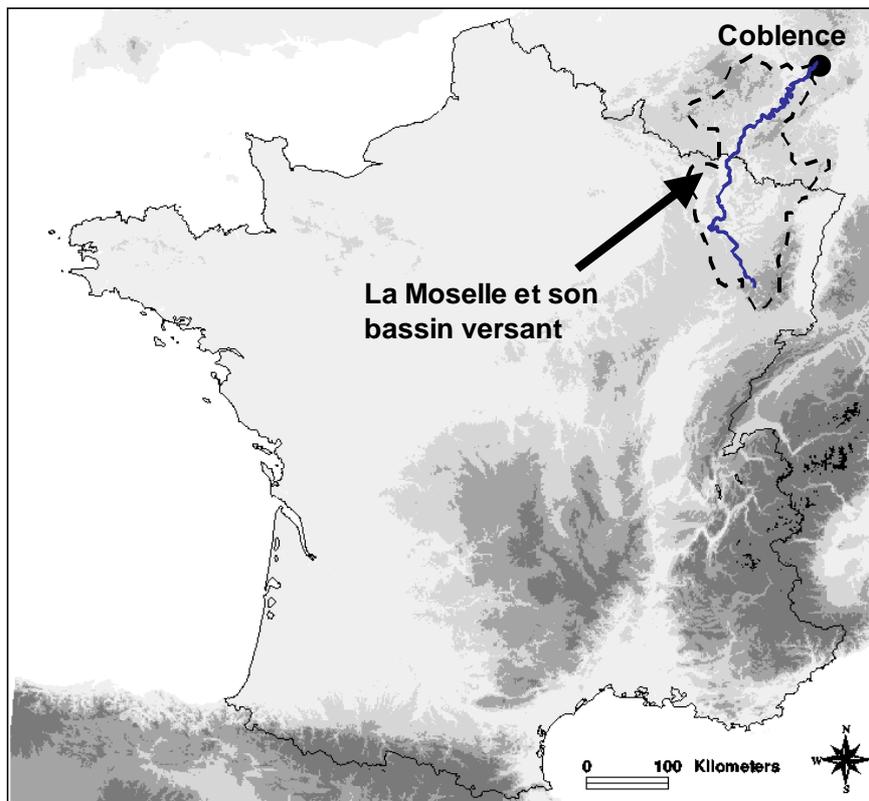


Figure 0-3 : Localisation du bassin hydrographique de la Moselle

Les chapitres suivants présentent plus en détail ces différentes données, au fur et à mesure de leur utilisation.

PARTIE I

Analyse Locale



Illustration 2 : Crue du 3 novembre 1859, place Sainte Claire, Grenoble
Source : Pilot J.-J.-A., Grenoble inondé, Maisonville & Jourdain, 1859. Dessin de D. Raoult

Chapitre 1

Description du régime hydrologique : introduction à l'approche débit-durée- fréquence

L'objectif de ce chapitre est de montrer comment décrire le régime hydrologique en général, et celui des crues en particulier. Curieusement, même si de nombreux travaux traitent « du régime », il n'existe pas ou peu de définitions valables pour l'ensemble des approches rencontrées.

Nous tentons donc, dans la première partie de ce chapitre de donner un cadre général à la suite de ce travail, en définissant ce qu'est le régime hydrologique, ainsi que les notions qui y sont rattachées, telles que les échelles temporelles.

La seconde partie illustre à partir d'exemples comment la durée est définie dans la plupart des analyses du régime. Cette notion de durée porte sur le pas de temps de la chronique de débit étudiée, ainsi que sur la durée de regroupement des différentes observations.

La troisième partie s'intéresse plus précisément à l'étude des crues et présente l'approche débit-durée-fréquence. En terme de régime, mais aussi de façon beaucoup plus pragmatique, pour ce qui concerne la prévention des inondations, les durées des crues, et donc les volumes écoulés, sont aussi importants à connaître que les débits de pointe. En réponse à ce besoin, l'approche débit-durée-fréquence intègre la notion de durée dans l'analyse fréquentielle. Cette approche est adoptée dans la suite de la thèse et sera développée plus en détail par la suite.

1. La notion de régime hydrologique

1.1. Définition du régime hydrologique

Etymologiquement, le mot *régime* a la même racine que le verbe *régir* : du latin *regimen* qui signifie *action de diriger*. Ainsi, le Petit Robert définit tout d'abord le régime de la façon suivante : « *Façon d'administrer, de gouverner une communauté* ». Concernant le régime hydrologique, il précise : « *Le régime d'un fleuve est caractérisé par les variations de son débit* ». De façon très générale, l'étude du régime hydrologique consiste donc à trouver des règles **régissant** la variabilité temporelle des écoulements.

Curieusement, il est difficile de rencontrer une définition plus précise du régime hydrologique dans la littérature. Très souvent il est fait exclusivement référence aux régimes définis par Pardé (1955), relatifs aux variations interannuelles des débits mensuels. Bien qu'incontournable, cette représentation du régime ne concerne que la fluctuation mensuelle qui ne constitue qu'une partie de la variabilité temporelle des écoulements. Un projet du Programme National de Recherche en Hydrologie du CNRS s'est récemment penché sur la notion de régime et s'est efforcé de rassembler un maximum de points de vue autour de concepts communs (Oberlin, 1999). Les réflexions suivantes ont tiré leur inspiration de ces travaux.

1.2. Régimes et échelles temporelles

A l'origine des écoulements se trouvent les phénomènes physiques qui les provoquent. Il s'agit des précipitations (pluie, neige), de la fusion commandée par l'énergie du soleil, des évaporations et évapotranspirations également dues à l'ensoleillement, mais aussi des phénomènes de stockage et de transfert dans les réservoirs souterrains. Il faut donc s'attendre à ce que ces différents phénomènes physiques imposent leurs rythmes aux variations des écoulements (Grésillon, 1999).

Les durées sur lesquelles chacun de ces rythmes est observable varient considérablement d'un phénomène physique à un autre. Ainsi, selon les cas, ces durées peuvent être de l'ordre :

- **de quelques heures à quelques jours** : pour des variations brutales du débit, lors des crues provoquées par de fortes pluies sur des petits bassins ;
- **du cycle diurne** : pour la fonte (le jour) et le regel (la nuit) responsables des fluctuations de débit sur des rivières glaciaires ;
- **du mois** : pour des crues lentes provoquées par des précipitations sur des grands bassins ou bien dues à la fonte nivale ;
- **de la saison** : pour les variations liées aux évaporations et évapotranspirations qui agissent sur les niveaux des nappes et indirectement sur les écoulements en rivière (réservoir sol...);

- **de l'année ou davantage**, s'il s'agit d'évolutions d'origine anthropique ou liées à des variations climatiques.

Un des problèmes rencontrés dans la description du régime hydrologique est lié au fait que les différentes échelles temporelles évoquées ci-dessus se superposent, ce qui rend difficile l'interprétation brute des débits observés. Par exemple, la Figure 1-1 montre l'évolution du débit instantané de la Nied Allemande à Faulquemont (département de la Moselle) sur la période 1984-1997. Certes, on distingue des périodes plus ou moins régulières de hautes et basses eaux. On apprend que la plus forte crue observée atteint les 50 m³/s, mais ces observations semblent noyées dans un bruit dont la logique nous échappe.

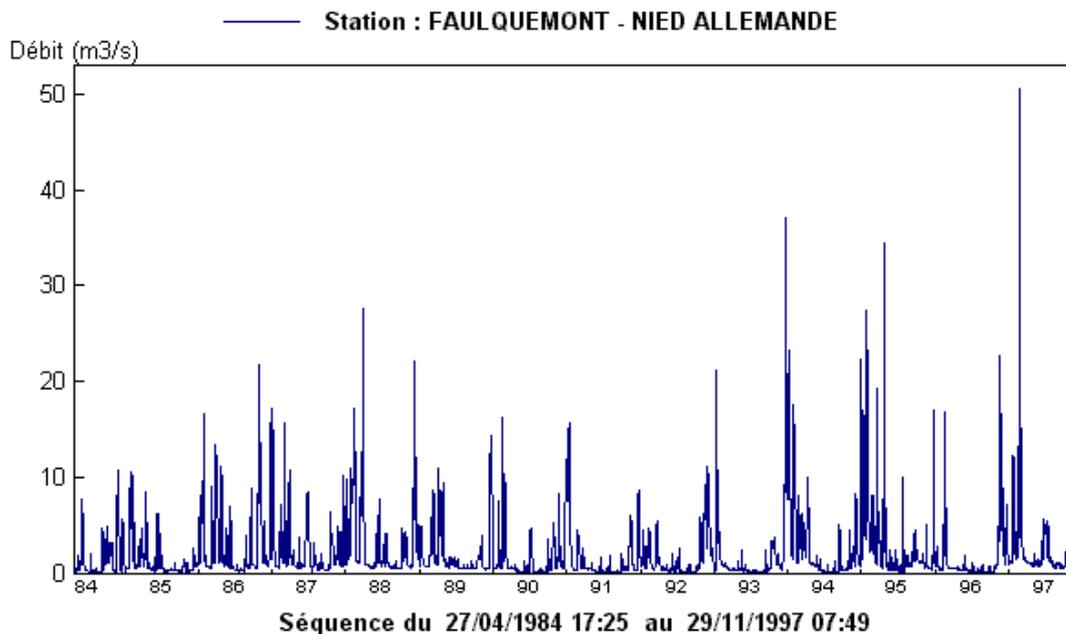


Figure 1-1 : Variations du débit de la Nied Allemande à Faulquemont, sur la période 1984-1997, département de la Moselle, (S=187 km²)

Par conséquent, la caractérisation du régime hydrologique passe nécessairement par une synthèse de la variabilité observée. Cette synthèse doit se faire en fonction de l'échelle temporelle des événements que l'on cherche à caractériser. On ne procède pas de la même façon si l'on désire étudier le régime des crues, des étiages ou encore celui des variations saisonnières du débit.

1.3. L'information de base : la chronique de débit

La chronique de débit constitue l'information de base pour caractériser les variations des écoulements. Il convient donc avant toute chose de bien définir ce que représentent ces débits. Comme nous le verrons, les chroniques enregistrées sont en fait presque toujours des chroniques de hauteur qui sont ensuite converties en débit grâce à ce qu'on appelle la courbe de tarage.

1.3.1. La courbe de tarage

La courbe de tarage donne pour une section d'une rivière la relation entre la hauteur du niveau de l'eau relevée sur une échelle limnimétrique et le débit. Le débit est un volume par unité de temps. Cependant, dans le cas d'une rivière, il n'est bien sûr pas possible de mesurer directement les volumes écoulés. Par contre, on peut mesurer la vitesse de l'écoulement en différents points de la section de la rivière. Le débit est alors obtenu en intégrant ces vitesses sur toute la section. Le débit calculé est quasi instantané, puisque les vitesses ont été calculées sur de courts instants, durant lesquels les grandeurs mesurées sont supposées constantes. Selon la taille de la rivière, différentes techniques existent pour mesurer les vitesses : moulinet, traceur chimique, courantomètre, effet doppler, etc. La courbe de tarage est obtenue en regroupant sur un même graphe les mesures de débit instantané effectuées pour différentes hauteurs. La Figure 1-2 donne l'exemple de la Sémine à Sermet, bassin géré par le Cemagref-Lyon (Ribot-Bruno et Breil, 1998)

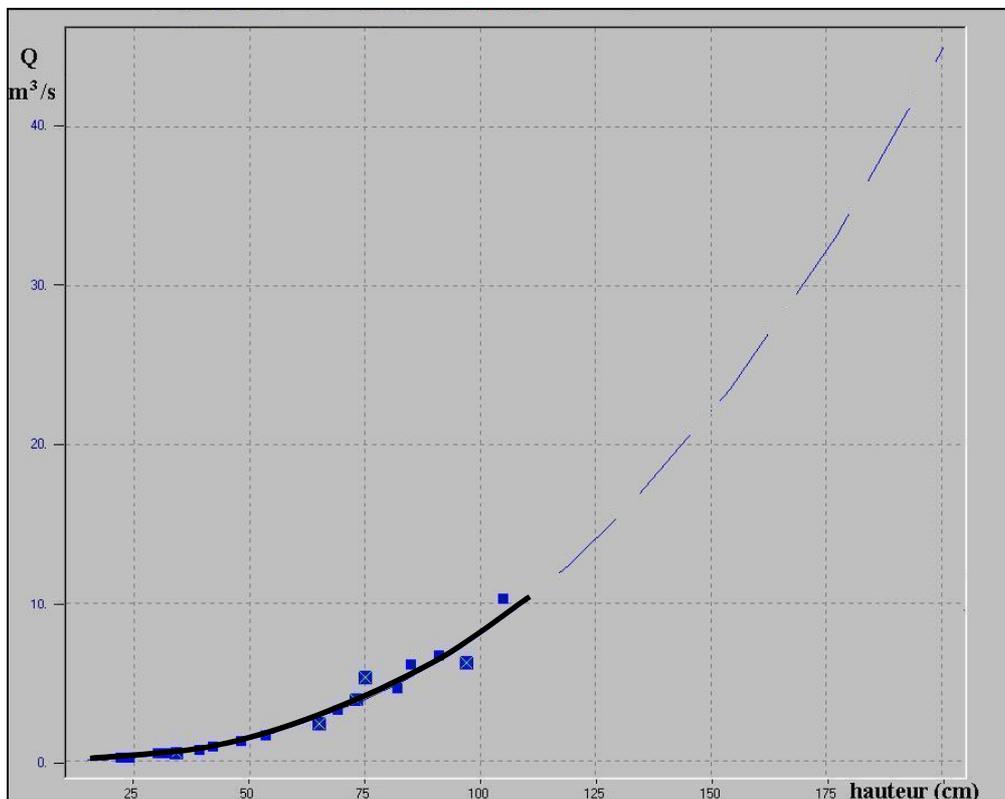


Figure 1-2 : Courbe de tarage. Exemple de la Sémine à Sermet, dans l'Ain (50 km²)

La courbe de tarage permet de ramener la mesure du débit à celle de la hauteur, beaucoup plus simple à obtenir. Les gestionnaires de chaque station de jaugeage doivent définir la courbe de tarage avec le plus grand soin, afin de limiter au maximum les incertitudes de mesure. En effet, les problèmes posés sont les suivants :

- problème de la stabilité dans le temps de la courbe de tarage : si la section d'écoulement est modifiée ou si les conditions d'écoulement sont changées en aval de la station (nouveau pont, embâcles, etc.), la relation entre la hauteur et le débit se trouve elle aussi modifiée (si l'écoulement est fluvial, ce qui est le plus fréquent) ;

- problème d'hystérésis : la relation entre le débit et la hauteur n'est pas forcément la même selon que l'on est en crue ou en décrue ;
- problème de l'extrapolation : lors de très fortes crues, le plus souvent il n'est pas possible d'effectuer une mesure du débit correspondant à la hauteur relevée. Le calcul du débit est alors effectué en extrapolant la courbe de tarage, ce qui a pour conséquence une augmentation des incertitudes (Figure 1-2).

1.3.2. Les chroniques de débit instantané

Dans la pratique, les chroniques enregistrées sont des chroniques de hauteurs, appelées limnigrammes, qui sont ensuite converties en chronique de débit instantané grâce à la courbe de tarage. Selon les méthodes d'enregistrement utilisées, le pas de temps entre deux mesures peut être constant ou variable. Dans le cas d'un pas de temps constant, une mesure de la hauteur, et donc du débit, est prise à intervalle régulier, par exemple 24h (Figure 1-3, graphique de droite). Pour le pas de temps variable, l'intervalle entre deux mesures s'adapte à l'écoulement réellement observé noté $Q(t)$: si les variations de l'écoulement sont lentes, le pas de temps augmente, si les variations sont rapides, le pas de temps diminue. En principe, les points remarquables tels les maxima et les minima sont toujours relevés dans ce cas (Figure 1-3, graphique de gauche).

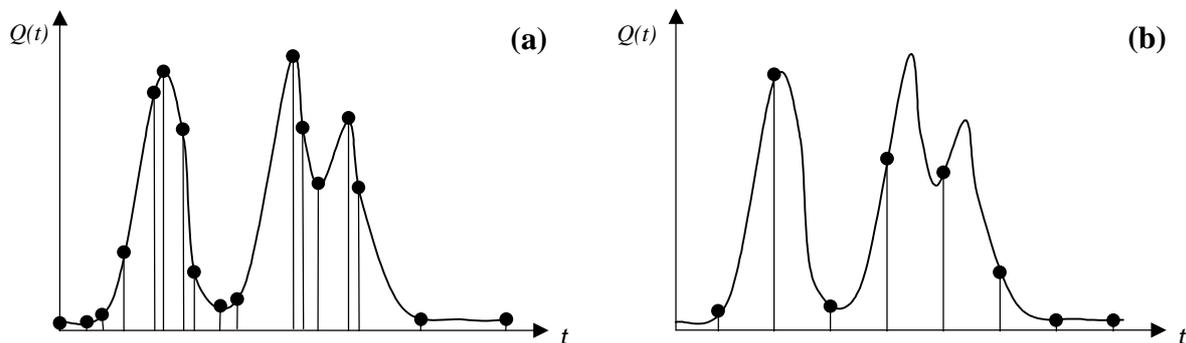


Figure 1-3 : Chronique de débit instantané à pas de temps variable (a) et à pas de temps constant (b)

L'intérêt des mesures au pas de temps variable est d'obtenir une quantité d'information optimale, au regard de la rapidité des variations mesurées. Comme le montre le graphique de droite de la Figure 1-3, les chroniques au pas de temps constant peuvent amener à une perte d'information importante si le pas de temps est trop grand par rapport à la dynamique de l'écoulement réellement observé $Q(t)$. Dans le cas inverse, on peut aussi avoir surabondance d'information si le pas de temps est trop fin, il suffit alors de compacter la chronique en supprimant les points redondants.

1.3.3. Les chroniques de débit moyen

Pour diverses applications, il est plus pratique de travailler sur des chroniques non plus de débit instantané, mais de débit moyen sur un pas de temps choisi. Lorsqu'une chronique de

débit instantané est disponible, les débits instantanés sont **interpolés linéairement** entre les différentes valeurs de débit mesurées à pas de temps constant ou variable, permettant de calculer par la méthode des trapèzes le volume écoulé entre chaque instant. En divisant les volumes obtenus par la durée sur laquelle ils ont été calculés, on obtient une chronique de débit moyen sur cette durée (Figure 1-4).

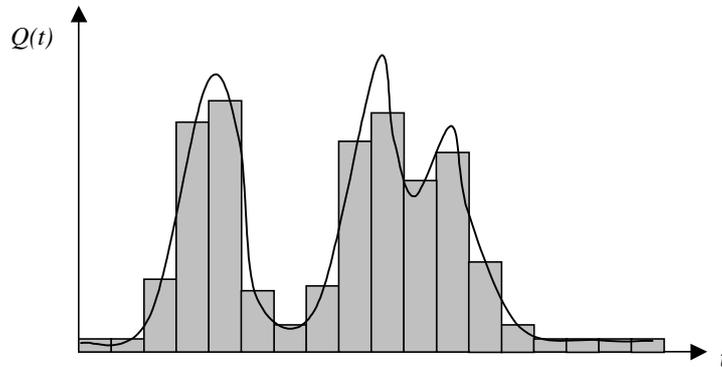


Figure 1-4 : Chronique de débit moyen à pas de temps constant

Il convient donc dès maintenant de bien distinguer :

- le pas de temps d'acquisition des débits instantanés (constant ou variable, cf. §1.3.2.) ;
- la durée des intervalles de temps, toujours constants, entre lesquels sont calculés les volumes écoulés, puis ensuite, les débits moyens.

Cette seconde durée étant en général la durée séparant deux valeurs de débit moyen, elle est aussi appelée pas de temps, mais elle ne doit pas être confondue avec le pas de temps d'acquisition des débits instantanés.

Dans certains cas, il peut arriver que l'on ne prenne qu'une mesure du débit instantané, avec un pas de temps constant Δt (par exemple tous les jours à midi), et que l'on considère que cette mesure obtenue à l'instant t est une approximation du débit moyen sur l'intervalle $[t-\Delta t ; t+\Delta t]$. Cette façon de faire suppose en fait que le débit varie **linéairement** sur l'intervalle $[t-\Delta t ; t+\Delta t]$. Si cette hypothèse n'est pas assez bien respectée, l'estimation des débits moyens peut comporter de grosses erreurs.

Pour des raisons pratiques et historiques, les chroniques de débit le plus souvent disponibles sont celles des débits moyens journaliers. Selon les applications, les pas de temps horaires, mensuels ou annuels peuvent aussi trouver leur utilité. Par contre, les chroniques de débit instantané à pas de temps variable restent peu utilisées, malgré leur intérêt. Le choix du pas de temps est déterminant pour toute la suite de l'analyse de la chronique. Le paragraphe suivant traite de ce point pour l'exemple des variations interannuelles des débits.

2. Variations saisonnières des débits interannuels : le choix du pas de temps

Traditionnellement, l'étude du régime hydrologique fait référence aux régimes définis par Pardé (1955). Il s'agit de l'étude des variations saisonnières des débits moyens interannuels pris au pas de temps mensuel. Mais pourquoi choisir ce pas de temps plutôt qu'un autre ? Quelle est l'influence du pas de temps sur le phénomène observé ?

2.1. Le cas du pas de temps mensuel : les régimes de Pardé

La représentation initiée par Pardé (1955) consiste à porter sur un graphique les moyennes interannuelles des 12 débits mensuels. Cette description peut être complétée par des quantiles exprimant aussi la variabilité interannuelle des débits (Figure 1-5). On trouvera dans Sauquet (2000) plus de détails sur les éléments présentés ci-après.

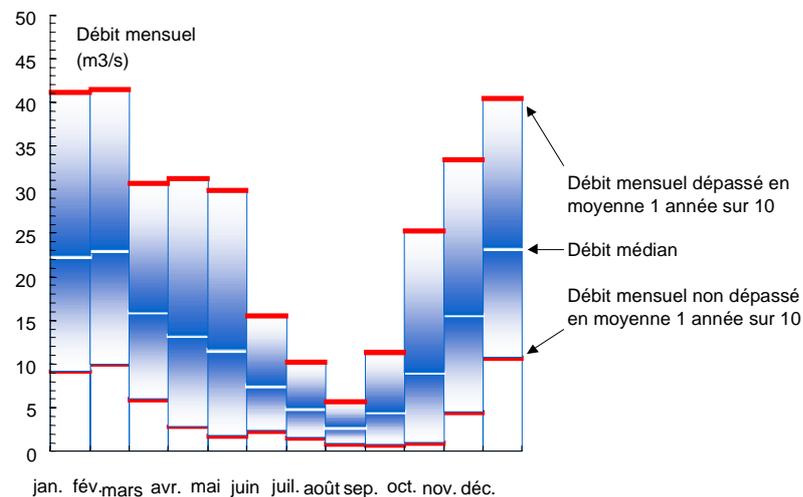


Figure 1-5 : Représentation des variations saisonnières du débit mensuel, Cas de la Seuille à Voiteur (Sauquet, 2000a)

Sur la base de cette description des débits mensuels, la classification de Pardé (1955) s'intéresse à la répartition des apports mensuels au cours de l'année ainsi qu'à l'origine des écoulements (neige, pluie, glacier). Trois principaux types de régimes sont distingués : les régimes simples, les régimes complexes originels, les régimes complexes changeants.

- Les régimes simples sont caractérisés par une seule provenance des écoulements. On distingue deux saisons hydrologiques : une pour les hautes eaux et une autre pour les basses eaux. Les quatre principaux régimes simples sont :
 - le régime glaciaire ;
 - le régime nival ;
 - le régime pluvial océanique ;

- *le régime pluvial tropical.*

- Les régimes complexes originels sont des compositions de régimes simples. Plusieurs modes d'approvisionnements sont recensés (par exemple pluie et neige), impliquant plusieurs saisons de hautes et basses eaux.
- Les régimes complexes changeant concernent les grands fleuves, traversant plusieurs climats, et alimentés par des affluents caractérisés par différents régimes simples et complexes originels. Dans ce cas le régime du fleuve peut être « changeant » d'amont en aval.

Cette classification reste une approche d'expert. Néanmoins Sauquet (2000ab) a pu définir des critères objectifs basés sur l'évolution des 12 débits moyens interannuels, permettant d'affecter un type de régime au bassin étudié. Ces travaux ont entre autres permis une cartographie spatialisée de ces régimes sur toute la zone Rhône-Méditerranée-Corse.

La description de Pardé a été beaucoup reprise de par le monde. Elle a été affinée lorsque les bassins étudiés étaient assez homogènes, comme par exemple en Suisse (Weingartner et Aschwanden, 1992), ou bien généralisée pour les besoins d'une description mondiale (Beckinsale, 1969). En Europe du Nord, une classification du régime des variations des débits mensuels inspirée des mêmes principes a aussi été élaborée (Tollan, 1975 ; Krasovskaia et Gottschalk, 1992 ; Krasovskaia, 1997). Les régimes y sont définis de façon objective, en fonction de la date d'occurrence des trois plus forts débits mensuels moyens interannuels (max1, max2 et max3, classés par ordre décroissant) et des deux plus faibles débits mensuels moyens interannuels (min1 et min2). Huit régimes ont ainsi été définis. Cette classification a été étendue à l'Europe, portant le nombre de régimes à 13 (Arnell *et al.*, 1993).

Ainsi, comme nous venons de le voir, lorsque l'on parle de régime hydrologique, la littérature fait largement écho aux régimes définis par Pardé, basés sur l'évolution des débits mensuels. Ce type de représentation donne une information sur les variations lentes du débit, qui correspondent pour une part importante à la réponse des réservoirs souterrains face au cycle des saisons : recharge lors des saisons humides, vidange lors des saisons sèches. De ce fait, les différents types de régimes hydrologiques identifiés reflètent largement les régimes climatiques auxquels ces bassins sont soumis.

Mais le pas de temps mensuel est-il le plus adéquat, ou bien est-il simplement choisi par convention ? Etudie-t-on le même phénomène si l'on change de pas de temps ? Les paragraphes suivants tentent d'apporter des réponses à ces questions.

2.2. Variations interannuelles au pas de temps journalier

Le type d'analyse décrit ci-dessus existe pour d'autres pas de temps que le mensuel. A titre d'exemple, nous présentons deux approches relatives au pas de temps journalier, l'une utilisée aux Etats-Unis, l'autre en France par EDF-DTG.

Aux Etats-Unis, l'USGS effectue les analyses suivantes sur les débits moyens journaliers. Chaque jour, des cartes indiquent comment se situe la valeur du débit moyen journalier par rapport aux débits mesurés le même jour les années précédentes. Un minimum de 30 années de mesure est considéré. Un code de couleur indique si le débit mesuré dépasse 0, 10, 25, 75, 90 ou 100% des débits mesurés le même jour des années précédentes (Figure 1-6). Ces cartes sont disponibles sur internet (<http://water.usgs.gov/dwc/>), avec la possibilité de voir en animation leur évolution jour après jour.

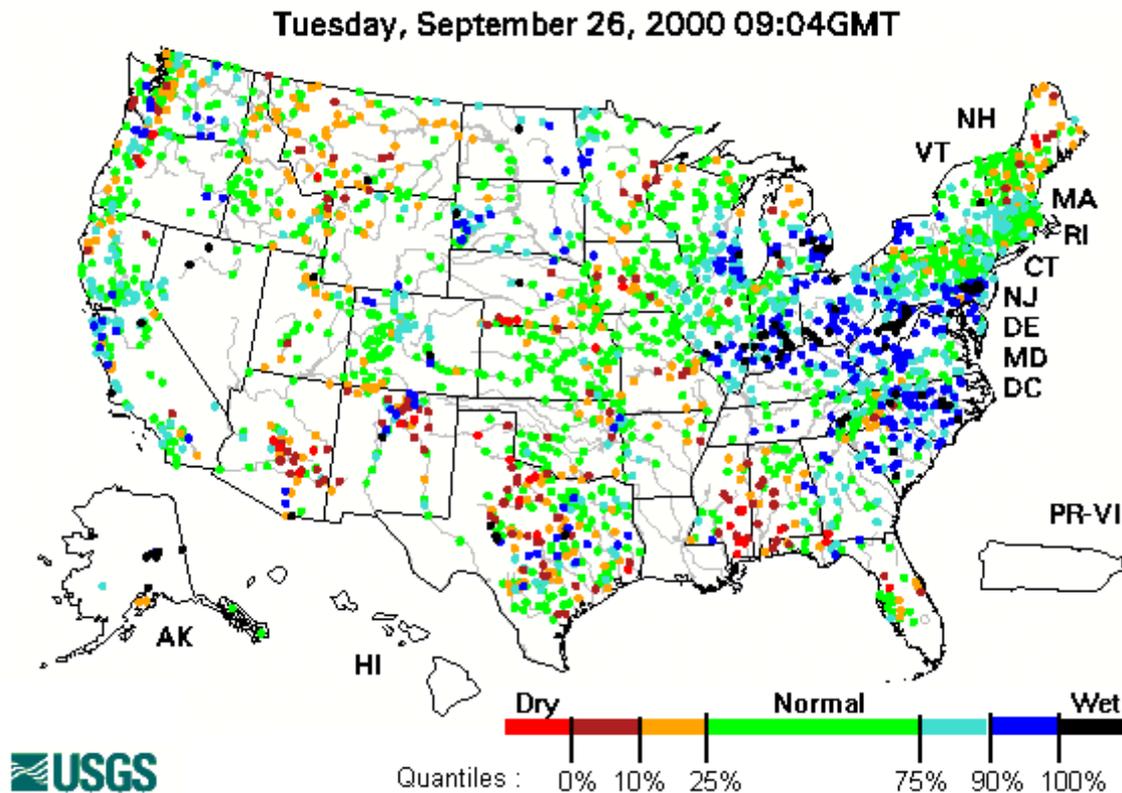


Figure 1-6 : Situation des débits moyens journaliers mesurés le 26 septembre aux Etats-Unis par l'USGS, par rapport aux « normales saisonnières »

Pour décrire les variations saisonnières du débit, EDF-DTG (Duband, 1992) représente les informations suivantes, sur un même graphe (Figure 1-7) :

- Courbes des valeurs journalières non-dépassées dans 10, 25, 50, 75 et 90% des cas, durant toute la période d'observation.
- Pour chaque mois, la valeur dépassée en moyenne une année sur 10 par le débit moyen journalier le plus fort du mois.
- Pour chaque mois, les crues au cours desquelles le débit moyen journalier a dépassé la valeur journalière décennale du mois (valeur précédente).

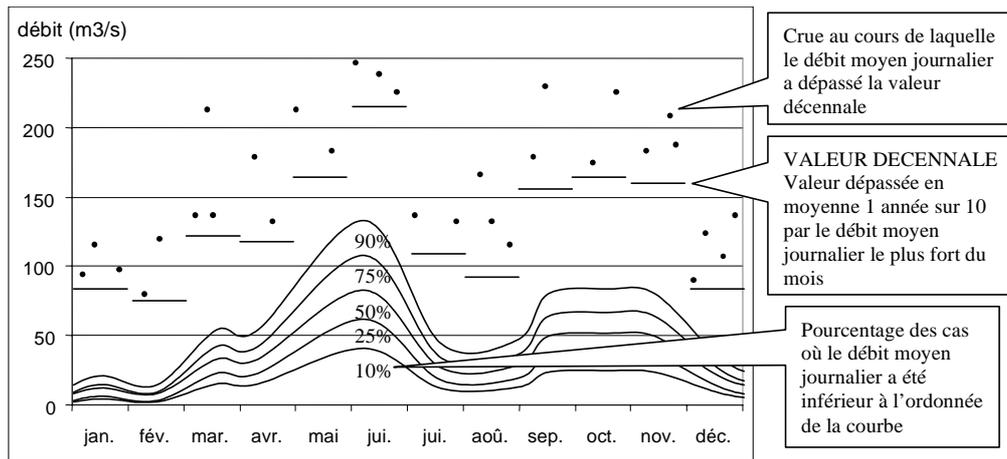


Figure 1-7 : Répartition des débits moyens journaliers suivant l'époque de l'année. Exemple, repris d'après Duband (1992)

Cette représentation associe donc des débits journaliers et leurs quantiles à chaque date mais aussi des grandeurs « hybrides » que sont le débit moyen journalier le plus fort du mois, ses quantiles et ses valeurs remarquables.

On associe donc ici un pas de temps journalier (pour la mesure du débit et son analyse fréquentielle chaque jour) et le pas de temps mensuel pour l'analyse fréquentielle des débits journaliers regroupés mois par mois. Evidemment d'autres combinaisons sont possibles (débits journaliers regroupés saisons pas saisons, années par années, etc.).

2.3. Exemple montrant l'importance du choix du pas de temps

Les exemples ci-dessus montrent qu'il est possible d'étudier les variations saisonnières des débits à différents pas de temps : mensuel au paragraphe 2.1. , journalier au paragraphe 2.2. La question posée est donc la suivante : pourquoi étudier un pas de temps plutôt qu'un autre ? Il est a priori aussi légitime de s'intéresser aux variations des débits journaliers qu'à celles des débits mensuels. Dans ce cas, le phénomène étudié n'est pas le même, selon le pas de temps que l'on considère. Comme nous l'avons noté au paragraphe 1.2. , si l'on étudie les crues sur des petits bassins versants (phénomènes observables sur quelques jours, selon les bassins), le pas de temps journalier est plus adapté que le pas de temps mensuel. En revanche, si l'on étudie la réponse des réservoirs de ces petits bassins face au cycle des saisons, le pas de temps mensuel semble plus adéquat. En revanche sur des bassins de grande taille, le pas de temps mensuel peut être adapté pour l'étude des crues.

Afin de vérifier cette hypothèse, nous nous sommes intéressés aux variations saisonnières du débit de la Rivière Nouvelle au Québec (1140 km²). Cette station fait partie de la base de données canadienne étudiée au chapitre 5 et a été choisie car soumise à un régime climatique très contrasté, avec des températures variant entre -40° en hiver et $+30^{\circ}$ en été.

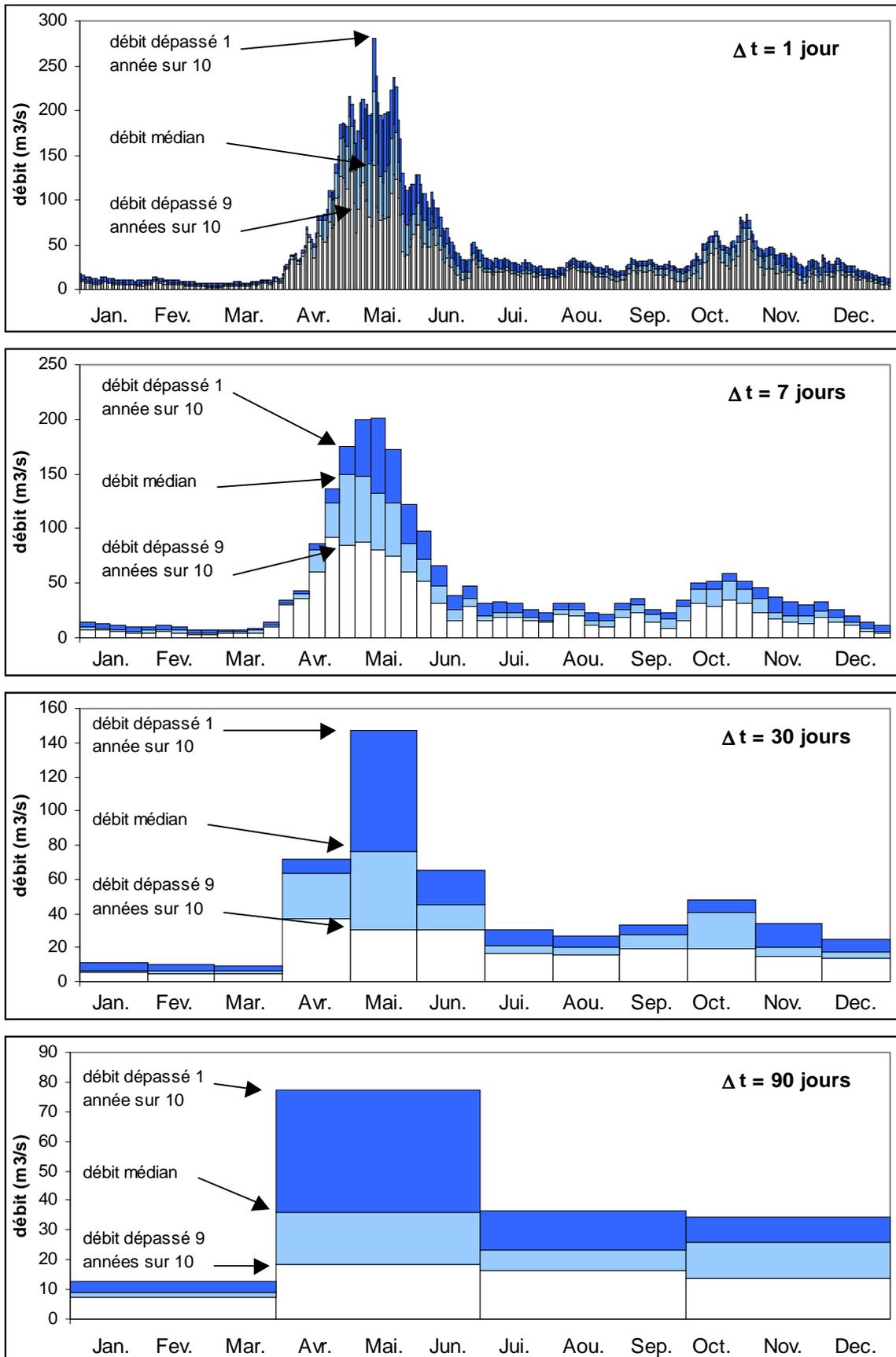


Figure 1-8 : Evolution des variations saisonnières du débit moyen en fonction de son pas de temps Δt . Exemple de la Rivière Nouvelle au Québec (1140 km²), sur la période 1965-1994

La Figure 1-8 donne le résultat des calculs réalisés. La chronique disponible de 30 années de débit moyen journalier a servi à construire d'autres chroniques de débit moyen, mais à des pas de temps Δt supérieurs. Quatre pas de temps ont ainsi été étudiés : $\Delta t = 1, 7, 30$ et 90 jours. Nous avons ensuite calculé pour les quatre chroniques obtenues les débits interannuels dépassés en moyenne 1 année sur 10, 5 années sur 10 (débit médian) et 9 années sur 10. Ces débits correspondent respectivement aux quantiles 90%, 50% et 10% pour chaque pas de temps.

Les résultats montrent que pour tous les pas de temps (sauf peut-être 90 jours), on observe la même évolution saisonnière des débits : deux saisons de hautes eaux, une au printemps suivie d'une autre moins marquée en automne et deux saisons de basses eaux durant les mois d'hiver et d'été. Cette évolution des débits s'explique évidemment par le contexte climatique dans lequel se trouve cette rivière :

- l'*hiver* très rude durant lequel le froid fait que tous les apports pluviométriques sont stockés sous forme de neige ;
- le *printemps* durant lequel tous ces apports stockés sont libérés sous l'effet du réchauffement ;
- l'*été* où les nappes alimentent les écoulements ;
- puis enfin, l'*automne* lors duquel l'augmentation des écoulements s'explique par un regain de l'activité pluviométrique, qui contrairement à l'hiver, n'est pas encore stockée sous forme de neige.

Concernant les différences entre les pas de temps, on observe naturellement que plus le pas de temps augmente, plus la variabilité saisonnière des débits est lissée. Le pas de temps de 90 jours nous empêche d'observer la saison de hautes eaux en automne. Mais cette constatation ne signifie pas qu'inversement, la diminution du pas de temps amène forcément une augmentation de l'information pertinente. En effet, selon le problème étudié, les variations supplémentaires qui apparaissent au pas de temps de 1 jour par rapport à 7 jours ne sont pas nécessairement significatives. On observe par exemple deux pointes distinctes en mai qui n'ont pas forcément de sens. En particulier, le débit journalier de 140 m³/s est jugé anormalement bas le jour correspondant à la plus forte de ces pointes, puisque dépassé en moyenne plus de 9 années sur 10. Puis, seulement trois jours plus tard, ce même débit est jugé tout à fait « de saison », car proche du débit médian. Cette remarque invite à s'interroger par exemple sur la pertinence des informations sur la carte présentée Figure 1-6 : les mêmes débits relevés simplement 3 jours plus tard peuvent donner dans certains cas des cartes complètement différentes !

2.4. Commentaire sur les différentes « durées » de l'analyse

Ces remarques, associées à celles que nous ont suggérées la Figure 1-6 et la Figure 1-7 permettent de distinguer plusieurs « durées » dont les rôles doivent être distingués :

- Tout d'abord, la **durée de l'année** : c'est elle qui rythme l'ensemble des mécanismes climatiques et hydrologiques, au travers du cycle des saisons établi par le temps que met la terre à faire le tour du soleil. Cette durée permet de définir la date et sert aussi d'unité pour exprimer la période de retour d'un événement, notion que nous expliciterons plus loin. La variabilité à laquelle nous nous intéressons pratiquement toujours est une variabilité interannuelle : on compare l'événement (ou les événements) de l'année n à celui (ou ceux) de toutes les autres années. Implicitement ou explicitement, l'année est presque toujours une durée de référence.
- Le pas de temps sur lequel la grandeur est moyennée (ici le débit, cf. §1.3.) : instantané, horaire, journalier, mensuel, etc. Selon la taille du bassin ce pas de temps peut prendre diverses significations :
 - Pour de très petits bassins le pas de temps instantané est nécessaire pour étudier les crues. Le pas de temps mensuel donne plutôt une information sur l'interaction entre pluie, évapotranspiration et capacité du sol à stocker ou à restituer l'eau.
 - Pour des bassins de grande taille, le pas de temps journalier a la signification d'un débit instantané, le pas de temps mensuel peut dans certains cas correspondre à celui d'une crue. Il faut éventuellement passer à un pas de temps plus important pour décrire les interactions entre précipitations, évaporations et réservoirs sols.

Ce « pas de temps » dépend de la « dynamique » du phénomène étudié. Il mérite de toute façon une approche moins empirique que celle consistant à le choisir arbitrairement. On décrira plus loin la méthode débit-durée-fréquence qui permet de considérer de nombreux pas de temps simultanément.

- La durée pendant laquelle sont regroupées les mesures prises au pas de temps précédent pour :
 - i. sélectionner des valeurs caractéristiques de chaque regroupement (selon les cas : moyenne, maximum ou minimum observé sur chaque intervalle) ;
 - ii. regrouper année par année (valeurs interannuelles) au sein d'un même échantillon les valeurs caractéristiques ci-dessus .On obtient ainsi un échantillon pour chaque intervalle défini par la durée de regroupement sur une année. Par exemple, on obtient respectivement 365, 12 et 1 échantillon(s), suivant la durée de regroupement (1 jour, 1 mois, 1 an) ;
 - iii. réaliser une analyse statistique sur chaque échantillon ci-dessus (selon les cas, simple moyenne, calcul des quantiles expérimentaux, ou encore ajustement d'une loi statistique).

La durée de regroupement permet de réaliser un échantillonnage saisonnalisé.

Parmi les approches présentées, le pas de temps et la durée de regroupement peuvent être égaux ou bien différents. Pour les approches dérivées de celle de Pardé (Figure 1-5), ceux-ci sont tous deux égaux à 1 mois. De même pour la Figure 1-8, pas de temps et durée de regroupement varient tous deux de 1 jour à 3 mois.

Par contre, pour l'exemple de la Figure 1-7, la durée de regroupement est le jour d'abord, le mois ensuite, alors que le pas de temps reste journalier. De même, dans l'analyse fréquentielle des crues présentée plus en détail par la suite, on a l'habitude de ne considérer que le pas de temps journalier ou instantané, et une durée de regroupement soit annuelle, soit de un ou quelques mois (par exemple le printemps). Pour les étiages l'analyse est similaire, toutefois le pas de temps journalier est souvent remplacé par un pas de temps plus long. Le regroupement est le plus souvent simplement annuel.

Ces différents exemples illustrent le fait qu'en général il ne doit pas y avoir identité entre pas de temps de mesure du débit et durée de regroupement pour l'analyse fréquentielle. Cette dernière durée jouant un rôle essentiel sur l'analyse de la variabilité, elle mérite une attention particulière qui manque dans beaucoup d'études. On voit désormais des analyses qui considèrent des durées de regroupement par « saisons » définies arbitrairement : par exemple les quatre saisons. Cette approche est parfaitement justifiée mais on peut imaginer une étude – que nous ne ferons pas ici – sur l'optimisation de ces durées de regroupement sur des critères plus rigoureux. Par exemple, Kieffer (1998) effectue l'échantillonnage des pluies maximales annuelles considérant différentes durées de regroupement (ou saisons). Ceci lui permet de définir pour chaque poste et chaque pas de temps étudié la saison à plus fort risque. Dans un autre ordre d'idée, la méthode d'échantillonnage « sup-seuil » dont il sera question par la suite permet de définir des durées de regroupement variables. Dans ce cas, l'intervalle durant lequel on sélectionne la plus forte valeur de débit observé n'est plus fixe, avec une durée et une date donnée, mais débute lorsque le débit passe au-dessus d'un seuil fixé, puis se termine lorsqu'il repasse au-dessous. Il s'agit là de formes de réponse à la question de la définition optimale de la durée de regroupement, sans pour autant en épuiser le sujet.

3. Analyse du régime des crues : introduction à l'approche débit-durée-fréquence

Comme nous l'avons vu au paragraphe 2, la représentation saisonnière des débits interannuels (pas de temps et durées de regroupement de l'ordre de la saison) informe sur les variations lentes des écoulements. Elle n'est pas adaptée à l'étude des crues, même si l'on descend au pas de temps journalier. Le but de ce paragraphe est de présenter une méthode pour décrire le régime des crues de façon complète, notamment en prenant en compte une notion essentielle pour les crues : la durée (le « pas de temps » selon le vocabulaire précédent).

3.1. Echantillonnage des valeurs maximales

Afin d'étudier le régime des crues en particulier, il faut extraire uniquement les événements qui nous intéressent. En analyse fréquentielle des crues, trois phases principales sont généralement distinguées :

- la définition de la variable d'intérêt : **le débit de pointe**, c'est à dire la plus forte valeur de débit atteinte pendant la crue. Selon le type de chronique analysée il faut bien préciser de quel débit il s'agit : débit de pointe instantané, horaire, journalier, ... ;
- l'**échantillonnage**, qui consiste à sélectionner les plus forts débits de pointe observés ;
- l'**ajustement** d'une loi de probabilité, permettant de définir la fréquence théorique de chaque événement, souvent exprimée en terme de période de retour T .

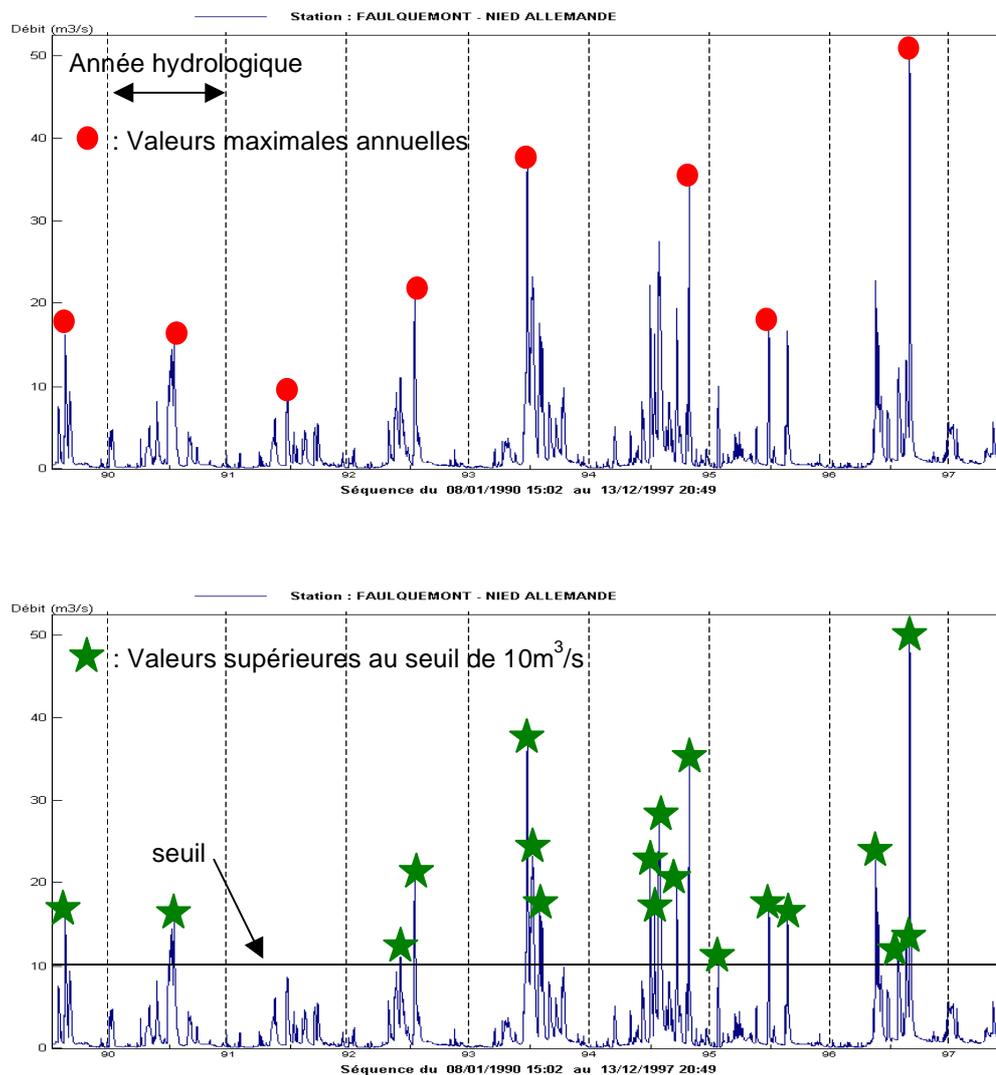


Figure 1-9 : Comparaison d'un échantillonnage par maximum annuel et d'un échantillonnage par valeurs supérieures à un seuil effectué sur une chronique de débit instantané. Exemple de la Nied Allemande à Faulquemont (187 km²)

On distingue deux méthodes d'échantillonnage : la méthode du maximum annuel et celle des valeurs supérieures à un seuil. Dans le premier cas, l'échantillon est composé du plus fort débit de chaque année hydrologique. Dans le second, l'échantillon est composé du plus fort débit de chaque événement, un événement commençant lorsque le débit passe au-dessus du seuil fixé, et s'arrêtant lorsqu'il repasse au-dessous. La Figure 1-9 illustre ces deux méthodes pour l'exemple de la Nied Allemande déjà présenté.

3.2. La prise en compte de la date de crues

Avant de nous intéresser à l'analyse fréquentielle des débits échantillonnés, ce paragraphe présente différentes méthodes pour prendre en compte une information importante sur le régime : la date d'apparition de chaque crue. En couplant cette information avec les conditions climatiques régnant sur le bassin aux mêmes dates, on a ainsi une idée sur l'origine des crues (fonte nivale au printemps, orages en automne, etc.).

Une méthode simple est de diviser l'année en un nombre de classes (par exemple 12 ou 24) et de compter le pourcentage d'événement ayant lieu dans chaque. Lecce (2000) utilise cette représentation pour étudier la répartition par mois de la crue annuelle (méthode du maximum annuel). Il en déduit différentes régions aux régimes bien distincts dans le sud-est des Etats-Unis. Magilligan et Graber (1996) ont utilisé cette approche en Nouvelle Angleterre. La Figure 1-10 montre la répartition du nombre de crues sélectionnées par mois, par la méthode sup-seuil, pour les deux exemples présentés Figure 1-13 (Nied Allemande et Jabron). Dans les deux cas, le seuil a été fixé de façon à sélectionner deux valeurs par an en moyenne, ce qui signifie que la pointe de la plus petite crue sélectionnée a une période de retour de 6 mois. Avec une forme de représentation équivalente, il est aussi possible d'illustrer la moyenne et l'écart type des valeurs observées.

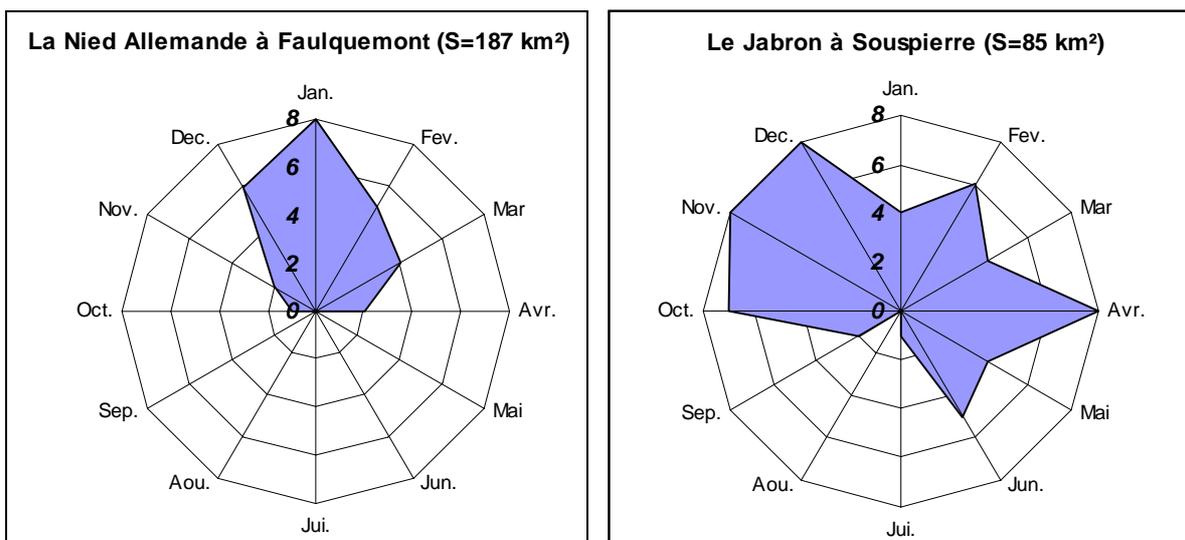


Figure 1-10 : Nombre d'événements sélectionnés par mois par la méthode sup-seuil.

Pour la Nied Allemande, on constate que les crues apparaissent principalement autour des mois de janvier et décembre. Au contraire, pour le Jabron, les crues sont beaucoup plus dispersées dans le temps, entre octobre et juin.

Une autre approche voisine de celle décrite ci-dessus consiste à utiliser les « statistiques circulaires ». Des exemples d'application en hydrologie se trouvent dans Magilligan et Graber (1996), ou encore dans Black et Werritty (1997). La théorie des statistiques circulaires est décrite par Mardia (1972). L'idée est de représenter la date de chaque crue (sélectionnée par la méthode « sup-seuil » ou « maxi-annuel ») par un angle θ :

$$\theta_i = j_i \left(\frac{2\pi}{365} \right) \quad (1-1)$$

avec : j_i le numéro du jour de la crue i , compris entre 1 et 365

Chaque date de crue peut être représentée par un vecteur unité, faisant un angle θ avec l'axe des abscisses. Il est alors possible de calculer le vecteur correspondant à la date moyenne des crues observées, noté V . Ses coordonnées x_{moy} et y_{moy} sont exprimées par :

$$x_{moy} = \frac{1}{n} \sum_{i=1}^n \cos(\theta_i) ; y_{moy} = \frac{1}{n} \sum_{i=1}^n \sin(\theta_i) \quad (1-2)$$

L'angle formé par ce vecteur et l'axe des x correspond à la date moyenne d'occurrence des crues :

$$\theta_{moy} = \tan^{-1} \left(\frac{y_{moy}}{x_{moy}} \right) \quad (1-3)$$

La longueur de ce vecteur est une mesure de la « saisonnalité » des crues. Si cette valeur est proche de 1, cela signifie que toutes les crues se produisent toujours à la même date. Au contraire, si cette valeur est proche de zéro, cela signifie que les crues se produisent tout au long de l'année.

$$\|\vec{V}\| = (x_{moy}^2 + y_{moy}^2)^{1/2} \quad (1-4)$$

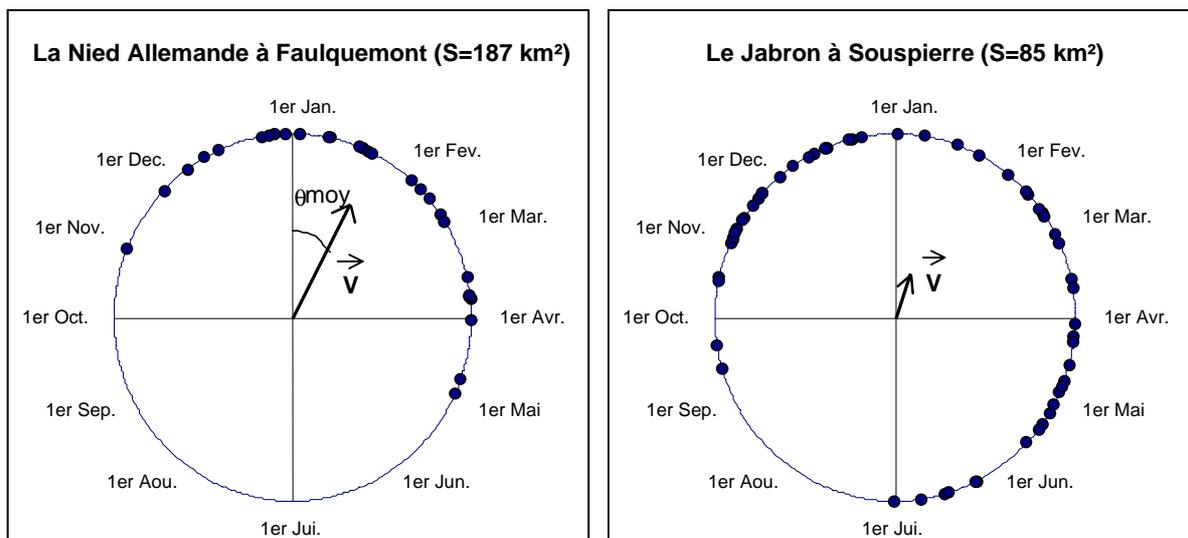


Figure 1-11 : Statistiques circulaires sur la date des pointes de crues sélectionnés par la méthode sup-seuil.

La Figure 1-11 permet de comparer les dates d'apparition des crues pour les deux exemples étudiés. En particulier, la taille du vecteur V indique que les crues de la Nied Allemande ont un caractère saisonnier très marqué. A l'inverse, dans le cas du Jabron, la petite taille du vecteur indique que le caractère saisonnier n'est pas net, l'orientation du vecteur n'est donc pas très significative. Nous touchons là à la principale limite de cette représentation. Elle ne fait pas la distinction entre un bassin pour lequel les crues n'ont pas de caractère saisonnier et un bassin pour lequel il y a deux saisons très marquées, mais à l'opposées l'une de l'autre. Dans ce cas, l'analyse fréquentielle doit être menée individuellement sur chacune des saisons.

L'avantage de cette représentation est d'être facilement cartographiée : chaque bassin est représenté par une flèche indiquant la date moyenne d'apparition des crues. La Figure 1-12 montre un exemple de carte pour le Royaume-Uni (Institute of Hydrology, 1999). Les crues étudiées ont été sélectionnées par valeurs supérieures à un seuil de façon à avoir trois crues en moyenne par an. On voit sur cet exemple que la date d'apparition des crues dépend de la localisation géographique des bassins.

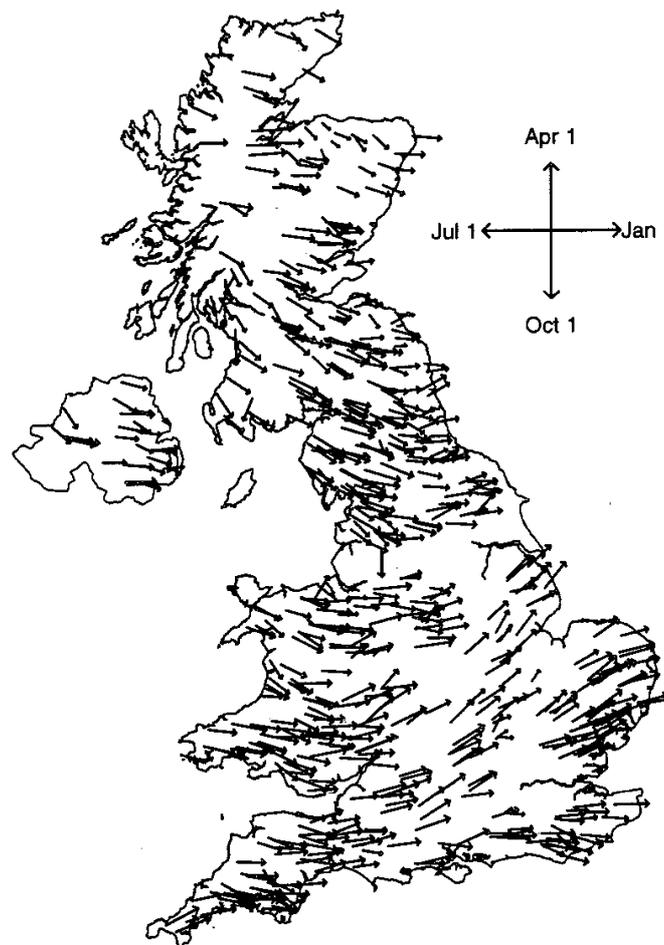


Figure 1-12 : Cartographie de la date moyenne d'apparition des crues au Royaume-Uni (Institute of Hydrology, 1999)

3.3. Les limites de l'analyse fréquentielle des pointes de crues

La Figure 1-13 montre le résultat d'une analyse fréquentielle réalisée sur les débits de pointe instantanés, échantillonnés par la méthode des valeurs supérieures à un seuil. Dans les deux cas, le seuil a été fixé de façon à sélectionner en moyenne deux valeurs par an. Nous reviendrons dans le détail au chapitre 2 cette technique d'échantillonnage. Deux bassins sont présentés : la Nied Allemande à Faulquemont, déjà présenté et située dans le département de la Moselle, et le Jabron à Souspierre, dans la Drôme. A première vue, les distributions obtenues sur ces deux bassins semblent très similaires. Par exemple, dans les deux cas, on trouve un débit de pointe décennal proche de 40 m³/s.

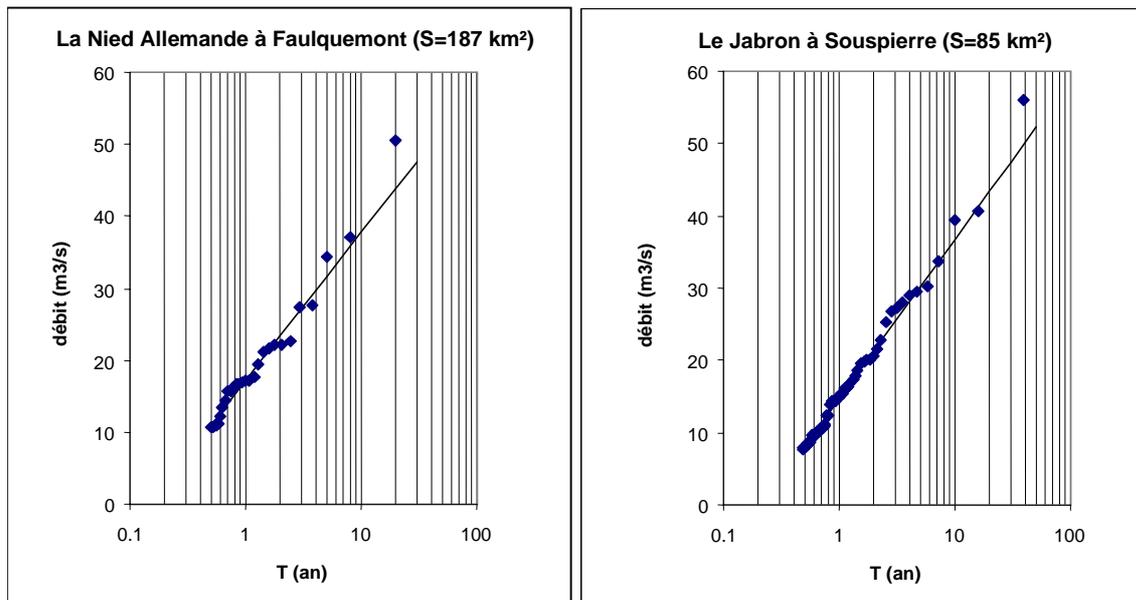


Figure 1-13 : Distributions des débits de pointe instantanés sélectionnés par la méthode seuil.

Comme le montre la Figure 1-14, bien que les pointes de crues observées sur ces deux bassins soient comparables, leur dynamique est complètement différente. Dans le cas de la Nied Allemande, les crues observées sont très lentes et très volumiques, avec une durée de l'ordre de la semaine. Pour le Jabron, les crues sont très rapides, peu volumiques, et ne durent que quelques heures. En terme d'aménagement, les mesures de prévention contre les inondations doivent tenir compte de cet aspect. Par conséquent, la simple étude fréquentielle des pointes de crues observées n'apporte qu'une information incomplète, ne tenant pas compte de la durée des crues.

Les paragraphes suivants proposent différentes approches pour prendre en compte cette notion de « dynamique » des crues.

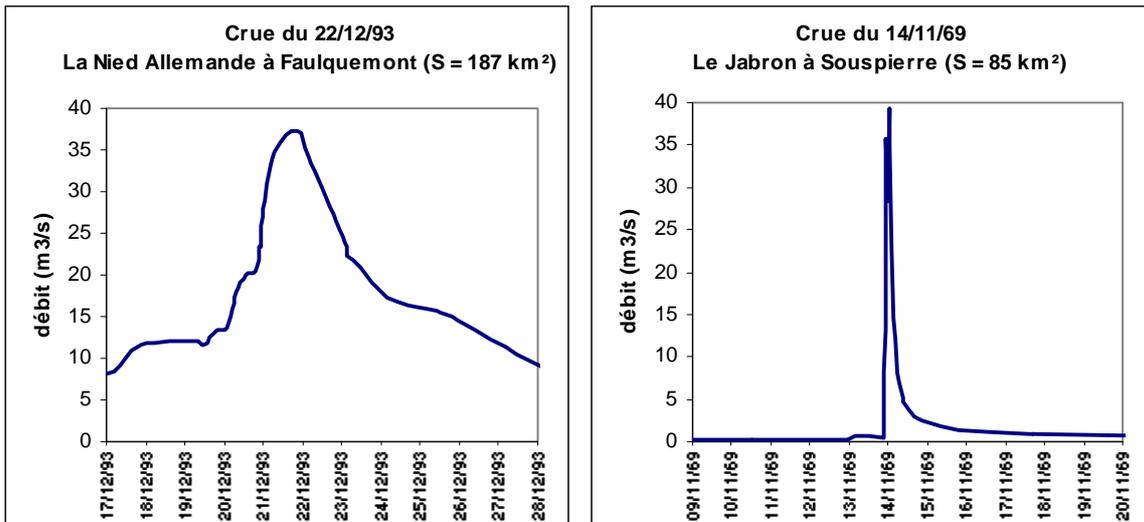


Figure 1-14 : Comparaison des hydrogrammes instantanés des crues dont le débit de pointe est proche de la valeur décennale

3.4. Analyse des durées de crue : définition de la durée de SOCOSE

3.4.1. Durées relatives à un seul événement de crue

Comme nous venons de le voir, la notion de durée est indispensable dans la description du régime des crues. C'est pourquoi de nombreuses durées ont été définies par les hydrologues. Les définitions suivantes sont issues de Roche (1963).

Les temps caractéristiques définis ci-dessous sont tous relatifs à un événement de crue. Ce sont des ordres de grandeurs, plutôt que des valeurs calculables avec précision. En effet, leur définition fait référence à une représentation très simplifiée de la transformation pluie-débit (en particulier les écoulements sont supposés être majoritairement des "ruissellements").

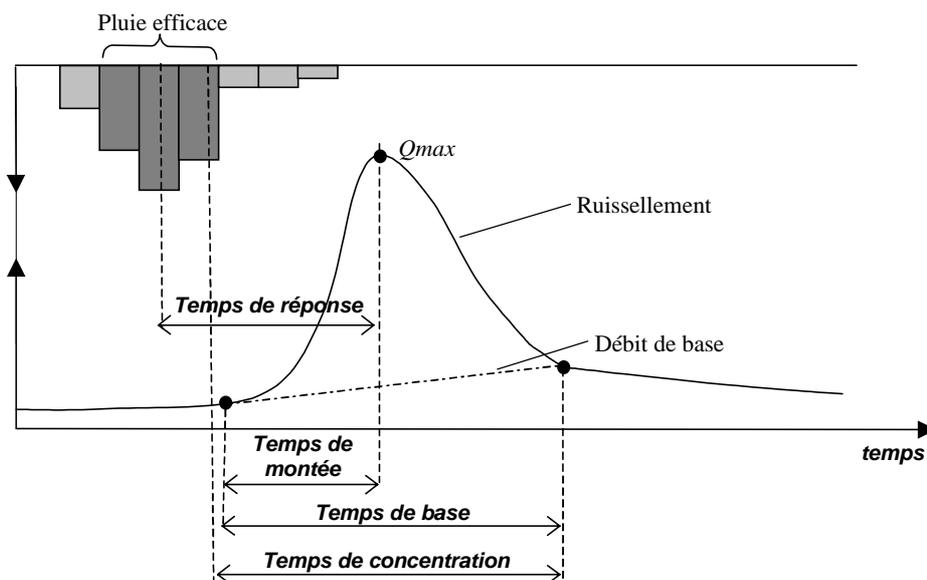


Figure 1-15 : Définition des durées caractéristiques relatifs à un hydrogramme de crue

Temps de réponse : Le temps de réponse est l'intervalle de temps qui sépare le centre de gravité de la pluie efficace de la pointe de l'hydrogramme.

Temps de montée : Il s'agit de la durée écoulée entre le début de l'arrivée à l'exutoire du « ruissellement » et le maximum de l'hydrogramme.

Temps de base : Le temps de base (ou durée de « ruissellement ») est la longueur sur l'abscisse de la base de l'hydrogramme de ruissellement.

Temps de concentration : Le temps de concentration est la durée que met une particule d'eau provenant de la partie du bassin hydrauliquement la plus éloignée de l'exutoire. On peut l'estimer en mesurant la durée comprise entre la fin de la pluie efficace et la fin du ruissellement.

3.4.2. Durée caractéristique de crue d'un bassin : la durée de Socose

Il s'agit maintenant de définir une durée caractérisant non plus un seul événement comme cela a été présenté ci-dessus, mais la **dynamique de crue**, caractéristique de l'ensemble des événements possibles. Pour ce faire, le Cemagref (CTGREF, 1980) propose dans un premier temps de calculer pour chaque hydrogramme de crue, la durée t_s pendant laquelle le débit dépasse la moitié du débit de pointe Q_s (Figure 1-16).

L'intérêt de la définition de cette durée t_s est de reposer sur un critère objectif et reproductible pour chaque hydrologue, à l'inverse des durées précédentes (§ 3.4.1) qui nécessitent un choix subjectif du début et de la fin de la crue.

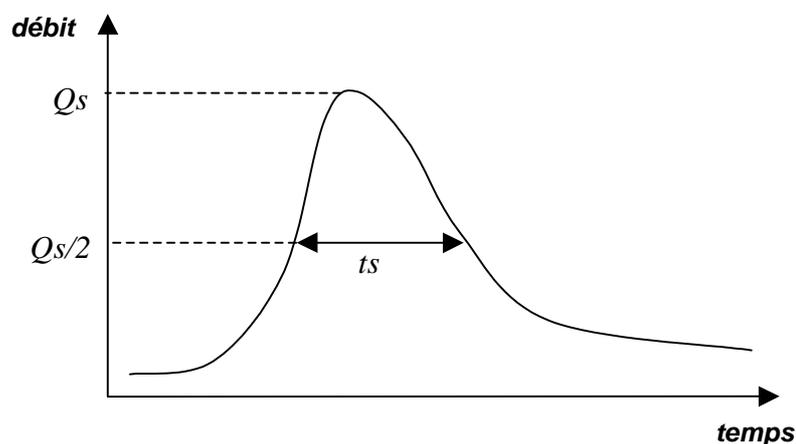


Figure 1-16 : Définition des variables Q_s et t_s pour une crue observée

Pour déterminer une durée caractéristique de crue relative non plus à une crue observée, mais au bassin étudié, la méthode *Socose* (CTGREF, 1980) propose de représenter sur un même graphe l'ensemble des couples (Q_s, t_s) obtenus pour les N plus fortes crues observées. La méthode consiste à définir plusieurs classes de Q_s , puis pour chaque classe, à calculer la durée d_s (médiane des valeurs de t_s dans la classe), et ensuite à tracer une courbe reliant ces

différentes valeurs médianes. Enfin, la valeur de la durée D , appelée durée de *Socose*, est lue sur cette courbe, pour un débit Q_s égal au débit instantané décennal, noté Q_{IXA10} .

La Figure 1-17 montre le calcul de la durée D pour les deux bassins précédemment étudiés, à savoir la Nied Allemande et le Jabron. Comme le laissait présager l'étude des deux hydrogrammes de crues présentés Figure 1-14, les durées D calculées sont très différentes, puisque l'on trouve respectivement pour ces deux bassins 2.5 jours (60 heures) et 0.2 jour (5 heures).

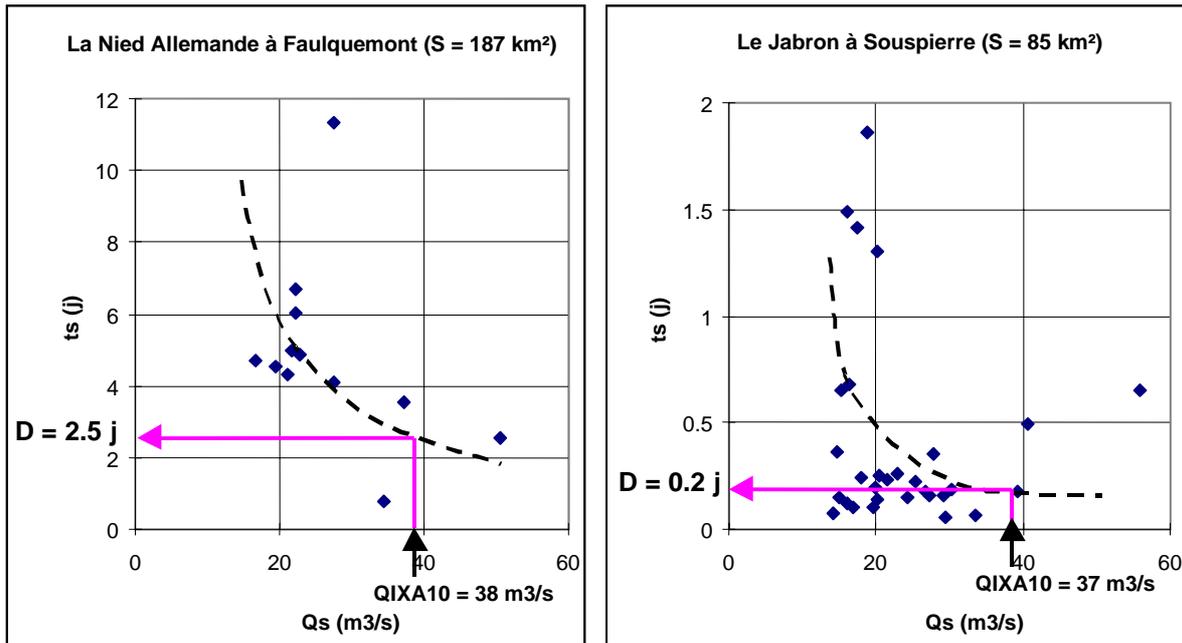


Figure 1-17 : Calcul de la durée caractéristique D de *Socose*

Si la détermination des couples (Q_s, ts) est rigoureuse et ne pose pas de problème (Figure 1-16), le calcul de la durée D à partir des nuages de points obtenus peut paraître délicat, voire hasardeux (Figure 1-17). Il est vrai que le résultat obtenu contient une part de subjectivité, notamment dans le choix des classes pour tracer la courbe reliant les différentes médianes. De plus, la dispersion des points autour de la courbe ainsi obtenue appelle à la prudence, quant à la valeur de D obtenue. Sa détermination contient donc une part d'incertitude mêlée à un brin d'expertise. Dans la pratique, l'idée est de se concentrer principalement sur les crues proches de la crue décennale, ce qui fait que souvent, on ne tient finalement compte que de ces quelques crues pour déterminer la valeur de D .

Cependant, la durée D de *Socose* a l'avantage de définir une valeur relative au bassin versant étudié, simple à calculer, au regard d'autres méthodes existantes, mettant par exemple en œuvre la théorie de l'hydrogramme unitaire. La valeur obtenue est à prendre comme un ordre de grandeur, donnant un bon indicateur de la durée des crues observées sur le bassin. Dans l'exemple traité ci-dessus, la durée D nous indique clairement que les deux bassins ont des dynamiques de crue très différentes. Ceci nous conduit donc à considérer cette durée

caractéristique comme une grandeur intéressante, tout en gardant à l'esprit qu'il s'agit bien d'un ordre de grandeur. Nous y ferons référence dans la suite de ce travail.

3.5. L'analyse combinée pointe-volume des crues

Pour palier au problème rencontré au paragraphe 3.3. , à savoir la non-prise en compte de la durée des crues par l'analyse fréquentielle des débits de pointe, certains auteurs ont décidé de caractériser les crues par leur pointe, mais aussi leur durée et leur volume (Figure 1-18). D'après Yue *et al.* (1999), cette façon de faire a été initiée par Ashkar (1980).

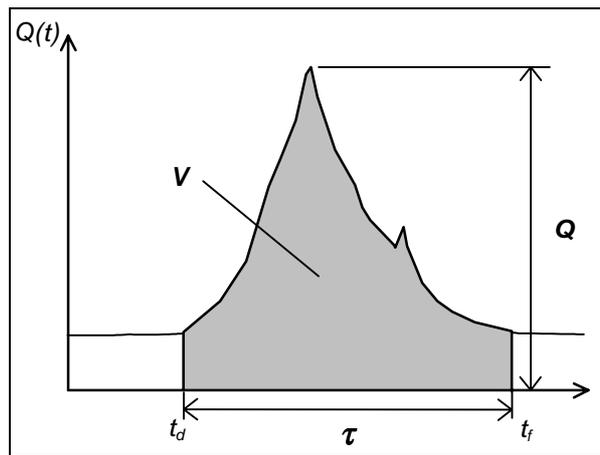


Figure 1-18 : Définition de Q, V et τ pour une crue

Pour chaque crue observée, on doit définir sa date de début et de fin (t_d et t_f). Il est possible de définir la pointe de la crue (Q), la durée (τ), et le volume écoulé (V). Q , τ , V sont alors considérés comme des variables aléatoires. L'analyse statistique consiste à étudier deux à deux les variables ci-dessus, au moyen de lois à deux variables. Par exemple, Yue *et al.* (1999) proposent une loi de Gumbel à deux variables.

Cette façon de faire est séduisante, puisqu'elle tient aussi compte des volumes de crue. Cependant, la méthode d'échantillonnage invite à la prudence. Tout d'abord, la définition des dates de début et de fin de crue (t_d et t_f) contient une part de subjectivité, et ne peut être faite de façon automatique. D'autre part, la méthode d'échantillonnage utilisée par les auteurs est celle du maximum annuel. Se pose alors la question suivante : doit-on sélectionner les crues ayant les plus fortes pointes, celles ayant duré le plus longtemps, ou celles ayant le plus fort volume ? En effet, il peut arriver que certaines années, ces trois conditions soient remplies par des événements différents. Ce point est soulevé par Cole et Tawn (1991). D'après Adamson *et al.* (1999), une réponse consiste à sélectionner la plus forte valeur annuelle de chaque variable aléatoire étudiée, même si les valeurs sélectionnées sont relatives à des événements différents.

3.6. L'approche débit-durée-fréquence

L'analyse débit-durée-fréquence, dite QdF , est une analyse fréquentielle des débits similaire à l'analyse en intensité-durée-fréquence (IdF) pour les pluies. Si l'on dispose d'une chronique de débit instantané, on peut caractériser chaque crue observée par son débit de pointe instantané (Q_{max}), mais aussi par le volume maximal qui s'est écoulé sur une certaine durée d . En divisant ce volume maximal par d , on obtient le débit moyen maximal que nous noterons Vd , en référence au volume et à la durée qui ont servi à son calcul. D'après les notations définies par Oberlin (1992), Vd est appelé $VCXd$. Si l'on fait de nouveau le lien avec l'approche intensité-durée-fréquence, la différence entre le volume et le débit moyen est la même qu'entre la hauteur de précipitation et l'intensité.

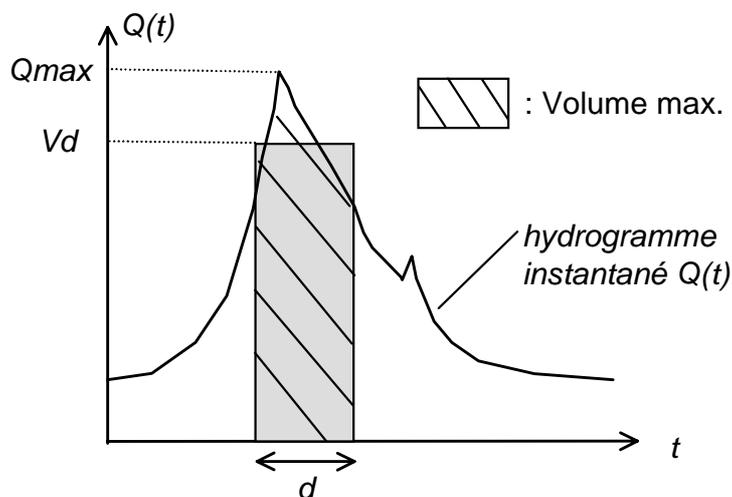


Figure 1-19 : Définition des caractéristiques de crue

Afin d'échantillonner les débits Vd sur l'ensemble de la chronique, une méthode pratique est de faire passer une moyenne mobile de largeur d sur la chronique de débit instantané, puis de sélectionner les valeurs maximales de cette nouvelle chronique, comme pour n'importe quelle autre chronique de débit (Lang, 1995). La Figure 1-20 donne un exemple de transformation d'une chronique de débits instantanés, en une chronique de débits moyens sur 1 jour et 5 jours.

Remarque

Les chroniques obtenues Figure 1-20 sont un peu différentes des chroniques de débits moyens décrites au paragraphe 1.3.3. Il s'agit ici de moyennes glissantes, ce qui signifie que d'une part, deux points voisins de la courbe utilisent des données communes, et d'autre part, que la durée entre deux points de la chronique n'est pas forcément constante, tout comme la chronique de débit instantané traitée (cf. §1.3.2.).

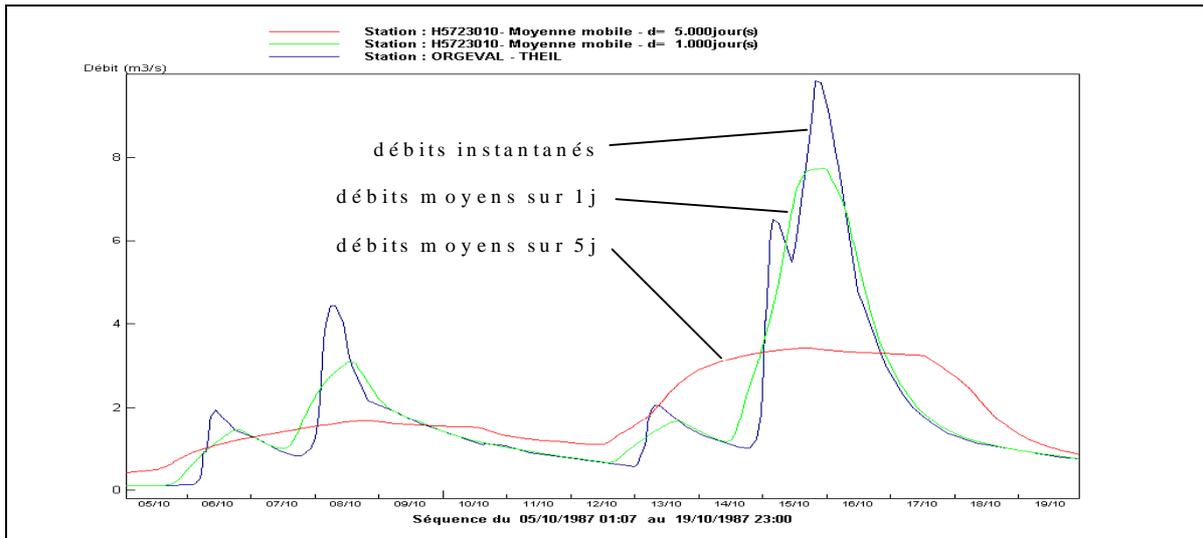


Figure 1-20 : Transformation de la chronique de débit instantané en débit moyen par moyennes mobiles

Les chroniques de débits instantanés de la Nied Allemande et du Jabron précédemment analysées (Figure 1-13) ont donc été moyennées pour des durées de 1, 2 et 5 jours. La Figure 1-21 donne le résultat de l'analyse fréquentielle effectuée sur ces nouvelles chroniques. Pour chacune des durées considérées, on obtient une nouvelle distribution de débits moyens, $d=0$ correspondant à la distribution des débits instantanés (chronique initiale). L'ensemble de ces courbes est appelé courbes débit-durée-fréquence, par analogie aux courbes intensité-durée-fréquence.

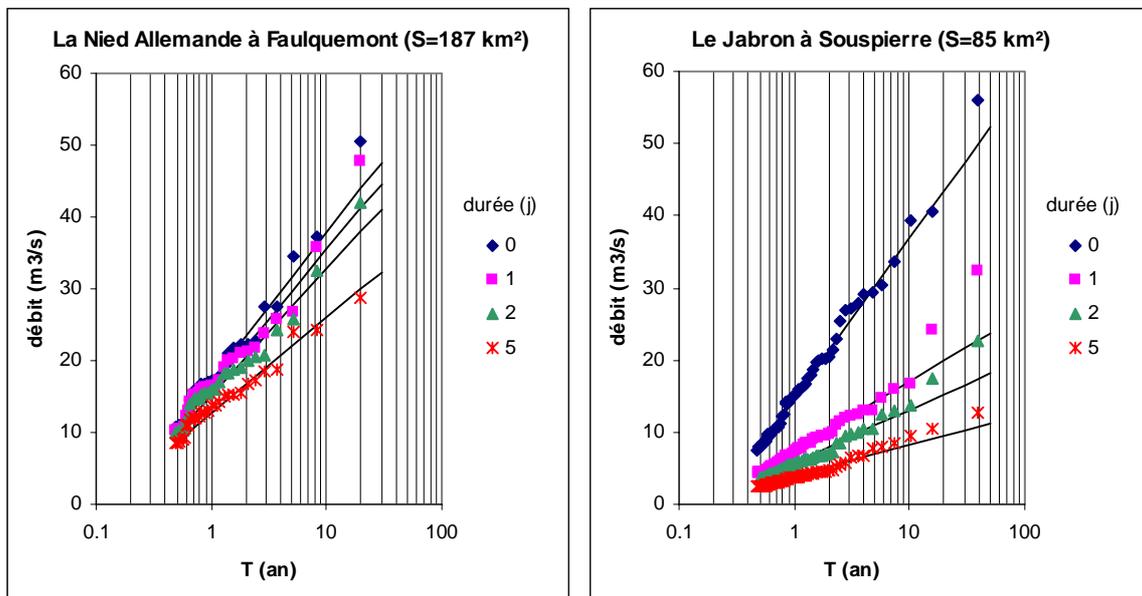


Figure 1-21 : Courbes débit-durée-fréquence des débits moyens sur d

Contrairement à la Figure 1-13, on observe maintenant une grande différence entre les faisceaux de distributions relatifs aux deux bassins. Celui de la Nied Allemande est bien plus

resserré. En effet, sur ce bassin, les crues étant beaucoup plus volumiques (Figure 1-14), la distribution des débits moyens sur 5 jours est au-dessus de celle du Jabron.

L'analyse débit-durée-fréquence permet donc de rendre compte de la dynamique des crues. Le fait de décrire la variabilité des débits moyens pour différentes durées fait écho au paragraphe 2.3. où nous avons étudié l'influence du pas de temps dans l'étude des variations saisonnières des débits moyens interannuels. Ici aussi, le choix des durées étudiées est important. Si l'on souhaite étudier le régime des crues, il ne faut pas que les pas de temps choisis dépassent la durée des crues observées. Il est possible de faire l'analyse sur de plus longues durées. Dans ce cas, la variabilité observée ne sera plus due aux crues, mais à des variations à plus long terme du débit.

Contrairement à l'approche intensité-durée-fréquence, l'analyse débit-durée-fréquence reste peu utilisée malgré son intérêt. Le Cemagref travaille depuis plusieurs années ce type d'approche (Moallemi, 1988), Prudhomme (1995), Galéa et Prudhomme (1997). Nous reviendrons en détail sur les modèles développés. En dehors de ces travaux, on trouve quelques références dans la littérature, souvent sous le terme anglais « *volume-duration-frequency curves* » (NERC, 1975 ; Sherwood, 1994 ; Balocki et Burges, 1994 ; Maione *et al.*, 2000).

Conclusion du chapitre 1

Il existe différentes formes de représentation des régimes hydrologiques. Chacune à sa manière constitue une synthèse de la variabilité observée des écoulements. Nous adopterons donc la définition simple suivante :

Décrire le régime hydrologique d'une rivière c'est décrire la variabilité temporelle de ses écoulements.

Cette conception du régime hydrologique repose sur la volonté de lui donner un contenu précis et d'éviter de faire rentrer dans l'analyse des régimes toute l'hydrologie de surface. Dans ce contexte, les recherches d'explications de cette variabilité temporelle constituent des commentaires, mais ne font pas partie de sa définition.

Comme nous l'avons noté, la notion de régime est liée à celle d'échelle temporelle. Ainsi, il n'y a pas un régime, mais plutôt des régimes associés à des phénomènes observables sur différentes échelles temporelles. Dans ce cadre, le choix du pas de temps ainsi que celui de la durée de regroupement définis au paragraphe 2.4. sont particulièrement importants. Ils permettent de se concentrer sur une échelle temporelle et sur un phénomène particuliers (crues par exemple, pour ce qui concerne la suite de ce travail).

Concernant le régime des crues, l'approche débit-durée-fréquence nous paraît bien adaptée. En effet, elle est capable de prendre en compte la durée, notion indispensable lorsque l'on parle de crues. Elle considère donc des « pas de temps variables ». En ce qui concerne les durées de regroupement de l'analyse fréquentielle, notre approche sera celle de la méthode de l'échantillonnage par valeurs supérieures à un seuil, qui considère des durées de regroupement variables. Nous allons donc dans la suite étudier cette approche plus en détail, avec une attention particulière pour les travaux menés ces dernières années par le Cemagref.

Chapitre 2

Echantillonnage par valeurs supérieures à un seuil

Le chapitre précédent a servi d'introduction à l'approche « débit-durée-fréquence » (QdF) et a permis à l'aide de deux exemples de montrer son intérêt : la prise en compte de la notion de durée dans l'analyse fréquentielle des crues. Avant d'entrer plus dans le détail de cette approche, ce qui sera fait au chapitre 3, ce chapitre propose de passer quelques instants sur la première étape de toute analyse fréquentielle : l'échantillonnage. En effet, la qualité des résultats ultérieurs dépend en premier lieu du soin apporté à cette première étape qui renvoie aux remarques faites au chapitre 1 sur les choix des pas de temps et des durées de regroupement. Il nous paraît donc essentiel d'en dire quelques mots. Comme cela a été rapidement présenté au chapitre 1, il existe en hydrologie deux principales méthodes d'échantillonnage des valeurs extrêmes : celle des valeurs maximales annuelles, et celle des valeurs supérieures à un seuil. Ce chapitre s'intéresse à la seconde, et explique pourquoi.

Tout d'abord, nous présentons quelques **éléments théoriques** sur la modélisation statistique des valeurs supérieures à un seuil.

Puis, nous indiquons une série d'**outils et de tests statistiques** aidant au choix des paramètres d'échantillonnage.

Enfin, nous présentons les **résultats obtenus** sur vingt bassins, pour lesquels nous nous intéressons entre autres à la sensibilité des résultats en fonction des paramètres d'échantillonnage.

1. Eléments théoriques sur la modélisation des valeurs supérieures à un seuil

L'échantillonnage par valeurs supérieures à un seuil consiste à sélectionner la valeur maximale de chaque événement, un événement commençant et finissant lorsque le débit passe au-dessus puis au-dessous d'un seuil fixé (Figure 1-9, chapitre précédent). L'intérêt de cette méthode, par rapport à un échantillonnage par valeurs maximales annuelles, est une estimation plus précise des quantiles de crue. En effet, Lang (1995) montre à partir de simulations de Monte-carlo que l'erreur quadratique entre les quantiles estimés et théoriques est deux à trois fois plus importante lorsqu'on effectue un échantillonnage par maxima-annuels, pour des échantillons de même taille. En contre partie, l'échantillonnage par valeurs supérieures à un seuil s'avère plus délicat à mettre en œuvre, notamment parce que l'échantillon obtenu dépend du choix des paramètres d'échantillonnage. Il s'agit du seuil choisi, mais aussi d'autres paramètres nécessaires pour garantir l'indépendance des événements successifs, et qui seront présentés par la suite. Une fois l'échantillon constitué, une autre difficulté liée à la modélisation des valeurs supérieures à un seuil, est due au fait que l'étude statistique doit être faite dans deux domaines :

- L'étude du *processus d'occurrence*, c'est à dire des dates d'apparition des événements supérieurs au seuil
- L'étude de la distribution fréquentielle des débits sélectionnés.

Ce n'est que par la combinaison de ces deux lois que l'on peut finalement estimer un quantile de crue en fonction de sa période de retour.

1.1. Etude du processus d'occurrence : le modèle du renouvellement

Les premières applications du modèle du renouvellement dans le domaine de l'hydrologie sont à mettre à l'actif de Borgman (1963), Shane et Lynn (1964) et Bernier (1967). La présentation qui suit est tirée de Lang (1995).

Nous définissons à partir d'un exemple illustré (Figure 2-1) les variables aléatoires servant à définir un processus d'occurrence par la méthode du renouvellement. Le graphique du haut représente une série d'événements supérieurs ou égaux au seuil S . Une façon de caractériser les dates d'apparitions des événements de cette série est de représenter en fonction de t , le nombre d'événements survenus entre 0 et t . Ce nombre d'événement, noté m_t , est représenté par la courbe en escalier sur le graphique du bas de la Figure 2-1. Si maintenant on considère que la série de valeurs supérieures à un seuil étudiée sur cet exemple est la réalisation d'un tirage aléatoire, m_t est alors aussi une variable aléatoire, et la courbe en escalier représente une de ses réalisations possibles.

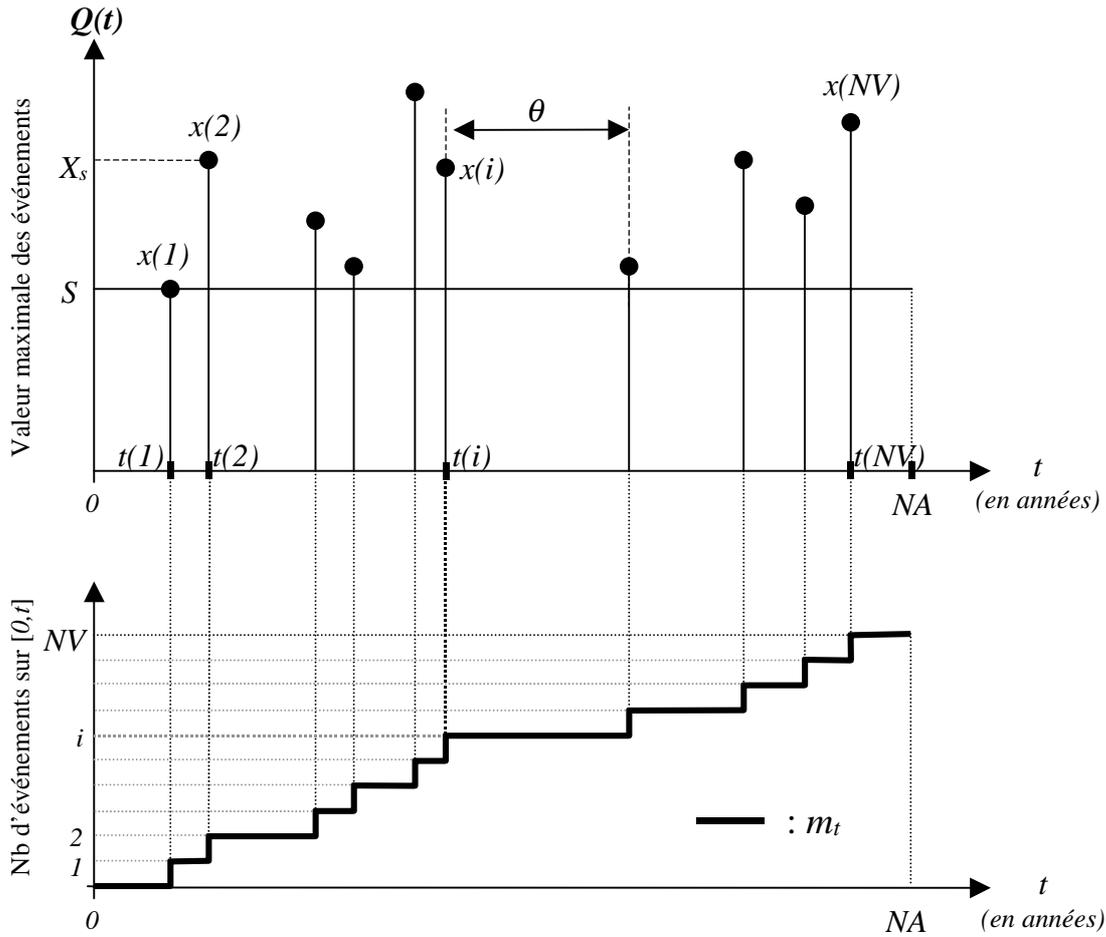


Figure 2-1 : Définition de la variable aléatoire m_t décrivant le processus d'occurrence

Dans ces conditions, la méthode du renouvellement consiste à déterminer la loi régissant la variable aléatoire m_t .

Notons $w_k[0, t]$, la probabilité que m_t soit égal à k , c'est à dire que l'on ait k événements sur l'intervalle $[0, t]$. Notre but est donc de définir la loi $w_k[0, t]$.

$$w_k[0, t] = \text{Prob}(m_t = k) \quad (2-1)$$

On note $N(t)$, le nombre moyen d'événements sur $[0, t]$:

$$N(t) = E(m_t) = \sum_{k=1}^{+\infty} k \cdot w_k[0, t] \quad (2-2)$$

Deux notions vont nous permettre de caractériser le processus d'occurrence. Il s'agit de son **intensité** notée $\mu(t)$:

$$\mu(t) = \frac{dN(t)}{dt} \quad (2-3)$$

ainsi que de l'**indice de dispersion** I_t :

$$I_t = \frac{Var(m_t)}{E(m_t)} = \frac{Var(m_t)}{N(t)} \quad (2-4)$$

Graphiquement, $\mu(t)$ correspond à la pente de la courbe $N(t)$, le nombre moyen d'événements sur $[0,t]$. $Var(m_t)$ indique l'écart quadratique moyen entre la courbe moyenne $N(t)$ et les réalisations de m_t .

Si maintenant, on suppose qu'il y a homogénéité des événements dans le temps, c'est à dire que $w_k[0,t] = w_k[t',t'+t]$, quel que soit t' , alors, le processus est dit **stationnaire**, c'est à dire qu'il ne dépend pas de t . Dans ce cas, l'intensité $\mu(t)$ est constante : $\mu(t) = \mu$. Par conséquent, $N(t)$ est représentée par une droite d'équation $N(t) = \mu t$.

Cas de la loi de Poisson

Une loi largement utilisée pour décrire un processus stationnaire est la **loi de Poisson**. La probabilité que l'on ait k événements sur un intervalle quelconque de largeur t , $w_k(t)$ (cette notation signifie que l'origine de l'intervalle n'a plus d'importance), est donnée par :

$$w_k(t) = \exp(-\mu t) \cdot \frac{(\mu t)^k}{k!} \quad (2-5)$$

La moyenne et la variance de cette loi sont toutes deux égales à $\mu.t$. Par conséquent l'indice de dispersion I_t est constant et égal à 1.

Les hypothèses vérifiées par cette loi sont :

- i. Homogénéité dans le temps des événements : $w_k[0,t] = w_k[t',t'+t]$
- ii. La probabilité d'avoir un événement pendant une courte durée dt est très faible, de même ordre que dt : $w_1[0,dt] = \mu \cdot dt$
- iii. $\sum_{k=2}^{\infty} w_k[0,dt]$, la probabilité d'avoir plus d'un événement pendant dt , est négligeable devant dt
- iv. Indépendance successive des événements

Un seul paramètre doit être ajusté sur l'échantillon : μ . En prenant l'année comme unité de temps, ce paramètre est estimé en divisant NV , le nombre de valeurs sélectionnées, par NA , le nombre total d'années d'observation :

$$\hat{\mu} = \frac{NV}{NA} \quad (2-6)$$

Nous présenterons dans la suite des tests statistiques permettant de savoir si la loi de Poisson ainsi ajustée donne une description acceptable ou pas de l'échantillon étudié.

Signification de l'indice de dispersion

L'indice de dispersion permet de caractériser la façon dont les événements sont regroupés. Par exemple, si l'on considère l'indice de dispersion relatif à l'année ($t=1an$), l'indice de dispersion expérimental est obtenu en comptant chaque année le nombre d'événements qui sont apparus, puis en calculant leur variance et leur moyenne. Comme nous venons de le voir, dans le cas d'un processus d'occurrence régi par une loi de Poisson, l'indice de dispersion est de 1. Si maintenant, pour le même nombre d'événements ceux-ci surviennent par « grappes », cela implique que la variance du nombre d'événement sur 1 an est plus importante que dans le cas précédent : certaines années, il n'y aura aucun événement, et d'autres, ils seront tous regroupés. Dans ce cas, l'indice de dispersion est donc supérieur à 1. Si maintenant, les événements arrivent plus régulièrement que dans un processus de Poisson, cela signifie qu'on a chaque année pratiquement le même nombre d'événements. Dans ce cas la variance est plus faible que pour Poisson, et donc, l'indice de dispersion inférieur à 1.

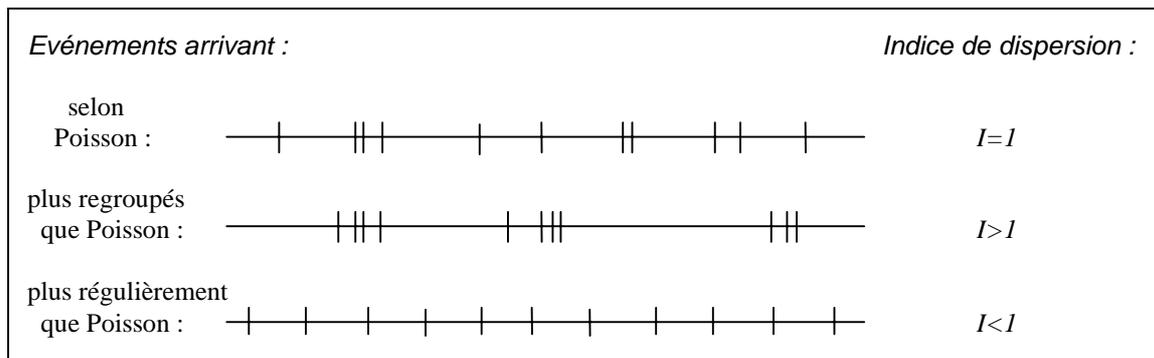


Figure 2-2 : Signification de l'indice de dispersion, d'après le Flood Estimation Handbook (Institute of Hydrology, 1999)

Lorsque l'indice de dispersion est significativement différent de 1, une alternative à la loi de Poisson est d'utiliser une loi binomiale négative ($I>1$) ou une loi binomiale ($I<1$). Dans le premier cas, les quatre hypothèses du processus de Poisson sont vérifiées, sauf l'hypothèse iii. On trouvera dans Lang (1999) une série de développements théoriques sur l'utilisation de la loi binomiale négative. A titre d'exemple, cette loi a été adoptée pour le Royaume-Uni par le Flood Estimation Handbook (Institute of Hydrology, 1999).

Cas d'un processus non stationnaire

Un autre cas particulier est celui d'un processus non stationnaire (hypothèse i. non vérifiée), mais pour lequel on est capable de décrire l'intensité $\mu(t)$ en fonction du temps. Par exemple, si toutes les crues arrivent pendant la même période de l'année, on peut imaginer que $\mu(t)$ a des variations saisonnières mais avec stabilité inter-annuelle. Ce cas a été étudié par Borgman (1963). Connaissant $\mu(t)$, il est alors possible de se ramener à un processus de Poisson stationnaire, par l'intermédiaire d'un changement d'échelle sur le temps. Dans un tout autre domaine, celui des pannes, l'intensité $\mu(t)$ croît avec le temps : au fur et à mesure que le

matériel étudié vieillit, la probabilité d'avoir une panne augmente. C'est d'ailleurs dans ce contexte que le modèle de renouvellement a initialement été développé.

La notion de période de retour

L'intérêt du modèle de renouvellement est de pouvoir calculer la période de retour d'un événement supérieur au seuil S . En effet, la durée entre deux événements successifs supérieurs à S (Figure 2-1) est une variable aléatoire notée θ . On peut définir sa densité de probabilité $h(y)$ et sa fonction de répartition $H(y)$ donnant la probabilité que θ soit inférieur à y :

$$h(y) \cdot dy = Prob(y < \theta < y + dy) \quad (2-7)$$

$$H(y) = \int_0^y h(y) \cdot dy = Prob(\theta < y) \quad (2-8)$$

Dans ces conditions, $T(S)$, la période de retour d'un événement supérieur à S , qu'il faudrait en toute rigueur appeler **période moyenne de retour**, est l'espérance de θ .

$$T(S) = E(\theta) = \int_0^{+\infty} y \cdot h(y) \cdot dy \quad (2-9)$$

La connaissance du processus d'occurrence décrit par $w_k[0, t]$ nous permet de calculer la période de retour définie ci-dessus. En effet, la probabilité pour que θ soit supérieur à y est égale à la probabilité qu'il n'y ait aucun événement sur l'intervalle $[t', t'+y]$, t' étant la date du premier des deux événements espacés de la durée θ (Figure 2-1) :

$$Prob(\theta > y) = w_0[t', t'+y] \quad (2-10)$$

En combinant la relation ci-dessus avec (2-8), on obtient :

$$Prob(\theta < y) = H(y) = 1 - w_0[t', t'+y] \quad (2-11)$$

Dans le cas du processus de Poisson, l'hypothèse d'homogénéité (i) indique que :

$$w_0[t', t'+y] = w_0[0, y] = w_0(y) \quad (2-12)$$

Sachant grâce à la relation (2-5) que $w_0(y) = \exp(-\mu y)$, on démontre finalement que la probabilité pour que θ soit inférieur à y suit une loi exponentielle simple :

$$H(y) = Prob(\theta < y) = 1 - \exp(-\mu \cdot y) \quad (2-13)$$

En dérivant $H(y)$ par rapport à y pour obtenir $h(y)$ et en l'intégrant dans l'équation (2-9), on obtient l'expression de la période de retour d'un événement supérieur à S :

$$T(S) = 1/\mu \quad (2-14)$$

Ce résultat peut paraître évident, et il est souvent supposé comme acquis. Cependant, comme nous venons de le voir, il suppose que le processus soit décrit par une loi de Poisson.

En effet, si le processus est décrit par d'autres lois, Lang *et al.* 1997 montrent que la relation entre T et μ n'est plus la même. Dans tous les cas, il est donc utile de vérifier qu'on est bien dans les conditions d'applications de la loi de Poisson. A cet effet, quelques tests seront présentés dans la suite de ce chapitre.

1.2. Etude de la distribution fréquentielle des débits maximaux

La période de retour définie au paragraphe précédent est donc relative au seuil fixé (Figure 2-1). Ainsi, plus on augmente le seuil, plus on estime la période de retour d'événements rares. Dans le cas limite où le seuil est égal à la plus forte valeur de la chronique, un seul événement est sélectionné et sa période de retour est estimée par NA , le nombre d'années d'observation.

Cette manière de faire pose deux problèmes. L'augmentation du seuil a aussi pour effet de diminuer la taille de l'échantillon servant à caler la loi $w_k(t)$. De ce fait l'estimation de la période de retour devient de plus en plus imprécise. D'autre part, il n'est pas possible d'extrapoler l'estimation de $T(S)$ à des événements supérieurs à la plus forte valeur observée sur la chronique.

Ces deux problèmes sont résolus par l'ajustement d'une deuxième loi statistique, cette fois-ci décrivant la distribution fréquentielle des débits. La variable aléatoire étudiée est alors notée X_s (Figure 2-1) et correspond au débit maximal atteint par l'événement. Elle est décrite par sa fonction de répartition $G(x)$:

$$G(x) = Prob(X_s < x) \quad (2-15)$$

Ainsi, il devient possible d'extrapoler la distribution fréquentielle des débits, et en y associant la méthode du renouvellement (loi $w_k(t)$), d'obtenir une estimation de la période de retour $T(x)$ quel que soit x . Lorsque $w_k[0,t]$ est décrit par la loi de Poisson, on a alors la relation suivante entre la période moyenne de retour $T(x)$ et $G(x)$ (Rosbjerg, 1985)

$$G(x) = 1 - \frac{1}{\mu \cdot T(x)} \quad (2-16)$$

avec μ le nombre moyen d'événements par an. En particulier, on retrouve la relation (2-14) en écrivant que la probabilité que X_s soit inférieur au seuil S est nulle.

Nous avons regroupé dans l'annexe 1 quelques rappels relatifs aux lois de distributions des débits et aux différentes méthodes d'ajustement. Il existe dans la littérature un grand nombre de lois. Cependant, dans le cas des valeurs supérieures à un seuil, les deux lois les plus souvent utilisées sont la loi exponentielle (à 2 paramètres) et la loi de Pareto Généralisée (à 3 paramètres). En effet, comme cela est expliqué en annexe 1, il y a équivalence entre le binôme Poisson/loi Exponentielle ajusté sur un échantillon de valeurs supérieures à un seuil et une loi de Gumbel (2 paramètres) ajustée sur un échantillon de valeur maximales annuelles.

De même il y a équivalence entre le binôme Poisson/Pareto et la loi *GEV* (Generalized Extreme Values, 3 paramètres).

Concernant le choix d'une loi plutôt qu'une autre, il faut savoir que la plupart donnent des résultats comparables pour des périodes de retour inférieures à 20 ans. Par contre, pour des fréquences plus rares, leurs résultats divergent sensiblement. A moins de disposer de très longues séries, il est donc difficile de départager les lois entre elles, surtout si l'on tient compte de l'erreur d'échantillonnage. Il peut en effet arriver que sur un enregistrement de 10 années de mesures, on soit tombé par chance (ou malchance) sur une crue centennale. Dans ce cas, la fréquence expérimentale de cette crue sera largement sous-estimée. Il est donc déconseillé de chercher une loi avec un grand nombre de paramètres pour pouvoir venir s'ajuster sur cette valeur exceptionnelle.

Dans la suite de ce travail, compte tenu de la taille des chroniques qui seront étudiées (de 9 à 33 ans), nous ajusterons simplement une loi exponentielle (2 paramètres) sur l'échantillon des valeurs supérieures à un seuil. De ce fait, les quantiles correspondants ne seront calculés que pour des périodes de retour inférieures à 20 ans, c'est à dire sans utiliser d'extrapolation.

Nous donnons ici l'expression de la fonction de répartition $G(x)$ des valeurs supérieures au seuil S pour la loi exponentielle. On pourra se reporter à l'annexe 1 pour plus de précisions :

$$G(x) = 1 - \exp\left(-\frac{x-S}{a}\right) \quad (2-17)$$

où S est le seuil et a le paramètre d'échelle, souvent appelé en France GRADEX (**GRAD**ient des valeurs **EX**trêmes), suite aux travaux de Guillot et Duband (1967).

Pour un échantillon de NV valeurs classées par ordre croissant : $x(1) < x(2) < \dots < x(NV)$, les paramètres de la loi peuvent être estimés par la méthode des moments :

$$\begin{cases} \hat{S} = x(1) \\ \hat{a} = \frac{1}{NV} \sum_{j=1}^{NV} x(j) - x(1) \end{cases} \quad (2-18)$$

Si maintenant, on associe le modèle de renouvellement (loi de Poisson), on peut exprimer x en fonction de sa période moyenne de retour T , en résolvant l'égalité entre (2-16) et (2-17) :

$$x(T) = x_0 + a \cdot \ln(T) ; \quad \text{avec : } x_0 = S + a \cdot \ln(\mu) \quad (2-19)$$

Dans ces conditions, trois paramètres doivent être estimés à partir de l'échantillon : x_0 (appelé paramètre de position) et a pour la loi exponentielle, et μ relatif à la loi de Poisson :

$$\begin{cases} \hat{\mu} = \frac{NV}{NA} \\ \hat{a} = \frac{1}{NV} \sum_{j=1}^{NV} x(j) - x(1) \\ \hat{x}_0 = x(1) + \hat{a} \cdot \ln(\hat{\mu}) \end{cases} \quad (2-20)$$

avec : NV , le nombre de valeurs échantillonnées et NA , le nombre d'années d'observation.

Comme nous l'avons vu, le modèle ci-dessus implique que l'on fasse l'hypothèse d'un processus d'occurrence décrit par une loi de Poisson. L'objet du paragraphe suivant est donc de proposer quelques tests afin de déterminer si l'échantillon étudié satisfait ou non à cette hypothèse.

2. Outils et tests aidant au choix des paramètres d'échantillonnage

Nous présentons dans ce paragraphe les outils et tests utiles dans l'étude des valeurs supérieures à un seuil. Leur application sera détaillée dans le paragraphe suivant, présentant les résultats obtenus.

2.1. Les critères d'indépendances *COEF1* et *COEF2*

La difficulté de l'échantillonnage par valeurs supérieures à un seuil est que selon le choix du seuil S , il se peut que des valeurs non indépendantes soient sélectionnées. Par exemple, pour l'événement de crue présenté Figure 2-3, le choix du seuil S_1 conduit à sélectionner deux valeurs non-indépendantes, au lieu d'une seule avec le seuil S_2 .

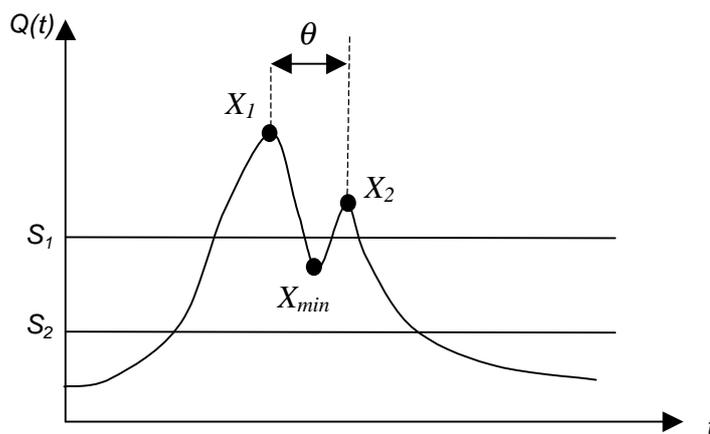


Figure 2-3 : Exemple de valeurs supérieures à un seuil non-indépendantes : X_1 et X_2

Pour palier à ce problème, des critères d'indépendances peuvent être rajoutés. Soit un critère de redescente noté *COEF1* faisant intervenir X_{min} , le débit minimal atteint entre les valeurs X_1

et X_2 , soit un critère sur le temps θ écoulé entre deux valeurs sélectionnées successives noté $COEF2$. Ainsi, la valeur X_2 sera rejetée si :

$$X_{\min} > COEF1 \cdot X_1 \text{ ou } \theta < COEF2 \quad (2-21)$$

Dans la pratique, le choix $COEF1$ et $COEF2$ dépend de la forme des crues observées et peut parfois demander une certaine expérience.

Lang *et al.* (1999) ont fait une synthèse des tests aidant au choix des paramètres d'échantillonnage, à savoir : le seuil S et les critères $COEF1$ et $COEF2$.

2.2. Etude du nombre moyen d'événements sélectionnés par an en fonction du seuil S

La Figure 2-4 montre l'évolution généralement observée de μ en fonction du seuil S .

Quatre domaines sont distingués :

1. Un seuil très faible, inférieur au plus faible débit observé sur la chronique conduit au cas limite où aucune valeur n'est sélectionnée.
2. Lorsque le seuil augmente, le nombre d'événements sélectionnés augmente lui aussi.
3. Après avoir atteint un maximum, le nombre d'événements sélectionnés diminue avec le seuil.
4. Lorsque le seuil appliqué est supérieur au maximum de la chronique étudiée, plus aucun événement n'est sélectionné.

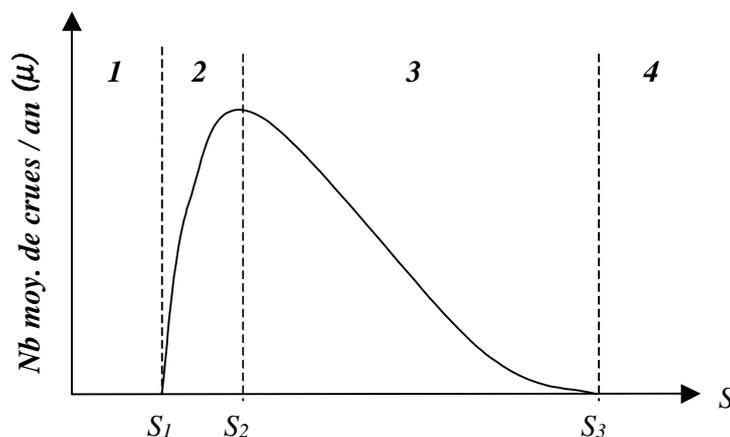


Figure 2-4 : Nombre moyen d'événements sélectionnés par an en fonction du seuil S

Le seuil doit être choisi dans le domaine 3 en cherchant un compromis entre les deux cas extrêmes S_2 et S_3 . Un seuil trop proche de S_2 conduit à la sélection de beaucoup de valeurs mais généralement liées (crues à plusieurs pointes), un seuil trop proche de S_3 conduit à la sélection de valeurs indépendantes, mais d'un nombre trop faible. Dans la pratique, μ dépend du bassin étudié. S'il s'agit d'un bassin rapide comportant de nombreuses crues indépendantes chaque année, on pourra prendre un μ élevé. Dans la littérature, des valeurs de μ comprises entre 2 et 5 sont généralement conseillées. Cunnane (1973) montre qu'il faut prendre un

nombre minimum d'événement ($\mu > 1.65$ dans le cas d'une loi de Poisson) pour que l'échantillonnage par valeurs supérieures à un seuil reste préférable à l'échantillonnage par valeur maximale annuelle.

2.3. Test vérifiant si l'indice de dispersion I_1 est égal à 1

Si le processus d'occurrence est modélisé par une loi de Poisson, alors l'indice de dispersion doit être égal à 1. Pour ce test, on calcule $I_{t=1}$, c'est à dire l'indice de dispersion expérimental calculé sur une base annuelle. Chaque année, on relève le nombre d'événements supérieurs à S , puis on calcule la moyenne et la variance de l'échantillon ainsi obtenu. Cunnane (1979) a défini l'intervalle de confiance autour de 1 dans lequel doit se situer la valeur expérimentale de I_1 . Par exemple, l'intervalle de confiance de 90% signifie que si la valeur de I_1 sort des bornes, on a 10% de chance de se tromper en affirmant que l'hypothèse de la loi de Poisson n'est pas respectée. En effectuant ce test pour différents seuils, on peut ainsi connaître les valeurs de S pour lequel l'hypothèse de la loi de Poisson est acceptable.

2.4. Test vérifiant la stationnarité du processus de Poisson

Lang (1995) propose un test pour étudier la stationnarité du processus d'occurrence modélisé par une loi de Poisson. Si l'on fait référence à la Figure 2-1 présentée en début de chapitre, on teste si à partir des données expérimentales, on peut faire l'hypothèse que $\mu(t)$ est une constante, calculée par la relation (2-6). Le test repose sur le calcul de l'intervalle de confiance autour de la droite théorique $N(t) = \mu t$ dans lequel doit se situer la courbe expérimentale. Par exemple, l'intervalle de confiance à 90% signifie que l'on a 10% de chance de se tromper en affirmant que si un point sort de l'ellipse, alors **autour de ce point**, il y a une rupture de la stationnarité.

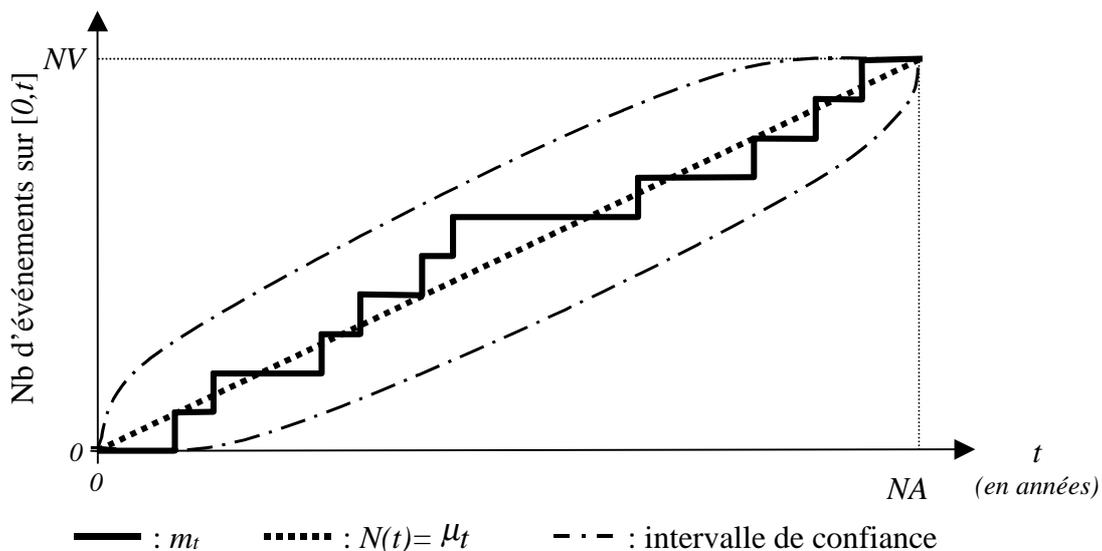


Figure 2-5 : Test de stationnarité du processus d'occurrence

2.5. Outils développés

Afin de faciliter les calculs, nous avons automatisé les opérations qui pouvaient l'être. Ainsi, nous avons programmé une chaîne de traitement reprenant les éléments présentés ci-dessus, en partant d'outils développés par Lang (1995).

Le principal ajout, par rapport à la chaîne de traitement de Lang est qu'il n'est plus nécessaire que l'utilisateur fixe le seuil. Il lui suffit d'indiquer le nombre moyen d'événements par an qu'il souhaite obtenir, $\mu_{objectif}$, et le programme trouve par itérations successives le seuil $S_{objectif}$ correspondant à son souhait. Cette procédure est résumée Figure 2-6. Dix seuils différents sont d'abord définis, compris entre les valeurs minimales et maximales de la chronique. Pour chacun des ces 10 seuils, un échantillonnage est réalisé, respectant les critères d'indépendance $COEF1$ et $COEF2$, et permettant de calculer le nombre moyen de crues sélectionnées par an (Figure 2-12). Deux nouveaux seuils S_1 et S_2 sont alors identifiés conduisant respectivement à un μ juste inférieur et juste supérieur à $\mu_{objectif}$. Afin d'être sûr que nous nous situons bien sur la partie 3 de la courbe $\mu(S)$ (Figure 2-4), S_1 et S_2 sont déterminés en considérant les différents seuils par ordre décroissant. Afin d'affiner la détermination du seuil, une seconde itération est effectuée : 10 nouveaux seuils sont testés, compris entre S_1 et S_2 , et de la même façon, on retient le seuil conduisant à un μ le plus proche de $\mu_{objectif}$. Il peut arriver que le μ maximal atteignable (Figure 2-4) soit inférieur à $\mu_{objectif}$. Dans ce cas, le seuil sélectionné correspond à ce μ maximal atteignable, et l'utilisateur en est informé.

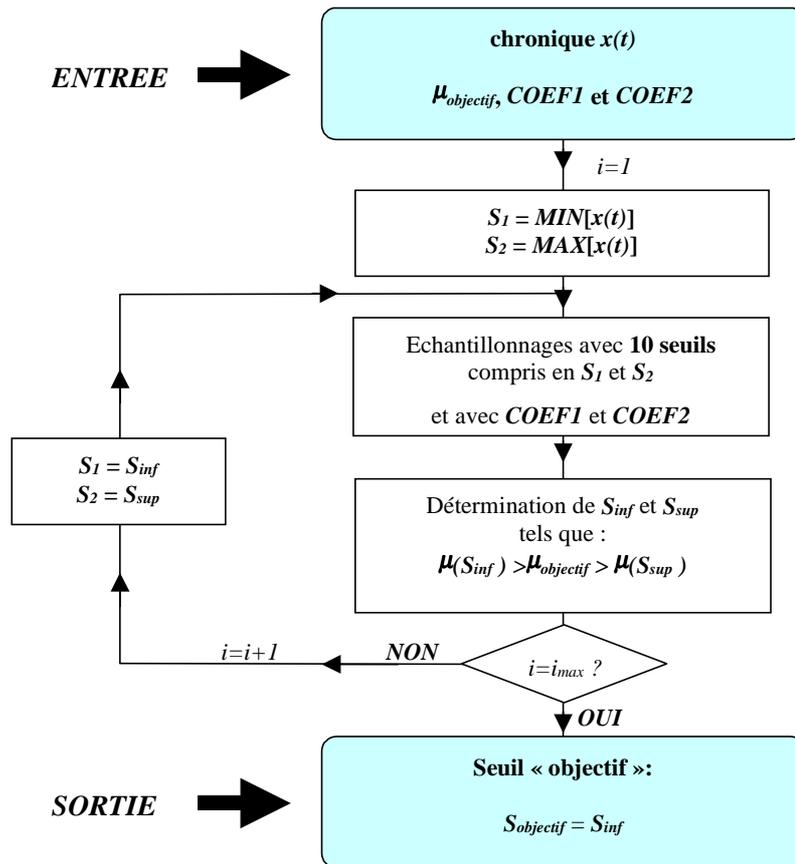


Figure 2-6 : Détermination du seuil $S_{objectif}$ tel qu'en échantillonnant avec $S_{objectif}$, $COEF1$ et $COEF2$ on obtienne un nombre moyen de valeur $\mu_{objectif}$ conforme à celui demandé par l'utilisateur

Il reste maintenant à déterminer les paramètres $\mu_{objectif}$, $COEF1$ et $COEF2$ tels que l'échantillonnage soit jugé satisfaisant. Cette procédure est résumée Figure 2-7 et n'a, dans un premier temps, pas été entièrement automatisée (une procédure entièrement automatique sera présentée au chapitre suivant, §3.1). En effet, il faut d'une part chercher à respecter les tests statistiques sur l'indice de dispersion et sur la stationnarité du processus de Poisson, et d'autre part vérifier visuellement sur la chronique que les valeurs sélectionnées ne sont pas aberrantes. En particulier, les valeurs sélectionnées ne doivent pas être liées, c'est à dire relatives au même événement de crue. D'autre part, il faut s'assurer que des événements pourtant indépendants n'ont pas été ignorés, ceci à cause de coefficients $COEF1$ et $COEF2$ trop sévères. Dans la pratique, on choisit de fixer le paramètre $\mu_{objectif}$ puis de tester différentes valeurs de $COEF1$ et $COEF2$ jusqu'à ce que l'échantillonnage paraisse satisfaisant, du point de vue des deux critères énoncés ci-dessus : tests statistiques satisfaisants **et** visualisation de la chronique avec les valeurs sélectionnées.

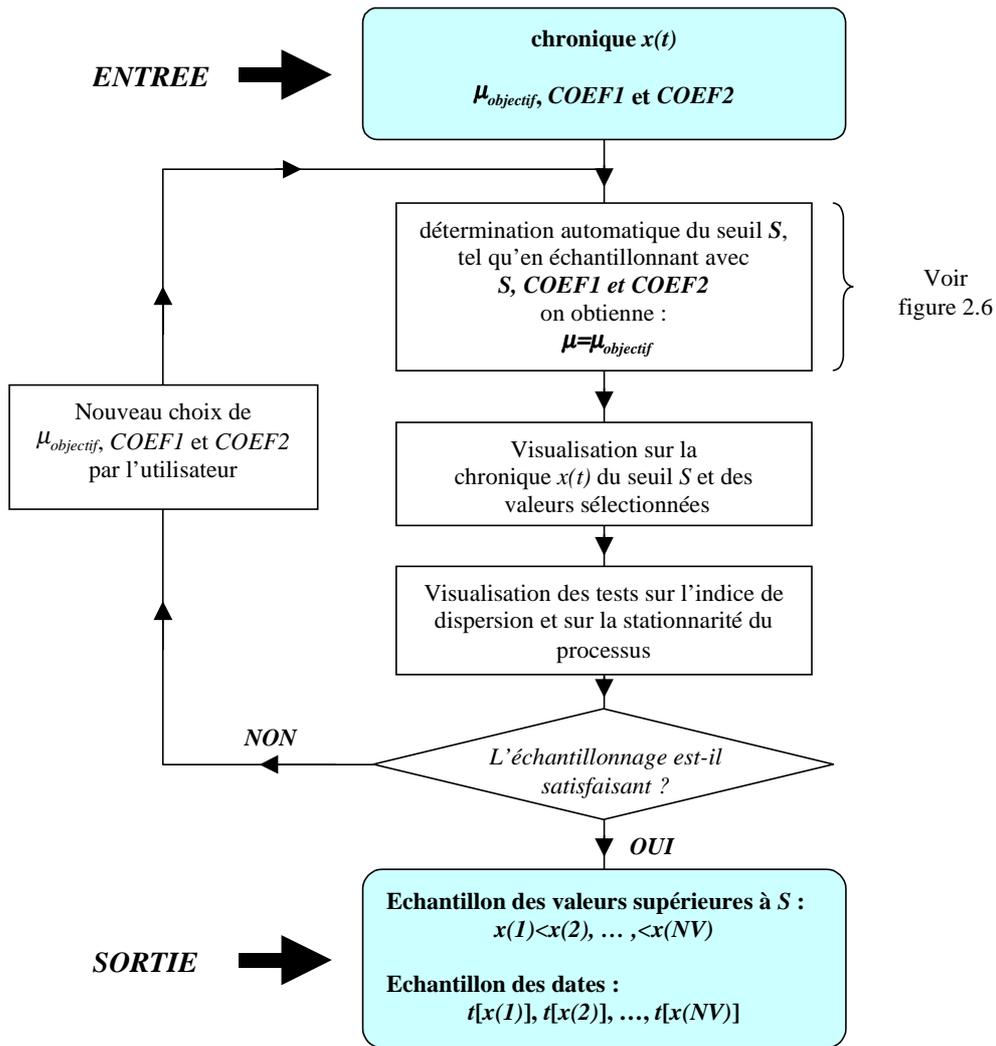


Figure 2-7 : Procédure “semi-automatique” pour l’échantillonnage d’une chronique $x(t)$ à partir des paramètres $\mu_{objectif}$, $COEF1$ et $COEF2$

3. Résultats obtenus

3.1. Bassins disponibles et critique des données

Nous présentons ici l’ensemble des bassins qui seront étudiés dans les chapitres ultérieurs. La base de données disponibles compte 107 stations situées en France (Figure 2-8). Les chroniques de débit étudiées sont des chroniques de débit instantané au pas de temps variable, achetées à la banque HYDRO par le Cemagref au cours de différentes études. Les surfaces de ces bassins varient entre 7 km² et 9400 km² (Figure 2-9).

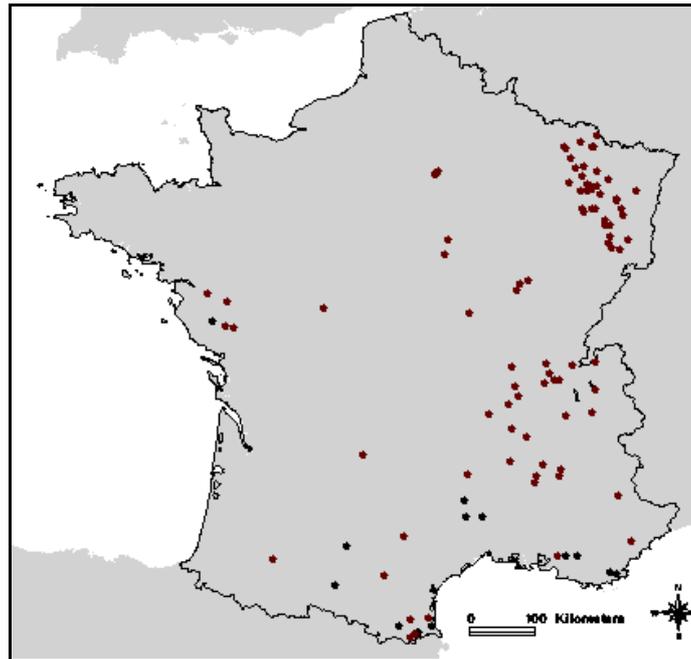


Figure 2-8 : Localisation des bassins étudiés

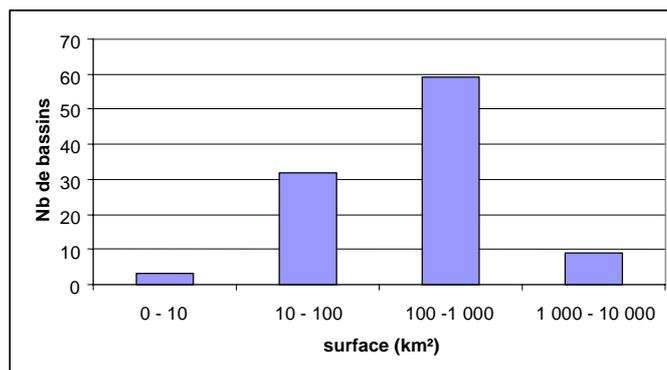


Figure 2-9 : Répartition du nombre de bassins étudiés en fonction de leur surface

L'analyse porte donc majoritairement sur des bassins dont les tailles sont comprises entre 100 et 1000 km². Le nombre d'années d'observation est quant à lui compris entre 9 et 33, avec en moyenne 19 années disponibles (Figure 2-10).

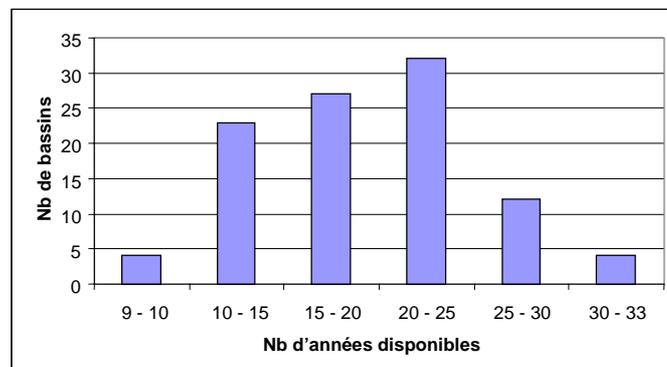


Figure 2-10 : Répartition du nombre de bassins étudiés en fonction du nombre d'années de mesures de débit disponibles

Toutes ces chroniques de débit ont été critiquées visuellement, afin de repérer les erreurs aberrantes. Trois types d'erreurs ont été trouvés en quantité non négligeable :

- valeur de débit aberrante, souvent due à un problème d'unité (par exemple, une valeur 1000 fois supérieure aux autres)
- erreur sur le codage de la date du débit mesuré comme sur l'exemple Figure 2-11 en 1982
- Oubli du codage d'une lacune, se traduisant par l'interpolation des données entre le dernier point mesuré et la reprise d'information (Figure 2-11 en 1985).

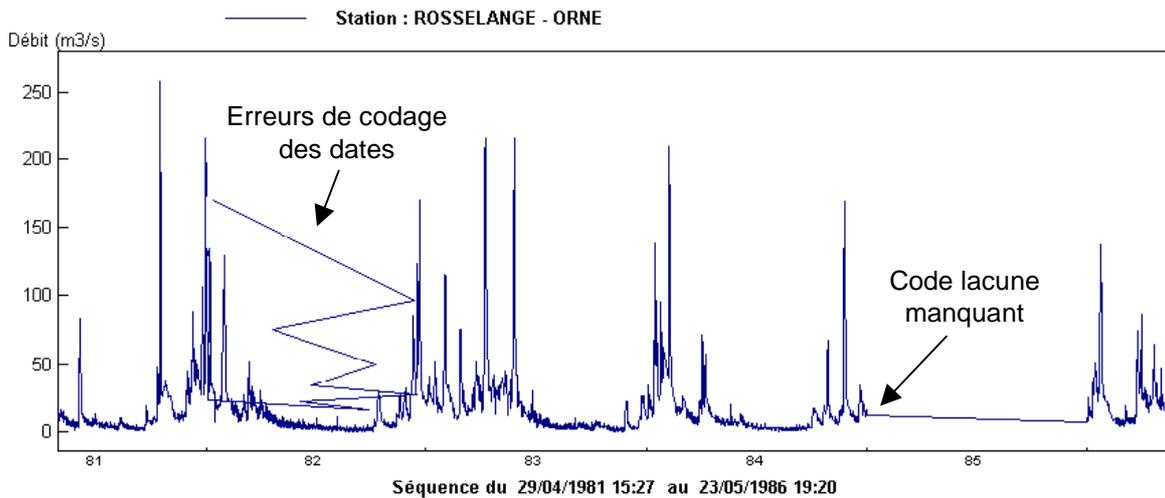


Figure 2-11 : Critique visuelle des données et correction des valeurs aberrantes

Ces erreurs ont été corrigées manuellement, soit en reconstituant la vraie chronique lorsqu'il s'agissait d'erreurs simples, soit en mettant en lacune les valeurs douteuses. Ces traitements, fastidieux car non automatisables, sont pourtant indispensables. A titre indicatif, environ un tiers des chroniques étudiées a dû être corrigé. Cette critique ne concerne que les erreurs facilement détectables. Mais d'autres types d'erreurs peuvent subsister, comme par exemple celles liées aux incertitudes de la courbe de tarage évoquées dans le chapitre précédent.

Comme nous l'avons vu aux paragraphes précédents, la procédure d'échantillonnage présentée requiert un contrôle expert à chaque étape. Pour cette raison, l'étude détaillée de l'ensemble des bassins disponible s'avèrerait beaucoup trop longue. Aussi, n'avons nous étudié que vingt bassins, choisis au hasard et présentés ultérieurement au Tableau 2-1. Les résultats obtenus serviront au chapitre suivant à définir une procédure cette fois-ci entièrement automatique, qui sera quant à elle appliquée à l'ensemble des bassins disponibles.

3.2. Influence du nombre d'événements sélectionnés $\mu_{objectif}$

Nous avons tout d'abord choisi de ne pas utiliser de critères d'indépendance $COEF1$ et $COEF2$, et de voir dans ces conditions l'influence du nombre d'événements sélectionnés. Pour ce faire, la procédure présentée Figure 2-7 a été appliquée en laissant $COEF1$ à 1 et $COEF2$ à 0, puis en fixant successivement $\mu_{objectif}$ à 1, 2 et 3.

Comme cela a été expliqué au paragraphe précédent, cette procédure détermine par itérations le seuil S permettant d'obtenir le nombre d'événements défini par $\mu_{objectif}$. A titre d'exemple, la Figure 2-12 montre l'évolution du nombre moyen de crues sélectionnées par an en fonction du seuil appliqué, pour la Nied Allemande à Faulquemont, déjà présenté au chapitre 1. L'application de la procédure avec un $\mu_{objectif}$ de 1, 2 et 3 indique qu'il faut choisir respectivement un seuil de 17.7, 14.5 et 11.7 m³/s.

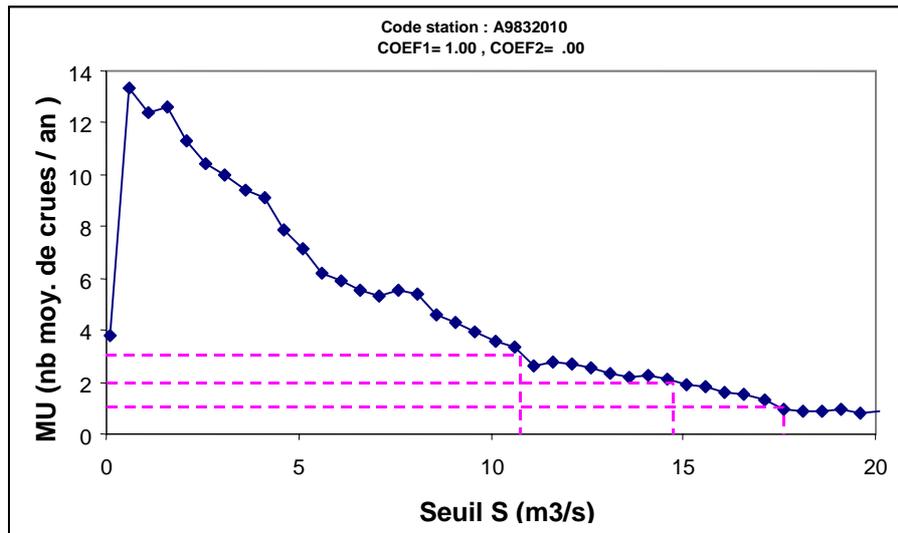


Figure 2-12 : Nombre moyen d'événements sélectionnés par an μ en fonction du seuil S ,
Exemple de la Nied Allemande à Faulquemont

Afin de tester la validité de la détermination automatique du seuil, la Figure 2-13 indique pour chacun des 20 bassins étudiés, le nombre moyen d'événements effectivement sélectionnés par an fonction de l'objectif demandé $\mu_{objectif}$. Comme nous l'avons vu Figure 2-6, cette procédure s'arrête au bout d'un nombre d'itération i_{max} fixé par l'utilisateur. Après différents essais, nous avons choisi de prendre $i_{max}=3$. La Figure 2-13 montre que les résultats obtenus sont acceptables. Il aurait été possible de les améliorer en augmentant le nombre d'itérations i_{max} , mais cela aurait alourdi les calculs pour un gain somme toute minime.

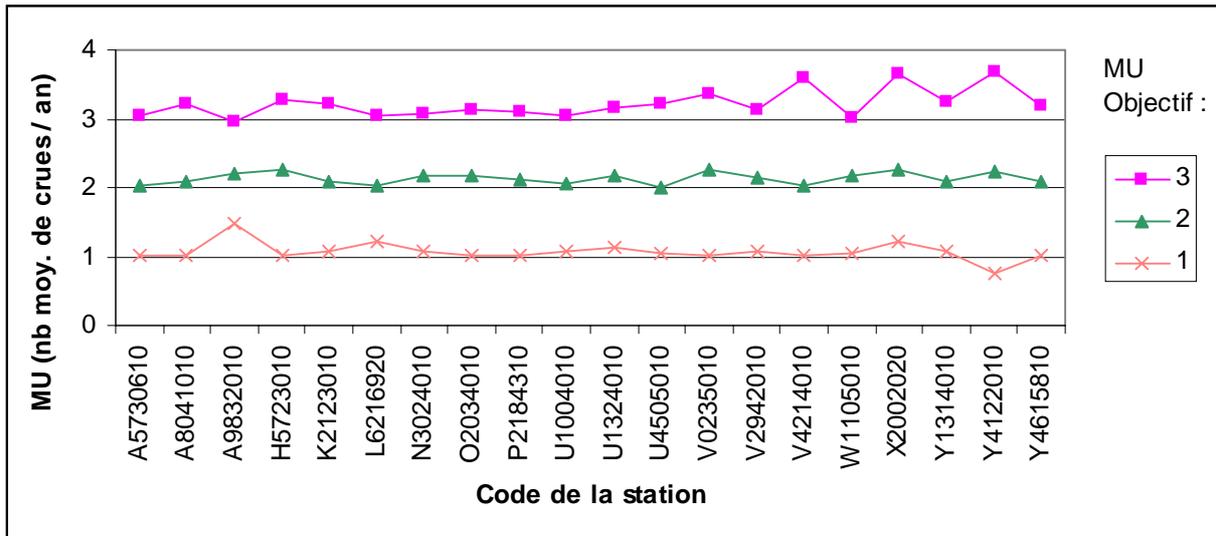


Figure 2-13 : Nombre moyen d'événements sélectionnés par an, fonction de l'objectif demandé $\mu_{objectif}$, pour chacun des 20 bassins étudiés

Pour chacun des bassins étudiés, nous avons effectué les deux tests statistiques présentés au paragraphe précédent. Les figures suivantes montrent les résultats obtenus pour l'exemple de l'Ouche à Plombière-les-Dijon dans le département de la Cote d'or (code Hydro : U1324010). Sur la Figure 2-14, le test sur l'indice de dispersion a été réalisé pour différents seuils, alors que pour la Figure 2-15, seul le seuil répondant à un $\mu_{objectif}$ de 2 a été étudié.

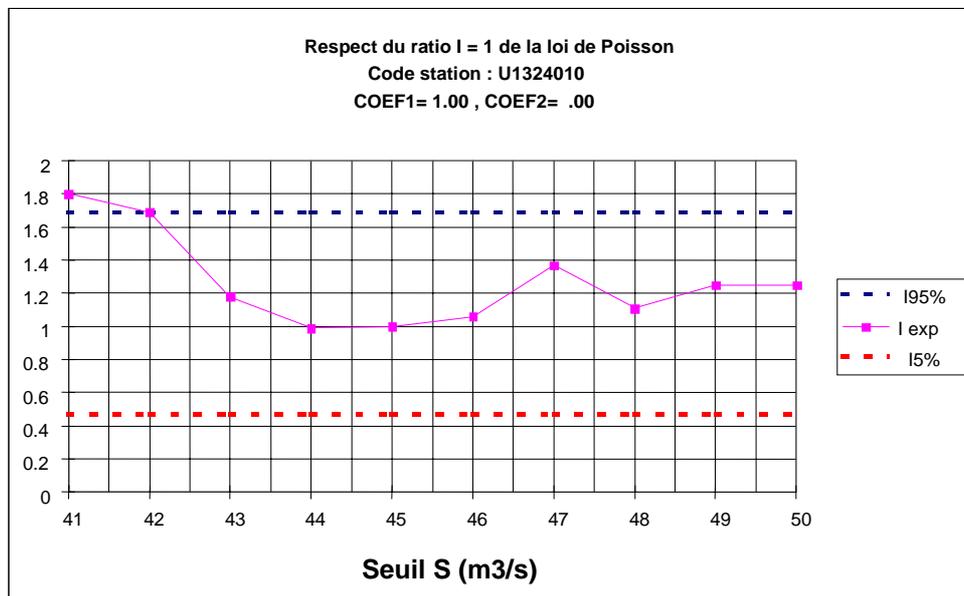


Figure 2-14 : Test sur l'indice de dispersion en fonction de la valeur du seuil. Exemple de l'Ouche à Plombière-les-Dijon

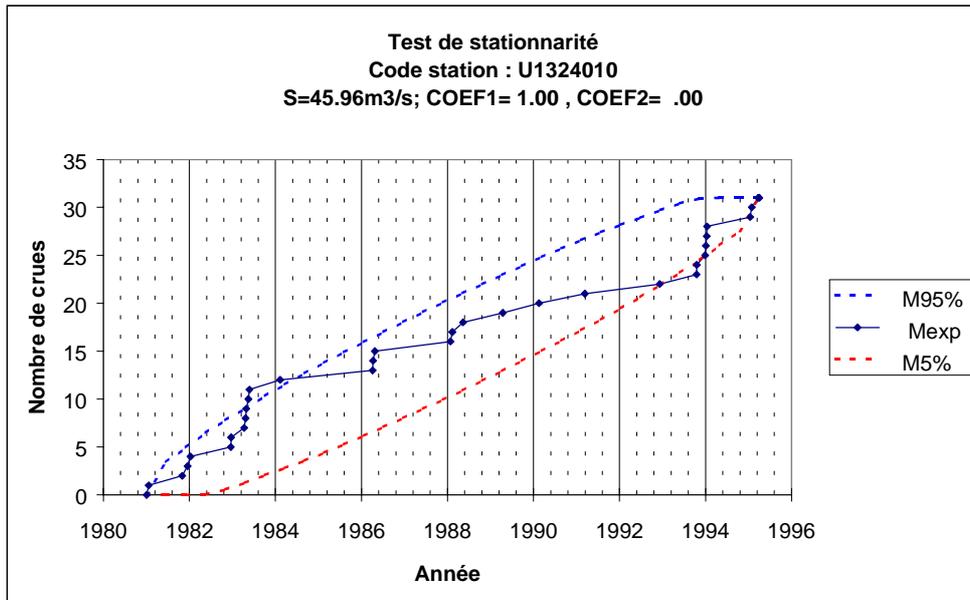


Figure 2-15 : Test de la stationnarité du processus de Poisson, Exemple de l’Ouche à Plombière-les-Dijon pour $\mu=2$

Pour le seuil retenu ($S = 46 \text{ m}^3/\text{s}$), le test sur l’indice de dispersion est acceptable. Par contre, celui sur la stationnarité indique un problème en 1983 et 1994. Nous reviendrons plus en détail sur cet exemple.

Les résultats suivants présentent une synthèse des tests réalisés sur les différents bassins, pour les 3 cas étudiés : $\mu_{objectif} = 1 ; 2 \text{ et } 3$. Ainsi, la Figure 2-16 indique le pourcentage de bassins pour lesquels le test sur l’indice de dispersion a été mis en défaut, en fonction du nombre moyen d’événements sélectionné par an, soit respectivement 40%, 60% et 75%.

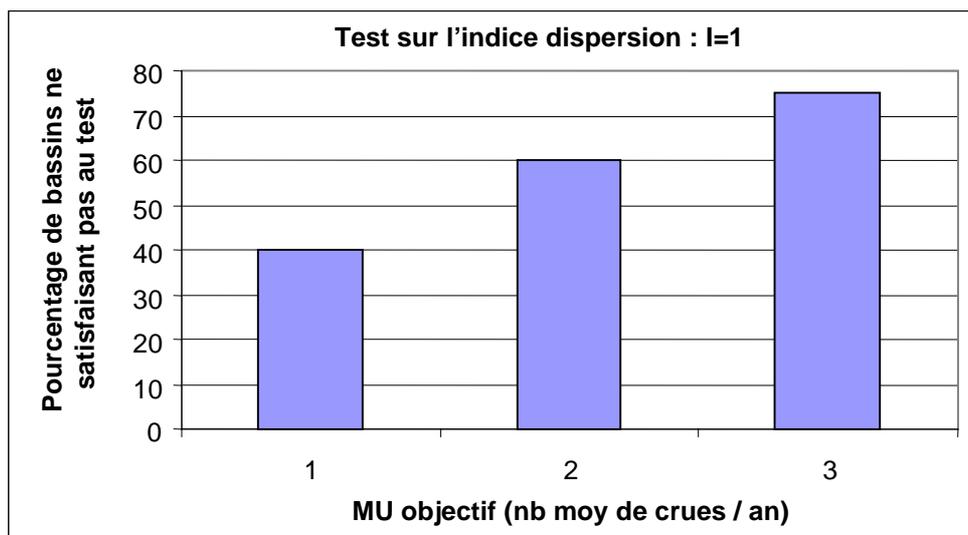


Figure 2-16 : Pourcentage de bassins ne respectant pas le test sur l’indice de dispersion $I = 1$ avec un intervalle de confiance à 90%, pour $\mu_{objectif}=1 ; 2 \text{ et } 3$ et $COEF1=1 ; COEF2=0$

Ce test est effectué avec un intervalle de confiance à 90%, ce qui signifie que l'on a 10% de chance de rejeter à tort l'hypothèse testée. Par conséquent, si les valeurs étaient extraites d'une loi de Poisson par une modélisation de Monte-Carlo, on devrait avoir un taux d'échec de 10%. Le taux d'échec constaté Figure 2-16 est bien plus élevé.

Concernant le test de stationnarité, l'axe des temps a été divisé en dix, et pour chaque dixième, nous avons compté le nombre de bassins pour lesquels la courbe expérimentale est en dehors de l'intervalle de confiance à 90% (Figure 2-15). C'est en effet de cette façon que l'on se trouve dans les conditions du test : **l'intervalle de confiance est relatif à chaque point, et non à la courbe totale**. En effet, ce problème est le même que celui rencontré pour un test similaire au notre, souvent appelé « test de l'ellipse de Bois » (Bois, 1976). Or, ce dernier montre en simulant 5000 séries de 100 valeurs (donc de données respectant statistiquement l'hypothèse testée) que pour un test effectué avec un risque d'erreur de 1%, il y a au moins un point en dehors de l'ellipse pour 8% des séries simulées (Bois, 1986).

La Figure 2-17 montre les résultats obtenus lorsqu'on sélectionne en moyenne 1, 2 et 3 crues par an. Ici aussi l'écart par rapport au risque d'erreur de 10% (intervalle de confiance à 90%) et augmente avec le nombre moyen d'événements pas an.

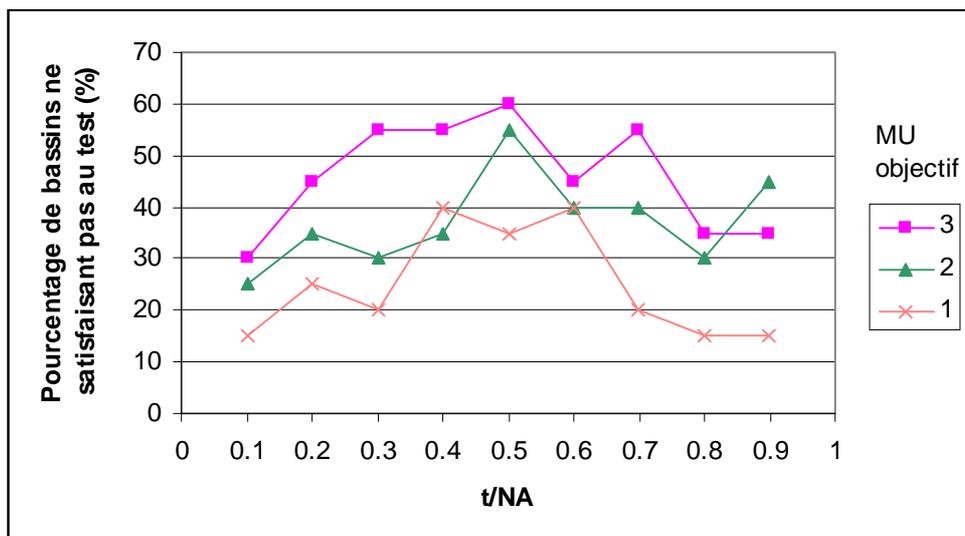


Figure 2-17 : Pourcentage de bassins ne respectant pas le test de stationnarité à l'instant t avec un intervalle de confiance à 90%, pour $\mu_{objectif}=1 ; 2$ et 3 et $COEF1=1 ; COEF2=0$

Les deux graphiques ci-dessus montrent donc clairement que plus on sélectionne un nombre élevé de crues, moins les tests sont respectés (dispersion et stationnarité). Ceci s'explique par le fait que nous n'avons pas utilisé de critères d'indépendance ($COEF1=1 ; COEF2=0$). Ainsi, plus on baisse le seuil afin de sélectionner un nombre important de crues (cf. Figure 2-4, partie 3 de la courbe), plus on a de chance que les valeurs obtenues soient liées.

Pour illustrer ce propos, revenons sur l'exemple de l'Ouche à Plombière-les-Dijon présenté Figure 2-15, pour lequel nous avons remarqué que le test de stationnarité n'était pas respecté. La courbe expérimentale sortait de l'ellipse en 1983 et fin 1994. Si l'on regarde pour cette dernière période les événements supérieurs au seuil étudié ($S=46 \text{ m}^3/\text{s}$), on voit clairement Figure 2-18 que les valeurs retenues ne sont pas indépendantes : les 2 valeurs sélectionnées en octobre font en fait partie de la même crue. De la même façon, les valeurs sélectionnées en janvier sont liées.

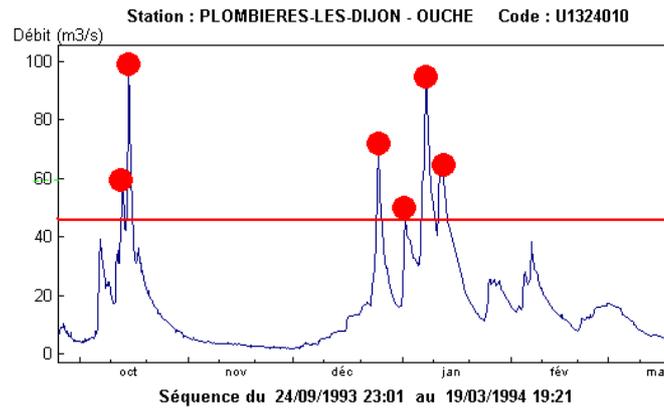


Figure 2-18 : Exemple de valeurs sélectionnées non-indépendantes

3.3. Echantillonnage utilisant les critères d'indépendance *COEF1* et *COEF2*

Les résultats présentés au paragraphe précédent montrent qu'il est préférable d'imposer des critères garantissant l'indépendance des événements sélectionnés (*COEF1* et *COEF2* définis au §2.1.).

Dans la suite de l'étude, nous avons donc fixé $\mu_{objectif}$ à 2, puis déterminé manuellement pour chaque bassin les valeurs de *COEF1* et *COEF2* optimales, selon la procédure définie Figure 2-7. Les valeurs de ces deux coefficients ont donc été déterminées de façon à respecter les deux tests statistiques (dispersion et stationnarité) mais aussi en veillant à ce que des événements jugés « visuellement » indépendants ne soient pas ignorés (cf. §2.5.)

La valeur de 2 pour $\mu_{objectif}$ est motivée par le fait que ce niveau est assez élevé pour que la méthode sup-seuil soit préférable à la méthode des max-annuels, tout en restant suffisamment bas pour ne pas avoir trop de valeurs liées.

Le Tableau 2-1 présente les coefficients *COEF1* et *COEF2* qui ont finalement été retenus pour chacun des 20 bassins étudiés, après l'examen visuel des chroniques ainsi que la vérification des tests statistiques.

Si l'on revient à l'exemple de l'Ouche à Plombière-les-Dijon (code U1324010), la Figure 2-19 montre que le test de stationnarité est maintenant respecté.

Code	Rivière	Station	Dép.	Surface (km ²)	NA (an)	COEF1	COEF2 (j)
A5730610	LA MOSELLE	TOUL	54	3350	17.8	0.2	10
A8041010	L' ORNE	BONCOURT	54	412	16.7	0.05	30
A9832010	LA NIED ALLEMANDE	FAULQUEMONT	57	187	13.6	0.2	10
H5723010	L' ORGEVAL	BOISSY-LE-CHATEL (LE THEIL)	77	104	30.5	0.2	10
K2123010	LE CHAPEAUROUX	CHATEAUNEUF-DE-RANDON	48	109	18.7	0.4	15
L6216920	LE BRIGNON	PAULMY	37	75	20.7	0.2	15
N3024010	LE LOUING	CHANTONNAY	85	120	20.6	0.4	30
O2034010	L' AUSSONNELLE	SEILH	31	192	22.6	0.2	5
P2184310	LA TOURMENTE	ST-DENIS-LES-MARTEL	46	202	19.0	0.4	30
U1004010	L' OGNON	SERVANCE (FOURGUENONS)	70	73.5	21.3	0.5	10
U1324010	L' OUCHE	PLOMBIERES-LES-DIJON	21	655	14.3	0.2	15
U4505010	L' ARDIERES	BEAUJEU	69	54.5	22.9	0.2	15
V0235010	LA MENOGE	VETRAZ-MONTHOUX	74	160	9.8	0.2	20
V2942010	L' AIN	CHAZEY-SUR-AIN	01	3630	15.0	0.2	30
V4214010	LA DROME	LUC-EN-DIOIS	26	194	14.8	0.2	5
W1105010	LE GELON	LA CROIX-DE-LA-ROCHETTE	73	65.4	13.3	0.2	10
X2002020	LE VERDON	ALLOS (LA FOUX)	04	10.1	10.6	0.2	30
Y1314010	LE FRESQUEL	VILLEPINTE	11	216	19.7	0.1	30
Y4122010	L' ARC	AIX-EN-PROVENCE (ST-PONS)	13	605	18.7	0.1	5
Y4615810	LE VALESCURE	COLLOBRIERES (AVAL)	83	9.22	23.9	0.2	5

Tableau 2-1 : Liste des bassins étudiés et valeurs des coefficients d'indépendance *COEF1* et *COEF2* déterminées manuellement

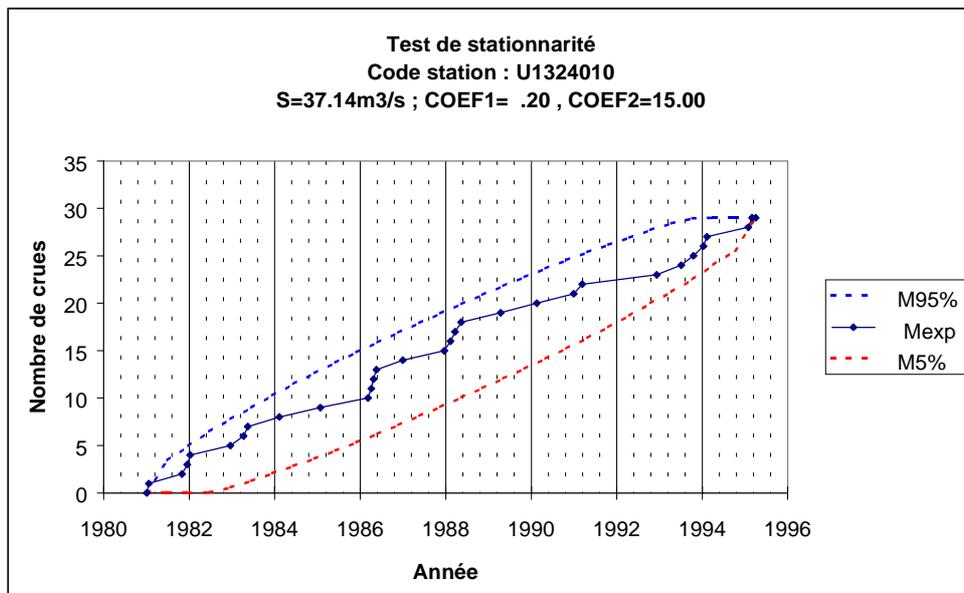


Figure 2-19 : Test de la stationnarité du processus de Poisson, Exemple de l'Ouche à Plombière-les-Dijon avec $\mu=2$; COEF1=0,2 et COEF2=15j

La Figure 2-20 et la Figure 2-21 comparent les nouveaux résultats avec ceux que nous avons déjà présentés Figure 2-16 et Figure 2-17, dans le cas où $\mu_{objectif}$ était égal à 2.

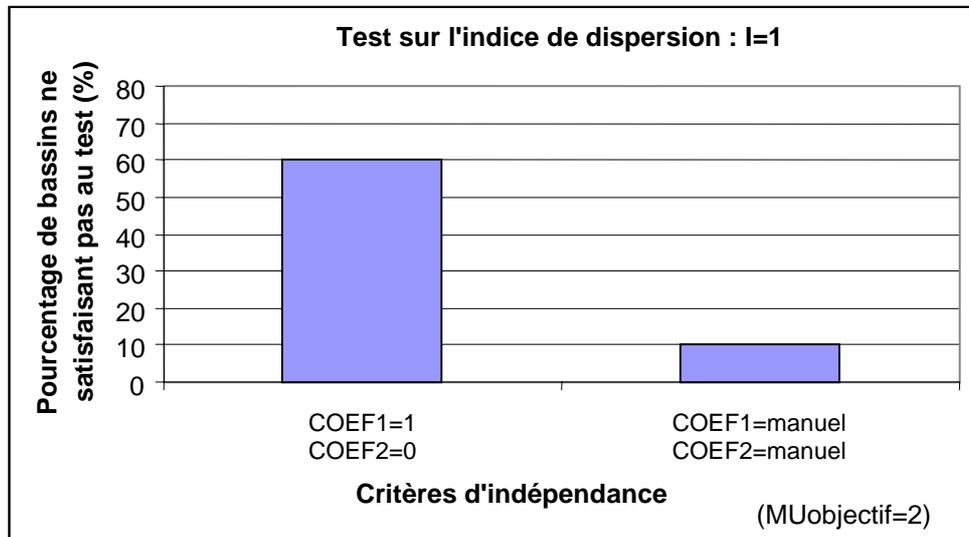


Figure 2-20 : Pourcentage de bassins ne respectant pas le test de l'indice de dispersion égal à 1, pour $\mu_{objectif}=2$; sans et avec critères d'indépendance $COEF1$ et $COEF2$

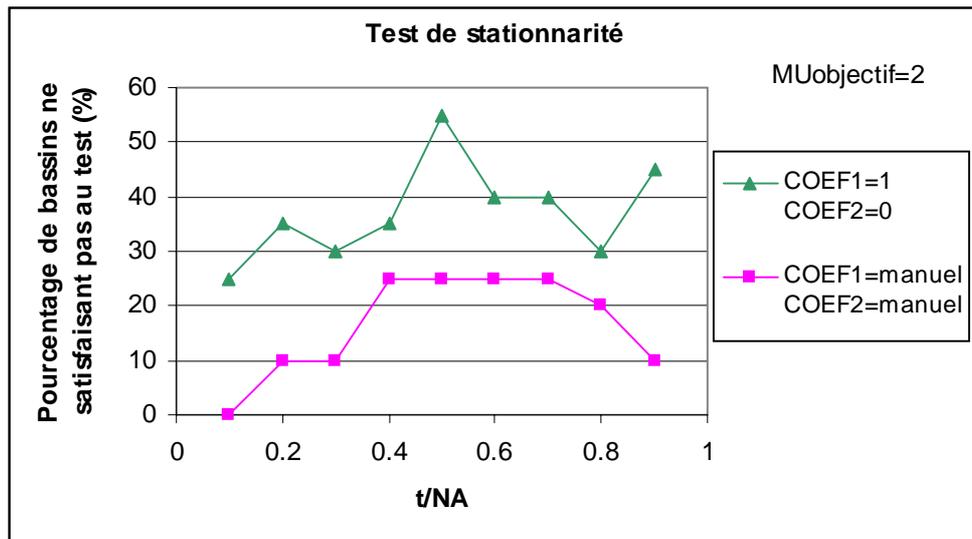


Figure 2-21 : Pourcentage de bassins ne respectant pas le test de stationnarité à l'instant t , pour $\mu_{objectif}=2$; sans et avec critères d'indépendance $COEF1$ et $COEF2$

On note que l'utilisation des critères d'indépendance améliore significativement les résultats des tests. Pour le test sur l'indice de dispersion, on passe de 60% à 10% de bassins rejetés. Pour la stationnarité, à l'instant t pour lequel le taux d'échec est maximal (ici, $t=NA/2$), on passe de 55% de bassins posant problème à 25%. Les résultats sont maintenant bien plus proches du taux d'erreur de 10% lié au fait que les tests ont été effectués avec un intervalle de confiance à 90%. Rappelons que si les valeurs étaient extraites d'une loi de Poisson par un procédé de simulation, alors 10% des valeurs devraient sortir de l'intervalle à 90%. Ceci indique que l'on peut considérer que les hypothèses de la loi de Poisson sont globalement

respectées sur l'ensemble des vingt bassins traités, sauf peut être pour un ou deux cas particuliers dont nous reparlerons.

3.4. Comparaison des quantiles obtenus avec et sans critères d'espacement

Comme nous venons de le constater, l'utilisation des critères d'indépendance permet de mieux respecter les hypothèses liées à la loi de Poisson. Ce paragraphe propose de tester si le fait d'utiliser ou non ces critères a une grande incidence sur le résultat final de l'analyse fréquentielle, à savoir le calcul des quantiles $x(T)$.

Pour cela, le modèle de dépassement a été ajusté avec la loi exponentielle (cf. §1.2.) sur les deux échantillons obtenus au paragraphe précédent, d'abord sans utiliser les critères ($COEF1=1$ et $COEF2=0$) puis en les déterminant manuellement. Dans le premier cas, les quantiles obtenus sont noté $x(T)_{sans}$, et dans le deuxième $x(T)_{avec}$. Afin de quantifier l'écart entre ces deux méthodes, nous avons ensuite calculé pour chaque bassin i l'erreur relative ε_i , en considérant comme valeur de référence le quantile obtenu avec les coefficients d'indépendance déterminés manuellement. Afin d'avoir une synthèse des résultats pour l'ensemble des bassins étudiés, nous avons calculé le $BIAIS$, c'est à dire la moyenne de ε , ainsi que l'erreur quadratique moyenne, notée $rRMSE$ (pour relative Root Mean Square Error), définie à partir de la moyenne ($BIAIS$) et de l'écart type de ε . L'erreur quadratique est donc un critère global rendant compte à la fois du biais et de la dispersion des erreurs autour de leur moyenne.

$$\left\{ \begin{array}{l} \varepsilon_i = \frac{x_i(T)_{sans} - x_i(T)_{avec}}{x_i(T)_{avec}} \\ BIAIS = \frac{1}{20} \sum_{i=1}^{20} \varepsilon_i \\ rRMSE = \sqrt{BIAIS^2 + \sigma(\varepsilon)^2} \end{array} \right. \quad (2-22)$$

On démontre facilement que la formulation suivante est équivalente à la relation (2-22) :

$$rRMSE = \sqrt{\frac{1}{20} \sum_{i=1}^{20} \left(\frac{x_i(T)_{sans} - x_i(T)_{avec}}{x_i(T)_{avec}} \right)^2} \quad (2-23)$$

Les valeurs du $BIAIS$ et de l'erreur quadratique moyenne ont été calculées pour différentes périodes de retour T . La Figure 2-22 illustre les résultats obtenus pour les périodes de retour observables sur l'échantillon, c'est à dire comprises entre 0.5 et 20 ans.

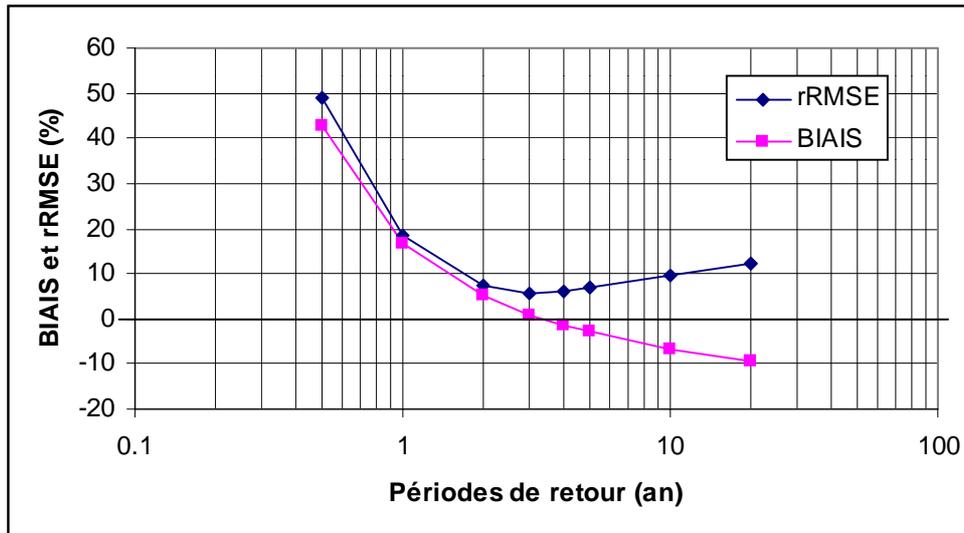


Figure 2-22 : Erreurs sur les quantiles obtenus sans utiliser de critères d'indépendance, les quantiles de référence étant ceux calculés en utilisant ces critères.

Aussi bien pour le biais que pour l'erreur quadratique, les écarts sont les plus importants pour la période de retour de 0,5 an : le biais atteint une valeur positive de 43% et l'erreur quadratique 48%. Lorsque T augmente les écarts se réduisent, jusqu'à devenir négligeables autour de 3 ans, puis ils augmentent à nouveau avec T . Le signe du biais indique que les quantiles estimés sans utiliser de critères d'indépendance sont en moyenne supérieurs à ceux estimés en utilisant les critères pour T inférieure à 3 ans. Lorsque T est supérieur à 3 ans ; la tendance s'inverse. Ce résultat peut s'expliquer de la façon suivante. En général, il y a peu de problèmes liés à l'indépendance des plus fortes valeurs échantillonnées. On retrouve donc ces valeurs que l'on utilise ou non les critères d'indépendance. Par contre, pour les faibles valeurs, le fait d'utiliser les critères d'indépendance a pour effet d'éliminer les valeurs liées. Ces valeurs sont remplacées par d'autres, cette fois-ci indépendantes mais plus faibles.

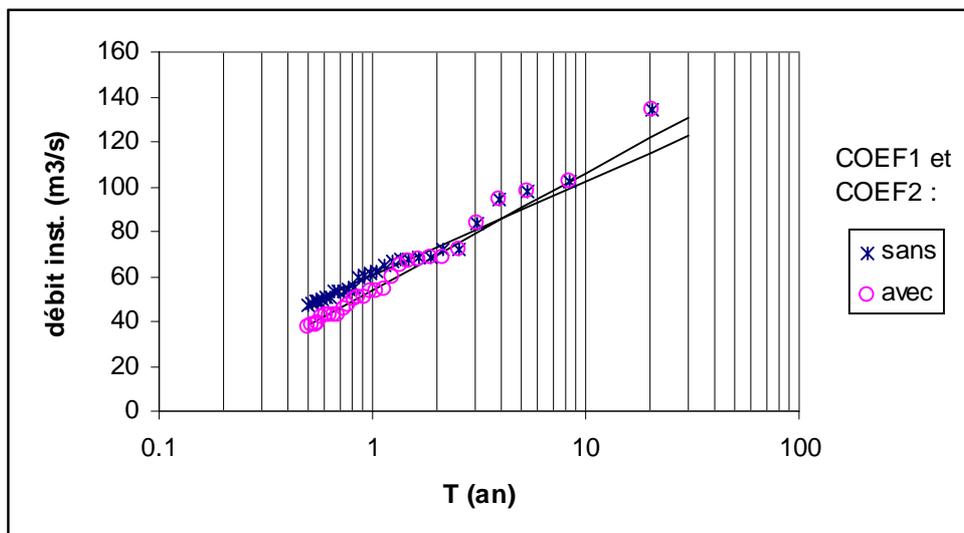


Figure 2-23 : Comparaison des distributions de débit obtenues avec et sans critères d'indépendance. Exemple de l'Ouche à Plombière-les-Dijon

Ce point est illustré pour l'exemple de l'Ouche à Plombière-les-Dijon par la Figure 2-23. L'utilisation des critères a donc pour effet de faire « pivoter » les distributions des débits vers le bas pour les faibles périodes de retour, tout en maintenant les plus fortes valeurs de l'échantillon à peu près intactes.

Ces résultats montrent qu'il est essentiel de veiller à ce que l'indépendance des valeurs échantillonnées soit respectée. L'erreur sur les quantiles peut être très importante pour les faibles périodes de retour ($T=0.5$ an), alors que pour des périodes de retour moyennes (2 à 20 ans), elle l'est moins. Cependant, elle peut redevenir forte pour des périodes de retour plus grandes.

3.5. Discussion sur le respect des hypothèses de la loi de Poisson

Comme nous l'avons vu, il peut arriver dans certains cas qu'il ne soit pas possible de respecter les tests statistiques (relatifs à la loi de Poisson), bien que l'étude visuelle de la chronique indique que les valeurs échantillonnées sont indépendantes. Le non respect d'une ou plusieurs des quatre hypothèses liées à la loi de Poisson (§1.1.) peut avoir différentes origines :

- Il se peut que la loi de Poisson ne convienne pas. Pour cela, le test sur l'indice de dispersion permet de confirmer ou non ce diagnostic. Par exemple, si l'indice de dispersion est supérieur à 1, une alternative à la loi de Poisson est la loi binomiale négative.
- Il peut s'agir d'une rupture de la stationnarité (hypothèse i.). Dans ce cas, l'intérêt du test de la Figure 2-5, par rapport à d'autres tests de stationnarité, est d'indiquer la date possible de cette rupture. Il faut alors découper la chronique en sous-périodes afin que chacune d'elles soit stationnaire et prendre la plus récente. Plusieurs raisons peuvent expliquer une rupture dans la stationnarité d'une chronique : des raisons "techniques" (changement de place de la station, changement du matériel de mesure, changement des conditions d'écoulement liés à la construction d'un ouvrage), ou des raisons plus profondes comme un éventuel changement climatique.
- Il peut s'agir du fait qu'on est en présence d'un phénomène saisonnier. Dans ce cas, deux solutions s'offrent à l'utilisateur : i) effectuer l'échantillonnage en ne considérant que la saison où surviennent majoritairement les événements, par exemple le printemps, ou bien, ii) ajuster une intensité $\mu(t)$ variant en fonction de la saison, et se ramener à un processus de Poisson grâce à une transformation sur l'échelle du temps.
- Enfin, il faut rappeler que la règle de décision des tests statistiques est basée sur un niveau de confiance donné. Il existe en particulier le risque d'erreur de seconde espèce, à savoir de rejeter à tort l'hypothèse testée (probabilité de 10%, pour les résultats présentés)

Pour les 20 bassins étudiés en utilisant les critères d'indépendance, le test sur l'indice de dispersion est mis en défaut pour seulement 2 bassins (U4505010 et V0235010), comme le montre la Figure 2-24. Sur cette figure, l'intervalle de confiance à 90% évolue en fonction des

bassins considérés, car il dépend du nombre d'années d'observation, ainsi que du nombre de crues sélectionnées. A priori, l'hypothèse que l'indice de dispersion doit être égal à 1 semble donc vérifiée. Les valeurs expérimentales se trouvent en grande majorité dans l'intervalle de confiance, sans qu'il y ait de tendance pour qu'elles soit plutôt supérieures ou plutôt inférieures à 1. Ce résultat, même s'il s'agit d'un nombre réduit de bassins, peut être comparé à ceux obtenus dans le Flood Estimation Handbook (Institute of Hydrology, 1999). Cette étude montre sur plusieurs centaines de bassins anglais qu'il y a une tendance pour que l'indice de dispersion soit supérieur à 1. De ce fait, l'hypothèse de la loi de Poisson a été abandonnée au profit de la loi binomiale négative. Cependant, on peut noter que cette étude n'a considéré que les valeurs expérimentales, et non les intervalles de confiance comme nous le faisons ici.

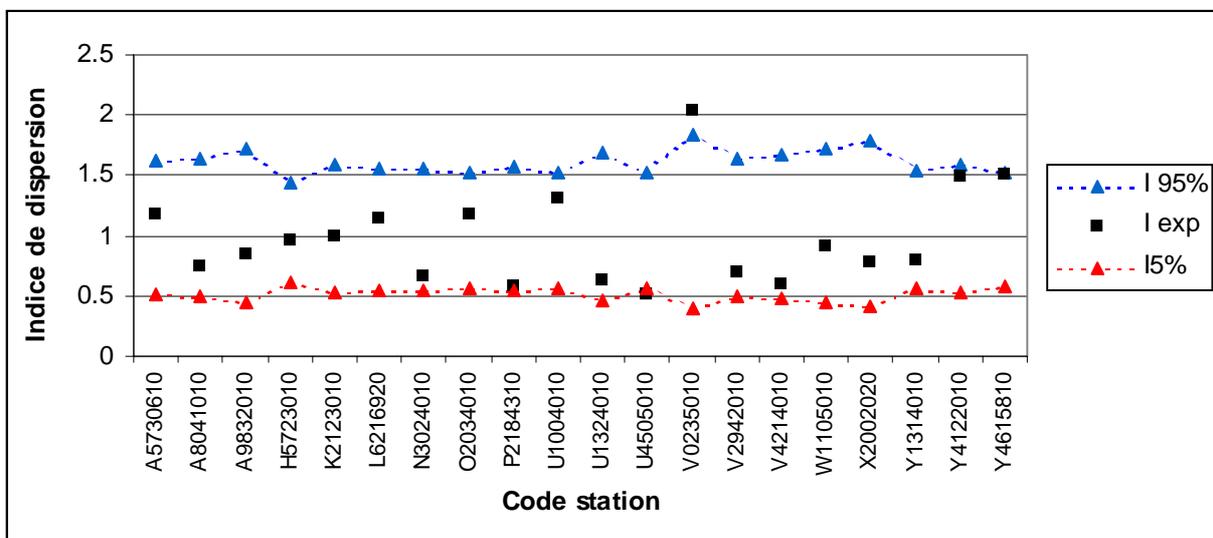


Figure 2-24 : Test sur l'indice de dispersion pour les 20 bassins étudiés

Le respect du test sur la stationnarité pose plus de problèmes. Pour 15 bassins sur 20, la courbe expérimentale sort au moins une fois de l'ellipse. Cependant, la rupture de stationnarité n'est manifeste que pour un seul cas parmi les vingt traités (Figure 2-25). En observant la chronique, il apparaît en effet que les crues survenant après 1983 sont bien moins fortes qu'avant. Il est donc probable qu'il soit survenu à cette date un événement venant modifier le régime des crues. Dans ce cas, il faudrait couper la chronique en deux et étudier individuellement chaque partie stationnaire.

Dans la plupart des autres cas, la rupture n'est pas manifeste. Même si la courbe sort, elle reste néanmoins très proche de la limite tolérée (Figure 2-26).

Les tests présentés sont à prendre avec un certain recul. Ils donnent un diagnostic objectif, plus ou moins précis. Cependant, l'hydrologue reste maître de la décision de se placer ou pas dans le cadre des hypothèses de la loi de Poisson.

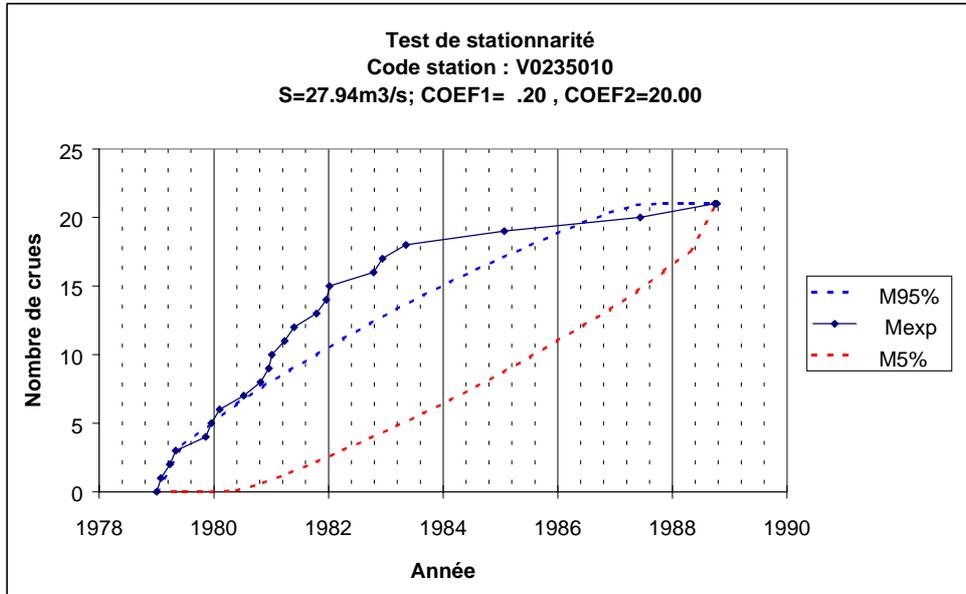


Figure 2-25 : Cas présentant une rupture manifeste de stationnarité

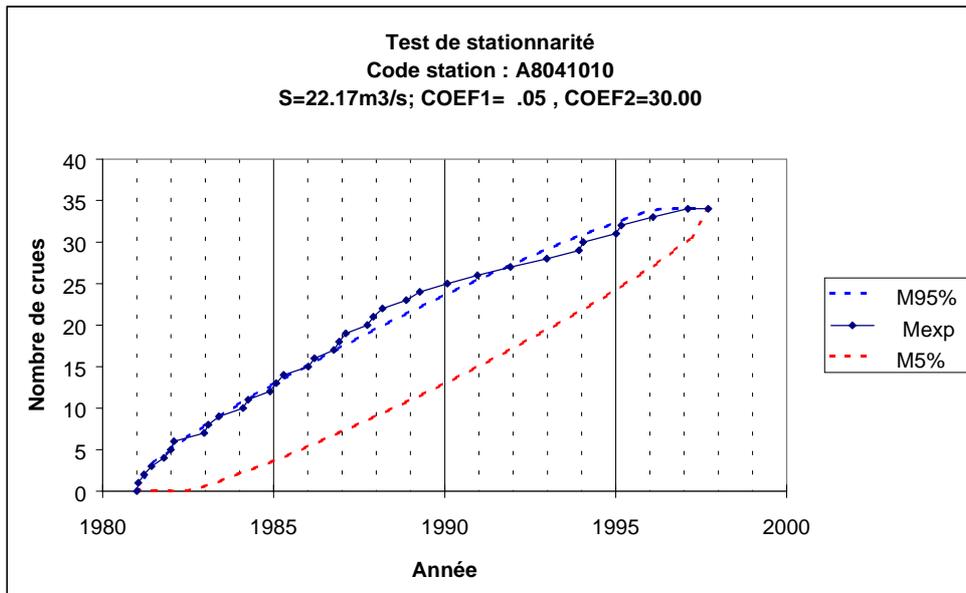


Figure 2-26 : Cas ne présentant pas de rupture manifeste de stationnarité

Si l'on considère que les hypothèses de Poisson ne sont pas vérifiées, et que donc on choisit un modèle plus complexe (par exemple une loi binomiale négative, ou une intensité saisonnière), alors le nombre de paramètres du modèle augmente, et avec lui, l'incertitude de leurs estimations. Comme toujours en statistique, il faut trouver le bon compromis entre un modèle simple mais robuste et un modèle plus détaillé mais qui demande de plus longues chroniques pour être ajusté de façon fiable. Dans le cadre des bassins que nous avons étudiés, le modèle de dépassement constitué du binôme « loi de Poisson/loi exponentielle » nous a paru être ce meilleur compromis.

Conclusion du chapitre 2

Ce chapitre a permis de présenter la modélisation des valeurs supérieures à un seuil. En particulier, nous avons montré qu'un tel modèle comprend deux parties : la modélisation du processus d'occurrence des événements (les dates) ainsi que celle des valeurs atteintes (distribution des débits). La combinaison de ces deux lois permet ensuite d'estimer les quantiles en fonction de leur période de retour. Dans notre cas, le modèle de dépassement constitué du binôme « loi de Poisson/loi exponentielle » a été détaillé, ainsi que deux tests permettant de vérifier si l'indice de dispersion et la stationnarité observés sur l'échantillon étaient conformes aux hypothèses de la loi de Poisson.

Des outils de traitement ont été développés permettant d'effectuer un échantillonnage par valeurs supérieures à un seuil de façon simple et rapide. Ces outils ont permis d'étudier en détail une vingtaine de bassins. Cette étude a montré l'intérêt de l'utilisation de critères d'indépendance. En effet, sans ces critères, de nombreuses valeurs relatives aux mêmes événements de crue sont sélectionnées et viennent fausser les résultats.

A condition d'utiliser des critères d'indépendance définis pour chaque bassin, l'étude montre que les hypothèses de la loi de Poisson sont bien vérifiées, même si dans certain cas, les tests statistiques sont mis en défaut, notamment celui sur la stationnarité. Ces tests ne sont qu'une indication. Leur non respect n'implique pas systématiquement le rejet de la loi de Poisson. De plus, le choix de lois plus complexes pour modéliser les données récalcitrantes à la loi de Poisson mérite des précautions. En effet, ces lois s'accompagnent d'une augmentation du nombre de paramètres à ajuster, et donc d'une plus grande incertitude.

Comme nous l'avons vu, le choix des critères d'indépendance, essentiels pour un échantillonnage correct, doit pour l'instant se faire manuellement. Lorsque le nombre de bassins à étudier est faible (par exemple 20), cette étude est tout à fait envisageable. Par contre, pour un plus grand nombre, le temps nécessaire devient considérable. Le chapitre suivant traite de la modélisation débit-durée-fréquence appliquée à une centaine de bassins. Une automatisation de l'échantillonnage est alors nécessaire.

Chapitre 3

Modélisation débit-durée-fréquence : le modèle local convergent

Après avoir étudié au chapitre précédent la méthode d'échantillonnage par valeurs supérieures à un seuil, nous allons maintenant nous intéresser dans le détail à l'approche débit-durée-fréquence brièvement introduite au chapitre 1.

Ce chapitre présente **les différentes étapes de l'approche débit-durée-fréquence** et en particulier la première d'entre elle : l'échantillonnage « multi-durées ».

Les valeurs obtenues sont ensuite modélisées afin de déterminer les quantiles de débits moyens sur une durée d , pour une période de retour T . Nous appellerons le modèle construit **modèle QdF convergent** en raison de la convergence observée des distributions des débits pour différentes durées.

La **validation de ce modèle** est ensuite effectuée sur une centaine de bassins français. Afin de traiter un aussi grand nombre de stations, nous avons été amenés à définir une procédure d'échantillonnage automatique, dont les résultats sont comparés à ceux de la procédure « rigoureuse » décrite au chapitre 2.

Enfin, une **interprétation hydrologique des propriétés du modèle** est proposée, afin d'expliquer les résultats obtenus.

1. Les différentes étapes de l'approche débit-durée-fréquence

1.1. Echantillonnage « multi-durées »

Comme pour toute analyse fréquentielle, la première étape de l'approche débit-durée-fréquence est l'échantillonnage. Ce paragraphe détaille la méthodologie adoptée.

Pour chaque durée d étudiée, une moyenne mobile de durée d est calculée sur une chronique de débit instantané au pas de temps variable $Q(t)$ (cf. chap. 1, §1.3.2). Les valeurs supérieures à un seuil de cette nouvelle chronique sont obtenues selon la méthodologie présentée au chapitre 2 (Figure 2-6), permettant une détermination automatique du seuil. Pour le cas particulier où $d=0$, la chronique initiale de débit instantané n'est pas transformée, et est directement échantillonnée. Il faut donc connaître pour chaque durée d étudiée : $\mu_{objectif}$, le nombre moyen d'événement pas an que l'on souhaite sélectionner ; $COEF1$ et $COEF2$ les critères d'indépendances. En sortie, on obtiendra pour les N durées d_i étudiées NV_i valeurs échantillonnées notées $v_{d_i}(j)$. Comme nous l'avons vu, il faut veiller pour chaque échantillon, à l'indépendance des valeurs sélectionnées et au respect des hypothèses liées à la loi de Poisson. Finalement, les quantiles $V_d(T)$ sont le résultat de l'ajustement du modèle de dépassement (chapitre 2) sur ces échantillons. La Figure 3-1 illustre la méthodologie proposée pour un échantillonnage « multi-durée ».

Un point essentiel de l'analyse concerne le choix des durées d étudiées. Si l'on veut que les débits moyens échantillonnés soient relatifs à des événements de crue, il faut que les durées d ne dépassent pas la durée des crues. En effet, au-delà on étudie des variations à plus long terme du débit, mais qui ne sont plus liées aux crues (voir au chapitre 1, le §1.2).

La Figure 3-2 présente des résultats obtenus pour trois bassins français choisis pour leur diversité : l'Orgeval au Theil (104 km²) en Seine-et-Marne ; le Valescure à Collobrière (9,2 km²) dans le Var, ainsi que la Moselle à Toul (3350 km²).

1.2. Objectif : une modélisation « continue »

A l'issue de l'échantillonnage « multi-durées » présenté ci-dessus, les distributions $V_d(T)$ donnent donc une description continue en T , grâce à l'ajustement d'une loi statistique (graphiques de gauche, Figure 3-2), mais discrète en d (graphiques de droite, même figure).

Dans ces conditions, l'objectif de la modélisation débit-durée-fréquence est de fournir une formulation continue en fonction de d et T , permettant de décrire les quantiles pour toute période de retour T et toute durée d . Ces quantiles seront alors notés $V(d,T)$, par opposition aux quantiles $V_d(T)$, définis uniquement pour les durées d étudiées*.

* Variable discrète en indice ; variable continue entre parenthèses . Ces notations seront conservées par la suite.

Outre la consolidation des données entre elles, l'intérêt d'une telle modélisation est de fournir une synthèse de la variabilité temporelle des débits. Ainsi, les paramètres capables de décrire la forme des courbes débit-durée-fréquence, reflèteront le régime des hautes eaux auquel est soumis la rivière étudiée. Nous reviendrons sur ce point dans la suite de ce chapitre.

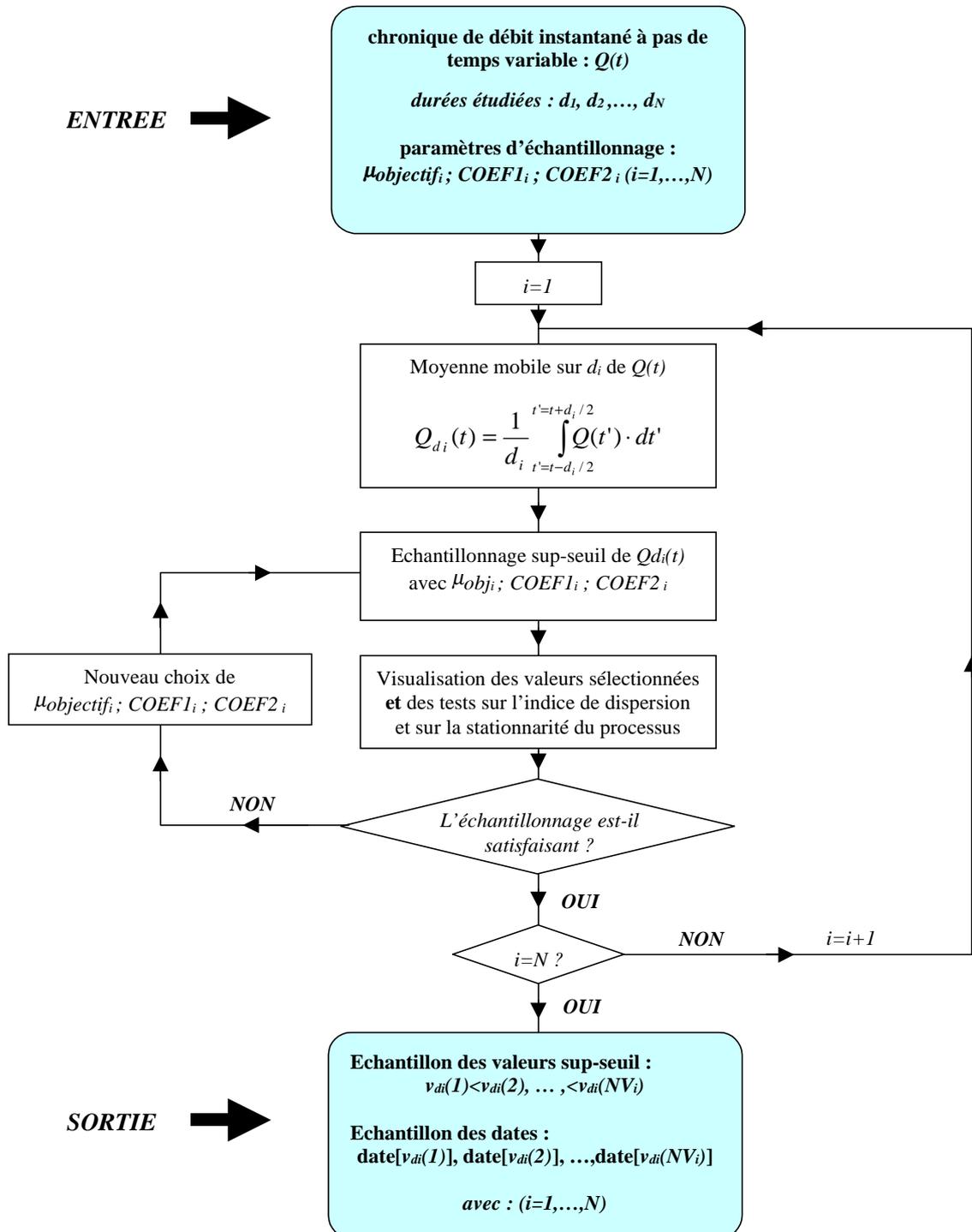
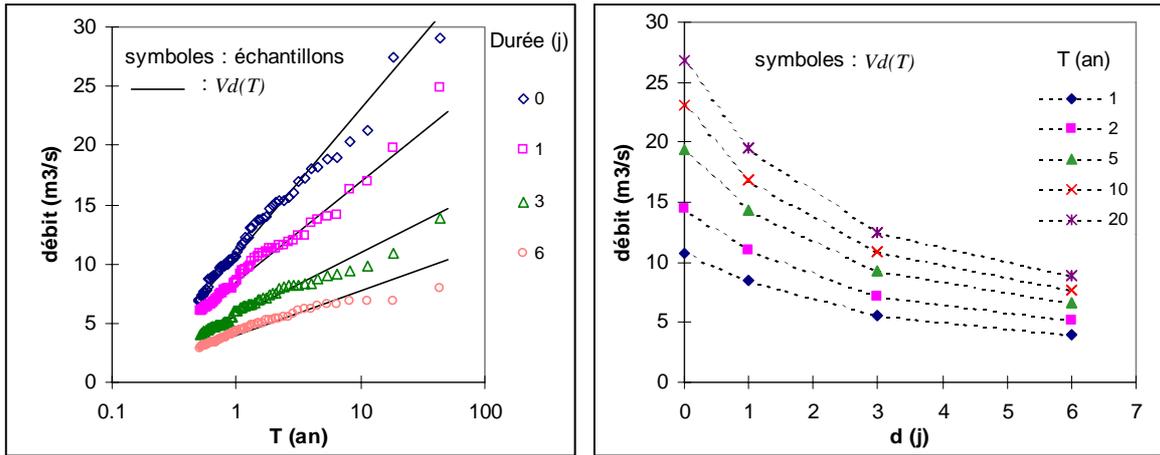
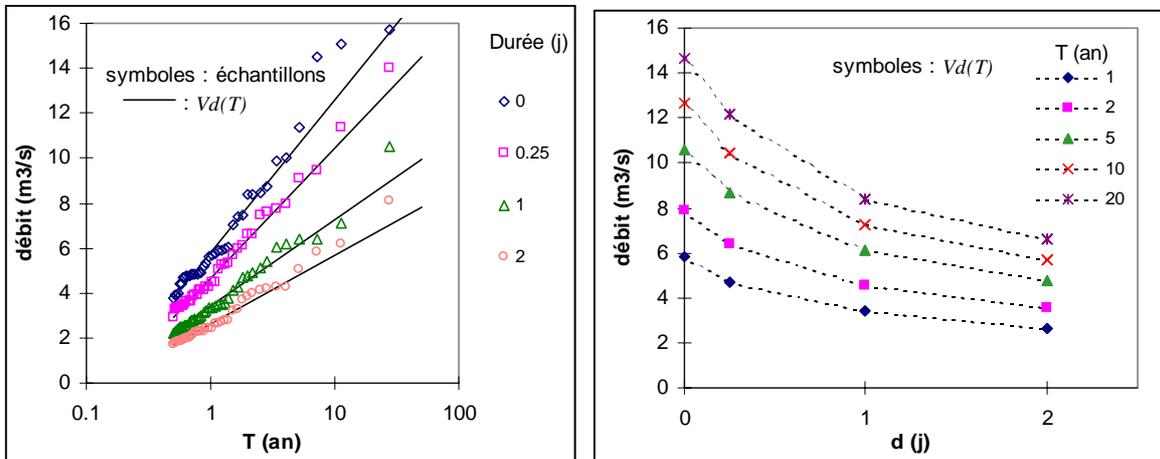


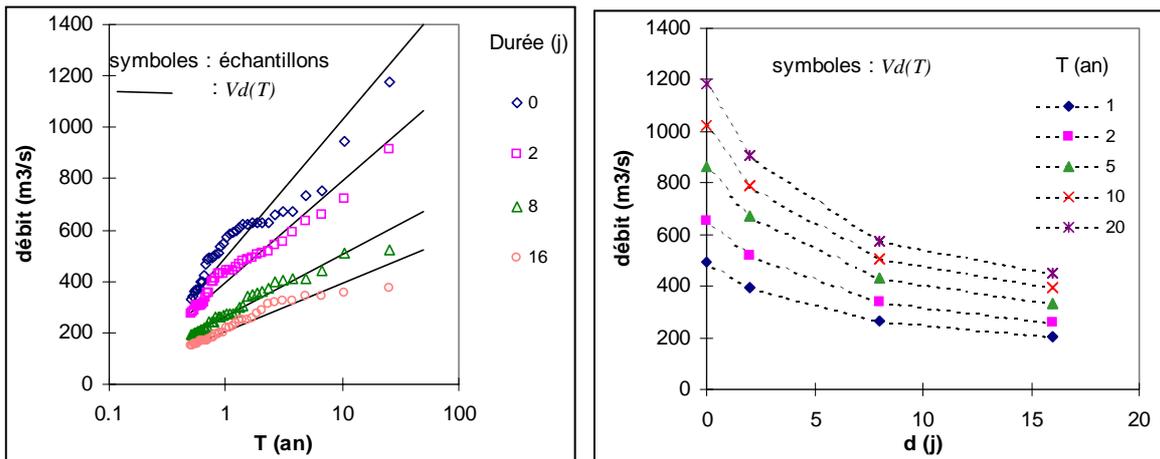
Figure 3-1 : Procédure d'échantillonnage "multi-durées"



a) L'Orgeval au Theil (104 km²)



b) Le Valescure à Collobrière (9,2 km²)



c) La Moselle à Toul (3350 km²)

Figure 3-2 : Représentation des quantiles $V_d(T)$ en fonction de T (graphiques de gauche) et en fonction de d (graphiques de droite)

1.3. Quelques modèles possibles

Modèles en fonction de d pour T fixé

Le problème posé ci-dessus est similaire à celui rencontré dans l'approche intensité-durée-fréquence bien connue pour les pluies. Dans ce cas, la formule de Montana permet, pour une période de retour fixée, de déterminer les quantiles pour toute durée d :

$$I_T(d) = a_T \cdot d^{b_T-1} \quad (3-1)$$

avec : $I_T(d)$ l'intensité de pluie moyenne durant la durée d , relative à la période de retour T
 a_T et b_T les coefficients de Montana, relatifs à la période de retour T .

Cette approche peut être transposée aux débits. Comme l'illustrent les exemples de la Figure 3-2, l'évolution des quantiles $V_d(T)$ en fonction de d peut être modélisée par une forme hyperbolique :

$$V_T(d) = \frac{1}{a_T + b_T \cdot d} + c_T \quad (3-2)$$

avec a_T , b_T et c_T les coefficients du modèle *QdF* continu, à ajuster pour chaque bassin, pour la période de retour T .

Cette approche a l'avantage d'être simple, mais elle présente l'inconvénient que le modèle doit être ajusté pour chaque période de retour étudiée. Ainsi, pour modéliser les courbes de droite de la Figure 3-2, il faut déterminer les paramètres a_T , b_T et c_T pour les périodes de retour 1, 5, 10 et 20 ans, ce qui fait pour cet exemple 12 paramètres à ajuster pour chaque bassin.

Modèle basé sur l'évolution des paramètres de la loi statistique

Une alternative proposée par Galéa et Prudhomme (1997) consiste à modéliser les paramètres de la loi statistique obtenus pour chaque durée. Ainsi, pour l'exemple de l'Orgeval au Theil, la Figure 3-3 représente l'évolution en fonction de d du paramètre d'échelle a (ou gradex) et de position x_0 , relatifs à la loi exponentielle ajustée sur les différentes distributions de débits :

$$V_d(T) = a_d \cdot \ln(T) + [x_0]_d \quad (3-3)$$

Les paramètres a_d et $[x_0]_d$ peuvent par exemple être modélisés par des fonctions hyperboliques à trois paramètres, permettant le calcul des quantiles $V(d,T)$ en fonction de d et T :

$$V(d,T) = \left[\frac{1}{x_1 + x_2 \cdot d} + x_3 \right] \cdot \ln(T) + \left[\frac{1}{x_4 + x_5 \cdot d} + x_6 \right] \quad (3-4)$$

avec x_1 à x_6 , les paramètres du modèle continu *QdF*, à ajuster sur les quantiles $V_d(T)$.

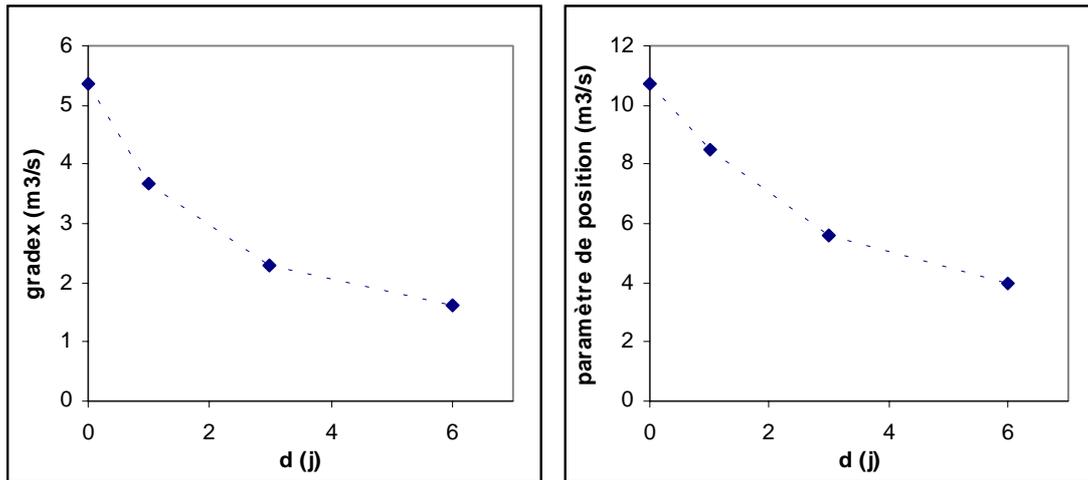


Figure 3-3 : Evolution du gradex et du paramètre de position (ex : l'Orgeval au Theil)

L'intérêt de ce modèle est qu'il permet, une fois les paramètres x_i ajustés de calculer les quantiles $V(d,T)$ pour toute période de retour T et toute durée d . Bien sûr, d et T doivent rester dans les limites imposées par les ajustements $V_d(T)$. Le nombre de paramètres du modèle est ainsi ramené à 6 par bassin étudié.

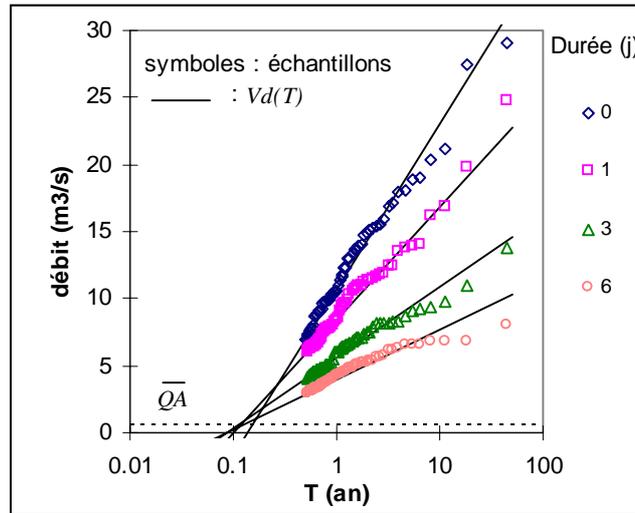
Cependant, l'ajustement de ces paramètres peut s'avérer compliqué. C'est pourquoi Galéa et Prudhomme (1997) ont choisi de ne déterminer ces six paramètres que sur quelques bassins, dits de *référence*, puis utilisent ces résultats pour déterminer les quantiles $V(d,T)$ sur un bassin quelconque. L'ajustement des quantiles $V(d,T)$ n'est donc pas effectué localement, ce qui peut entraîner quelques erreurs. Nous reviendrons dans le détail sur ce point au chapitre 5.

2. Présentation du modèle *QdF* convergent

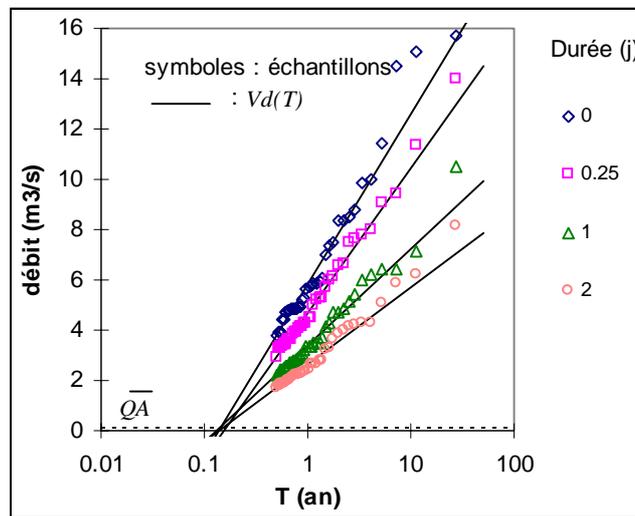
2.1. Hypothèses et formulation du modèle

Les modèles présentés ci-dessus n'étant pas totalement satisfaisants, une nouvelle approche a été développée. L'idée est de mettre en avant des propriétés structurantes permettant de faire des hypothèses simplificatrices. La Figure 3-4 reprend les exemples de courbes débit-durée-fréquence déjà présentés Figure 3-2, mais cette fois en prolongeant volontairement les ajustements vers les faibles période de retour. Ceci fait apparaître une propriété observée sur plus d'une centaine de bassins français : **les distributions semblent converger vers un même point.**

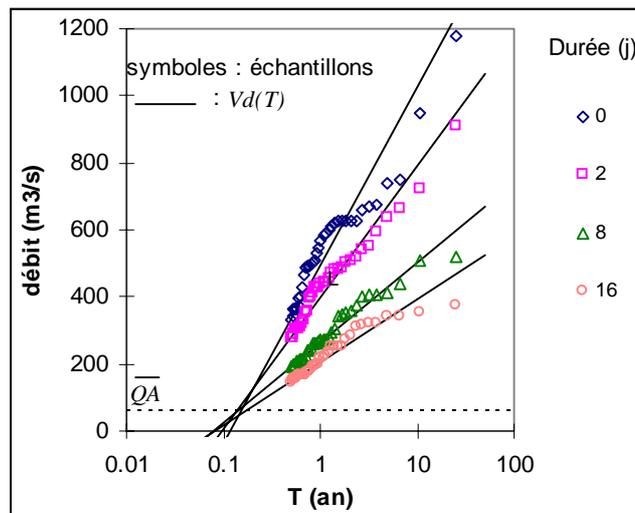
Nous avons choisi de faire l'hypothèse que cette convergence observée est parfaitement vérifiée, donnant ainsi son nom à la nouvelle approche développée : le modèle *QdF* convergent. Comme nous le verrons, cette hypothèse simplificatrice vient réduire le nombre de paramètres du modèle.



a) L'Orgeval au Theil (104km²)



b) Le Valescure à Collobrière (9,2 km²)



c) La Moselle à Toul (3350 km²)

Figure 3-4 : Propriété observée de convergence du faisceau des courbes débit-durée-fréquence

Les hypothèses de base du modèle convergent sont les suivantes :

Hypothèse 1 : Convergence des courbes QdF en fonction de T

La convergence des différentes distributions de débits est schématisée Figure 3-5. Cette figure fait le choix de la loi exponentielle mais comme nous le verrons au chapitre suivant, l'hypothèse 1 peut être généralisée à tout type de loi, en parlant non plus de convergence mais d'affinité des distributions.

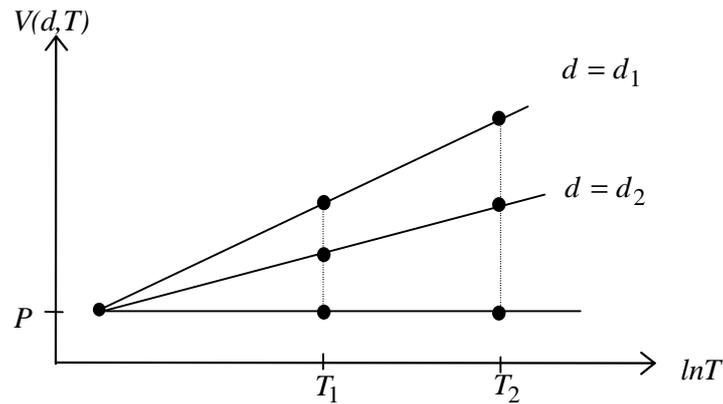


Figure 3-5 : Hypothèse 1 : Convergence des distributions exponentielles des débits

Hypothèse 2 : Forme hyperbolique des courbes QdF en fonction de d

La seconde hypothèse concerne l'évolution des quantiles $V(d,T)$ en fonction de d . On admet que celle-ci est donnée par une forme hyperbolique, conduisant à définir les fonctions $V(0,T)$, $L(T)$ et $\Delta(T)$:

$$V(d,T) = \frac{V(0,T) - L(T)}{1 + d / \Delta(T)} + L(T) \tag{3-5}$$

Comme l'illustre la Figure 3-6, $V(0,T)$ est la valeur de $V(d,T)$ quand $d=0$, $L(T)$: la limite de $V(d,T)$ quand d tend vers l'infini et $\Delta(T)$, une fonction de la dimension d'une durée décrivant la courbure de l'hyperbole relative à la période de retour T .

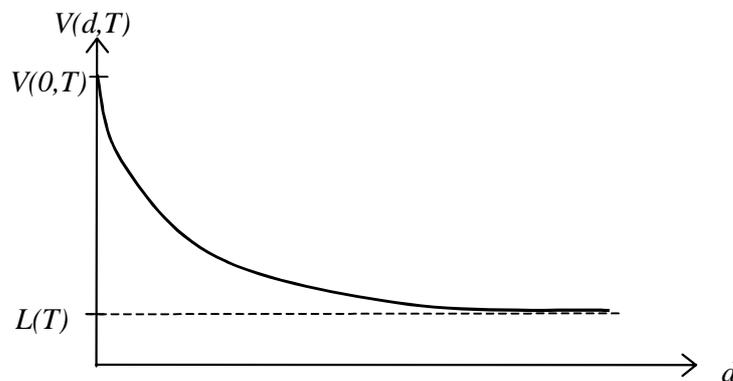


Figure 3-6 : Hypothèse 2 : Forme hyperbolique des courbes QdF en fonction de d .

Formulation du modèle

En adoptant l'hypothèse de convergence (hypothèse 1), la formulation (3-5) du modèle est simplifiée. P étant l'ordonnée du point de convergence, la Figure 3-5 indique que les rapports suivants ne dépendent pas de T :

$$\frac{V(d_1, T_1) - P}{V(d_2, T_1) - P} = \frac{V(d_1, T_2) - P}{V(d_2, T_2) - P} \quad (3-6)$$

L'application directe de cette égalité des rapports est la suivante. Si d_1 est une durée d quelconque et que d_2 est une constante, que nous prendrons égale à zéro pour simplifier la suite des calculs, alors le rapport ci-dessous n'est plus fonction que de d .

$$\frac{V(d, T) - P}{V(d=0, T) - P} = f(d) \quad (3-7)$$

D'autre part, l'équation (3-5) peut se mettre sous la forme :

$$\frac{V(d, T) - L(T)}{V(0, T) - L(T)} = \frac{1}{1 + d / \Delta(T)} \quad (3-8)$$

En comparant (3-7) et (3-8), on obtient par identification :

$$L(T) = P = cste \quad (3-9)$$

$$\Delta(T) = \Delta = cste' \quad (3-10)$$

Finalement, on obtient :

$$V(d, T) = \frac{V(0, T) - P}{1 + d / \Delta} + P \quad (3-11)$$

$V(0, T)$ est le quantile du débit de pointe ($d=0$) ; P est l'ordonnée du point de convergence des distributions et Δ est un paramètre décrivant la courbure des hyperboles. La signification de ce dernier paramètre est évoquée au paragraphe 2.4.

L'adoption de l'hypothèse de convergence des distributions a donc permis de simplifier l'équation générale du modèle (3-5) : les fonctions $L(T)$ et $\Delta(T)$ sont des constantes.

Comme nous le verrons au chapitre suivant, l'intérêt de la formulation (3-11) est qu'elle n'impose pas de loi statistique particulière. Si maintenant, on choisit de travailler avec la loi exponentielle pour décrire $V(0, T)$, on obtient :

$$V(d, T) = \frac{a(0) \cdot \ln(T) + x_0(0) - P}{1 + d / \Delta} + P \quad (3-12)$$

Dans ce cas, le modèle contient donc quatre paramètres : Δ et P précédemment présentés, ainsi que $a(0)$ et $x_0(0)$, respectivement le gradex et le paramètre de position de la distribution exponentielle du débit de pointe. Le paragraphe suivant explique comment ces paramètres sont ajustés sur les valeurs observées.

2.2. Ajustement des paramètres du modèle

2.2.1. Choix de l'ordonnée du point de convergence

Plusieurs possibilités s'offrent à nous, quant à la détermination de la valeur de P , l'ordonnée du point de convergence des distributions.

La première est de considérer P comme un paramètre du modèle, au même titre que Δ et les paramètres décrivant le quantile $V(0,T)$. Dans ce cas, P doit être ajusté.

Une deuxième possibilité consiste à fixer la valeur de P . Pour cela, on peut s'intéresser à la signification de ce paramètre. En effet, d'après (3-11), on a :

$$P = \lim_{d \rightarrow \infty} (V(d,T)) \quad (3-13)$$

Si l'on revient au mode de calcul de la variable Vd , celle-ci a été obtenue en effectuant une moyenne mobile sur la chronique $Q(t)$ (cf. chapitre 1, Figure 1-19). En moyennant sur une durée de plus en plus grande, les valeurs maximales sélectionnées deviennent de plus en plus faibles, jusqu'à atteindre une valeur limite. En raisonnant sur une chronique de longueur infinie, cette valeur limite tend nécessairement vers la moyenne de la chronique. Ainsi, on peut imposer à P la valeur du débit moyen interannuel de la chronique étudiée, noté \overline{QA} .

$$P = \overline{QA} \quad (3-14)$$

Nous avons calculé la valeur du débit moyen interannuel pour les trois bassins étudiés. Pour l'Orgeval, le Valescure et la Moselle, les valeurs obtenues sont respectivement de 0.65 ; 0.13 et 66 m³/s. La Figure 3-4, représente ces valeurs par une ligne pointillée horizontale. On constate que celles-ci sont proches de la convergence « observée ». Cette vérification « graphique » a aussi été faite sur d'autres bassins et confirme cette affirmation.

Il faut cependant noter une limite à ce raisonnement. Comme nous l'avons noté en début de chapitre, au §1.1. , le fait que nous étudions les crues nous oblige à travailler sur des durées compatibles avec la durée de ce phénomène, propre à chaque bassin étudié. Donc, le cas des longues durées, et a fortiori de la « durée limite infinie » sort de notre domaine d'étude.

Enfin on peut constater que les trois exemples présentés, l'ordonnée de ce point de convergence est très inférieure aux débits échantillonnés. Ceci se vérifie aussi sur les autres cas que nous avons traités. Une solution simple est donc supprimer le paramètre P , c'est à dire de le fixer $P=0$.

Comme nous le verrons par la suite, les résultats obtenus en fixant $P=0$ ou $P = \overline{QA}$ ont été jugés satisfaisants. Il n'a donc pas été jugé nécessaire de tester la première méthode envisagée, à savoir l'ajustement de P . Par contre, nous étudierons l'incidence sur les résultats du choix fait entre $P=0$ ou $P = \overline{QA}$.

2.2.2. Méthode des ajustements consolidés

La démarche que nous adoptons ici comporte deux étapes : un ajustement « consolidé » qui consiste à déterminer un faisceau de distributions parfaitement convergent, puis un ajustement en fonction de d , qui consiste à déterminer la relation entre les différentes distributions. Cette méthode a été présentée par Javelle *et al.* (1999).

Dans la suite nous noterons les valeurs échantillonnées pour les N durées d_i : $v_{d_i}(j)$, j étant le rang de la valeur dans l'échantillon. A l'issue de l'échantillonnage, chacune des valeurs est repérée par i le numéro de l'échantillon relatif à une durée, et j son rang. Le but est d'ajuster une fonction à deux variables $V(d,T)$ sur l'ensemble de ces valeurs.

- **Ajustement consolidé**

La relation (3-6) peut aussi se mettre sous la forme :

$$\frac{V(d_1, T_1) - P}{V(d_1, T_2) - P} = \frac{V(d_2, T_1) - P}{V(d_2, T_2) - P} \quad (3-15)$$

Tout comme les rapports de la relation (3-6) ne dépendaient pas de T , ceux de la relation (3-15) ne dépendent pas de d . On peut ainsi définir une variable adimensionnelle V^* , dont l'expression du quantile ne dépend pas en théorie de d :

$$V^*(T) = \frac{V_{d_i}(T) - P}{V_{d_i}(T = T_{ref}) - P} \quad (3-16)$$

Pour mettre en lumière cette propriété, les débits normés $v_{d_i}^*(j)$ sont définis à partir des débits échantillonnés $v_{d_i}(j)$, j étant le rang de la valeur échantillonnée

$$v_{d_i}^*(j) = \frac{v_{d_i}(j) - P}{V_{d_i}(T = T_{ref}) - P} \quad (3-17)$$

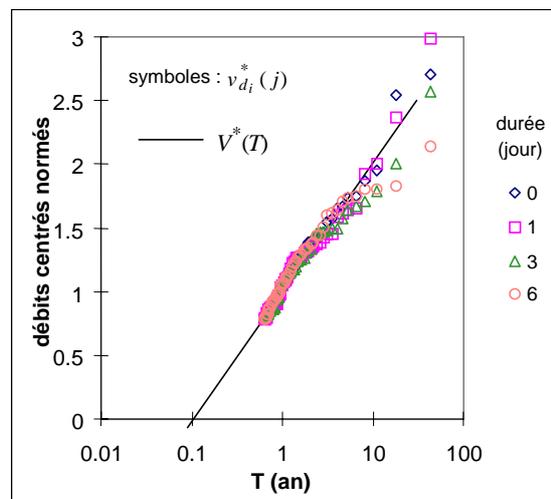


Figure 3-7 : Echantillons adimensionalisés ($T_{ref} = 1 \text{ an}$, $P = \overline{QA}$), ex : l'Orgeval au Theil

La Figure 3-7, relative à la rivière Orgeval, montre que les débits adimensionnalisés $v_{d_i}^*(j)$ sont pratiquement confondus, quelle que soit la durée d_i à laquelle ils se réfèrent. Ceci illustre sur cet exemple l'hypothèse d'affinité (hypothèse 1). Ici, T_{ref} est fixée arbitrairement à 1 an. Il est ensuite possible d'ajuster les paramètres A^* et B^* d'une loi statistique (ici, la loi exponentielle) sur toutes les valeurs $v_{d_i}^*(j)$, avec la relation :

$$V^*(T) = A^* \cdot \ln T + B^* \quad (3-18)$$

Une fois dénormé pour chaque durée d_i , cet unique ajustement permet d'obtenir des ajustements parfaitement convergents dont les quantiles sont notés $V_{d_i}^c(T)$, l'exposant c pour « consolidation » (Figure 3-8). En effet, chaque quantile $V_{d_i}^c(T)$ relatif à une durée d_i a été calculé, et donc consolidé, à l'aide des échantillons observés $v_{d_i}(j)$ relatifs à l'ensemble des durées $d_i, i=1, \dots, N$.

Les quantiles consolidés :

$$V_{d_i}^c(T) = A_{d_i}^c \cdot \ln T + B_{d_i}^c \quad (3-19)$$

sont définis par les paramètres d'échelle et de position $A_{d_i}^c$ et $B_{d_i}^c$ dont les valeurs s'obtiennent en combinant les relations (3-16) et (3-18) :

$$A_{d_i}^c = A^* (V_{d_i}(T = T_{ref}) - P) \quad (3-20)$$

$$B_{d_i}^c = B^* (V_{d_i}(T = T_{ref}) - P) + P \quad (3-21)$$

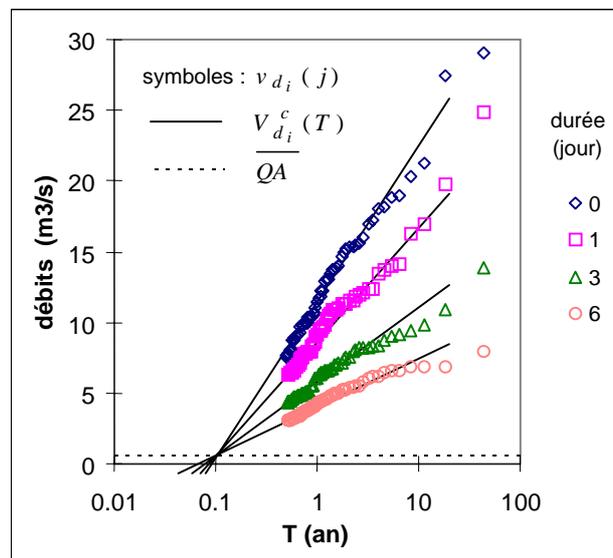


Figure 3-8 : Ajustements consolidés obtenus en imposant $P = \overline{QA}$, ex : l'Orgeval au Theil

L'ajustement consolidé consiste donc à faire "pivoter" chacune des distributions relatives à la durée d_i (Figure 3-4), afin de les rendre parfaitement convergentes. Par construction, le choix de la période de retour T_{ref} détermine l'abscisse aux environs de laquelle s'effectue cette

rotation. Dans la pratique, nous avons choisi de fixer T_{ref} à 1 an, car pour cette période de retour, l'intervalle de confiance sur les quantiles est faible.

• **Ajustement en fonction de d**

Nous allons effectuer une seconde normalisation, inspirée de la relation (3-6). Le but est de permettre cette fois-ci un ajustement en fonction de la durée d . Nous définissons une nouvelle variable adimensionnelle, dont le quantile ne dépend pas en théorie de d :

$$V_{d_i}^{**}(T) = \frac{V_{d_i}(T) - P}{V_{d_{ref}}^c(T) - P} \tag{3-22}$$

Afin d'illustrer cette propriété, nous normons les échantillons $v_{d_i}(j)$ de la façon suivante :

$$v_{d_i}^{**}(j) = \frac{v_{d_i}(j) - P}{V_{d_{ref}}^c(T = T_j) - P} \tag{3-23}$$

Cette normalisation fait référence au quantile consolidé $V_{d_{ref}}^c(T)$, relatif à une durée d_{ref} choisie arbitrairement parmi les durées d_i étudiées. La période de retour T_j est la période de retour expérimentale, relatives aux valeurs $v_{d_i}(j)$ de rang j .

Ces débits centrés normés ne dépendent plus de T . Cette fois-ci l'ensemble des distributions relatives aux quantiles consolidés $V_{d_i}^c(T)$ étant parfaitement convergent (Figure 3-8), le choix de d_{ref} n'a aucune incidence sur le résultat final, à savoir l'ajustement des paramètres $a(0)$, $x_0(0)$ et Δ . Dans l'exemple de la Figure 3-9, la distribution consolidée de référence est relative aux débits instantanés ($d_{ref} = 0$).

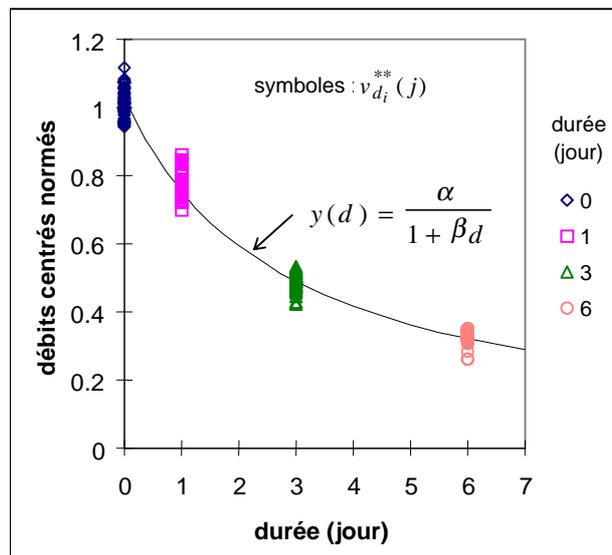


Figure 3-9 : Echantillons adimensionalisés ($d_{ref} = 0$), ex : l'Orgeval au Theil

L'étalement vertical des points (Figure 3-9) est dû à la dispersion expérimentale des valeurs $v_{d_i}(j)$ autour des quantiles $V_{d_i}^c(T)$ issus des distributions consolidées (Figure 3-8). L'évolution en fonction de d de la variable $V_d^{**}(T)$ est modélisée par une fonction du type :

$$y(d) = \alpha / (1 + \beta d) \quad (3-24)$$

Les paramètres α et β sont calés en minimisant la somme S des carrés des écarts entre $y(d)$ et les valeurs expérimentales $v_{d_i}^{**}(j)$ notée ε :

$$S = \sum_{i=1}^N \sum_{j=1}^{NV} [y(d = d_i) - v_{d_i}^{**}(j)]^2 \quad (3-25)$$

N est le nombre de durées étudiées, et NV le nombre de valeurs dans chaque échantillon. Les quantiles $V(d, T)$ en écrivant l'égalité entre les quantiles définis par la relation (3-22) et la fonction $y(d)$ définie par la relation (3-25).

sont obtenus en dénormalisant puis décentrant la fonction $y(d)$ (Figure 3-10) :

$$V(d, T) = \frac{\alpha}{1 + \beta d} \cdot (V_{d_{ref}}^c(T) - P) + P \quad (3-26)$$

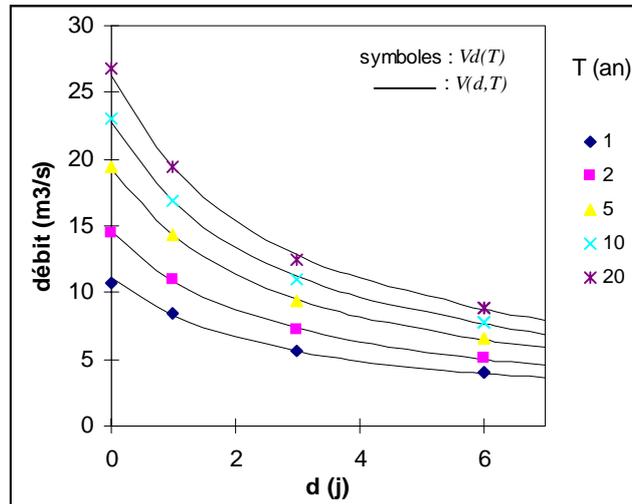


Figure 3-10 : Dénormalisation des valeurs modélisées par $y(d)$, ex : l'Orgeval au Theil

En comparant (3-26) et (3-12), on trouve par identification :

$$\Delta = 1 / \beta \quad (3-27)$$

$$a(0) = \alpha A_{d_{ref}}^c \quad \text{et} \quad x_0(0) = \alpha (B_{d_{ref}}^c - P) + P \quad (3-28)$$

2.2.3. Méthode simplifiée

Une fois les propriétés du modèle admises, à savoir la convergence des distributions en fonction de T et la forme hyperbolique de $V(d, T)$ en fonction de d , un allègement de la démarche ci-dessus est possible (Meunier 1999, 2001). L'équation (3-11) peut se mettre sous la forme suivante :

$$V(0, T) = [V(d, T) - P][1 + d / \Delta] + P \quad (3-29)$$

Elle signifie que chaque distribution $V(d,T)$ peut être ramenée à la distribution $V(d=0,T)$, connaissant Δ et P . Le principe est donc de transformer chaque valeur $v_d(j)$ de l'échantillon en une valeur notée $x_d(j)$:

$$x_{d_i}(j) = [v_{d_i}(j) - P] \cdot [1 + d / \delta] + P \quad (3-30)$$

La valeur de P étant fixée ($P=0$ ou $P = \overline{QA}$), ces distributions transformées ne dépendent que de δ . D'après l'équation (3-29), dans le cas où $\delta=\Delta$, les distributions expérimentales $x_d(j)$ doivent se confondre. Le principe est donc de faire varier δ et de déterminer la valeur δ_{opt} pour laquelle l'écart entre les valeurs échantillonnées transformées $x_d(j)$ est minimal. Un estimateur du paramètre Δ est donné par la valeur δ_{opt} . Afin de déterminer l'écart entre les distributions expérimentales $x_d(j)$, on calcule d'abord une distribution « moyenne », notée $xmoy(j)$:

$$xmoy(j) = \frac{1}{N} \sum_{i=1}^N x_{d_i}(j) \quad (3-31)$$

puis un critère ε rend compte de l'écart entre les différentes distributions $x_d(j)$.

$$\varepsilon = \frac{1}{NV} \frac{1}{N} \sum_{j=1}^{NV} \sum_{i=1}^N \left(\frac{x_{d_i}(j) - xmoy(j)}{xmoy(j)} \right)^2 \quad (3-32)$$

avec NV : le nb de valeurs dans chaque échantillon, N le nombre de durée d_i étudiées

Le cas où ε est égal à zéro correspond l'identité de toutes les distributions $x_d(j)$. La valeur optimale de Δ est donc obtenue en minimisant ε . Dans le calcul de ε , les écarts calculés sont divisés par $xmoy(j)$, ceci afin de donner autant de poids aux faibles et fortes valeurs des échantillons.

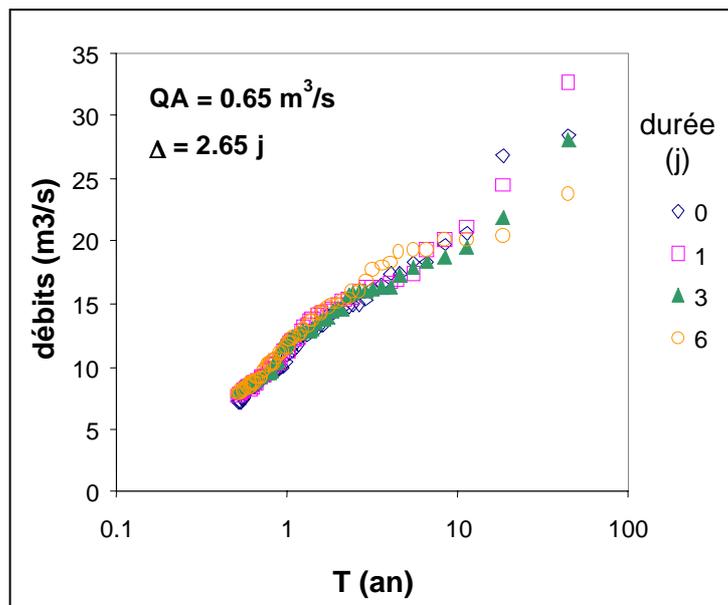


Figure 3-11 : Optimisation de Δ en minimisant l'écart entre les échantillons transformés $x_d(j)$, exemple avec $P = \overline{QA}$ (bassin : L'orgeval au Theil)

La Figure 3-11 donne un exemple d'optimisation de Δ pour l'exemple de la rivière Orgeval, déjà présenté. La dernière étape consiste à calculer les paramètres de la distribution $V(d=0, T)$. Ceux-ci sont ajustés à partir de la distribution expérimentale $xmoy(j)$. Si on choisit de travailler avec la loi exponentielle, la méthode des moments donne :

$$a(0) = \frac{1}{NV} \sum_{j=1}^{j=NV} xmoy(j) - xmoy(1) \quad (3-33)$$

$$x_0(0) = xmoy(1) + a(0) \ln\left(\frac{NV}{NA}\right) \quad (3-34)$$

avec NV le nombre de valeurs dans chaque échantillon et NA le nombre d'années d'observation.

Le résultat de cet ajustement est présenté Figure 3-12. Les paramètres $a(0)$, $x_0(0)$ et Δ étant maintenant estimés, il est maintenant possible de calculer les quantiles $V(d, T)$ grâce à la formule (3-12).

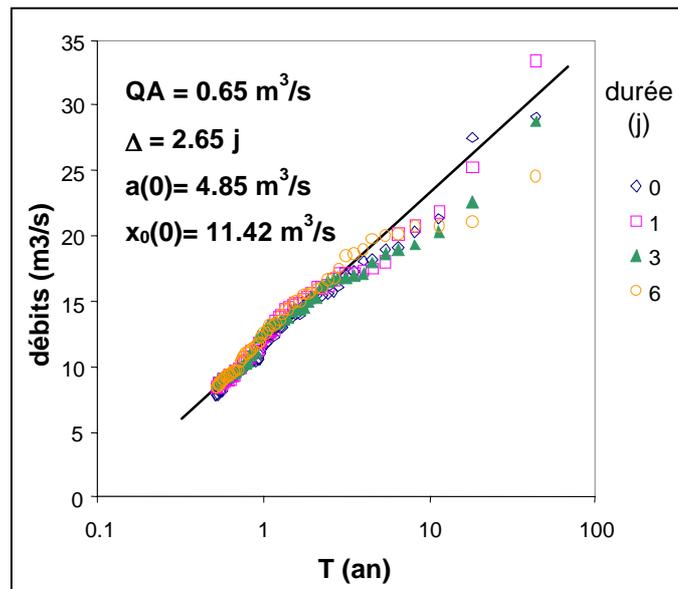


Figure 3-12 : Ajustement de la distribution $V(d=0, T)$ à partir de la distribution expérimentale $xmoy(j)$, exemple avec $P = \overline{QA}$. (bassin : L'Orgeval au Theil)

Par la suite, nous opterons pour cette méthode simplifiée, qui admet d'emblée les deux hypothèses du modèle. Son intérêt est qu'elle est facilement transposable pour ajuster d'autres types de loi statistique. En effet, l'ajustement de Δ ne dépend d'aucun choix préalable quant à la loi statistique utilisée. Il est réalisé directement à l'aide des valeurs échantillonnées $v_d(j)$. Ce n'est qu'une fois la valeur de Δ déterminée que le choix de la loi est effectué. La méthode a ici été présentée pour l'ajustement d'une loi exponentielle, mais nous verrons dans le chapitre suivant des applications à d'autres types de lois.

2.3. Validation du modèle sur les trois exemples étudiés

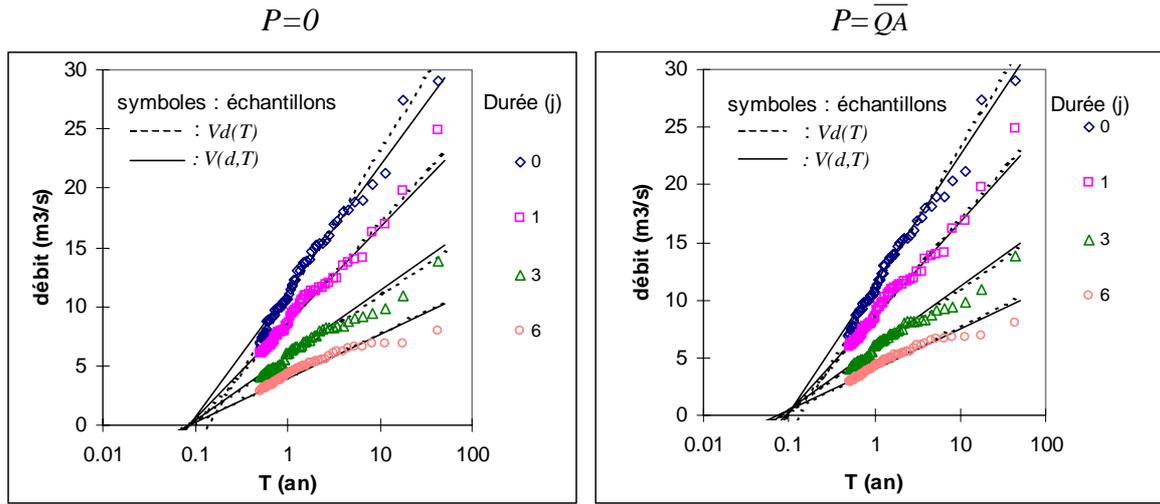
L'ajustement du modèle convergent, décrit ci-dessus a été appliqué aux trois exemples déjà traités. Nous avons considéré les deux cas : l'ordonnée du point de convergence fixée à zéro, puis à \overline{QA} . La Figure 3-13 compare les résultats obtenus en ajustant le modèle avec $P=0$ et $P=\overline{QA}$. On constate que dans les deux cas, les quantiles observés $V_d(T)$ sont bien décrits par les quantiles $V(d,T)$.

Le choix de l'ordonnée de convergence semble avoir peu d'incidence sur les résultats. Simplement, comme le montre le Tableau 3-1, les paramètres du modèle sont légèrement différents. En particulier, lorsqu'on fixe $P=0$, la durée Δ est plus grande. Cela s'explique Figure 3-6 : si l'on ramène l'asymptote des courbes à zéro, le paramètre Δ augmente de façon à ce que l'hyperbole décrite garde la même forme.

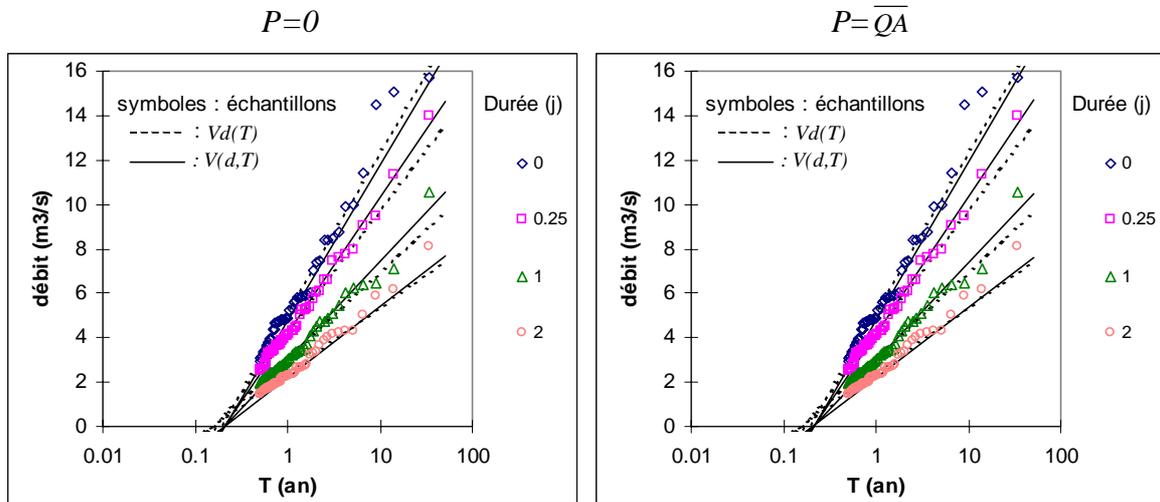
		P (m3/s)	$x_0(0)$ (m3/s)	$a(0)$ (m3/s)	Δ (jour)
L'Orgeval au Theil	$P=0$	0.00	11.29	4.60	3.25
	$P=\overline{QA}$	0.65	11.42	4.85	2.65
Le Valescure à Collobrière	$P=0$	0.00	4.81	3.05	1.70
	$P=\overline{QA}$	0.13	4.82	3.10	1.60
La Moselle à Toul	$P=0$	0.00	501.65	194.46	9.02
	$P=\overline{QA}$	66.00	511.11	216.59	6.40

Tableau 3-1 : Paramètres du modèle convergent en fixant $P=0$ et $P=\overline{QA}$

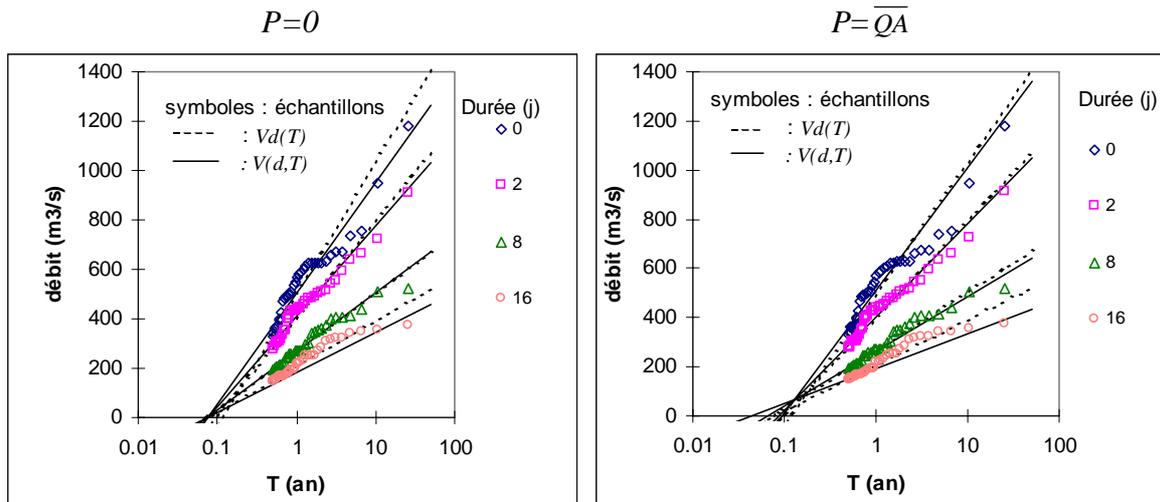
Ces résultats ont été obtenus sur simplement trois exemples et doivent être testés sur un plus grand nombre de bassins. Cette validation est l'objet du paragraphe 3. Les résultats qui y sont présentés concernent une centaine de bassins.



a) L'Orgeval au Theil (104km²)



b) Le Valescure à Collobrière (9,2 km²)



c) La Moselle à Toul (3350 km²)

Figure 3-13 : Comparaison du modèle convergent ajusté avec $P=0$ et $P=\overline{QA}$ *

* $V_d(T)$ correspond à l'ajustement réalisé pour une seule durée, $V(d,T)$ à l'ajustement du modèle convergent.

2.4. Signification et propriété du paramètre Δ

Le paramètre Δ donne la forme des hyperboles définissant les quantiles $V(d,T)$ pour T fixé. Si Δ est faible, les hyperboles sont très incurvées. Inversement, si Δ est fort, les hyperboles sont très aplaties. Or, comme le schématise la Figure 3-14, la forme des hyperboles est liée à celle des crues observées. En effet, plus l'ensemble des crues étudiées est rapide, plus la différence entre les débits de pointe et les débits maximaux moyens sur une durée d (par exemple 1 jour, sur la Figure 3-14) est grande. Ce point avait déjà été souligné au chapitre 1, §3.5, à l'aide d'un exemple. Cette différence entre les débits de pointes et les débits moyens se traduit pour les courbes QdF (en fonction de d , pour T fixé) par une plus ou moins forte courbure. Par conséquent, un bassin caractérisé par des crues rapides a des courbes QdF très incurvées. Inversement, s'il est caractérisé par des crues lentes, ses courbes QdF sont plus aplaties. Le paramètre Δ servant à décrire la forme des hyperboles (3-11), sa valeur nous renseigne donc sur la dynamique des crues étudiées. Δ , qui a la dimension d'un temps, peut donc être considéré comme une durée caractéristique de crue du bassin étudié. A ce titre, nous effectuerons dans la suite une comparaison entre ce paramètre et la durée caractéristique de crue de la méthode Socose, présentée au chapitre 1 (§3.2).

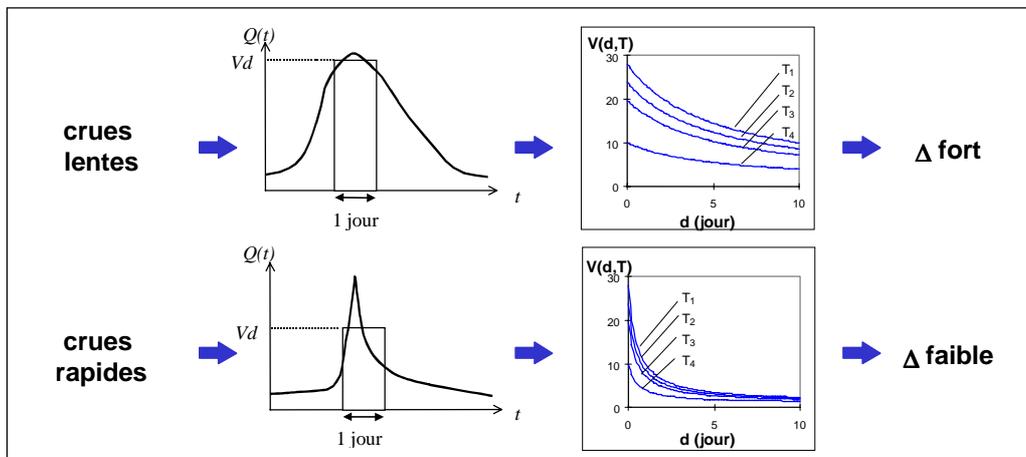
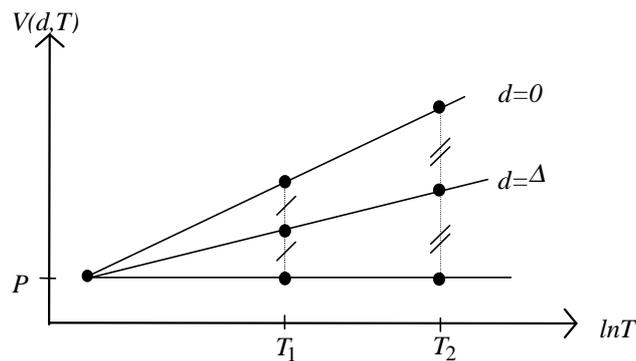


Figure 3-14 : Signification du paramètre Δ

D'autre part, la durée Δ peut aussi être interprétée de la façon suivante. La distribution des débits moyens relative à cette durée se trouve être à mi-distance entre la distribution des débits instantanés $V(0,T)$ et la droite $y=P$, P étant l'ordonnée du point de convergence. En effet, l'équation (3-11) nous donne la relation suivante:

$$\frac{V(d = \Delta, T) - P}{V(d = 0, T) - P} = \frac{1}{2} \quad (3-35)$$

Cette relation est illustrée graphiquement par la Figure 3-15 et est vérifiée quel que soit T grâce à la propriété de convergence des distributions.


 Figure 3-15 : Propriété de la durée Δ

D'autre part, toujours d'après la relation (3-11), l'égalité suivante est vérifiée quel que soit T :

$$\frac{V(d, T) - P}{V(d = 0, T) - P} = \frac{1}{1 + d / \Delta} \quad (3-36)$$

Cette égalité signifie que les quantiles de débit normés comme l'indique la relation ci-dessus s'expriment tous uniquement en fonction de d/Δ . Cette invariance d'échelle ne dépend ni de la période de retour considérée, ni du bassin, puisqu'à chaque fois, les valeurs sont normées par les paramètres $V(0, T)$, P et Δ propres à chaque bassin (Tableau 3-1). Afin d'illustrer cette propriété, nous avons repris les trois exemples précédents ainsi que les résultats du modèle convergents ajusté avec $P = \overline{QA}$. Les quantiles « observés » $V_d(T)$ (c'est à dire les quantiles non consolidés obtenus Figure 3-2) et la durée d , ont été normés pour obtenir les valeurs adimensionnelles $V_d^*(T)$ et d^* :

$$V_d^*(T) = \frac{V_d(T) - P}{V(d = 0, T) - P} \quad ; \quad d^* = d/\Delta \quad (3-37)$$

Pour les trois exemples étudiés, la Figure 3-16 représente les quantiles non consolidés $V_d(T)$ en fonction de d pour les périodes de retour de 1 et 20 ans (graphique de gauche), ainsi que les mêmes données, mais cette fois-ci normées selon la relation ci-dessus (graphique de droite). A cause de la grande différence d'échelle entre les bassins étudiés, les graphiques de gauche ont dû être représentés en échelle logarithmique pour les débits.

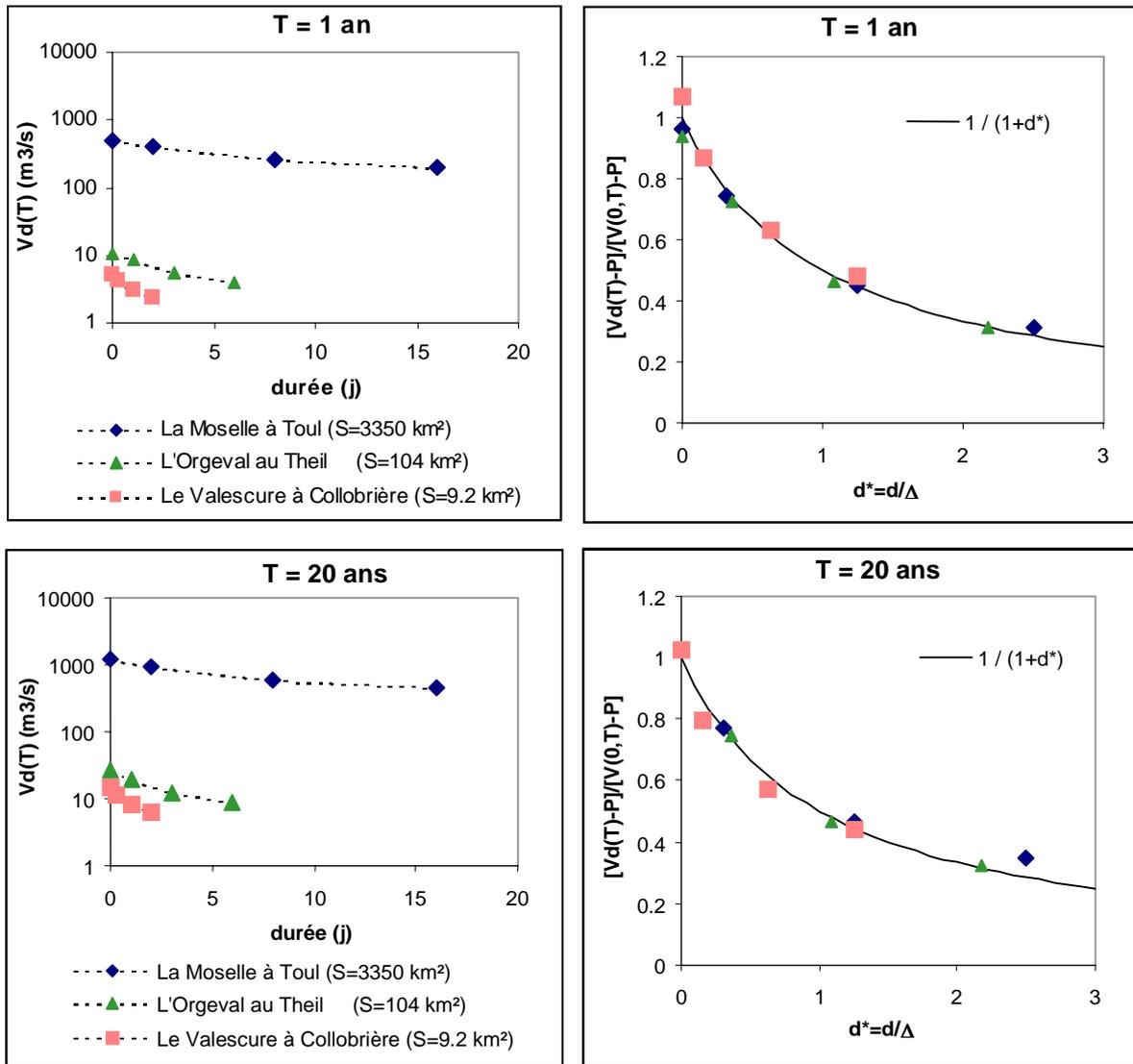


Figure 3-16 : Illustration de la propriété d'invariance d'échelle

Les résultats montrent qu'une fois normés (graphiques de droite), les quantiles $V_d(T)$ de chaque bassin s'ajustent tous sur la courbe d'équation $y=1/(1+d^*)$, quelle que soit la période de retour considérée, comme le prévoyait la relation (3-36).

La durée Δ , et de façon plus générale le modèle *QdF convergent*, mettent donc en lumière des propriétés d'invariance d'échelle. Ces résultats sont d'autant plus intéressants que les trois bassins étudiés ont des échelles spatiales et temporelles complètement différentes. Ces observations doivent maintenant être vérifiées sur un plus grand nombre de bassins.

3. Validation du modèle QdF convergent sur 105 bassins français

Ce paragraphe propose donc de valider le modèle *QdF* convergent sur la centaine de bassins français présentés au chapitre 2, §3.1. Pour cela, quelques problèmes techniques sont à résoudre, notamment pour ce qui concerne l'échantillonnage d'un aussi grand nombre de stations.

3.1. Automatisation de l'échantillonnage

En effet, nous avons vu au chapitre 2 que l'échantillonnage « multi-durées » par valeurs supérieures à un seuil requiert une certaine attention, et cela pour deux raisons :

- les valeurs sélectionnées doivent être indépendantes (critères *COEF1* et *COEF2*) ;
- les durées d étudiées ne doivent pas être supérieures aux durées des crues, car si tel est le cas, on n'étudie plus les crues, mais des variations du débit à plus long terme (ce qui est tout à fait possible, mais sort de notre sujet).

Jusqu'à présent, ce problème a été résolu en étudiant une à une les stations, et pour chacune d'elles en choisissant les durées d_i appropriées, ainsi que les critères $COEF1_i$ et $COEF2_i$ (cf. Figure 3-1). Cette manière de faire devient vite laborieuse pour peu que l'on veuille faire un échantillonnage multi-durées, sur de nombreux bassins, comme cela va être notre cas.

Il nous faut donc trouver un moyen de déterminer de façon **automatique** les critères *COEF1* et *COEF2*, ainsi que les durées d étudiées. Après plusieurs essais, la méthodologie résumée sur la Figure 3-17 a été adoptée :

Un premier échantillonnage « multi-durées » est effectué de manière standard, en étudiant les durées 0, 10, 20 et 30 jours, avec les critères suivants : $COEF1=1$ et $COEF2=0$ (aucune contrainte). Le nombre moyen de crues par an souhaité, $\mu_{objectif}$, est fixé à 2, et le restera tout au long des calculs. Sur ce premier essai (imparfait puisque l'on n'a ni utilisé de critères d'indépendance, ni choisi les durées d en fonction de la spécificité du bassin étudié) le modèle *QdF* convergent est ajusté, permettant de définir les paramètres $a(0)$, $x_0(0)$, P et surtout la durée Δ .

Cet échantillonnage est alors répété, mais cette fois-ci en indexant les durées d sur Δ ($d = 0 ; 0.125\Delta ; 0.25\Delta ; 0.5\Delta ; \Delta ; 1.5\Delta$) et en fixant le critère d'espacement *COEF2* à 2Δ . En effet, on suppose que, même si l'échantillonnage précédent n'est pas parfait, la durée Δ obtenue donne un ordre de grandeur sur la dynamique des crues observées sur le bassin (cf. §2.4.). Plusieurs itérations sont alors effectuées, chacune d'elles affinant l'estimation des paramètres $a(0)$, $x_0(0)$, Δ .

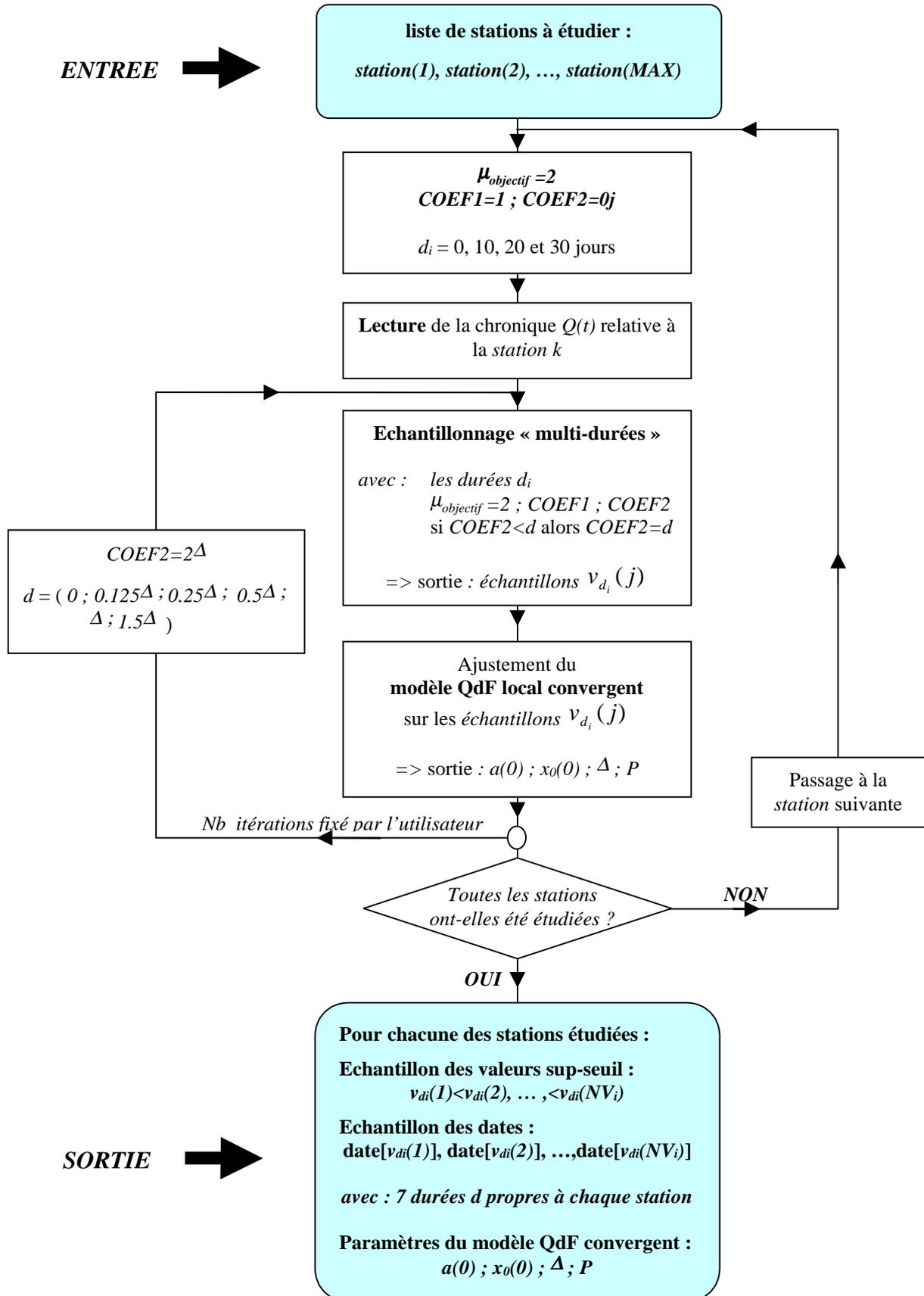


Figure 3-17 : Traitement automatique de l'ensemble des stations étudiées

Nous avons étudié, pour les trois exemples précédemment traités, la vitesse de convergence de ces paramètres. Trente itérations ont été effectuées, le paramètre P étant quant à lui fixé à 0. La Figure 3-18 présente l'écart relatif entre les paramètres $a(0)$, $x_0(0)$, Δ obtenus à la $k^{\text{ième}}$ itération et ceux obtenus à la $30^{\text{ième}}$. Pour ces exemples, dès la quatrième itération, l'écart relatif maximal est inférieur à 3%. Dans la suite des calculs, le nombre d'itérations nécessaires à donc été fixé à quatre.

Pour valider cette procédure d'échantillonnage automatique, nous avons comparé les résultats avec ceux obtenus pour les vingt bassins du chapitre 2. Dans les figures suivantes, les résultats de la procédure automatique sont notés « $COEF1=1$, $COEF2=2\Delta$ », en référence aux critères d'indépendance utilisés (Figure 3-17). De même, pour les résultats du chapitre 2, nous notons « $COEF1=1$, $COEF2=0$ » lorsqu'aucun critère n'est utilisé, et « $COEF1=manuel$, $COEF2=manuel$ » lorsque les critères ont été déterminés manuellement pour chaque bassin (chapitre 2, Tableau 2.1). Dans tous les cas, le nombre moyen de crues par an, est fixé à 2.

Les Figure 3-19 et Figure 3-20 donnent la synthèse des deux tests statistiques présentés au chapitre 2, à savoir celui sur l'indice de dispersion et celui sur la stationnarité du processus de Poisson. Les résultats présentés concernent les échantillons de débits instantanés ($d=0$). Pour ces deux tests, les résultats montrent que l'échantillonnage automatique ($COEF1=1$, $COEF2=2\Delta$) ne donne pas d'aussi bons résultats que l'échantillonnage manuel, mais reste néanmoins nettement préférable au cas n'utilisant pas de critère d'indépendance ($COEF1=1$, $COEF2=0$).

La Figure 3-21 indique l'erreur sur les quantiles $V_{d=0}(T)$ en prenant comme référence les quantiles obtenus à l'issue de l'échantillonnage « manuel ». Deux indicateurs sont calculés pour des périodes de retour comprises entre 0,5 et 20 ans : le biais (BIAIS) et l'erreur quadratique moyenne ($rRMSE$), définis au chapitre 2, formule 2-20. Tout comme au chapitre 2, Figure 2-21, on constate que l'erreur est la moins sensible pour des périodes de retour voisines de 3 ans, avec en-deçà un biais positif (surestimation des quantiles) et au-delà un biais négatif (sous-estimation des quantiles). L'erreur liée à l'échantillonnage automatique ($COEF1=1$, $COEF2=2\Delta$) est tout à fait acceptable, contrairement à ce qui avait été obtenu au chapitre 2 ($COEF1=1$, $COEF2=0$) : le biais et une erreur quadratique sont désormais réduits de moitié.

Ces résultats montrent que la procédure automatique résumée Figure 3-17 peut être considérée comme opérationnelle. Nous pouvons donc désormais l'appliquer à la centaine de bassins disponibles afin de valider le modèle QdF convergent.

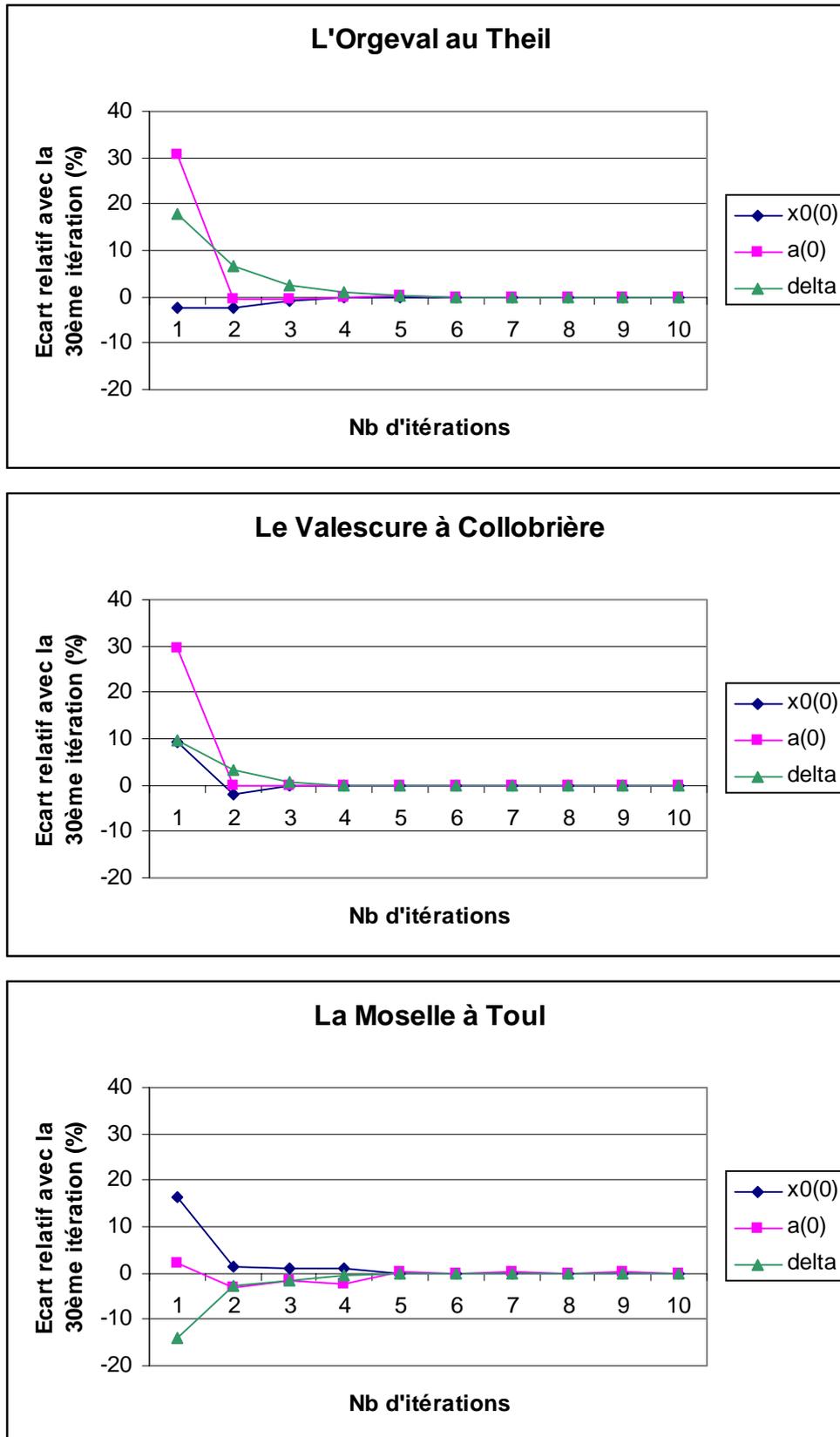


Figure 3-18 : Etude de la convergence des paramètres du modèle QdF convergent ajusté avec avec $P = \overline{QA}$

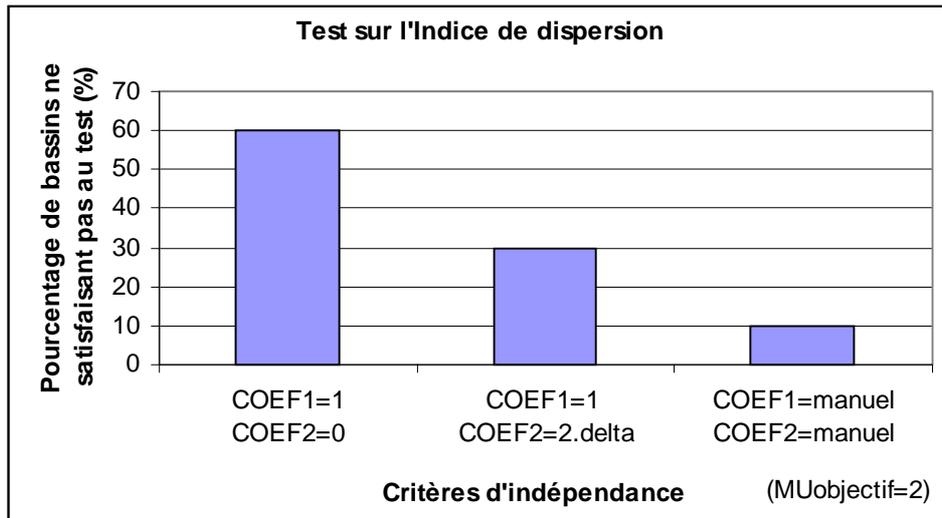


Figure 3-19 : Test sur l'indice de dispersion pour $d=0$

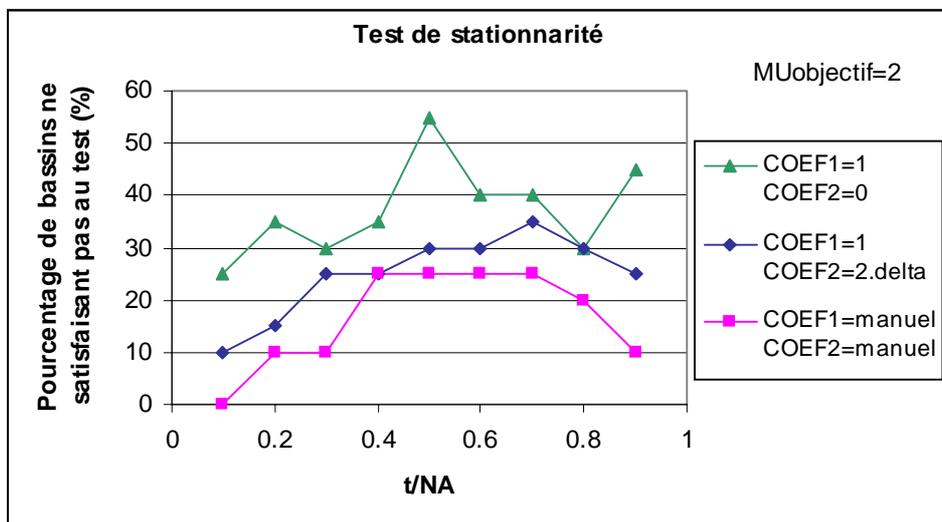


Figure 3-20 : Test de stationnarité pour $d=0$

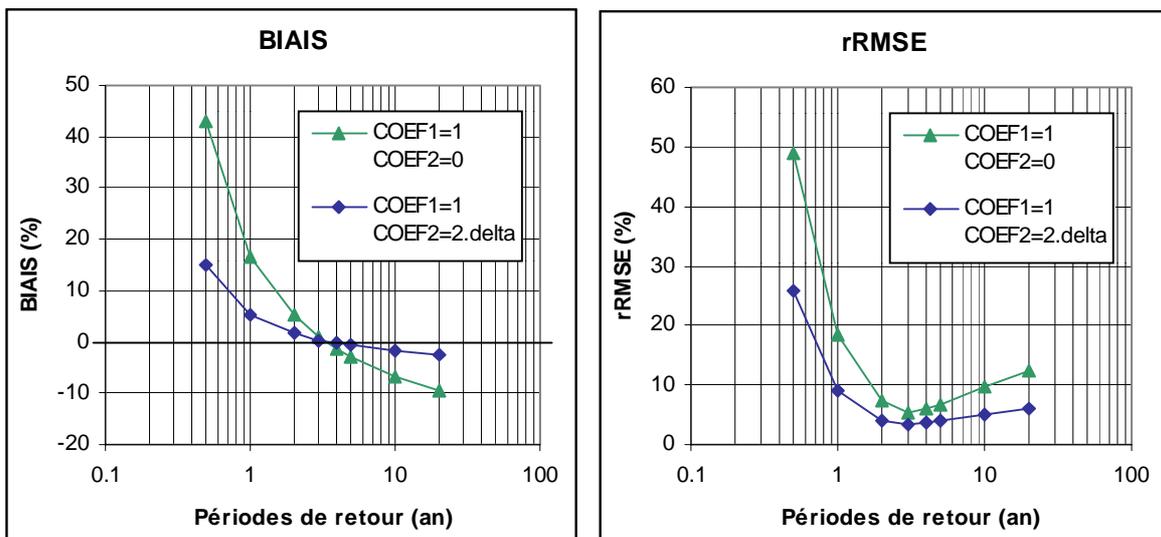


Figure 3-21 : Ecart avec les quantiles issus de l'échantillonnage manuel (pour $d=0$)

3.2. Validation du modèle QdF convergent

Pour chaque bassin, les débits moyens ont donc été échantillonnés pour six durées d différentes, permettant d'obtenir les quantiles $V_d(T)$ selon la procédure définie en Figure 3-17.

Rappelons que la totalité des chroniques étudiées a été critiquée visuellement : détections d'éventuelles erreurs sur le codage du débit ou bien de la date ; oubli d'un code lacune (voir chapitre 2, §3.1 pour plus de détails). Les chroniques présentant des erreurs flagrantes ne pouvant être corrigées manuellement ont par la suite été éliminées des traitements.

Cependant, malgré cela, sur les 107 chroniques traitées, la procédure d'échantillonnage automatique a été mise en défaut pour deux bassins. Dans le premier cas, la détermination de la durée Δ n'a pas convergé et a abouti à une valeur aberrante. Dans l'autre cas, le programme n'est pas parvenu à déterminer les seuils permettant de ne sélectionner que deux valeurs par an en moyenne. Ces deux cas particuliers aboutissent à un échantillonnage complètement erroné, ceci à cause d'un problème purement numérique. Par conséquent, nous avons choisi d'éliminer ces deux stations des traitements.

Finalement, le modèle convergent a été validé sur 105 bassins, dont la liste figure en annexe 2. De même qu'au §2.3. , nous avons étudié les deux cas suivants : P fixé à 0 et à P fixé à \overline{QA} , afin de voir l'incidence de ce choix. Les détails des calculs figurent en annexe 3 et 4. La Figure 3-22 compare de façon globale les quantiles non consolidés $V_d(T)$ (notés « observés » sur le graphique) et les quantiles $V(d,T)$ calculés par le modèle convergent avec $P=0$ et $P=\overline{QA}$, pour les périodes de retour de 0.5, 1 et 20 ans, et pour les six durées d étudiées.

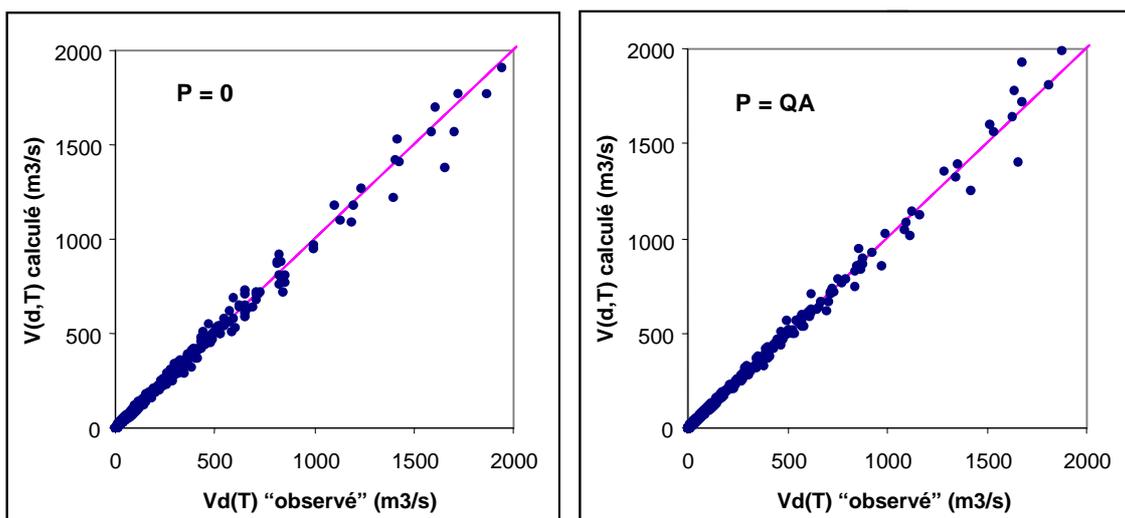


Figure 3-22 : Comparaison globale des quantiles « observés » $V_d(T)$ et des quantiles $V(d,T)$ calculés avec le modèle convergent ajusté pour $P=0$ et $P=\overline{QA}$

Ces résultats sont ensuite détaillés Figure 3-23 et Figure 3-24 où nous avons distingué chaque période de retour et chaque durée étudiée.

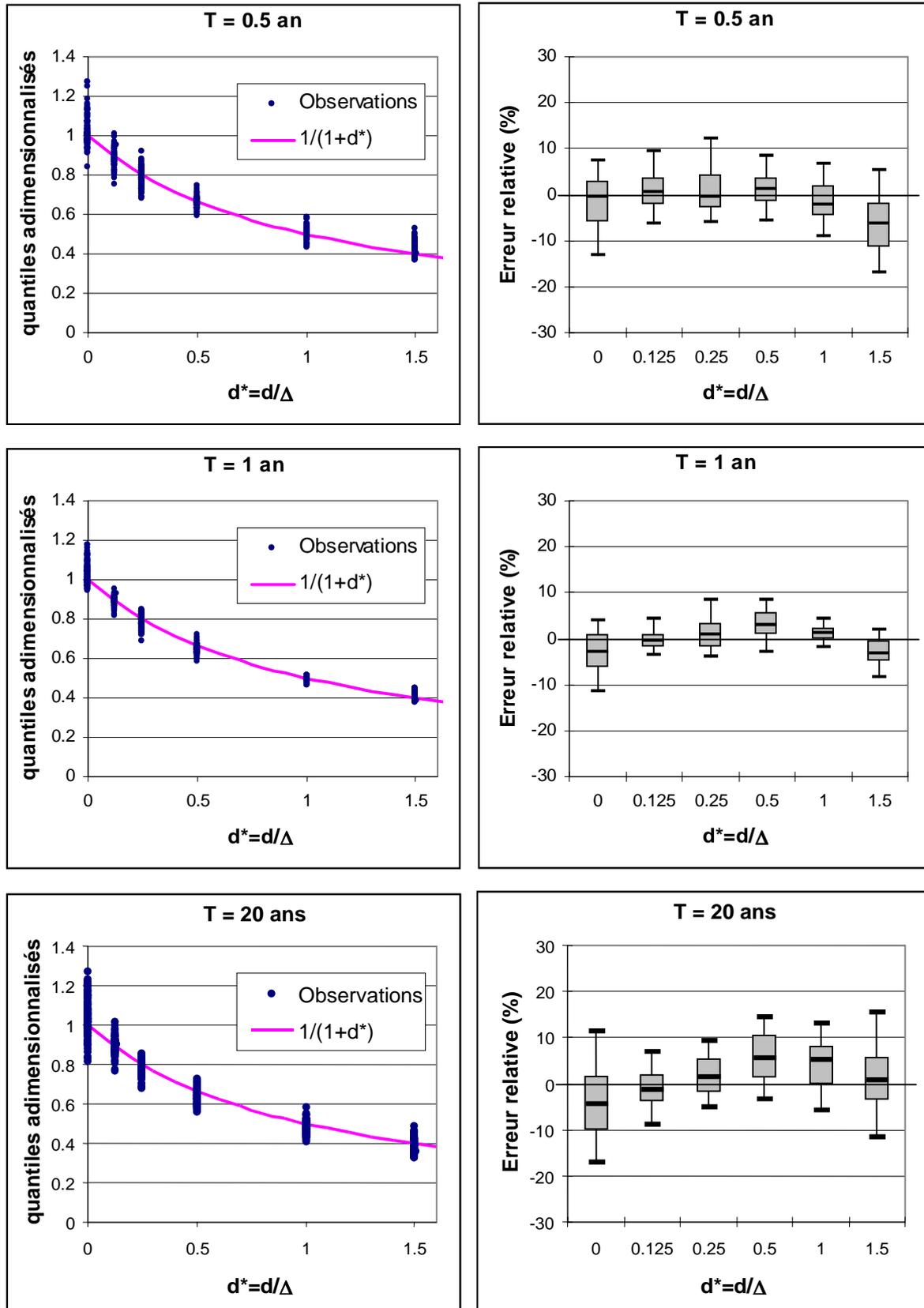


Figure 3-23 : Validation du modèle convergent ajusté avec $P=0$ *
 (le terme « observations » renvoie aux quantiles non consolidés $V_d(T)$)

* Les « boîtes » (plotting box) représentent l'erreur relative dépassée dans 95%, 75%, 50%, 25% et 5% des cas.

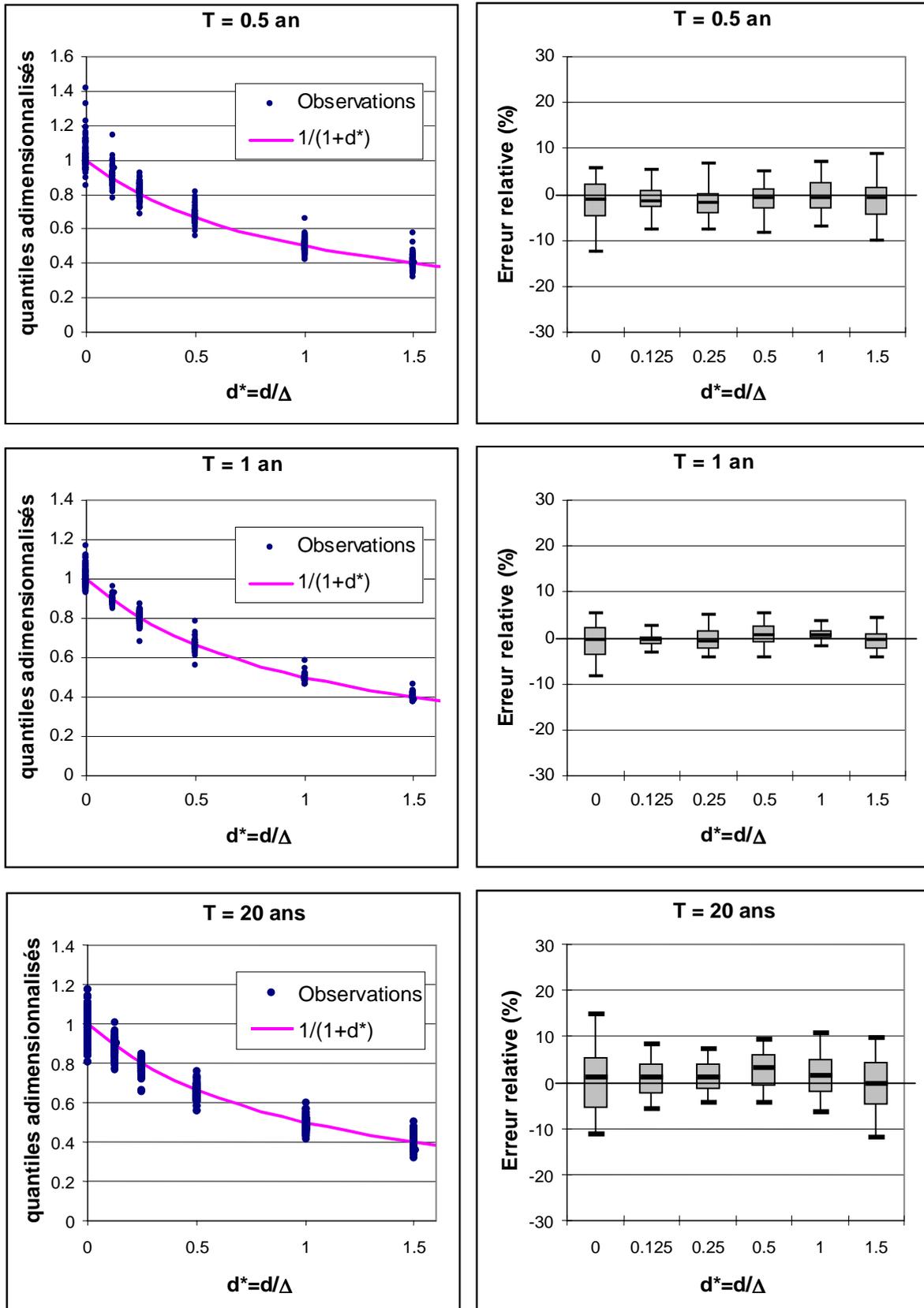


Figure 3-24 : Validation du modèle convergent ajusté avec $P = \overline{QA}$.*

(le terme « observations » renvoie aux quantiles non consolidés $V_d(T)$)

* Les « boîtes » (plotting box) représentent l'erreur relative dépassée dans 95%, 75%, 50%, 25% et 5% des cas.

Les graphiques de gauche de la Figure 3-23 et de la Figure 3-24, représentent les quantiles non-consolidés $V_d(T)$ adimensionnalisés selon la relation (3-37), en fonction des durées d normées par Δ . Comme nous l'avons vu en Figure 3-16, grâce à cette transformation, les quantiles calculés $V(d,T)$ se disposent, après transformation adimensionnelle, sur la courbe d'équation $y=1/(1+d^*)$. La différence avec la Figure 3-16 est qu'ici, les durées d ont été indexées sur Δ (Figure 3-17), ce qui explique la disposition verticale des points expérimentaux pour $d^*=0, 0.125, 0.25, 0.5, 1, 1.5$. Que ce soit en ajustant le modèle convergent avec $P=0$ (Figure 3-23) ou $P=\overline{QA}$ (Figure 3-24), on constate que la propriété d'invariance d'échelle est bien vérifiée, puisque les points expérimentaux sont groupés autour de la courbe théorique. Cette représentation permet donc de valider le choix de la forme hyperbolique pour décrire l'évolution des quantiles en fonction de d .

Les graphiques de droite représentent à l'aide de « boîtes » (plotting box) l'erreur relative entre les quantiles non-consolidés $V_d(T)$ et calculés $V(d,T)$. Chacune de ces boîtes synthétise les résultats pour une valeur de d^* (graphique de gauche) : elle indique l'erreur dépassée dans 95%, 75%, 50%, 25% et 5% des cas. Il est ainsi possible de connaître la dispersion des erreurs en fonction des durées étudiées, pour une période de retour fixée. Les résultats montrent que les erreurs observées sont faibles, puisque dans tous les cas, les bornes inférieures et supérieures des boîtes (quantiles 95% et 5%) sont toujours comprises entre $\pm 15\%$.

Afin d'avoir un critère de comparaison plus global, nous avons calculé l'erreur quadratique pour l'ensemble des 105 bassins et des six durées étudiés, pour différentes périodes de retour. Afin que les résultats puissent être comparables, il faut que les durées étudiées soient les mêmes selon que l'on fixe P à zéro ou à \overline{QA} . Or comme nous l'avons vu au Tableau 3-1, ce choix a une légère incidence sur le calcul de Δ , et donc sur les durées d étudiées, puisqu'indexées sur Δ . Pour effectuer la comparaison, nous avons donc choisi d'ajuster le modèle avec $P=0$ avec les mêmes durées d étudiées dans le cas où $P=\overline{QA}$.

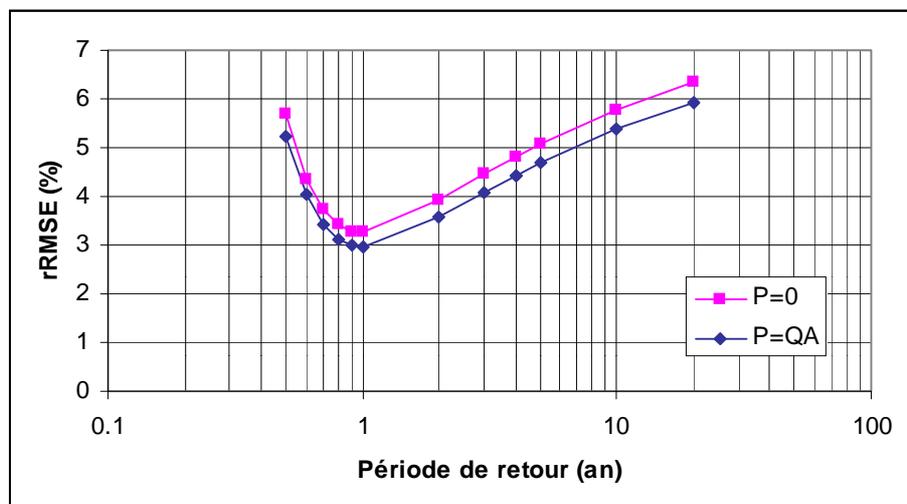


Figure 3-25 : Erreur quadratique du modèle QdF convergent, ajusté avec $P=0$ et $P=\overline{QA}$

Les résultats sont présentés Figure 3-25 et donnent un très léger avantage au modèle ajusté en fixant $P = \overline{QA}$. Dans les deux cas, l'erreur quadratique est minimale pour la période de retour de 1 an (3%), et maximale pour les périodes de retour de 0.5 et 20 ans. C'est pour cette raison que nous avons choisi de représenter ces périodes de retour à la Figure 3-23 et à la Figure 3-24.

Le modèle *QdF* convergent appliqué avec une valeur de P fixée (0 ou \overline{QA}) donne donc de bons résultats. Il n'a pas été jugé nécessaire de pousser plus en avant nos investigations, c'est à dire d'ajuster la valeur de P , au même titre que les autres paramètres du modèle. Compte tenu que les résultats obtenus en fixant P à 0 sont très proche de ceux obtenus avec $P = \overline{QA}$, nous préférons ajuster le modèle avec $P=0$, sa formulation s'en trouvant ainsi simplifiée.

3.3. Corrélation entre Δ et la durée de Socose

Le §2.4. a montré l'intérêt du paramètre Δ pour décrire la dynamique des crues. Un autre paramètre donne aussi cette information, il s'agit de la durée D définie à la méthode *Socose*, décrite au chapitre 1, (§3.2). De nombreux travaux du Cemagref utilisent cette durée, notamment l'approche *QdF* régionale dite « bassin de référence » développée par Galéa et Prudhomme (1997), dont nous parlerons au chapitre 7. Aussi, afin de montrer la cohérence de notre démarche avec des travaux antérieurs, il nous a semblé intéressant de comparer les durées Δ et D pour l'ensemble des 105 bassins étudiés. La Figure 3-26 montre une forte liaison entre ces deux durées. Sur cet exemple, Δ est issue du modèle convergent ajusté en fixant P à 0. Des résultats similaires sont obtenus en fixant P à \overline{QA} .

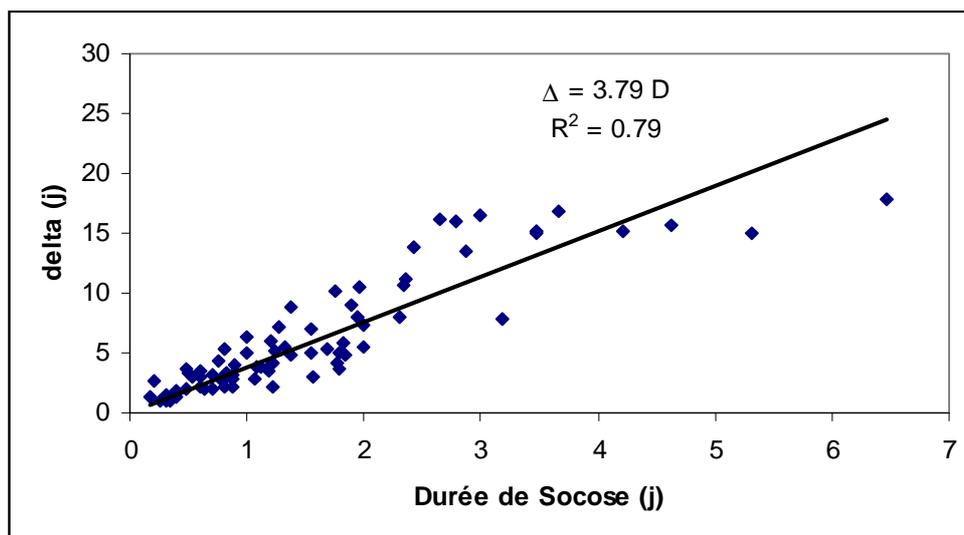


Figure 3-26 : Corrélation entre Δ (ajusté avec $P=0$) et D , la durée de Socose

Cette liaison entre D et Δ a aussi été observée sous des régimes climatiques très différents, comme en Martinique (Meunier, 1999, 2001) et en Guadeloupe (Galéa et Javelle, 2000). Dans

ces deux cas, on relève aussi un rapport Δ/D de l'ordre de 3, ce qui tend à montrer que cette liaison ne dépend pas de la région considérée.

Les durées Δ et D dépendent donc toutes les deux de la dynamique des crues observées. Cependant, la durée Δ a l'avantage de pouvoir être déterminée de façon précise et automatique, alors que D est déterminée manuellement, et doit plutôt être considérée comme un ordre de grandeur (cf. chapitre 1 §3.2).

3.4. Application : désagrégation du pas de temps

Ce paragraphe propose d'ajuster le modèle QdF convergent sur des distributions relatives à des durées d supérieures ou égales à 1 jour, puis d'extrapoler les résultats du modèle pour la durée $d=0$ (débit instantané). En ce sens, il s'agit d'une désagrégation du pas de temps : on part d'une information journalière pour estimer les débits de pointe instantanés (Javelle *et al.*, 1998).

Pour chaque bassin étudié précédemment, l'échantillonnage a donc été effectué sur des durées de 1, 2, ..., N jours, en veillant bien à ce que N ne dépasse pas la durée des crues du bassin. Puis le modèle QdF convergent a été ajusté (avec $P=0$), permettant d'obtenir les paramètres $x_0(0)$, $a(0)$ et Δ . Enfin, en posant $d=0$ dans l'équation du modèle, on obtient la distribution $V(0,T)$ relative aux débits instantanés. Afin d'estimer la validité des résultats ainsi trouvés, nous avons comparé la valeur du quantile décennal de débit de pointe, obtenu soit à partir du modèle convergent calé à partir de données journalières, soit à partir de données de débit instantané. On note :

$V(0,T=10)$: Quantile instantané décennal calculé à partir du modèle local convergent ajusté sur les distributions expérimentales des débits moyens maximaux, relatives aux durées 1, 2, ... N jours

$V_{d=0}(T=10)$: Quantile instantané décennal calculé à partir de l'ajustement sur la distribution expérimentale relative aux débits instantanés maximaux.

La Figure 3-27 compare les débits $V_{d=0}(T=10)$ et $V(0,T=10)$, ainsi que l'erreur relative obtenue, en fonction de la durée Δ .

Remarque : les données « journalières » utilisées sont en fait des chroniques de débits instantanés moyennées sur un jour à l'aide d'une fenêtre mobile (cf. chap. 1, Figure 1-20). Il ne s'agit donc pas de chronique de débit journalier à pas de temps fixe. Dans ce cas, une correction doit être effectuée, équivalente à celle du coefficient de Weisse pour les pluies.

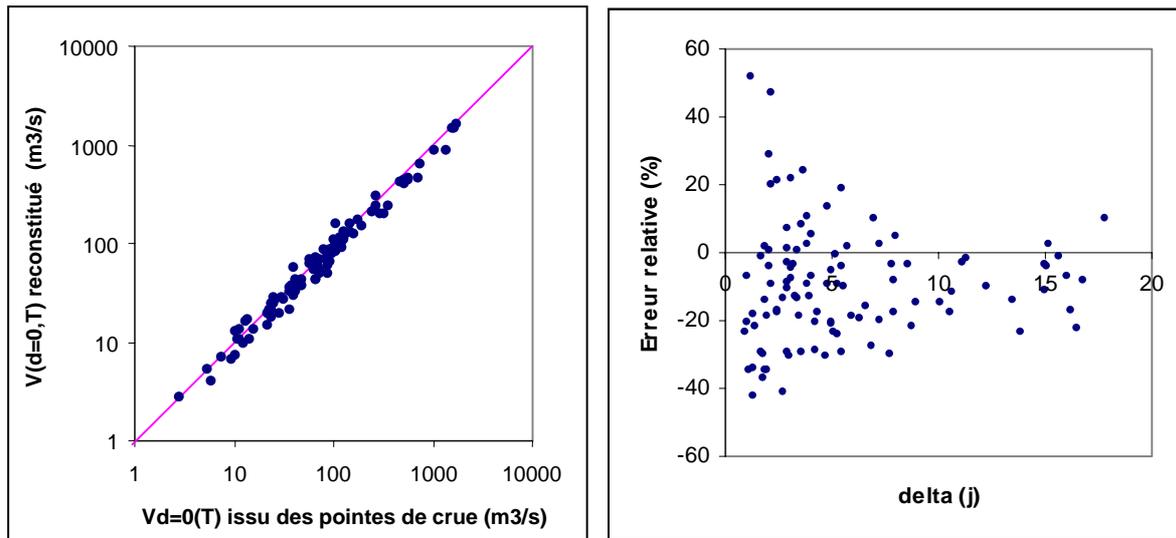


Figure 3-27 : Comparaison du débit de pointe instantané décennal obtenu à partir des pointes de crues, avec celui calculé à partir de durées supérieures ou égales à 1 jour, pour la $T=10$ ans.

Comme le montrent les résultats, l'erreur sur l'estimation des quantiles instantanés dépend de la durée Δ , et donc de la dynamique des bassins. Pour des bassins rapides (Δ inférieure à 5 jours), l'estimation du débit décennal peut comporter de fortes erreurs relatives (jusqu'à 40 à 50%). Par contre pour les bassins lents, l'estimation s'améliore. L'explication est simple : la perte d'information causée en moyennant les débits sur un jour est d'autant plus importante que ces débits ont une grande variabilité dans le temps. Lorsque les crues sont lentes, la différence entre le débit de pointe et le débit journalier est faible.

4. Tentative d'interprétation des propriétés

Les paragraphes précédents ont donc présenté puis validé le modèle QdF convergent. Ses bons résultats nous amènent à nous interroger sur la signification des propriétés du modèle. D'un point de vue purement théorique, la propriété d'invariance d'échelle illustrée Figure 3-16 peut faire penser à la théorie fractale. En ce sens, un projet du PNRH (Hubert *et al.*, 2000) a comparé la formulation du modèle QdF convergent avec les modèles multi-fractals utilisés par exemple pour décrire les courbes intensité-durée-fréquence des précipitations (Bendjoudi *et al.*, 1997). Il en ressort que la formulation du modèle, en son état actuel, n'est pas compatible avec la théorie multi-fractale. Dans la suite de ce paragraphe, nous ne rentrerons pas dans ces considérations théoriques, et chercherons simplement à donner une interprétation « géométrique » des propriétés du modèle.

4.1. Hypothèse explicative : l'affinité des hydrogrammes de crue

De façon très schématique, la propriété de convergence des distributions des débits moyens $V_d(T)$ (Figure 3-5) peut trouver une explication dans la forme des crues observées. L'hypothèse avancée est la suivante : les différents hydrogrammes de crue se déduisent les

uns des autres par **affinité orthogonale** par rapport à la droite d'équation $y=P$, comme c'est le cas pour les distributions des débits moyens $V_d(T)$ (Figure 3-5). Cette affinité s'exprime de la façon suivante (Figure 3-28):

$$Q_3(t) - P = k_2[Q_2(t) - P] = k_1[Q_1(t) - P] \quad (3-38)$$

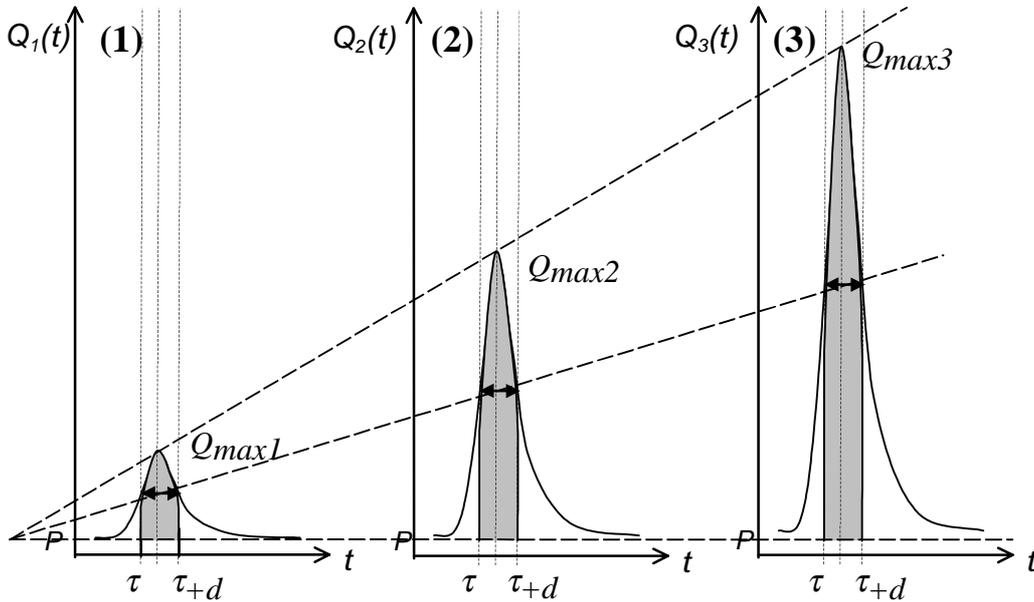


Figure 3-28 : Affinité verticale des hydrogrammes de crue

La variable Vd étant définie sur une crue par le volume maximal écoulé pendant d , divisé par

d , c'est à dire : $Vd = \frac{1}{d} \int_{\tau}^{\tau+d} Q(t)dt$ (cf. chap. 1, §3.5) on déduit de (3-38) que :

$$[Vd]_3 - P = k_2([Vd]_2 - P) = k_1([Vd]_1 - P) \quad (3-39)$$

En écrivant cette égalité successivement pour deux durées d_1 et d_2 , puis en divisant la première par la seconde, on obtient :

$$\frac{[Vd_1]_3 - P}{[Vd_2]_3 - P} = \frac{[Vd_1]_2 - P}{[Vd_2]_2 - P} = \frac{[Vd_1]_1 - P}{[Vd_2]_1 - P} \quad (3-40)$$

Cette égalité signifie que si les hydrogrammes (translatés de $-P$) se déduisent les uns des autres, à un simple rapport d'affinité près, alors le rapport $(Vd_1 - P)/(Vd_2 - P)$ ne dépend pas de l'hydrogramme considéré. De plus, si l'on accepte cette représentation simplifiée de la réalité, pour une même période de retour les valeurs échantillonnées Vd sont toutes issues du même événement. Ce point autorise alors à faire le parallèle entre l'égalité (3-40) définie pour des événements et l'égalité (3-6) définie pour des quantiles.

Cette ordonnée P prendrait alors une certaine signification hydrologique. Considérant le schéma de la Figure 3-28, cette valeur peut être interprétée comme celle d'un débit de base « moyen » durant les hautes eaux, au-dessus duquel toutes les crues sont proportionnelles les

unes aux autres. Dans les résultats précédents, nous avons considéré les deux cas suivants : $P=0$ et $P=\overline{QA}$, le module interannuel. Dans le premier cas, il s'agirait alors d'une estimation « faible » du débit de base, le deuxième d'une estimation un peu trop « forte ». Mais dans les deux cas, la valeur réelle de ce débit de base est très inférieure aux débits de crue observés.

4.2. Vérification de l'hypothèse sur l'exemple de l'Orgeval au Theil

Nous avons choisi de vérifier l'hypothèse ci-dessus sur le BVRE de l'Orgeval au Theil. Ce bassin a été choisi car ses courbes débit-durée-fréquence vérifient très bien les propriétés étudiées, à savoir la convergence des distributions des débits moyens $V_d(T)$. (Figure 3-4).

Si l'on étudie la forme des six plus fortes crues, on remarque une certaine affinité. En effet, la Figure 3-29 représente ces crues centrées sur le débit de pointe (graphique du haut) ainsi que leurs hydrogrammes translattés de \overline{QA} puis normés par $Q_{max}-\overline{QA}$ (graphique du bas).

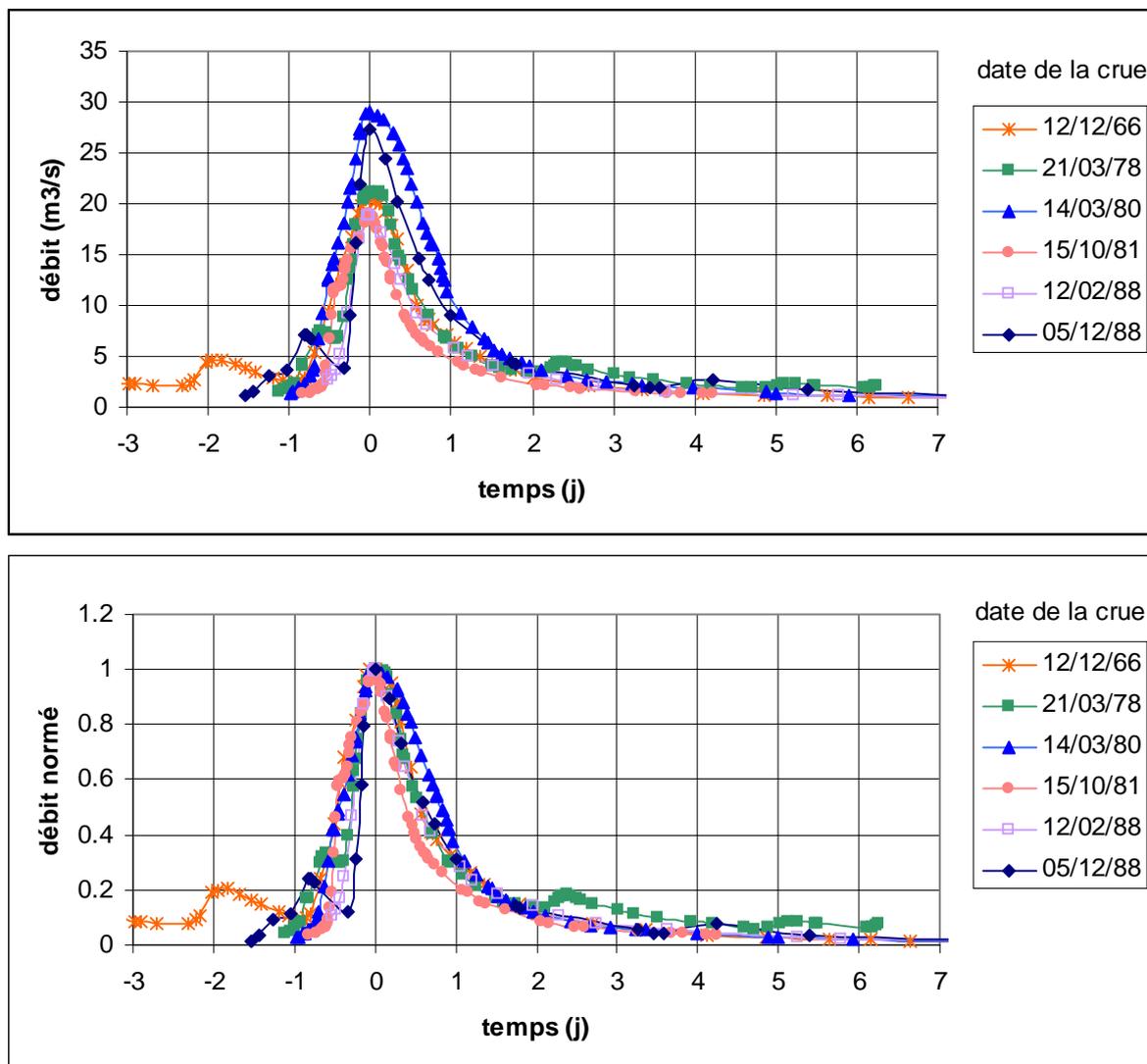


Figure 3-29 : Hydrogrammes des 6 plus fortes crues enregistrées, centrées sur le débit de pointe (haut), et normées par le débit de pointe (bas), Exemple de l'Orgeval au Theil

On constate un étalement relativement faible de ces hydrogrammes normés, ce qui vérifie assez correctement l'affinité supposée. Toutefois, cette affinité n'est vérifiée que pour les plus forts événements (les six plus forts en terme de débits de pointe, pour l'exemple ci-dessus).

Cette affinité des hydrogrammes, observée sur les plus fortes crues, est en accord avec la théorie de l'hydrogramme unitaire (Sherman, 1932). L'hydrogramme unitaire est la réponse du bassin à une pluie unitaire. La réponse du bassin à toute une averse est alors calculée par la convolution de l'hydrogramme unitaire. Ainsi, comme le montre la Figure 3-30, deux crues générées par deux pluies de même durée sont alors proportionnelles.

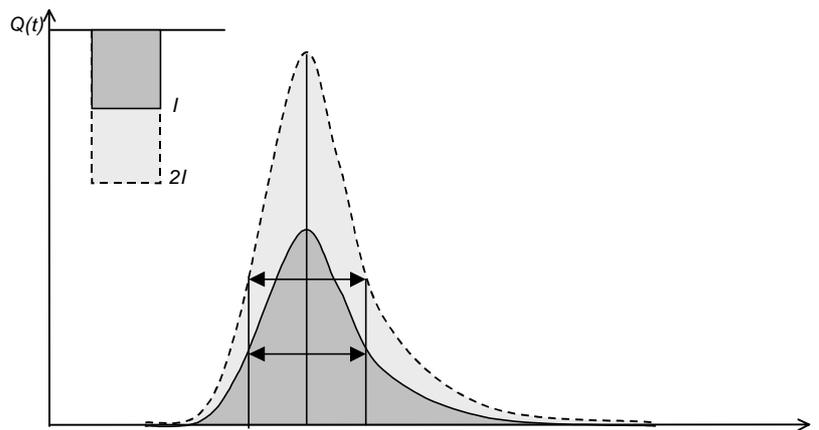


Figure 3-30 : Réponses à deux pluies de même durée, d'après la théorie de l'hydrogramme unitaire

L'affinité des plus fortes crues, Figure 3-29, peut donc s'expliquer par la théorie de l'hydrogramme unitaire, en supposant que les crues sont générées par des pluies de formes et de durées équivalentes.

Comme le prouvent les résultats du modèle QdF convergent sur la centaine de bassins étudiés, ces hypothèses paraissent cohérentes. Cependant, nous sommes bien conscient qu'il s'agit d'hypothèses très simplificatrices, la réalité étant bien sûr beaucoup plus compliquée. Nous reviendrons sur ce point au chapitre 4, dans le paragraphe 2.4.1.

Conclusion du chapitre 3

Ce chapitre est entré dans le détail de l'approche débit-durée-fréquence et a présenté une modélisation des courbes obtenues : le modèle convergent. Tirant parti de propriétés observées, ce modèle fait deux hypothèses : d'une part, pour différentes durées fixées, les distributions des débits en fonction de la période de retour convergent vers un même point, et d'autre part, pour une période de retour fixée, les courbes débit-durée-fréquence sont données par une forme hyperbolique en fonction de d . La formulation fait intervenir un nouveau paramètre, de la dimension d'un temps : la durée Δ . Cette durée ajustée définit la concavité des hyperboles. Elle témoigne de la plus ou moins grande rapidité de l'ensemble des crues observées sur un même bassin.

Afin de valider ce modèle QdF convergent sur une centaine de bassins français, une procédure automatique d'échantillonnage « multi-durée » a été mise au point. Cette procédure donne des résultats comparables à l'échantillonnage rigoureux défini au chapitre 2. L'intérêt de cette nouvelle méthode est un gain de temps et une absence de subjectivité dans le choix des paramètres d'échantillonnage, notamment pour les critères d'indépendance.

L'extension du modèle convergent à un grand nombre de bassins montre que les hypothèses du modèle sont pratiquement toujours bien vérifiées. Deux cas ont été testés pour le choix de P , l'ordonnée du point de convergence : $P=0$ et pour $P=\overline{QA}$, le module interannuel. Les résultats obtenus sont équivalents, ce qui nous fait préférer la formulation avec $P=0$, plus simple.

On avance une explication quant à la convergence observée des distributions. Celle-ci peut être due à une affinité des hydrogrammes de crue que prévoit la théorie de l'hydrogramme unitaire, pour des pluies de durées unitaires. Cette explication simplifie grandement la réalité. Cependant, force est de constater les bons résultats du modèle convergent, sur la centaine de bassins testés.

Chapitre 4

Autres développements du modèle *QdF* local convergent

Le précédent chapitre a présenté le modèle débit-durée-fréquence (*QdF*) que nous avons appelé « local convergent », puisqu'il est ajusté localement sur les données observées, et que parmi ses hypothèses, il y a celle de la convergence du faisceau des distributions vers les faibles périodes de retour. Ce modèle a été appliqué dans le cas de l'étude des débits moyens maximaux, et a été développé pour des distributions décrites par la loi exponentielle.

L'objet de ce chapitre est de montrer des développements possibles de ce modèle, en gardant comme deux points communs : l'ajustement local du modèle, et la convergence des distributions en fonction de T .

Les développements ont été effectués dans trois directions :

- application à d'autres lois statistiques que la loi exponentielle ;
- application à d'autres variables de crue que les débits moyens : « les débits seuils » ;
- application à d'autres événements que les crues : les étiages

Les paragraphes suivants présentent chacun de ces développements.

1. Utilisation d'une loi statistique à trois paramètres sur de longues séries

Ce paragraphe montre que le modèle convergent présenté au chapitre précédent pour la loi exponentielle (à deux paramètres) peut être généralisé à d'autres types de lois, notamment à trois paramètres. De façon générale, l'ajustement d'une loi à trois paramètres est moins robuste que celui d'une loi à deux paramètres, car plus sensible aux valeurs extrêmes de l'échantillon. L'utilisation de ce type de loi est donc justifiée dans les deux cas suivants :

- lorsqu'on a une longue chronique de débit (avec éventuellement des données historiques),
- lorsqu'on a une courte chronique, mais que l'on peut utiliser de l'information régionale pour consolider l'ajustement local.

Ce paragraphe concerne le cas de longues chroniques, nous reviendrons en détail sur l'analyse régionale au chapitre 5.

1.1. Présentation de la loi *GEV* (Generalised Extreme Value)

La loi *GEV* (Generalised Extreme Value) a été choisie car elle est très utilisée pour décrire la distribution des débits maximaux annuels (Bobée, 1993). Cette loi exprime la probabilité au non dépassement par :

$$F(x) = \exp \left[- \left(1 - k \frac{x - x_0}{a} \right)^{1/k} \right] \quad ; \quad k \neq 0 \text{ et } a > 0 \quad (4-1)$$

En utilisant la relation $F=1-1/T$, on peut exprimer les quantiles en fonction de la période de retour T :

$$x(T) = x_0 + \frac{a}{k} \left[1 - (-\ln(1-1/T))^k \right] \quad (4-2)$$

avec : $x(T)$: quantile de période de retour T ; x_0 : paramètre de position ; a : paramètre d'échelle ; et k : paramètre de forme

La loi *GEV* est définie pour des valeurs de k différentes de 0. Lorsque k tend vers 0, la loi *GEV* tend vers la loi de Gumbel. Pour cette raison, on représente souvent les distributions dans un papier de Gumbel, c'est à dire en fonction de la variable réduite u :

$$u = -\ln(-\ln(1-1/T)) \quad (4-3)$$

Ajustement par la méthode des moments pondérés (Hosking et al., 1985) :

De manière générale, les moments pondérés théoriques introduits par Greenwood *et al.* (1979), s'écrivent :

$$M(p, r, s) = E \left[X^p F^r (1-F)^s \right] = \int_0^1 x(F)^p F^r (1-F)^s dF \quad (4-4)$$

Dans la pratique, on fixe $p=1$ et $s=0$, les moments pondérés sont alors plus simplement définis par :

$$M(1,r,0) = \beta_r = \int_0^1 x(F)F^r dF \quad (4-5)$$

Les moments pondérés empiriques sont quant à eux estimés par la relation suivante, pour un échantillon de NV valeurs classées par ordre croissant, $x(1) < \dots < x(NV)$:

$$b_r = \frac{1}{NV} \sum_{j=1}^{NV} \frac{(j-1)!(NV-r-1)!}{(j-r-1)!(NV-1)!} x(j) \quad (4-6)$$

La méthode des moments pondérés consiste à résoudre $\beta_r = b_r$ pour $r=1,2,3$ où β_r est fonction des paramètres x_0 , a et k de la loi, et b_r est calculé sur l'échantillon. Hosking *et al.*, (1985) proposent la solution approchée suivante, pour le calcul des estimations de x_0 , a et k :

$$\hat{k} = 7.8590c + 2.9554c^2 \quad (4-7)$$

$$\hat{a} = \frac{(2b_1 - b_0)\hat{k}}{\Gamma(1 + \hat{k})(1 - 2^{-\hat{k}})} \quad (4-8)$$

$$\hat{x}_0 = b_0 + \hat{a} \frac{\Gamma(1 + \hat{k}) - 1}{\hat{k}} \quad (4-9)$$

avec : $c = \frac{2b_1 - b_0}{3b_2 - b_0} - \frac{\ln 2}{\ln 3}$ et : Γ : la fonction gamma.

La méthode des moments pondérés est souvent recommandée, car elle est réputée robuste, c'est à dire peu sensible aux valeurs extrêmes (*outlier* ou horsain). Cependant, cet avantage peut devenir un inconvénient dans le cas d'une valeur élevée observée et validée, puisqu'on accorde un faible poids à cette information importante (Bobée, 1999).

Nous avons traité deux bassins, la Loire à Montjean (115 000 km²) et le Danube à Orsova (576 000 km²), pour lesquels 128 et 152 années sont respectivement disponibles. Ces données ont été traitées dans le cadre d'un projet du Programme National de Recherche en Hydrologie (Hubert *et al.*, 2000). Même si le chapitre 2 prend position pour l'échantillonnage par valeurs supérieures à un seuil, nous avons ici utilisé la méthode des valeurs maximales annuelles. En effet, pour ces deux bassins, du fait de leur très grande taille, on n'observe qu'une seule crue par an. Dans ce cas, la méthode "sup-seuil" n'est pas adaptée. D'autre part, rappelons que la loi *GEV* ajustée ici sur des valeurs maximales annuelles correspond à la loi de Pareto Généralisée ajustée sur des valeurs supérieures à un seuil (*cf.* annexe 1). Pour ces deux bassins, la chronique de débit journalier a été transformée par une moyenne mobile de durée d , puis la valeur maximale de chaque année hydrologique a été retenue. Pour les deux cas, nous avons fait démarrer l'année hydrologique en septembre, les crues ayant lieu majoritairement autour du mois de mars. La plus petite durée étudiée est $d=1$ jour, puisque les chroniques traitées sont des chroniques de débits moyens journaliers. La Figure 4-1 montre le résultat de l'ajustement de la loi *GEV* sur chacun des échantillons obtenus.

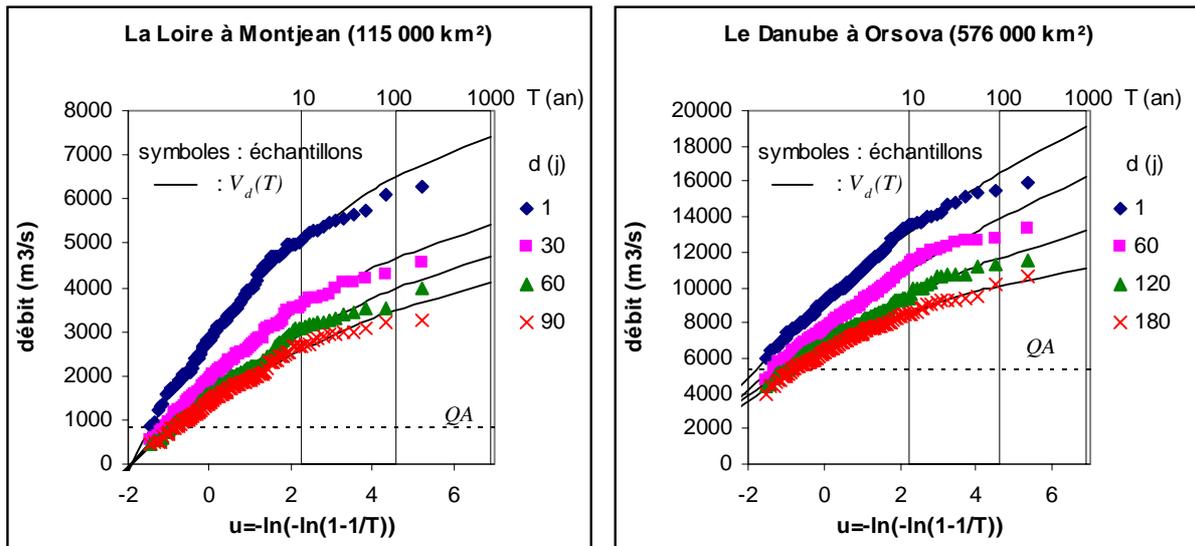


Figure 4-1 : Ajustement d'une loi GEV par la méthode des moments pondérés sur chacune des distributions expérimentales des débits moyens maximaux

1.2. Généralisation des hypothèses du modèle convergent

Au chapitre précédent nous avons fait l'hypothèse que les distributions convergeaient vers un même point. Pour la loi exponentielle, les distributions sont représentées par une droite lorsque l'axe de la période de retour est en échelle logarithmique. Ceci nous a permis de déduire des propriétés d'invariance, indépendantes de la période de retour.

Dans le cas général, toutes les lois statistiques ne peuvent pas être représentées par une droite lorsqu'on les représente sur une échelle appropriée ; aussi nous exprimerons de façon plus générale la propriété de convergence des distributions :

« **Toutes les distributions se déduisent les unes des autres par une affinité orthogonale ayant pour axe la droite d'équation $V(d,T)=P$.** »

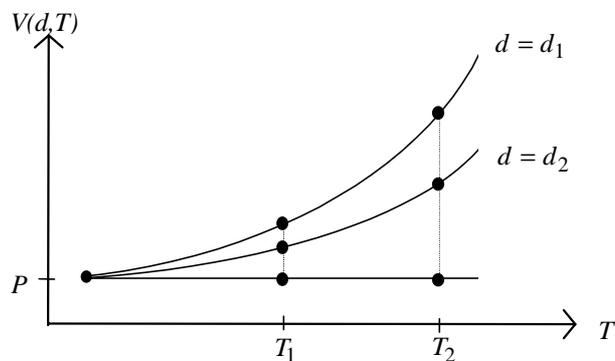


Figure 4-2 : Généralisation de l'hypothèse de convergence à tout type de loi : affinité des distributions

La Figure 4-2 illustre cette propriété pour une loi statistique quelconque. Dans ces conditions, la propriété d'invariance avec la période de retour exprimée au chapitre 3 reste valable :

$$\frac{V(d,T) - P}{V(d=0,T) - P} = f(d) \quad (4-10)$$

Et si l'on suppose que pour une période de retour fixée, les quantiles s'expriment en fonction de d selon une forme hyperbolique, la formulation générale du modèle est toujours :

$$V(d,T) = \frac{V(d=0,T) - P}{1 + d/\Delta} + P \quad (4-11)$$

avec : $V(d=0,T)$, la distribution relative aux débits instantanés ;

Δ : paramètre de forme des hyperboles, de la dimension d'un temps ;

P : ordonnée du point de convergence des distributions (qui peut, comme nous l'avons vu, simplement être fixé à 0).

De même, la méthode d'ajustement utilisée pour déterminer la durée Δ est toujours valide (chap. 3, §2.2.3). Celle-ci repose sur le fait que chaque distribution relative à une durée d peut être « reliée » à la distribution des débits instantanés ($d=0$) par l'équation :

$$V(d=0,T) = (1 + d/\Delta) \cdot (V(d,T) - P) + P \quad (4-12)$$

Les deux paragraphes suivants illustrent sur quelques exemples l'ajustement du modèle pour différents types de loi, à partir des propriétés énoncées ci-dessus.

1.3. Application au modèle local convergent

Si l'on intègre la loi GEV dans le modèle local convergent, celui-ci est alors donné par :

$$\left. \begin{aligned} V(d,T) &= \frac{V(d=0,T) - P}{1 + d/\Delta} + P \\ \text{avec :} \\ V(d=0,T) &= x_0(0) + \frac{a(0)}{k(0)} \left\{ 1 - [-\ln(1 - 1/T)]^{k(0)} \right\} \end{aligned} \right\} \quad (4-13)$$

P : l'ordonnée du point de convergence, en m^3/s

Δ : la durée caractéristique de crue, en jours

$x_0(0)$, $a(0)$, $k(0)$: paramètres de position, d'échelle et de forme de la distribution relative à $d=0$

Pour les deux cas traités Figure 4-1, l'ordonnée P du point de convergence a été fixée à 0. En effet, comme le montrent ces exemples, l'ordonnée du point de convergence est ici clairement inférieure au module interannuel noté \overline{QA} (en pointillé sur la figure). Ces deux cas sont particuliers. Il s'agit en effet de très grands bassins, avec de très lentes crues. Et comme le

montre la Figure 4-1, il peut arriver que certaines années, le débit journalier maximum annuel soit du même ordre que \overline{QA} .

P étant fixé, il reste maintenant à ajuster les paramètres Δ , $x_0(0)$, $a(0)$ et $k(0)$. La même méthode d'ajustement décrite au chapitre 2 §2.2.3 est utilisée. On effectue d'abord l'ajustement de Δ , puis l'ajustement des paramètres de la distribution $V(d=0, T)$.

Le modèle devant vérifier la relation (4-12), Δ est déterminé de façon à ce que toutes les distributions expérimentales « équivalentes » à l'instantané se confondent (cf. chapitre 3, §2.2.3). La Figure 4-3 montre les distributions expérimentales transformées selon (4-12) après que Δ ait été optimisé. Le regroupement des valeurs est moins bon pour les fréquences rares. Mais ceci peut s'expliquer par l'incertitude liée à l'échantillonnage qui augmente avec la période de retour.

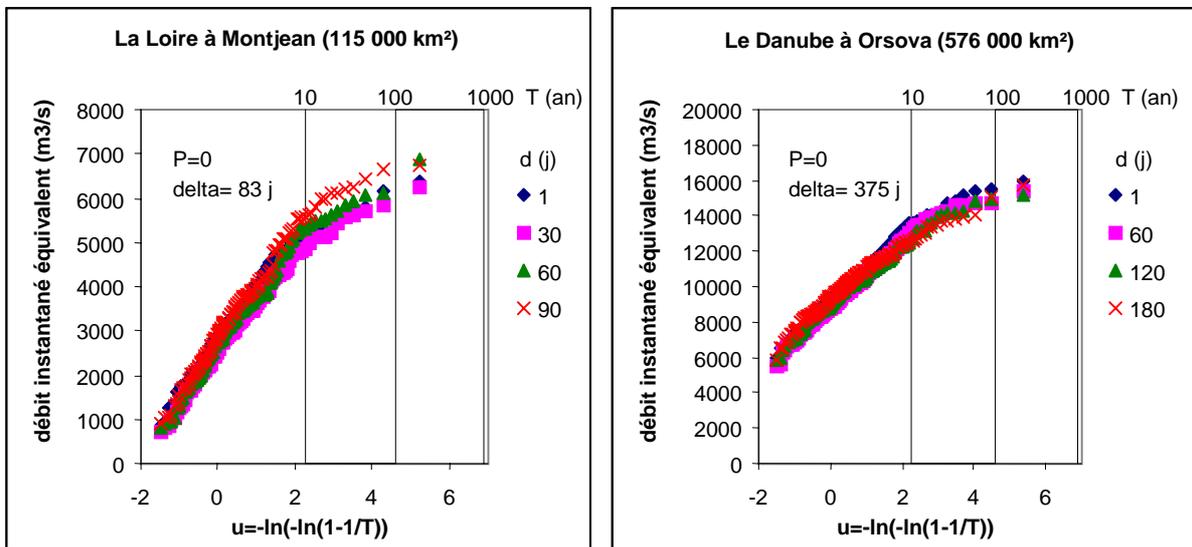


Figure 4-3 : Ajustement de la durée Δ , P étant fixé à 0, en minimisant l'écart entre les distributions instantanées « équivalentes » selon (4-12)

Les paramètres de la distribution $V(d=0, T)$ du modèle local convergent sont ensuite ajustés sur l'échantillon constitué de l'ensemble des débits instantanés « équivalents ». Pour cela, une distribution expérimentale « moyenne » est calculée en prenant pour chaque fréquence empirique la moyenne des valeurs (cf. chapitre 3, §2.2.3). En utilisant la relation (4-13) il est maintenant possible d'obtenir les distributions $V(d, T)$ pour toute durée d . La Figure 4-4 compare ces distributions avec les distributions $V_d(T)$ ajustées sur chaque échantillon, Figure 4-1.

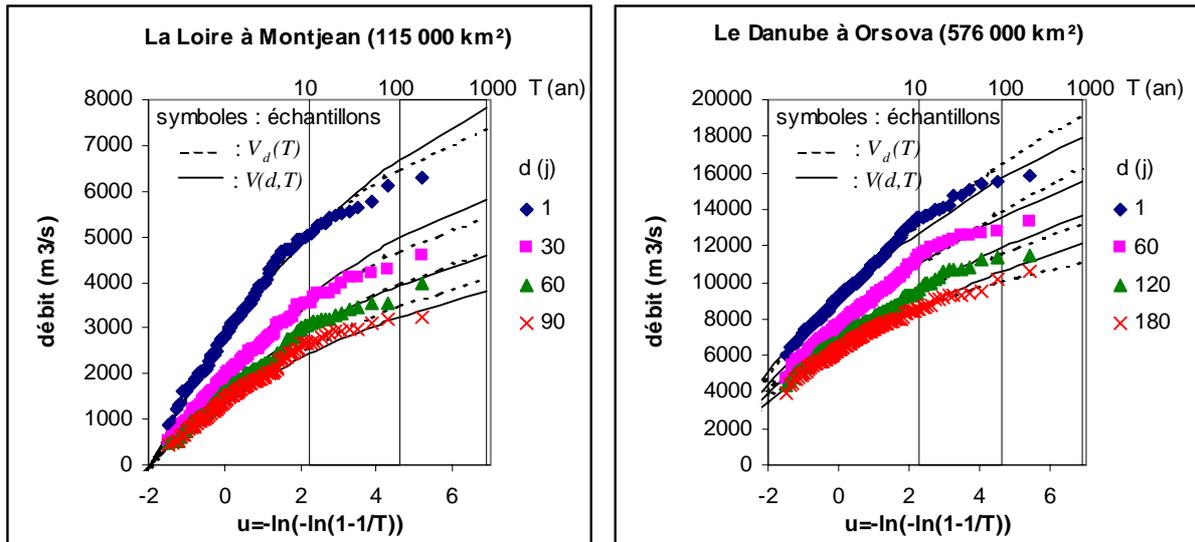


Figure 4-4 : Comparaison des quantiles $V_d(T)$ non consolidés et des quantiles $V(d,T)$ calculés par le modèle local convergent

1.4. Conclusion partielle

Au travers l'exemple de la loi GEV , il apparaît que le modèle présenté au chapitre 2 avec la loi exponentielle peut être généralisé à tout type de loi. En effet, la méthode d'ajustement retenue consiste à convertir les distributions expérimentales des débits moyens sur la durée d en un seul échantillon de débits instantanés « équivalents », puis à ajuster le paramètre Δ de façon à ce que ce dernier échantillon soit le plus homogène possible. Ce n'est qu'ensuite que le choix de la loi est effectué pour déterminer la distribution théorique $V(d=0,T)$. On se ramène ainsi au problème classique de l'ajustement d'une loi statistique sur un échantillon : il est donc possible de choisir la loi statistique, mais aussi la méthode d'ajustement (maximum de vraisemblance, moments, moment pondérés...). Bien entendu, les résultats présentés demandent à être vérifiés sur un plus grand nombre de bassins. C'est ce que propose le chapitre 5, dans lequel la loi GEV est utilisée pour une centaine de bassins situés au Québec et en Ontario, cette fois-ci non plus dans le cadre d'une analyse **locale**, mais **régionale**.

2. Etudes des « débits seuils » maximaux

2.1. Définition des débits seuils

Jusqu'à présent, ce travail, n'a concerné que l'étude des débits moyens sur une durée d . Oberlin (1992) définit une deuxième variable descriptive des crues, fonction d'une durée d fixée. Il s'agit du débit continûment dépassé sur la durée d , maximal lors de la crue, noté $QCXd$. De même que pour les débits moyens $VCXd$ que nous avons noté Vd , nous appellerons Qd la variable qu'il appelle $QCXd$, ceci afin de ne pas alourdir les notations (Figure 4-5).

L'intérêt des débits seuil est qu'elle permet de mieux formuler une demande de protection que ne le fait le débit de pointe, ou les débits moyens Vd . Par exemple :

*« Je ne veux pas être inondé **plus de 2 jours consécutifs**, plus d'une fois
tous les dix ans en moyenne »*

Le quantile de débit seuil correspondant à cette demande est $Q_{d=2j}(10ans)$.

Cette variable est aussi utilisée par la méthode Inondabilité pour construire des crues de projet, appelées Hydrogramme Synthétiques Mono-Fréquence (Gilard, 1998).

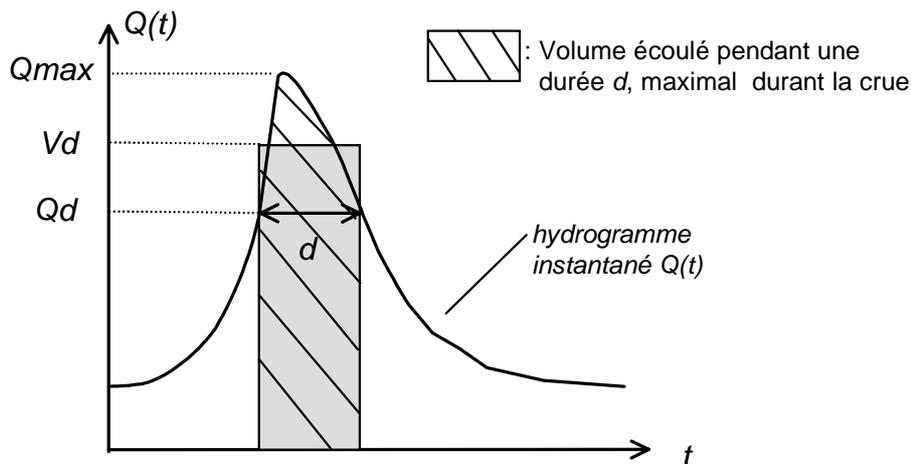


Figure 4-5 : Définition du débit seuil Qd

Signalons deux points communs entre les débits moyens Vd et les débits seuils Qd :

- ces deux variables tendent toutes les deux vers le débit de pointe instantané $Qmax$ lorsque d tend vers zéro
- sur des crues simples (c'est à dire avec une seule pointe), il y a concomitance entre les débits moyens Vd et les débits seuils Qd , comme cela est représenté sur la Figure 4-5.

Pour se convaincre du second point ci-dessus, la Figure 4-6 montre l'évolution du volume $S(t)$ écoulé entre t et $t+d$. lorsque t évolue. Ce volume peut être exprimé par la somme des volumes $S_1(t)$ et $S_2(t)$, respectivement situés au-dessus et au-dessous du débit seuil Qd . Graphiquement, on note que $S_1(t)$ et $S_2(t)$ sont tous deux maximaux pour $t=t_{Qd}$. Par conséquent, le volume $S(t)$ est lui aussi maximal pour $t=t_{Qd}$.

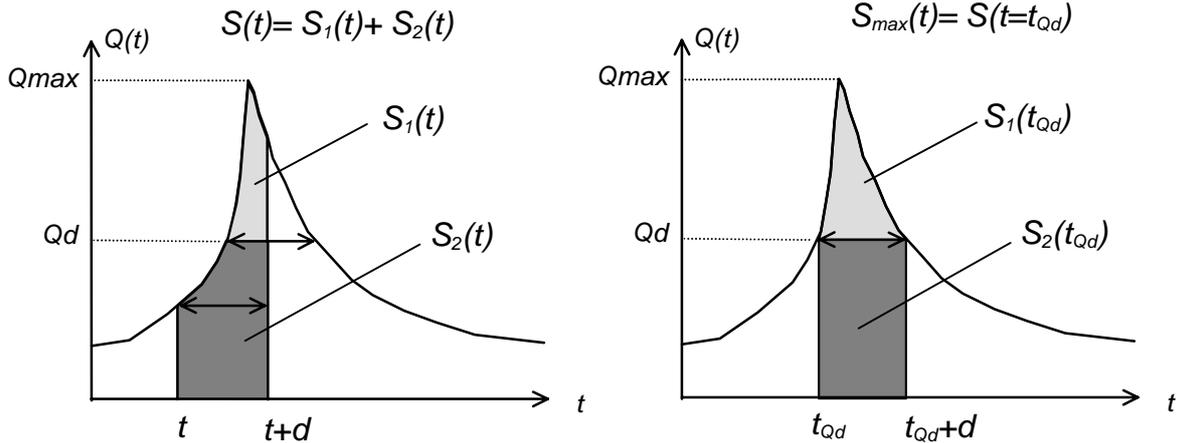


Figure 4-6 : Détermination du volume maximal compris entre t et $t+d$

Remarque :

Cette propriété n'est vraie que sur le cas simple illustré par la Figure 4-6. En particulier, on doit avoir : $Q(t) \leq Qd$ pour $t < t_{Qd}$ et pour $t > t_{Qd}+d$. Ceci exclut les crues avec plusieurs pointes.

De la même façon que pour les débits moyens $V_d(T)$, on échantillonne les débits seuils Qd à partir d'une chronique de débit instantané. Cette fois-ci l'échantillonnage ne se fait plus à l'aide d'une moyenne mobile, comme pour les débits moyens, mais en transformant la chronique à l'aide d'un « minimum mobile » : une fenêtre de largeur d est déplacée sur toute la chronique, et à chaque instant, on relève le débit minimum contenu dans la fenêtre. Cette procédure est détaillée dans Lang (1995). Enfin, grâce à l'ajustement d'une loi statistique, on obtient les quantiles notés $Q_d(T)$. Dans notre cas, comme pour les débits moyens $V_d(T)$, nous avons procédé à un échantillonnage par valeurs supérieures à un seuil, et ajusté une loi exponentielle sur la distribution des débits obtenue. La Figure 4-7 montre les courbes QdF associées aux débits moyens $V_d(T)$, et aux débits seuils $Q_d(T)$. Dans les deux cas, la loi statistique ajustée sur les échantillons est la loi exponentielle.

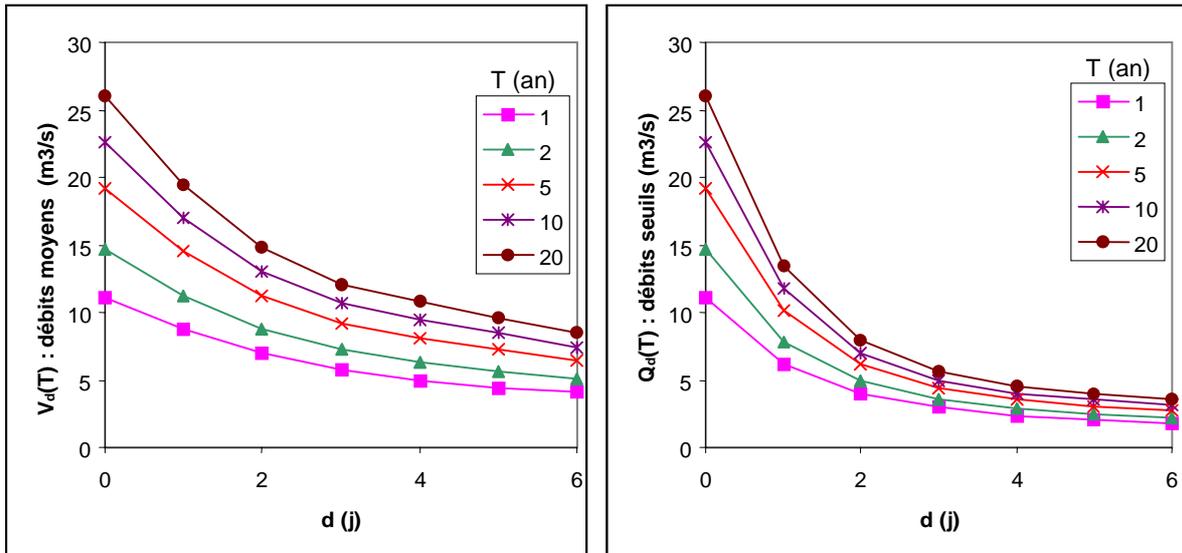


Figure 4-7 : Courbes QdF associées aux débits moyens $V_d(T)$, et aux débits seuils $Q_d(T)$
Exemple de l'Orgeval au Theil (104 km²)

2.2. Propriété expérimentale observée entre les courbes QdF en débits seuils et en débits moyens

Comme nous venons de le voir, les définitions du débit moyen et du débit seuil, bien que différentes, possèdent une certaine analogie. L'objet de ce paragraphe est de vérifier une propriété expérimentale remarquée par Prudhomme (1995), reliant les courbes débit-durée-fréquence associées aux débits moyens et celles associées aux débits seuils. Cette propriété est illustrée Figure 4-8, nous la formulons de la façon suivante :

Le volume maximal écoulé pendant la durée d avec une période de retour donnée, c'est à dire $V_d(T) \times d$, est égal au volume calculé entre 0 et d sous la courbe $Q_d(T)$ des débits seuils de même période de retour, et ceci, quelle que soit la durée d .

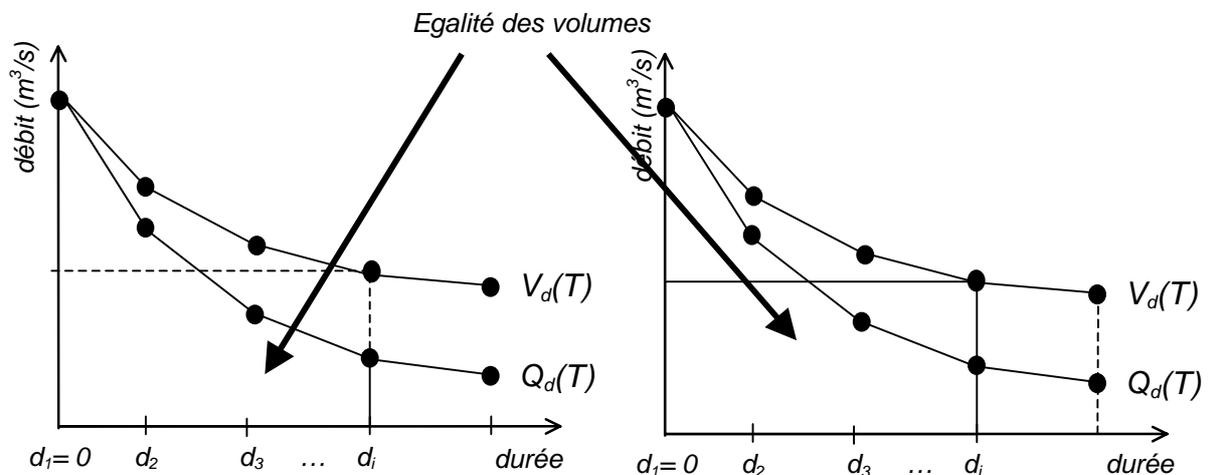


Figure 4-8 : Relation expérimentale entre les courbes QdF en débits moyens $V_d(T)$ et en débits seuils $Q_d(T)$ de même période de retour

Afin de vérifier cette propriété sur un grand nombre de bassins, nous avons repris les résultats obtenus au chapitre 3, concernant l'étude des débits moyens $V_d(T)$ pour la centaine de bassins étudiés. La même analyse fréquentielle a été effectuée, cette fois-ci pour les débits seuils $Q_d(T)$. Ainsi, pour chaque bassin, nous disposons des valeurs des débits moyens $V_d(T)$ et des débits seuils $Q_d(T)$ calculées pour six durées. Rappelons que ces durées ne doivent pas dépasser la durée observée des crues. Elles sont donc propres à chaque bassin étudié. Pour chacune de ces 6 durées, et pour différentes périodes de retour T fixées, nous avons défini le débit $Q_{moy_d}(T)$:

$$Q_{moy_{d_n}}(T) = \frac{1}{d_n} \sum_{i=2}^n \left(\frac{Q_{d_i}(T) + Q_{d_{i-1}}(T)}{2} (d_i - d_{i-1}) \right) \quad (4-14)$$

$Q_{moy_d}(T)$ représente la somme de chaque élément de surface contenu sous la courbe $Q_d(T)$ en fonction de d , pour T fixé, le tout divisé par d . De cette façon, si la relation expérimentale décrite ci-dessus est vérifiée, on doit avoir égalité entre $Q_{moy_d}(T)$ et $V_d(T)$.

La Figure 4-9 présente les résultats obtenus pour l'Orgeval au Theil, déjà étudié précédemment. Des périodes de retour de 1, 5, 10 et 20 ans ont été considérées, ainsi que des durées de 0 (débits instantanés), 1, 2, 3, 4, 5 et 6 jours.

On constate effectivement sur cet exemple pour chacune des périodes de retour et des durées considérées une bonne adéquation entre les débits moyens $V_d(T)$ et les débits $Q_{moy_d}(T)$ calculés à partir des débits seuils par la relation (4-14). Afin d'avoir une synthèse des résultats obtenus pour les 105 bassins étudiés, nous avons calculé pour chacun d'eux l'écart relatif entre $V_d(T)$ et $Q_{moy_d}(T)$ obtenu pour les 7 durées étudiées et pour 4 périodes de retour (1, 5, 10 et 20 ans). La Figure 4-10 présente ces résultats en fonction de d/Δ afin d'avoir une idée de l'évolution de l'erreur relative ε en fonction de la durée d considérée.

$$\varepsilon = \frac{Q_{moy_d}(T) - V_d(T)}{V_d(T)} \quad (4-15)$$

Les résultats montrent que l'écart relatif entre $V_d(T)$ et $Q_{moy_d}(T)$ est tout à fait acceptable, puisque 90% des erreurs observées sont comprises entre $\pm 10\%$, sauf pour $T=0,5$ ans et $d^*=1.5$, où la borne inférieure de la boîte (erreur dépassée dans 95% des cas) est de -15% , ce qui reste correct. La relation expérimentale entre les débits moyens $V_d(T)$ et les débits seuils $Q_d(T)$, énoncé au début de ce paragraphe, est donc confirmée.

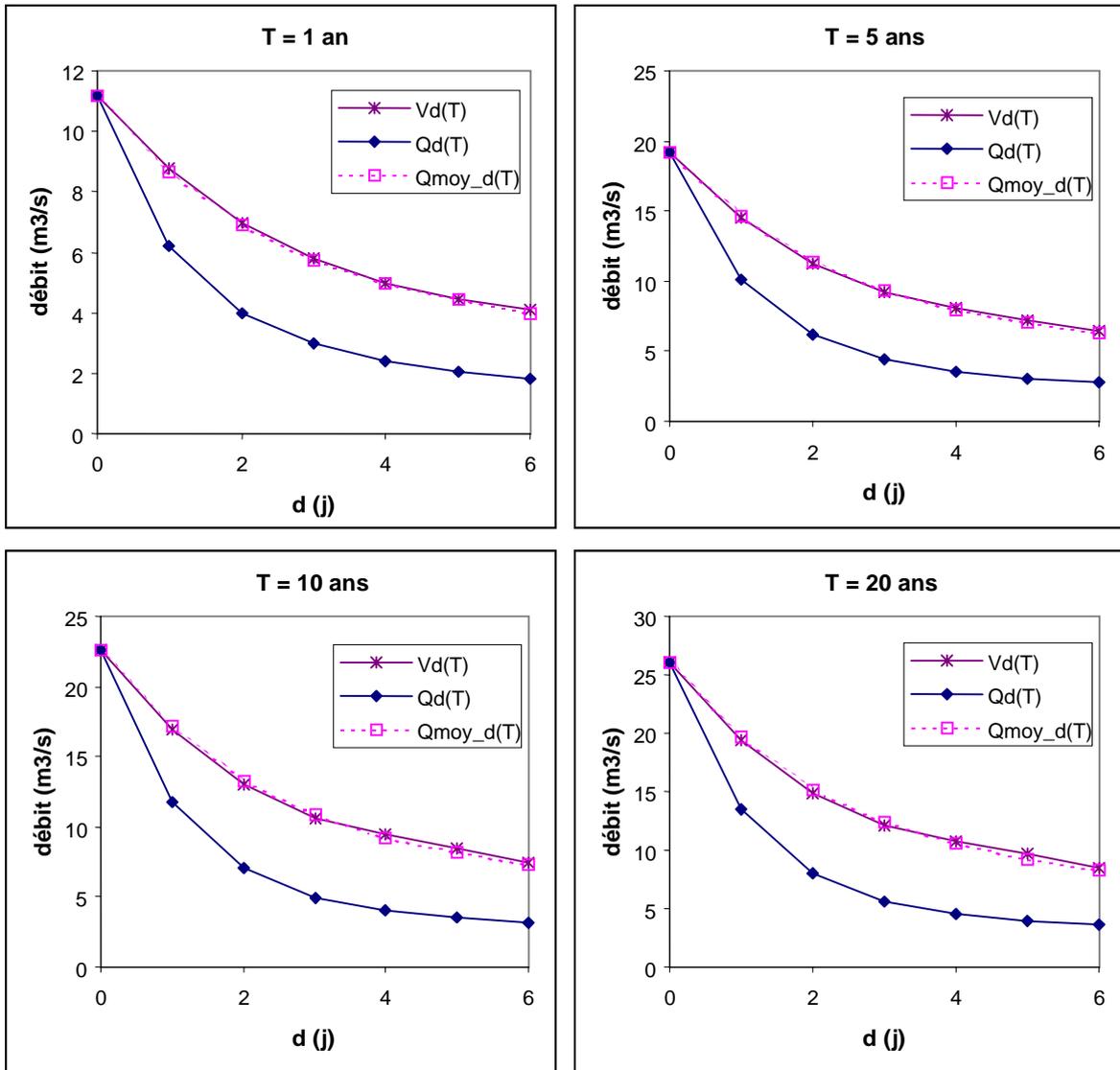


Figure 4-9 : Comparaison des quantiles $V_d(T)$ et $Q_{moy_d}(T)$, l'intégrale divisée par d de $Q_d(T)$ entre 0 et d . Exemple de l'Orgeval au Theil (104 km²)

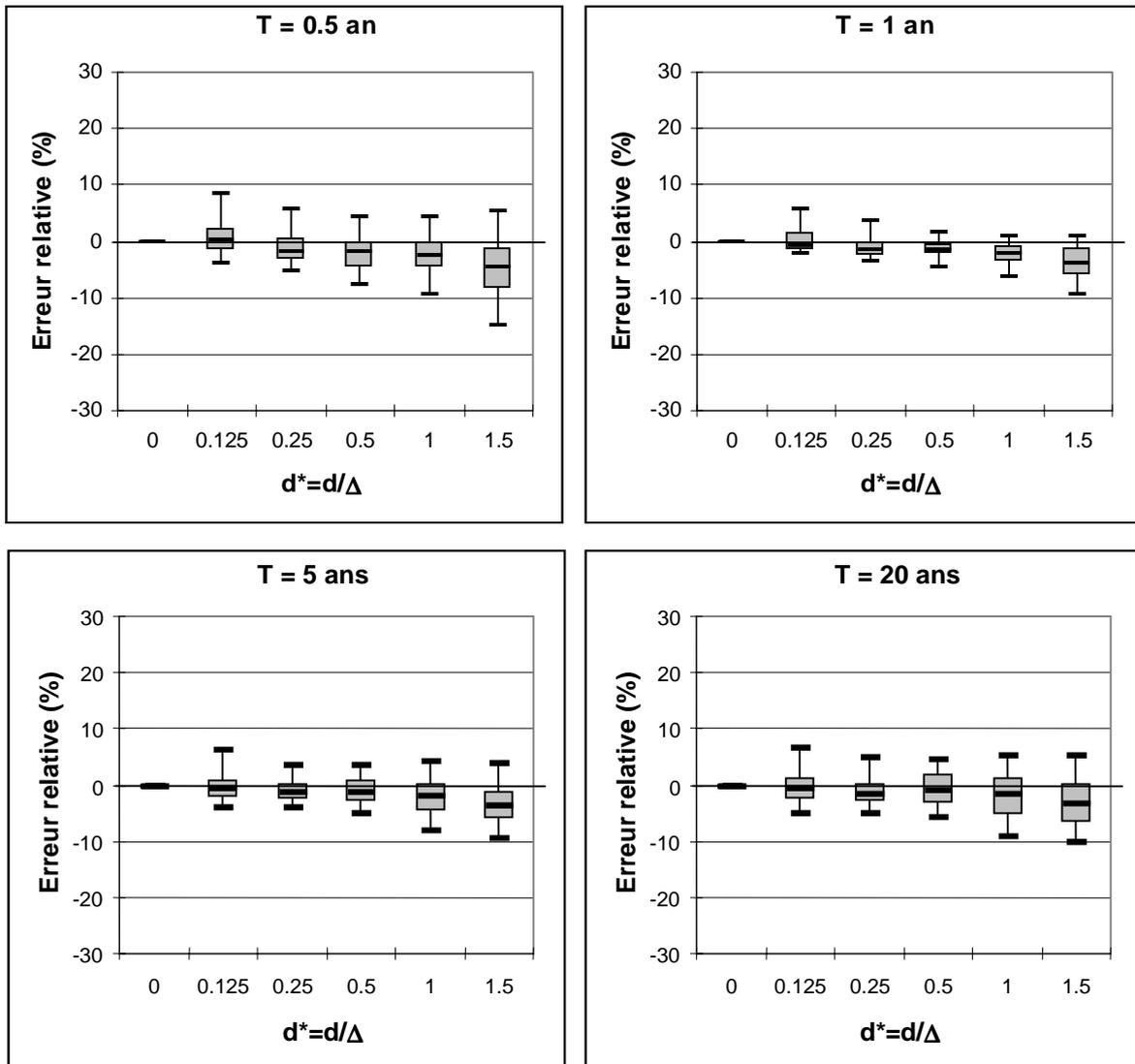


Figure 4-10 : Erreur relative entre $Q_{moy_d}(T)$ et $V_d(T)$ en fonction de d/Δ *

* Les boîtes (plotting box) représentent l'erreur relative dépassée dans 95%, 75%, 50%, 25% et 5% des cas.

2.3. Application au modèle QdF local convergent

2.3.1. Formulation de $Q(d,T)$ déduite de celle de $V(d,T)$

Comme nous l'avons vu, le modèle local convergent développé au chapitre 3 permet de passer des débits moyens $V_d(T)$, à une description continue en d et T : les débits moyens $V(d,T)$. Une démarche tout à fait analogue et indépendante pourrait être conduite afin de passer des débits seuils $Q_d(T)$ aux débits $Q(d,T)$. Plutôt que de refaire ce qui a déjà été fait, mais simplement sur une autre variable, nous avons choisi d'intégrer dans le modèle relatif aux débits moyens la propriété expérimentale vérifiée au précédant paragraphe. L'intérêt est l'obtention d'un modèle décrivant à la fois les débits moyens $V(d,T)$ et les débits seuils $Q(d,T)$.

Nous allons donc faire l'hypothèse que la propriété vérifiée expérimentalement au paragraphe 2.2. est vraie. Cette hypothèse se traduit par la relation suivante :

$$V(d,T) = \frac{1}{d} \int_0^d Q(\delta,T) d\delta \quad (4-16)$$

En remplaçant $V(d,T)$ par la formulation donnée par le modèle local convergent (4-11), on obtient :

$$\frac{1}{d} \int_0^d Q(\delta,T) d\delta = \frac{V(0,T) - P}{1 + d/\Delta} + P \quad (4-17)$$

$$\frac{1}{d} \left(-dP + \int_0^d Q(\delta,T) d\delta \right) = \frac{V(0,T) - P}{1 + d/\Delta} \quad (4-18)$$

$$\frac{1}{d} \int_0^d (Q(\delta,T) - P) d\delta = \frac{V(0,T) - P}{1 + d/\Delta} \quad (4-19)$$

$$\frac{1}{d} \int_0^d \frac{Q(\delta,T) - P}{V(0,T) - P} d\delta = \frac{1}{1 + d/\Delta} \quad (4-20)$$

on pose :

$$Q^*(d,T) = \frac{Q(d,T) - P}{V(0,T) - P} \quad (4-21)$$

L'équation devient :

$$\int_0^d Q^*(\delta,T) d\delta = \frac{d}{1 + d/\Delta} \quad (4-22)$$

En dérivant par rapport à d , on obtient :

$$Q^*(d,T) = \frac{1(1 + d/\Delta) - d(1/\Delta)}{(1 + d/\Delta)^2} \quad (4-23)$$

Après simplifications, on a :

$$Q^*(d,T) = \frac{1}{(1 + d/\Delta)^2} \quad (4-24)$$

Et finalement, on obtient :

$$Q(d,T) = \frac{V(0,T) - P}{(1 + d/\Delta)^2} + P \quad (4-25)$$

Cette équation signifie que tout comme les quantiles de débits moyens $V(d,T)$, chaque distribution $Q(d,T)$ peut se déduire de la distribution $V(0,T)$ des débits instantanés, connaissant P , l'ordonnée du point de convergence des distributions et Δ , la durée caractéristique de crue définie par le modèle local convergent (cf. chapitre 3).

Considérant les équations (4-11) et (4-25) on obtient finalement une relation entre les quantiles $V(d,T)$ et $Q(d,T)$:

$$Q(d,T) = \frac{V(d,T) - P}{1 + d/\Delta} + P \quad (4-26)$$

2.3.2. Application : estimation des débits seuils à partir des débits moyens

L'application testée est la suivante. Nous avons cherché à voir dans quelle mesure, les quantiles $V(d,T)$ ajustés au chapitre 3 pouvaient permettre de calculer les débits seuils grâce à la relation (4-26). Nous avons donc repris les résultats du chapitre 3, concernant le calcul de $V(d,T)$, Δ et P . Les deux cas testés au chapitre 3 ($P=0$ et $P=\overline{QA}$) donnant des résultats équivalents, nous avons considéré le cas où P était fixé à zéro, pour simplifier les calculs. La relation (4-26) nous permet donc de calculer les débits seuil « modélisés » $Q(d,T)$ que nous comparons aux quantiles $Q_d(T)$ non consolidés obtenus dans les paragraphes précédents.

La Figure 4-11 présente les résultats obtenus pour les deux bassins déjà étudiés : le Valescure à Collobrière et l'Orgeval au Theil. La Figure 4-12 présente quant à elle une synthèse de ces résultats sur l'ensemble des 105 bassins étudiés.

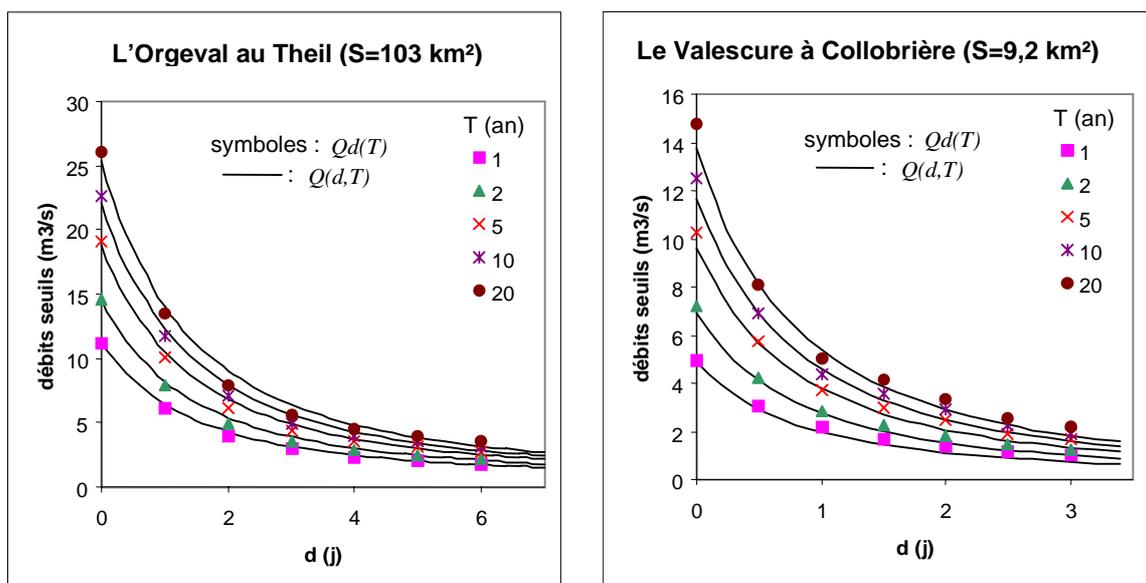


Figure 4-11 : Comparaison entre les débits seuil observés $Q_d(T)$ et les débits seuil $Q(d,T)$ déduits des débits moyens $V(d,T)$ par la relation (4-26)

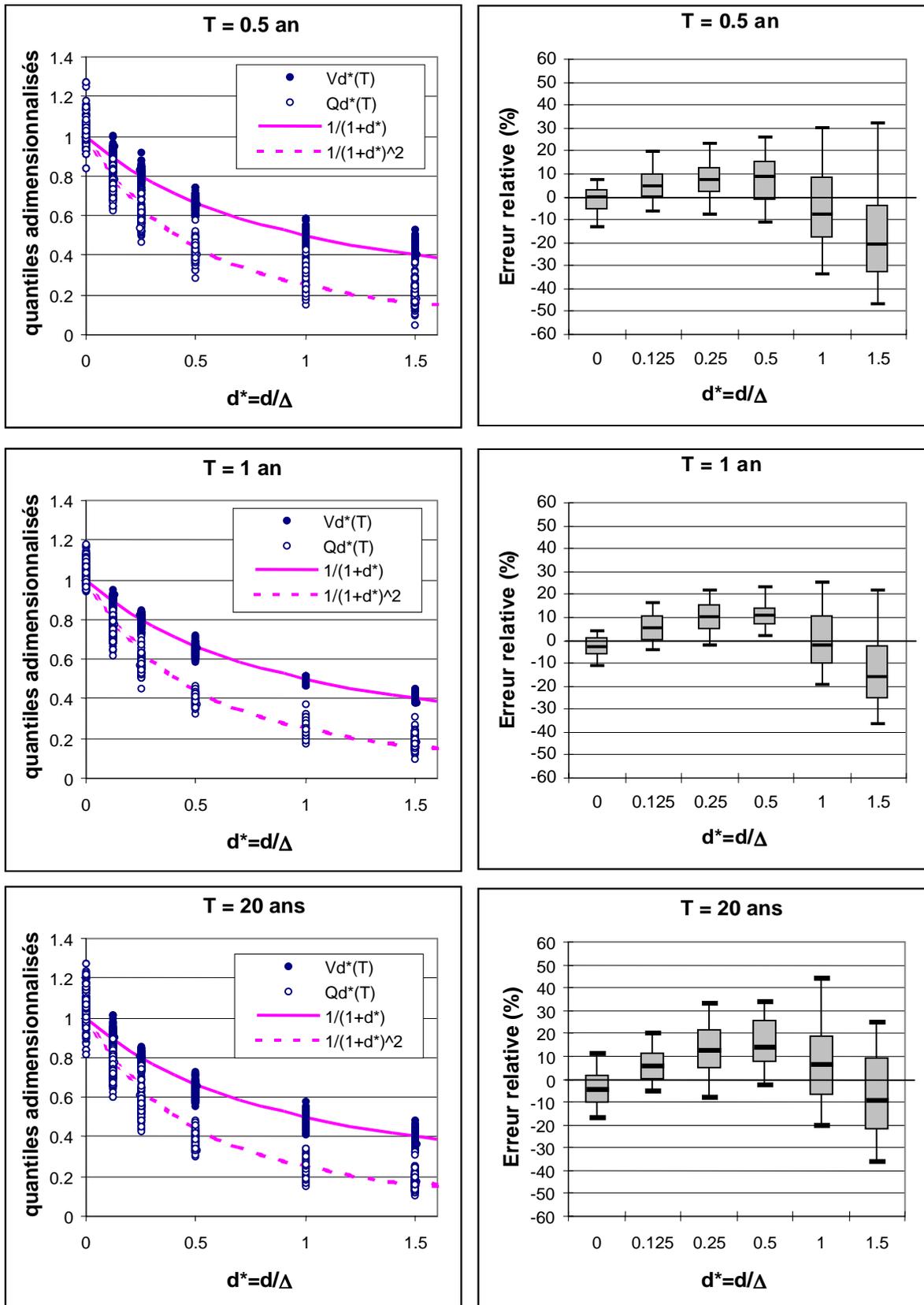


Figure 4-12 : Ecart relatif entre les débits seuils $Q(d,T)$ déduits des débits moyens $V(d,T)$ et les débits seuils observés $Q_d(T)$, en fonction de d/Δ *

* Les boîtes (plotting box) représentent l'erreur relative dépassée dans 95%, 75%, 50%, 25% et 5% des cas.

Afin de représenter l'ensemble des résultats sur une même figure (graphiques de gauche, Figure 4-12), nous avons procédé de la même façon qu'au chapitre 3, Figure 3-23 : les quantiles observés sont transformés selon la relation suivante (dans notre cas, $P=0$) :

$$X^* = \frac{X - P}{V(0,T) - P} \quad (4-27)$$

L'intérêt est que les quantiles modélisés transformés $V^*(d,T)$ et $Q^*(d,T)$ s'expriment simplement par :

$$V^*(d,T) = \frac{1}{1 + d^*} \quad \text{et} : \quad Q^*(d,T) = \frac{1}{(1 + d^*)^2} ; \quad \text{avec} : \quad d^* = d / \Delta \quad (4-28)$$

D'autre part, les graphiques de droite, Figure 4-12, indiquent la distribution de l'erreur relative entre les quantiles de débit seuil modélisés $Q(d,T)$ et non consolidés $Q_d(T)$:

$$\varepsilon = \frac{Q(d,T) - Q_d(T)}{Q_d(T)} \quad (4-29)$$

Les deux exemples de la Figure 4-11 présentent de bons résultats. Rappelons que les courbes $Q(d,T)$ ne sont pas ajustées sur les observations $Q_d(T)$ mais déduites des quantiles $V(d,T)$ par la relation (4-26). Par contre, ceux obtenus sur l'ensemble des bassins étudiés sont plus mitigés (Figure 4-12). On constate que l'estimation des débits seuils $Q(d,T)$ à partir des débits moyens $V(d,T)$ se détériore lorsque la durée augmente. Pour des durées supérieures à Δ ($d/\Delta > 1$) le calcul de $Q(d,T)$ devient trop imprécis, avec des erreurs relatives pouvant dépasser les 40%. Ces résultats s'expliquent par le fait que l'on combine deux types d'erreur : celles liées aux hypothèses de la modélisation des débits moyens $V(d,T)$ et celles liées à l'hypothèse donnant la relation (4-16) entre $V(d,T)$ et $Q(d,T)$. Prises séparément, ces hypothèses ont été validées : au chapitre 3, Figure 3-24, pour la modélisation $V(d,T)$ et au paragraphe 2.2. de ce chapitre (Figure 4-10) pour la relation expérimentale entre $Q_d(T)$ et $V_d(T)$. La Figure 4-12 montre que lorsqu'on combine ces différentes hypothèses pour obtenir les débits seuil $Q(d,T)$ à partir des débits moyens $V(d,T)$, les résultats se dégradent significativement. Il faut cependant noter que l'étude des erreurs relatives durée par durée est assez sévère. Si l'on regarde sur les graphiques de gauche, Figure 4-12, simplement l'évolution globale des quantiles $Q(d,T)$ on constate que les ordres de grandeurs sont quand même bien respectés. Rappelons que sur ces graphiques, les valeurs non consolidées et modélisées ont simplement été transformées par la même grandeur, puisque $P=0$, relation (4-27).

2.4. Interprétation des propriétés observées

L'interprétation des propriétés du modèle exposées au chapitre 3 §4, implique que les débits échantillonnés Qd et Vd sont tous issus du même événement. Ceci permettrait aussi d'expliquer la propriété reliant les quantiles de débits moyens $V_d(T)$ et de débits seuils $Q_d(T)$ vérifiée ci-dessus.

En effet, le raisonnement suivant, valable pour une crue, pourrait être transposé aux quantiles. Le graphique de gauche (Figure 4-13), représente un hydrogramme de crue $Q(t)$ sur lequel on a relevé les débits seuils Qd pour différentes durées d . Le graphique de droite représente ces mêmes débits seuils mais cette fois-ci en fonction de d , notés $Q_{seuil}(d)$.

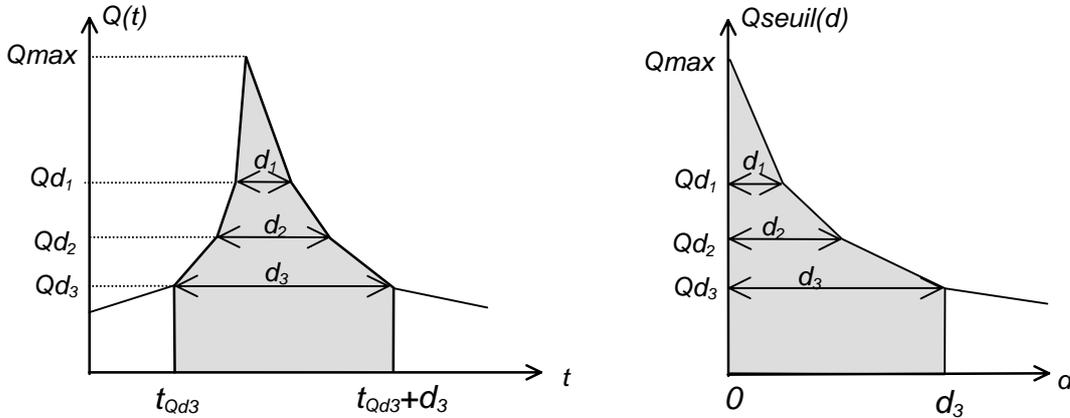


Figure 4-13 : Relation entre $Q(t)$ et $Q_{seuil}(d)$

Si maintenant, on calcule le débit moyen Vd maximal on a vu, grâce à la propriété de concomitance entre les débits seuils et les débits moyens (Figure 4-6) que Vd est égal à :

$$Vd = \frac{1}{d} \int_{t_{Qd}}^{t_{Qd}+d} Q(t) dt \quad (4-30)$$

Graphiquement, il y a égalité entre les surfaces grisées sur les deux graphiques de la Figure 4-13, la surface de chacun des trapèzes ne dépendant que de leur base et de leur hauteur. En raisonnant sur des trapèzes de hauteur infiniment fine, on peut écrire l'égalité suivante :

$$Vd = \frac{1}{d} \int_0^d Q_{seuil}(\tau) d\tau \quad (4-31)$$

Cette égalité montre que la relation expérimentale vérifiée entre les quantiles de débits moyens $V_d(T)$ et de débits seuils $Q_d(T)$ (Figure 4-8) est vérifiée pour un hydrogramme de crue. L'affinité des différents hydrogrammes de crue observée au chapitre 3 §4, implique que pour une même période de retour expérimentale, les valeurs échantillonnées proviennent du même événement de crue. Si elle est vérifiée, cette propriété permet donc de transposer ce raisonnement aux quantiles.

2.4.1. Discussion du caractère mono-fréquence des événements

Nous avons choisi de vérifier l'hypothèse ci-dessus sur le bassin de l'Orgeval au Theil. Ce bassin a été choisi car ses courbes QdF vérifient très bien les propriétés étudiées, à savoir l'affinité des distributions des débits moyens Vd (chapitre 3) et la relation entre les débits moyens et les débits seuils (Figure 4-9). La Figure 4-14 présente les distributions des débits seuils, associées aux dates des crues dont ils ont été issues. Les six plus fortes crues ont été représentées.

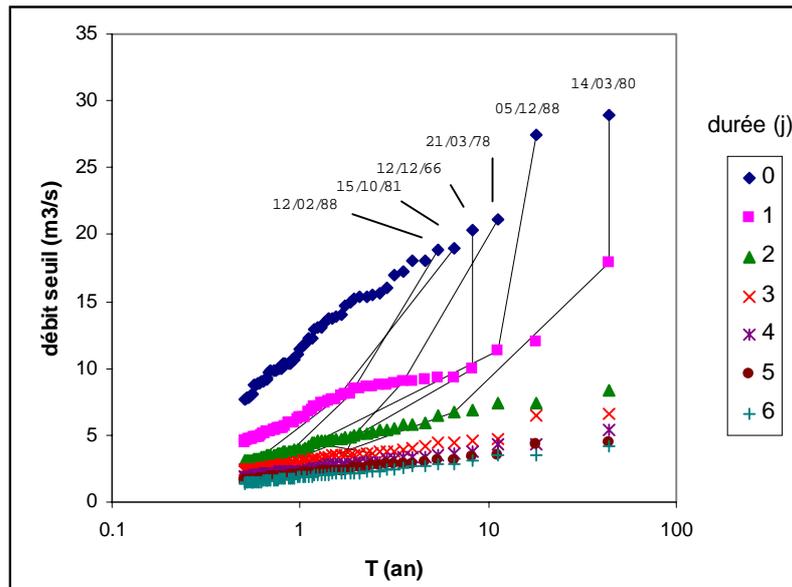


Figure 4-14 : Distribution des débits seuils associée à la date des crues, Exemple de l'Orgeval au Theil (103 km²)

La Figure 4-14 montre clairement, que pour les plus forts événements en terme de débits de pointe, la période de retour diminue lorsque la durée sur laquelle le débit seuil est calculé augmente. Ceci indique que les plus fortes crues sont moins étalées dans le temps. Ceci avait déjà été remarqué par Schober (1994) qui a étudié les crues du Réal Collobrier (Var).

En toute rigueur, le caractère « mono-fréquence » des événements n'est pas respecté (Figure 4-14) : les événements les plus rares sont moins étalés dans le temps. L'affinité des hydrogrammes évoquée au chapitre 3 (Figure 3-29) est donc à considérer comme une **première approximation**, néanmoins suffisante pour expliquer la convergence des distributions.

Remarque sur le caractère mono-fréquence des hydrogrammes

La tentative d'explication, d'abord présentée au chapitre 3 puis développée ici, repose sur une approximation : le caractère monofréquence des hydrogrammes observés. Lorsqu'on cherche à définir des scénarios hydrologiques de référence, par exemple pour la prévention du risque d'inondation, il est probable que cette approximation « mono-fréquence » est insuffisante. Des travaux de recherche sont d'ailleurs en cours pour justement essayer de caractériser la forme des hydrogrammes de crues exceptionnelles, et non plus simplement du débit de pointe ou du débit continûment dépassé pendant une durée d (cf. le programme Risque Inondation (1999-2000) associant le Cemagref et EDF/DTG et la thèse en cours de Le Clerc (2000-2003)). La propriété que nous mettons en évidence sur les distribution mériterait d'être approfondie dans le cadre de ces recherches.

3. Etudes des étiages

L'analyse fréquentielle du régime présentée jusqu'à présent a concerné les événements de crues. L'approche que nous avons développée a permis d'appliquer la même démarche aux étiages. Les résultats suivants ont été obtenus avec l'aide de Nathalie Chaput (1999) et Vincent Dussaux (2000), puis publiés par Galéa *et al.* (2001). Ils constituent une généralisation intéressante du modèle local convergent. Notre travail concernant principalement l'étude des crues, nous ne donnons ici qu'un bref aperçu de ces résultats. On trouvera en annexe 5 plus de détails, notamment la validation du modèle obtenu sur une trentaine de bassins situés sur le bassin hydrographique de la Moselle.

3.1. Variables étudiées et échantillonnage

De façon analogue aux débits moyens maximaux VCX et aux débits seuils QCX relatives aux crues, Oberlin (1992) définit des variables fonction de la durée caractérisant les étiages (Figure 4-15). Il s'agit du :

- débit moyen minimum : $VCNd$

L'intégration d'un hydrogramme continu sur une durée continue d permet d'obtenir le volume écoulé sur cette durée et par conséquent le débit moyen correspondant à cette durée. Il est ainsi possible d'obtenir une chronique de moyennes mobiles de débits moyens sur une durée d et d'en extraire la valeur minimale

- débit-seuil non dépassé : $QCNd$

Le débit-seuil non dépassé correspond à la valeur minimale du débit qui n'est pas dépassé pendant une durée d . Cette notion est complémentaire de celle de débit moyen minimum et très utile dans certains cas (pompages en rivières, tirant d'eau pour la protection des écosystèmes aquatiques).

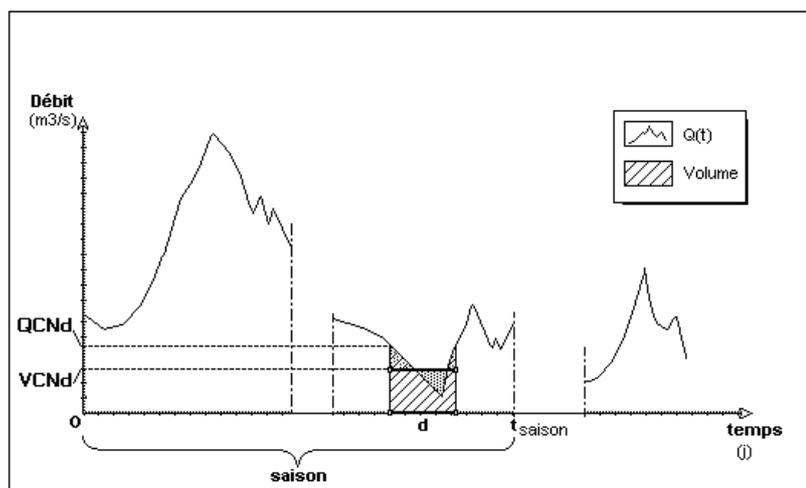


Figure 4-15 : Variables caractéristiques d'étiage

Les débits caractéristiques $VCNd$ ou $QCNd$ sont échantillonnés à partir des chroniques de débits journaliers. Un intervalle mobile de durée $d = N$ jours est déplacé jour après jour sur la saison étudiée en débordant de $N/2$ jours de part et d'autre de la saison. Pour chacune des périodes de N jours balayées, sont retenues la valeur maximale du débit journalier et la moyenne des débits journaliers. Les minima de ces valeurs donnent respectivement le $QCNd$ et le $VCNd$.

Pour s'assurer de l'indépendance des événements sélectionnés, on ne retient qu'un seul événement d'étiage par an, que l'on suppose encadré par une ou plusieurs saisons humides pendant lesquelles les réserves souterraines se rechargent (Figure 4-16). Pour cette raison, contrairement aux crues, nous ne considérons que l'échantillonnage par la méthode des minima annuels. L'échantillonnage par valeurs inférieures à un seuil est en effet très délicat, du fait de la difficulté (ou de l'impossibilité) de sélectionner plusieurs valeurs par an indépendantes.

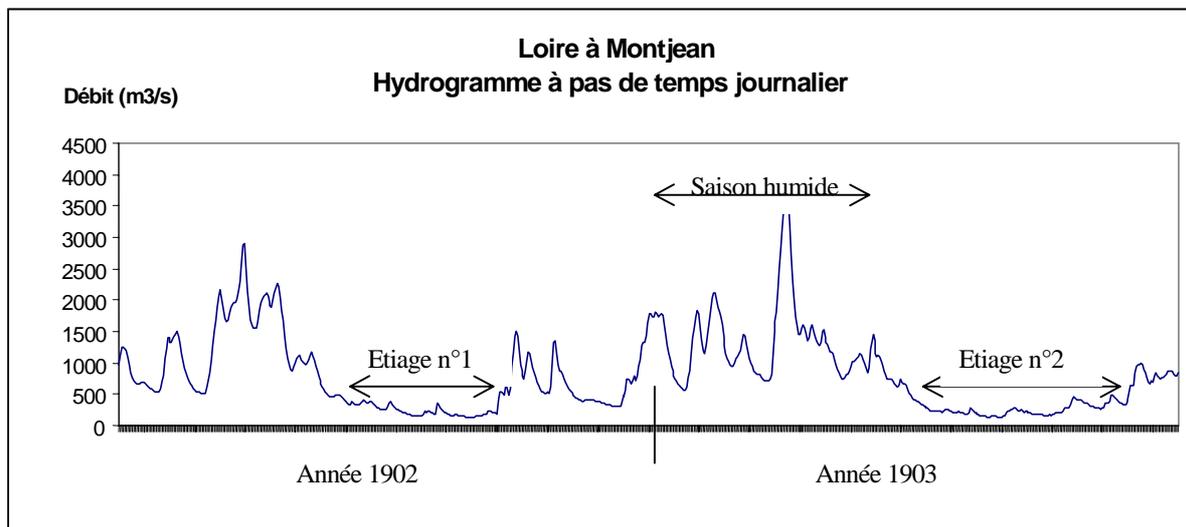


Figure 4-16 : Hydrogramme journalier sur deux ans, d'après Dussaux (2000)

Une fois l'échantillon (VCN ou $QC�N$) obtenu, il est classé par ordre croissant et un indice de rang i est affecté à chaque valeur. De même que pour les crues, on adopte la fréquence empirique de non dépassement suivante, i étant le rang, N le nombre de valeurs (expression de Chegodayev) :

$$F_i = \frac{i - 0.3}{N + 0.4} \quad (4-32)$$

Différentes lois statistiques sont utilisées dans la littérature. On trouvera dans Bobée et Abi-Zeid (1999) une revue bibliographique sur les plus utilisées. Parmi elles, se trouve la loi log-normale à deux paramètres, qui a l'avantage d'être robuste et simple à utiliser. Dans le cas des débits moyens $VCNd$, elle s'écrit :

$$\ln[VCNd(u_F)] = s_d \cdot u_F + m_d \quad (4-33)$$

u_F est la variable réduite de Gauss ; en particulier $u_F=0$ correspond à $F=0.5$
 m_d et s_d sont estimés à partir de l'échantillon des valeurs $vcn_d(j)$ de la façon suivante :

$$m_d = \frac{\sum_{j=1}^N \ln[vcn_d(j)]}{N} ; \quad s_d = \sqrt{\frac{\sum_{j=1}^N [\ln(vcn_d(j)) - m_d]^2}{N-1}} \quad (4-34)$$

N étant le nombre de valeurs dans l'échantillon.

Il est parfois utile d'exprimer la probabilité en terme de période de retour.

Lorsque $F < 0.5$ le quantile est dit « sec » et on lui associe la période de retour :

$T=1/F$, correspondant à une probabilité de non-dépassement.

Lorsque $F > 0.5$ le quantile est dit « humide » et on lui associe la période de retour :

$T=1/(1-F)$, correspondant à une probabilité de dépassement.

Le Tableau 4-1 donne la correspondance entre U_F la variable réduite de Gauss et T la période de retour définie ci-dessus, pour les quantiles dits “secs” et “humides”.

	Années sèches					Années humides			
T (an)	50	20	10	5	2	5	10	20	50
U_F	-2.06	-1.64	-1.28	-0.84	0	0.84	1.28	1.64	2.06

Tableau 4-1 : Correspondance entre la période de retour T et la variable réduite de Gauss U_F .

3.2. Illustration de la propriété d'affinité des hydrogrammes d'étiages

Il est pratique de représenter les quantiles en échelle logarithmique. Dans le cas d'une loi log-normale ceux-ci s'alignent sur une droite. La Figure 4-17 montre un exemple d'ajustement pour des quantiles $VCN(d; U_F)$ pour l'exemple de la Loire à Montjean (115 000 km²), déjà traité pour les crues (Figure 4-4).

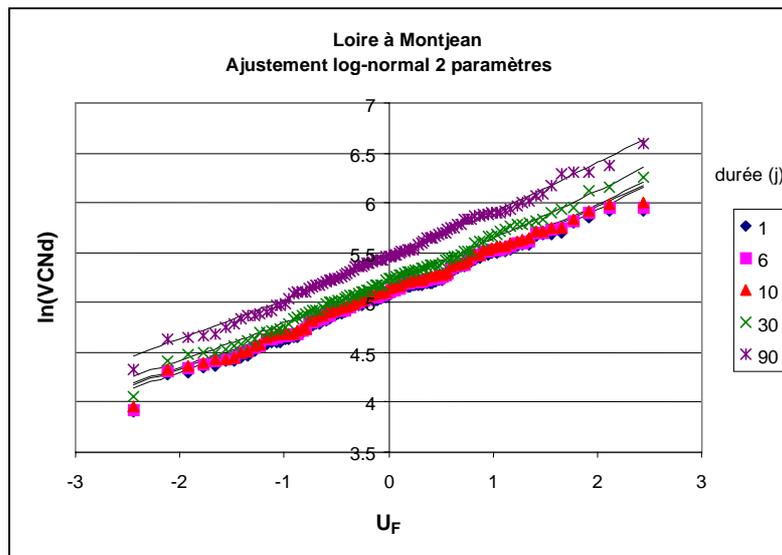


Figure 4-17 : Parallélisme observé des distributions des débits moyens minimaux $VCNd$ en échelle logarithmique (Dusseaux, 2000)

La Figure 4-18 montre le même exemple que précédemment, mais cette fois-ci en valeurs naturelles. Si maintenant on norme chacune de ces distributions par sa valeur pour $U_F=0$, on s'aperçoit qu'elles se confondent (Figure 4-19). Nous avons donc montré sur un exemple **l'affinité des distribution**. Cette affinité pouvait déjà être remarquée dès la Figure 4-17, en remarquant que les distributions sont parallèles, lorsqu'on exprime les débits sur une échelle logarithmique.

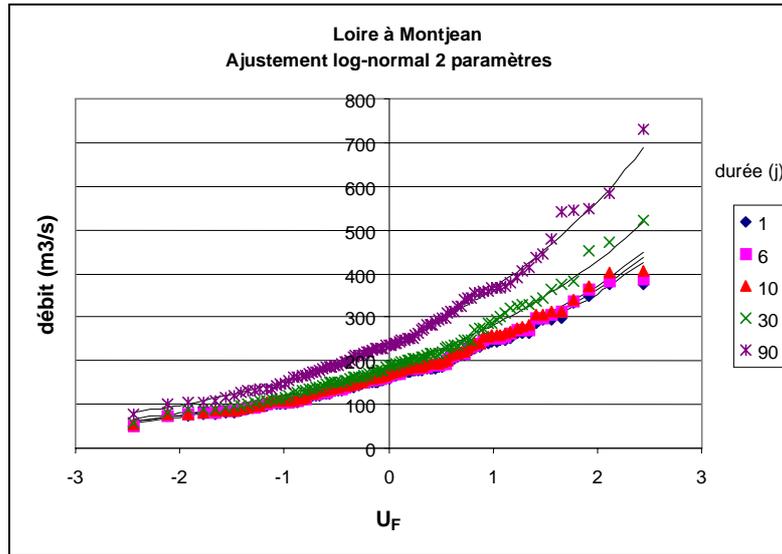


Figure 4-18 : Ajustement d'une loi log-normal à 2 paramètres sur les débits moyens minimaux $VCNd$

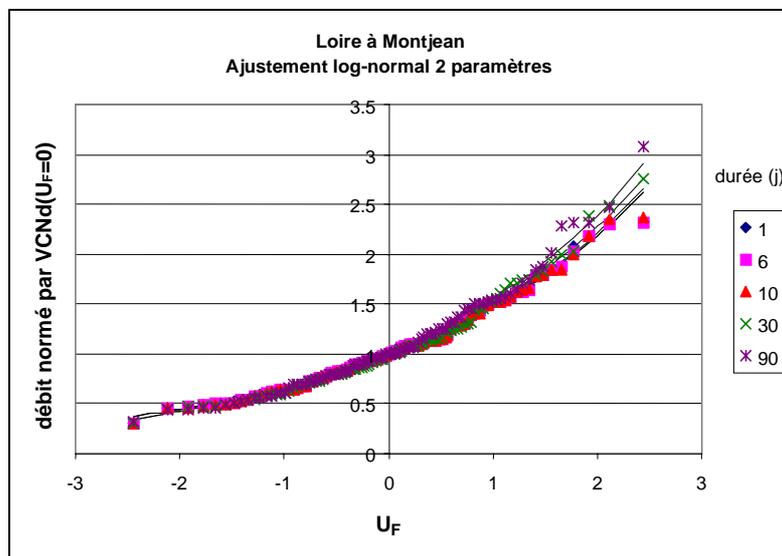


Figure 4-19 : Vérification de l'affinité sur les débits moyens minimaux $VCNd$

La propriété d'affinité est schématisée Figure 4-20. Le parallélisme observé Figure 4-17 se traduit par une convergence des distributions vers un point fictif situé en $(0, -\infty)$.

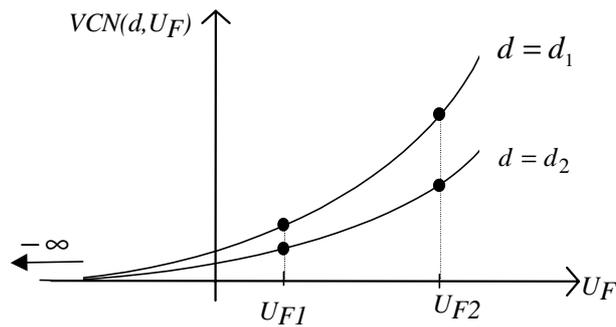


Figure 4-20 : Propriété d'affinité des distributions appliquée aux étiages

De façon analogue aux crues, la propriété d'affinité permet d'établir la relation suivante, indépendante de U_F .

$$\frac{VCN(d_2, U_{F1})}{VCN(d_1, U_{F1})} = \frac{VCN(d_2, U_{F2})}{VCN(d_1, U_{F2})} \quad (4-35)$$

Cette propriété a été utilisée pour construire un modèle débit-durée-fréquence analogue à celui des crues, mais cette fois-ci adapté aux étiages. Ces développements sont présentés en annexe 5.

3.3. Interprétation des propriétés en terme de forme des hydrogrammes

De même que pour les crues, il est possible d'interpréter l'affinité des distributions des débits (Figure 4-20) comme étant due à l'affinité des hydrogrammes. Ceci tend donc à généraliser l'hypothèse que nous avons faite au chapitre 3, Figure 3-28, et qui ne concernait que les hydrogrammes de crue. De la même façon, cette affinité des hydrogrammes est illustrée Figure 4-21 et s'exprime par :

$$Q_3(t) = k_2 Q_2(t) = k_1 Q_1(t) \quad (4-36)$$

La seule différence par rapport au chapitre 3, est que nous avons fixé P à zéro.

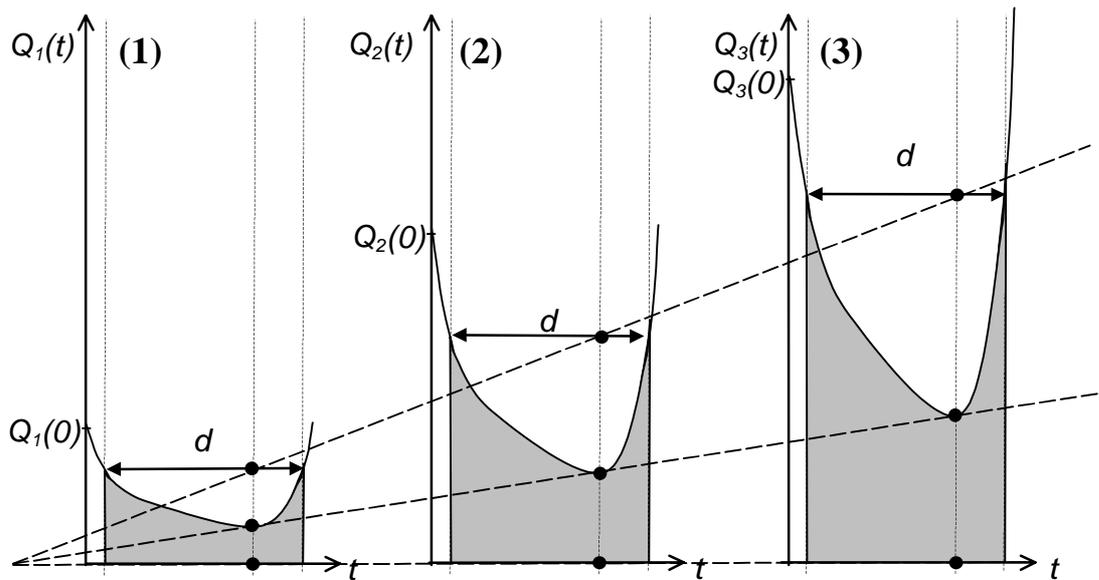


Figure 4-21 : Affinité verticale des hydrogrammes d'étiages

Cette hypothèse d'affinité, même si elle simplifie la réalité, n'est pas dénuée de tout fondement. En effet, elle va dans le sens d'une autre approximation souvent faite en hydrologie, et qui concerne les courbes de récession, pour lesquelles on suppose que le débit décroît exponentiellement, selon la loi de Maillet (Roche, 1963) :

$$Q(t) = Q(0) \cdot e^{-\alpha t} \quad (4-37)$$

α étant le coefficient de tarissement, de la dimension inverse d'un temps.

Si, comme cela est souvent supposé, le coefficient α est constant, alors la loi de Maillet (4-37) est un cas particulier de l'affinité supposée. En effet, si l'on écrit cette loi pour trois courbes de récessions $Q_1(t)$, $Q_2(t)$ et $Q_3(t)$ et que l'on pose $k_1 = Q_1(0)/Q_2(0)$ et $k_2 = Q_2(0)/Q_3(0)$, on retombe alors sur l'égalité (4-36).

Reste maintenant à discuter du bien fondé de l'hypothèse d'affinité des hydrogrammes, mais cette fois-ci lors de la remontée du débit à la fin de l'étiage. Cette remontée est de toute façon beaucoup plus rapide que la période de récession. On peut raisonnablement penser que le problème de l'affinité ou de la non affinité des courbes de remontée passe au second plan par rapport à celle des courbes de récession. Graphiquement, cela revient à faire l'approximation qu'à l'échelle de l'étiage (plusieurs mois), la courbe de remontée est presque verticale. Nous allons donc par la suite vérifier l'affinité des hydrogrammes simplement sur les courbes de récession, et supposer que la courbe de remontée ne change pas grand chose au résultat.

Comme nous l'avons vu, l'affinité des courbes de récessions est respectée si α est constant. Cette simplification est souvent faite, et pour calculer α , on a l'habitude de représenter toutes les courbes de récession sur un même graphe, avec une échelle logarithmique pour les débits, et une échelle naturelle pour le temps. Les courbes obtenues sont alors des droites ayant un coefficient directeur de $-\alpha$, si la loi de Maillet est respectée. Donc, si α est constant, ces

droites doivent être parallèles. La Figure 4-22 illustre cette propriété pour la Seille à Nomeny, bassin situé en Moselle. On constate en effet, sauf pour une courbe durant laquelle un épisode pluvieux a fait remonter le débit, que les courbes obtenues sont des droites parallèles. Cette observation a aussi été faite pour l'ensemble des bassins traités (Galéa *et al.*, 2001).

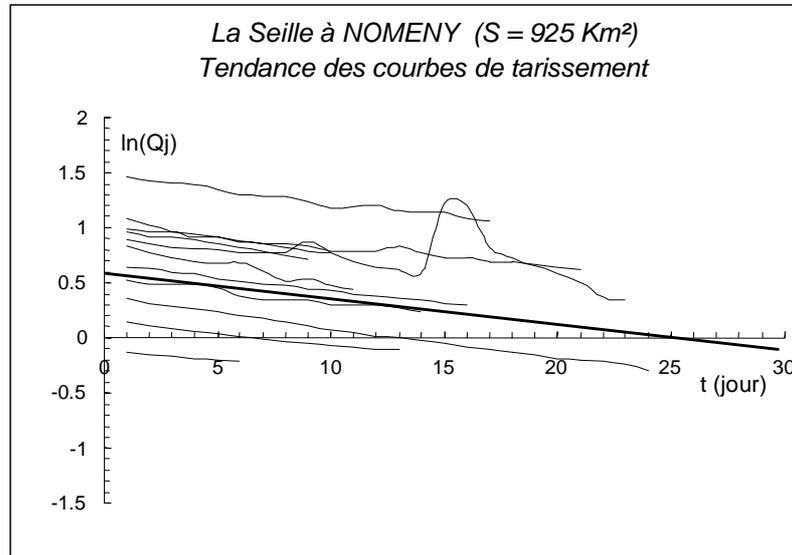


Figure 4-22 : Illustration de la constance du coefficient de récession α , pour l'exemple de la Seille à Nomeny (Galéa *et al.*, 2001).

3.4. Conclusion

Ce paragraphe a permis de montrer l'applicabilité aux étiages du modèle local convergent défini pour les crues. La propriété d'affinité des distributions a été conservée, et le modèle a été adapté en adoptant une autre loi pour décrire l'évolution en fonction des fréquences (loi log-normal à deux paramètres), ainsi qu'une évolution linéaire en fonction de la durée d . Ces résultats obtenus sont encourageants. Toujours dans le même esprit que ce qui a été fait pour les crues, le modèle a aussi été adapté à d'autres types de lois : log-normal à trois paramètres, weibull à deux et trois paramètres (Dussaux, 2000), donnant aussi pour les exemples étudiés des résultats satisfaisants.

Conclusion du chapitre 4

Ce chapitre complète le précédent sur la modélisation débit-durée-fréquence locale.

Nous avons montré que le modèle local convergent développé pour les débits moyens V_d n'était pas limité à la seule loi exponentielle. Il peut être adapté à d'autres types de lois, comme par exemple la loi GEV à trois paramètres. Cette souplesse quant à la loi utilisée est intéressante, car elle permet au modèle de s'adapter à des contextes différents.

Un autre type de variable utile dans la description des crues a été étudié. Il s'agit des débits continûment dépassés sur la durée d , aussi appelés débits seuils. Une propriété expérimentale, reliant les courbes débit-durée-fréquence des débits moyens et celle des débits seuil a été vérifiée. Cette propriété comme la propriété d'affinité sur les distributions évoquée au chapitre précédent trouve son explication dans la forme des hydrogrammes analysés. L'analyse détaillée d'un bassin, l'Orgeval au Theil, donne une explication partielle à cette propriété ouvrant des pistes de recherche à développer plus en détail. Quoi qu'il en soit, il apparaît que l'analyse *QdF* appuyée sur la propriété d'affinité mise en évidence ici, est un outil intéressant permettant une synthèse structurée de la variabilité des débits.

Une autre composante du régime hydrologique a été étudiée : les étiages. On observe aussi une affinité sur les distributions des débits minimaux relatifs à différentes durées. Comme pour les crues, cette propriété est interprétée par une affinité sur les hydrogrammes.

Les développements présentés jusqu'à présent n'ont concerné que l'analyse locale. Tout un domaine n'a pas encore été traité : celui de l'analyse régionale. Nous l'aborderons dans la partie suivante.

PARTIE II

Analyse Régionale



Illustration 3 : La Porte Créqui le 3 novembre 1859 au matin, Grenoble
Source : Pilot J.-J.-A., Grenoble inondé, Maisonville & Jourdain, 1859. Dessin de D. Raoult

Chapitre 5

Présentation du modèle QdF régional convergent

Le travail présenté dans les chapitres précédents n'a concerné que l'**analyse fréquentielle locale**. Or le plus souvent à l'endroit où l'on souhaite connaître le régime des crues, par exemple pour faire un aménagement, aucune donnée n'est disponible, où lorsqu'elles le sont, le nombre d'années de mesures est insuffisant pour déterminer avec précision les quantiles qui nous intéressent, c'est à dire relatifs aux fortes périodes de retour.

Les trois derniers chapitres de cette thèse tentent d'apporter des réponses au problème évoqué ci-dessus. Le domaine abordé est celui de l'**analyse fréquentielle régionale** : le manque d'information locale est pallié en utilisant l'information hydrologique disponible sur d'autres sites « similaires » au site d'étude. Pour simplifier, on peut distinguer deux niveaux dans l'analyse :

Le but de ce chapitre est d'appliquer les concepts de l'analyse régionale au modèle QdF **local** convergent présenté dans la première partie de la thèse. Le modèle obtenu est appelé modèle QdF **régional** convergent.

Le premier paragraphe effectue quelques rappels sur les méthodes régionales, en particulier sur la méthode de l'indice de crue.

Puis, nous présentons le modèle *régional* convergent, défini en adaptant la méthode de l'indice de crue à l'approche débit-durée-fréquence.

Enfin, cette nouvelle approche est appliquée sur plus d'une centaine de bassins situés au Québec et en Ontario.

1. Quelques rappels sur l'analyse régionale

1.1. Qu'est-ce que l'analyse régionale ?

Nous entendons par analyse régionale toute analyse pour laquelle l'estimation d'une caractéristique en un point utilise l'information disponible en d'autres points. En ce sens, les techniques de spatialisation de la pluie sont des analyses régionales, puisqu'elles mettent à profit la structure spatiale existante entre les caractéristiques étudiées en plusieurs points. Cependant, ce type d'approche est difficilement applicable aux débits, du fait que la cohérence entre les débits observés en différents points n'est pas due à la distance euclidienne entre les points de mesures mais entre autres choses à la notion de réseau hydrographique (Cunnane, 1988). On note toutefois que cette difficulté n'est pas insurmontable, puisque Sauquet *et al.* (2000a) spatialisent le module interannuel grâce à une technique de krigeage adaptée, prenant en compte la topologie du réseau hydrographique.

Concernant les crues, les méthodes rencontrées dans la littérature sont regroupées sous le terme d'analyse fréquentielle régionale des crues, traduction de *Regional Flood Frequency Analysis*, souvent noté *RFFA*. Deux contextes différents d'application peuvent être distingués :

- Le premier est celui des sites non jaugés, pour estimer des quantiles de crues.
- Le second est celui des sites jaugés, mais pour lesquels on souhaite calculer des quantiles dont les périodes de retour sont largement supérieures à la taille de la chronique disponible.

En France, l'information régionale est utilisée pour estimer des quantiles de crue sur des sites non jaugés, mais en général pas pour l'extrapolation aux fréquences rares. On préfère utiliser l'information pluviométrique : la méthode du Gradex (Guillot et Duband, 1967) ou l'une de ses variantes : le Gradex esthétique (Michel, 1982) ou le modèle Agrégée (Margoum *et al.*, 1994). Ainsi, les méthodes présentées ci-après n'ont connu que peu d'écho en France, jusqu'à présent.

Un projet canadien regroupant différents laboratoires sous le nom de **Groupe de Recherche en HYdrologie Statistique** a réalisé une étude comparative de différentes méthodes régionales (GREHYS, 1996ab). Celui-ci distingue deux étapes communes à toutes les méthodes régionales :

1. La définition de régions hydrologiques homogènes.
2. Le transfert de l'information donnée par l'ensemble des sites jaugés vers un site cible (*non jaugé ou peu jaugé*), à l'intérieur d'une même région homogène. Deux principales méthodes existent : la méthode des régressions directes, et celle de l'indice de crue.

1.2. Constitution d'une région hydrologique homogène

La définition d'une région hydrologique homogène n'est pas très précise (Rasmussen *et al.*, 1994). Il s'agit d'un regroupement de sites ayant le même comportement par rapport à certaines variables hydrologiques. Nous reviendrons au paragraphe 1.5. sur les caractéristiques hydrologiques testées pour juger de l'homogénéité d'une région.

Au fil du temps, la notion de « région » hydrologique a évolué. Il a d'abord été question de régions géographiques classiques. Puis on a constitué des groupes de stations sans nécessairement qu'il y ait un lien géographique entre elles : il s'agit de régions non contiguës. Enfin, pour chaque site étudié, on peut constituer un ensemble de stations qui lui ressemblent le plus : il s'agit des régions de type voisinage.

- ***Régions géographiques***

Le concept de région homogène regroupant des sites géographiquement voisins a été largement utilisé en hydrologie. Ces approches se justifient si les facteurs déterminant le régime hydrologique ont bien une cohérence spatiale, comme par exemple la pluie, l'altitude le type de sol. Ainsi le NERC (1975) définit pour l'Angleterre et l'Irlande onze régions homogènes. Pour le Québec et l'Ontario, neuf régions géographiques homogènes ont été définies par Gingras *et al.* (1994).

- ***Régions non-contigües***

Cependant, certains facteurs de l'écoulement ne sont pas liés nécessairement à la position du bassin dans l'espace géographique. Il s'agit en particulier de la superficie et de la forme des bassins versants. C'est pourquoi sont apparues des approches préférant définir l'homogénéité dans un espace constitué de variables hydrologiques, « physiques » (c'est à dire liée à la géomorphologie et à l'occupation du sol du bassin) et climatiques. Ces techniques utilisent les analyses en composantes principales et des méthodes de classification automatique (cluster analysis) (Benzecri, 1973 ; Jambu, 1976, d'après GREHYS, 1996).

- ***Régions du type voisinage***

Les deux types de régions ci-dessus se heurtent au problème suivant. Un bassin peut très bien se trouver à la limite de deux régions. Laquelle choisir ? Ce problème est résolu par la notion de voisinage : chaque station cible définit sa propre région. Deux approches peuvent être distinguées, celle des régions d'influence, et l'analyse des corrélations canoniques.

Burn (1990ab et 1997) définit la méthode dite des "régions d'influence" à partir de la méthode de classification du « diamètre critique » présentée par Jambu (1978). Il calcule dans un espace multidimensionnel, dont les axes sont des caractéristiques hydrologiques, « physiques » (géomorphologie, occupation du sol) et climatiques des bassins, une distance entre le site étudié et les autres sites. La région d'influence du site cible est définie par

l'ensemble des sites dont la distance au site cible est inférieure à un seuil (distance critique). Le seuil d'inclusion de stations représente un compromis entre le nombre de stations dans le voisinage et son homogénéité hydrologique. Burn (1990a) définit également une fonction de pondération des sites du voisinage reflétant la proximité hydrologique relative des différents sites du site cible. Une variante de cette méthode utilisant la théorie des ensembles flous a été présentée par Fortin *et al.* (1995) et Bargaoui *et al.* (1998).

Une seconde méthode de définition des voisinages est celle de l'analyse des corrélations canoniques. Cette méthode se rapproche de l'analyse en composantes principales sauf qu'ici, deux espaces de variables « canoniques » sont définis : un espace des variables hydrologiques et un autre des variables physiques/climatiques des bassins. L'utilisation des variables canoniques permet de décrire la relation de dépendance qui existe entre les deux ensembles de variables aléatoires « hydrologiques » et « physiques/météorologiques ». Il est alors possible d'inférer sur les variables canoniques d'un ensemble connaissant les variables canoniques de l'autre ensemble. On peut également calculer une distance (Mahalanobis, 1936) entre deux variables canoniques, ce qui permet de déterminer des voisinages hydrologiques. L'application de l'analyse des corrélations canoniques dans le domaine de l'hydrologie est due à Cavadias (Cavadias, 1990 ; Ribeiro-Corréa *et al.*, 1995). Plus récemment, Ouarda *et al.* (1997, 2000ab), améliorent cette approche en y associant un critère d'appartenance à la région (distance minimale dans l'espace des caractéristiques physiques/climatiques) définie de façon plus objective que dans l'approche de Cavadias. Ouarda *et al.* (2000ab) présentent également une version formalisée de la méthode pour la régionalisation dans le cas où le site cible est complètement non jaugé, et appliquent la méthode pour la régionalisation conjointe de plusieurs caractéristiques hydrologiques telles que la pointe et le volume de crue (*cf.* chap. 1, §3.4 pour la définition de ces variables).

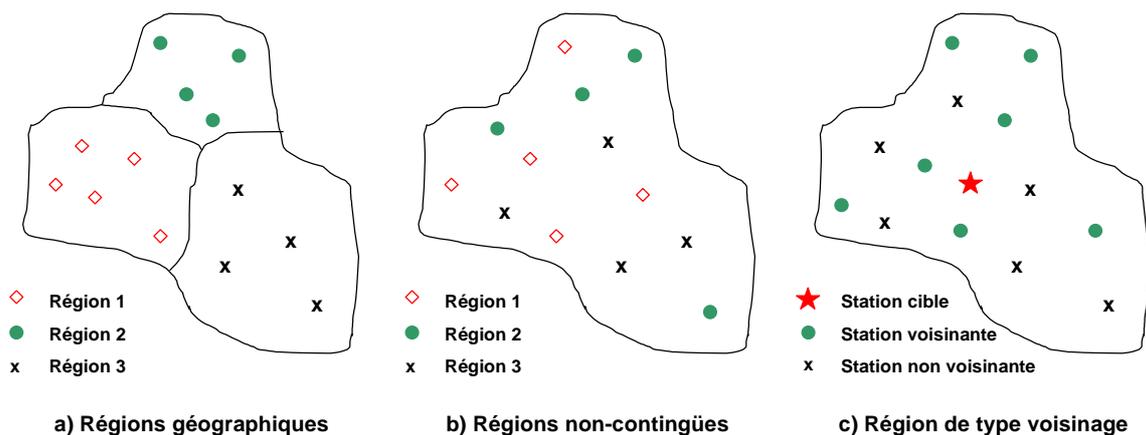


Figure 5-1 : Les différents types de régions hydrologiques homogènes

La Figure 5-1 schématise les trois types de régions décrits ci-dessus. Quel que soit le type de région considéré, les paramètres rentrant en compte dans la constitution des groupes sont de nature hydrologique (coefficient de variation, débit maximum annuel moyen, quantiles de crue...), « physique » (latitude, longitude, surface du bassin, pourcentage des lacs ou marais,

pourcentage de forêt, altitude moyenne du bassin, pente moyenne du bassin...), et climatique (pluie annuelle, autres statistiques sur la pluie ou la neige). Dans le cas où la station cible est non jaugée, les paramètres hydrologiques de cette station ne sont pas utilisés, puisque inconnus. Il faut donc que le mode d'attribution de la région au site ne prenne pas en compte les caractéristiques physiques et climatiques du bassin.

1.3. Estimation des quantiles de crue par régressions multiples

La méthode des régressions multiples permet d'établir une relation directe entre les variables hydrologiques (quantiles de crue) et les variables explicatives physiques et climatiques. Elle a l'avantage d'être simple, rapide et de pouvoir prendre en compte une certaine hétérogénéité au sein de la région. Le plus souvent, les quantiles de crues sont exprimés sous la forme :

$$Q(T) = \alpha_0 P_1^{\alpha_1} P_2^{\alpha_2} \dots P_n^{\alpha_n} e^\varepsilon \quad (5-1)$$

Avec : $Q(T)$, le quantile de période de retour T , α_i les paramètres à estimer, P_i les caractéristiques physiques et climatiques et ε l'erreur résiduelle.

L'estimation des paramètres se fait le plus souvent par la linéarisation de la relation ci-dessus, en effectuant une transformation logarithmique.

$$\log Q(T) = \log \alpha_0 + \alpha_1 \log P_1 + \alpha_2 \log P_2 + \dots + \alpha_n \log P_n + \varepsilon \quad (5-2)$$

Les paramètres α_i peuvent alors être estimés par la méthode des moindres carrés. Certains auteurs critiquent cette façon de faire estimant qu'elle est biaisée, et préfèrent des méthodes non linéaires, en minimisant un terme d'erreur (Pandey et Nguyen, 1999).

$$Q(T) = \alpha_0 P_1^{\alpha_1} P_2^{\alpha_2} \dots P_n^{\alpha_n} + \varepsilon \quad (5-3)$$

1.4. La méthode de l'indice de crue

La méthode de l'indice de crue a été introduite par Dalrymple (1960). Cette méthode fait l'hypothèse qu'au sein d'une région homogène, toutes les distributions locales normées par un indicateur de tendance, l'indice de crue, se confondent et permettent ainsi de définir une distribution régionale adimensionnelle. Les principales étapes de la méthode sont les suivantes.

1.4.1. Standardisation des données

A chaque site i , les valeurs échantillonnées sont adimensionnalisées en les divisant par l'indice de crue. Selon la notation que nous avons déjà employée, les valeurs échantillonnées sont notées par une lettre minuscule. Dans toute la suite, le symbole * signifiera que l'on travaille à partir de l'échantillon adimensionnalisé.

$$q_i^*(j) = q_i(j) / \mu_i \quad (5-4)$$

où $q_i(j)$ est la valeur échantillonnée au site i , de rang j ; $q_i(j)^*$, la valeur adimensionnalisée ; et μ_i , l'indice de crue.

Le plus souvent, l'indice de crue μ_i correspond à la moyenne ou bien à la médiane des valeurs $q_i(j)$, échantillonnées par maximum annuel, ce qui équivaut à un quantile de période de retour d'environ 2 ans, selon la loi utilisée. On considère que, étant relatif à une période de retour assez faible, cet indice est plus robuste et plus facilement estimable que d'autres quantiles plus rares.

1.4.2. Calage d'une distribution régionale adimensionnelle

Une distribution adimensionnelle régionale notée $Q_R^*(T)$ est alors établie à partir des données $q_i(j)^*$ de tous les sites i de la région. En général, une loi à trois paramètres est ajustée car on considère que, contrairement à l'analyse locale, la taille de l'échantillon des crues de la région permet l'extrapolation aux fréquences rares.

- **Méthode des stations années**

Cette méthode consiste à regrouper les valeurs échantillonnées adimensionnalisées $q_i(j)^*$, au sein d'un seul échantillon. Dans le cas de valeurs maximales annuelles, la taille de cet échantillon est alors égale à la somme des années d'observations toutes stations confondues, d'où le nom de cette méthode : « stations années ». Puis la distribution $Q_R^*(T)$ est ajustée sur ce nouvel échantillon. Cette méthode est généralement déconseillée car il se peut que cet échantillon regroupe des événements non-indépendants. C'est le cas par exemple pour une même crue affectant plusieurs sous-bassins emboîtés.

- **Moyenne régionale des paramètres de la loi locale adimensionnalisée**

Pour chacun des sites i , une distribution $Q_i^*(T)$ est ajustée localement sur les échantillons adimensionnalisés $q_i(j)^*$. Cette loi s'exprime à l'aide de ses n paramètres $p_i^{1*}, \dots, p_i^{n*}$. Puis, les paramètres $p_R^{1*}, \dots, p_R^{n*}$ relatifs quant à eux à la loi régionale $Q_R^*(T)$ sont obtenus en faisant la moyenne des paramètres locaux p_i^{k*} . Cette moyenne peut être pondérée par le nombre d'années d'observation au site i . En effet, plus le nombre d'observation est grand, plus on considère que l'échantillon est représentatif de la population, et donc, un poids plus important lui est attribué :

$$p_R^{k*} = \frac{\sum_{i=1}^m NA_i \cdot p_i^{k*}}{\sum_{i=1}^m NA_i} \quad (5-5)$$

avec m : le nombre de sites dans la région, NA_i le nombre d'années de mesure disponible au site i , p_i^{k*} : le $k^{\text{ième}}$ paramètre de la loi ajustée sur l'échantillon adimensionnalisé du site i .

- **Moyenne régionale des moments (classiques ou pondérés) locaux adimensionnalisés**

Cette méthode est voisine de la précédente, sauf que l'on travaille sur les moments locaux de chacun des échantillons adimensionnalisés $q_i(j)^*$. Pour une loi à p paramètres, on calcule la moyenne des p premiers moments locaux adimensionnalisés :

$$m_R^{k*} = \frac{\sum_{i=1}^m NA_i \cdot m_i^{k*}}{\sum_{i=1}^m NA_i} \quad (5-6)$$

avec m : le nombre de site dans la région, NA_i le nombre d'années de mesure disponible au site i , m_i^{k*} : le moment d'ordre k calculé sur l'échantillon adimensionnalisé du site i .

Puis la loi régionale adimensionnelle $Q_R^*(T)$ est calculée par la méthode des moments, en utilisant les moments régionaux m_R^{k*} .

Une variante consiste non plus à travailler avec les moments simples, mais sur les moments pondérés, présentés au paragraphe 1.1 du chapitre 4. Les moments pondérés régionaux b_R^{k*} sont obtenus à partir des moments pondérés locaux b_i^{k*} :

$$b^{k*} = \frac{\sum_{i=1}^m NA_i \cdot b_i^{k*}}{\sum_{i=1}^m NA_i} \quad (5-7)$$

De même, la loi régionale adimensionnelle $Q_R^*(T)$ est calculée par la méthode des moments pondérés (Hosking *et al.*, 1985), décrite au chapitre 4 (§1.1) pour le cas de la loi *GEV*. La méthode des moments pondérés régionaux est plus robuste que les deux méthodes précédemment présentées. Pour cette raison, nous l'avons utilisée dans la suite de ce chapitre, au paragraphe 3. Cependant, comme le note Bobée (1999), cette méthode accorde moins de poids aux valeurs rares, même s'il s'avère qu'elles sont exactes.

1.4.3. Estimation d'un quantile au site cible

Lorsque l'on cherche maintenant à connaître la distribution $Q_s(T)$ à un site peu jaugé ou non jaugé, appelé « site cible », il suffit de connaître la distribution régionale $Q_R^*(T)$ de sa région homogène, ainsi que son indice de crue μ_s .

$$Q_s(T) = Q_R^*(T) \cdot \mu_s \quad (5-8)$$

Si le site cible est jaugé, l'indice est calculé à partir des mesures de débit disponibles. Si le site n'est pas jaugé, il est estimé à partir de caractéristiques physiques, grâce à des formules de régressions établies au sein de la région homogène (*cf.* 1.3.).

1.5. Remarques sur la notion de région hydrologique homogène

Lors de la définition des régions hydrologiques homogènes (§1.2.) nous n'avons pas détaillé les critères définissant cette homogénéité. En effet ils dépendent de la méthode appliquée : méthode des régressions ou celle de l'indice de crue, par exemple.

Pour la méthode des régressions multiples, il s'agit de définir la région qui donne les meilleurs résultats. Par exemple Ouarda *et al.* (1999ab) définissent des régions d'influence par l'analyse des corrélations canoniques puis appliquent la méthode des régressions multiples.

La distance critique au delà de laquelle ils considèrent qu'un bassin ne fait pas partie de la région d'influence du bassin cible est optimisée en minimisant l'erreur d'estimation des régressions. Selon Bernier (Rasmussen *et al.*, 1995), le découpage en région homogène à l'aide de variables physiques/climatiques, puis l'établissement de régressions au sein de chacune de ces régions à partir de ces mêmes variables n'est pas toujours logique. En effet, leur signification est moins importante dans les régressions, puisqu'elles ont déjà été utilisées dans la phase de constitutions des régions. Par exemple, si on sélectionne tous les bassins ayant exactement la même surface, il est clair que cette variable ne sera plus significative.

Pour ce qui est de la méthode de l'indice de crue, la notion de région hydrologique homogène repose sur une motivation un peu différente. Il s'agit de regrouper les bassins ayant la même distribution des débits de pointes, une fois que celle-ci a été normée par l'indice de crue. Pour ce faire, on étudie souvent les coefficients de variations, calculés sur les débits maximaux annuels ($C_v = \text{écart type de l'échantillon} / \text{moyenne}$). L'écart type se trouve être proportionnel au paramètre d'échelle a (aussi appelé gradex) de la loi de Gumbel ($a = 0,78\sigma$, voir Annexe 1 pour plus de détails sur la loi de Gumbel). La moyenne de l'échantillon est souvent prise comme indice de crue. Donc la valeur de C_v est proportionnelle à la pente de la distribution normée, représentée sur un graphique de Gumbel.

Dans ces conditions, chercher tous les bassins ayant la même distribution normée revient à chercher les bassins dont les échantillons de crues ont le même C_v . Si l'on affine la comparaison, on peut aussi considérer les coefficients d'asymétrie C_s qui sont calculés à partir d'un moment de plus (moments d'ordre 1, 2 et 3). Ceci explique pourquoi l'homogénéité de la région est souvent basée sur la comparaison des deux caractéristiques C_v et C_s .

Différents tests existent pour savoir si l'hypothèse que ces deux coefficients sont constants au sein de la région est acceptable ou pas (Wiltshire 1986). Avec l'introduction des *L-moments*, plus robustes que les moments classiques tout en gardant la même signification statistique ($L-C_v$ et $L-C_s$), Hosking et Wallis (1993) proposent trois tests : une mesure de discordance d'un site particulier, une mesure de l'hétérogénéité de la région et une mesure de la qualité d'ajustement (goodness-of-fit) pour déterminer le choix de la distribution régionale.

Comme le note le GREHYS (1996b), les résultats de la méthode de l'indice de crue sont très sensibles à l'homogénéité de la région. A l'inverse, la méthode des régressions multiples permet de prendre en compte une certaine hétérogénéité ; par contre, elle ne permet pas d'extrapoler la distribution des débits.

2. Généralisation de la méthode de l'indice de crue : le modèle « régional » convergent

2.1. Présentation du modèle régional convergent

Le modèle local convergent défini au chapitre 3 peut maintenant être ajusté de façon « régionale », de la même façon que la distribution régionale relative à la méthode de l'indice de crue.

Les étapes sont :

- définir une région hydrologique homogène pour un site cible ;
- ajuster localement en chaque site i de la région le modèle convergent *QdF*. Le chapitre 3 a montré que le modèle pouvait s'écrire sous sa forme simplifiée ($P=0$), cependant nous restons ici dans le cas général :

$$V_i(d, T) = \frac{V_i(0, T) - P_i}{1 + d / \Delta_i} + P_i \quad (5-9)$$

Les paramètres à ajuster sont la durée caractéristique Δ_i , ainsi que ceux définissant la distribution $V_i(0, T)$, lorsque d est égale à zéro.

- calculer la distribution adimensionnelle régionale $V_R^*(0, T)$ à partir de toutes les stations i de la région, excepté le site cible s'il est non-jaugé ; ainsi que le paramètre régional P_R^* .

- calculer la distribution $V_s(0, T)$, au site cible en dénormant la distribution régionale $V_R^*(0, T)$ et le paramètre régional P_R^* :

$$V_s(0, T) = V_R^*(0, T) \cdot \mu_s ; \quad P_s = P_R^* \cdot \mu_s \quad (5-10)$$

- déduire les quantiles $V_s(d, T)$ à partir de la distribution $V_s(0, T)$ grâce au modèle *QdF* convergent :

$$V_s(d, T) = \frac{V_s(0, T) - P_s}{1 + d / \Delta_s} + P_s \quad (5-11)$$

Ainsi, les paramètres régionaux à connaître sont :

- $V_R^*(0, T)$: la distribution régionale adimensionnelle, relative à la durée instantanée, déterminée à l'aide des sites i de la région.
- P_R^* : le paramètre donnant la limite de $V(d, T)$ lorsque d tend vers l'infini.

Les paramètres locaux, relatifs au site cible, à estimer à partir de ses caractéristiques climato-physiographiques, sont :

- Δ_s : la durée caractéristique de crue du site cible
- μ_s : l'indice de crue du site cible

Remarque

Le modèle régional convergent défini ci-dessus revient à calculer pour toute une région donnée les courbes débit-durée-fréquence adimensionnelles suivantes :

$$V_R^*(d^*, T) = \frac{V_R^*(0, T) - P_R^*}{1 + d^*} + P_R^* \quad (d^* = d/\Delta) \quad (5-12)$$

et à les dénormer au site cible par ses caractéristiques μ_s et Δ_s :

$$V_R(d = d^* \cdot \Delta_s, T) = \mu_s \cdot V_R^*(d^*, T) \quad (5-13)$$

comme l'illustre la Figure 5-2.

En ce sens, le modèle régional convergent est une généralisation à l'approche débit-durée-fréquence de la méthode de l'indice de crue, décrite précédemment.

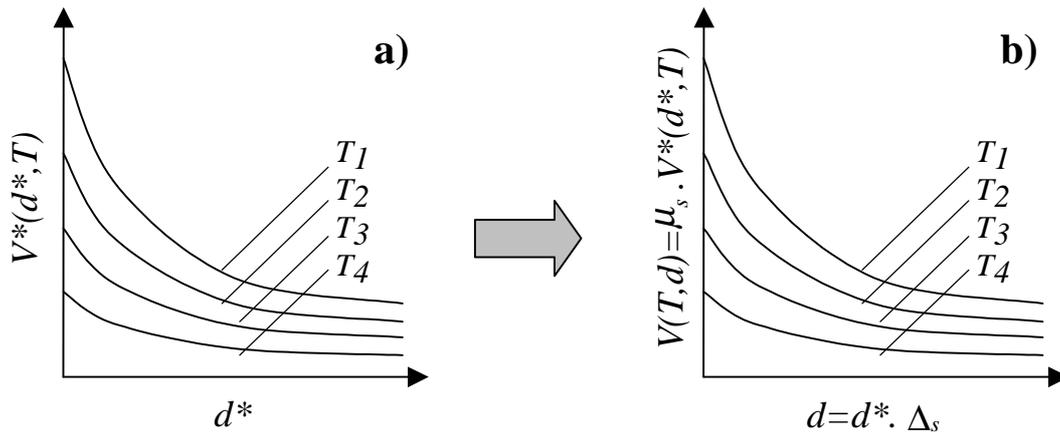


Figure 5-2 : Courbes QdF adimensionnelles régionales (a) dénormées localement à partir des caractéristiques μ_s et Δ_s du bassin cible (b)

2.2. Ajustement de la distribution $V_R^*(0, T)$ régionale : application à la loi GEV

Comme nous le verrons au paragraphe 3. , ce modèle régional est développé sur des données du Québec et de l'Ontario, pour lesquelles les régions hydrologiques homogènes ont déjà été définies et la méthode de l'indice de crue déjà été appliquée (GREHYS, 1996ab). Pour cette raison, nous présentons ici le modèle convergent appliqué à la loi *GEV*, présenté au chapitre 4. En effet, c'est cette loi qui a été utilisée lors de l'étude du GREHYS, et nous souhaitons rester cohérent avec ce choix. De même la norme μ adoptée pour l'indice de crue est relative à la moyenne des échantillons. Ce paragraphe explicite le calcul de ces différents paramètres.

2.2.1. Rappels sur l'ajustement du modèle local convergent

Comme nous l'avons vu au chapitre 4, si l'on fait le choix de la loi *GEV* dans le modèle convergent (5-9), alors la distribution locale $V_i(0, T)$ du site i , que nous noterons simplement $V(0, T)$, est donnée par :

$$V(0, T) = x_0(0) + \frac{a(0)}{k(0)} \left\{ 1 - [-\ln(1 - 1/T)]^{k(0)} \right\} \quad (5-14)$$

$$V(d, T) = \frac{V(0, T) - P}{1 + d/\Delta} + P \quad (5-15)$$

Les paramètres du modèle local à ajuster sont donc :

Δ : la durée caractéristique de crue, en jours ; $x_0(0)$, $a(0)$, $k(0)$: respectivement paramètres de position, d'échelle et de forme de la distribution relative à $d=0$. Le chapitre 3 a montré que le paramètre P pouvait être fixé à zéro. Nous adopterons cette valeur pour simplifier la suite des calculs.

Comme nous l'avons vu au chapitre 3, l'ajustement se fait en deux étapes : Δ , puis les paramètres de la distribution $V(d=0, T)$. Δ est déterminé de façon à ce que l'échantillon global constitué de l'ensemble des débits instantanés « équivalents », donné par la relation :

$$V(d = 0, T) = (1 + d / \Delta)V(d, T) \quad (5-16)$$

soit le plus homogène possible.

Les paramètres de la distribution $V(0, T)$ du modèle local convergent sont ensuite ajustés sur la moyenne des distributions expérimentales des débits instantanés « équivalents ». Pour cela, une distribution expérimentale $xmoy(j)$ est calculée en prenant pour chaque rang la moyenne des valeurs des distributions « ramenées à l'instantané », relatives aux N durées d :

$$xmoy(j) = \frac{1}{N} \sum_{i=1}^N [v_{d_i}(j) \cdot (1 + d_i / \Delta)] \quad (5-17)$$

Les paramètres $x_0(0)$, $a(0)$, $k(0)$ sont ensuite déterminés à partir de l'échantillon des $xmoy(j)$. Dans la suite de ce travail, nous avons utilisé la méthode des moments pondérés, décrite au chapitre 4.

2.2.2. Ajustement de la distribution régionale $V_R^*(0, T)$

L'adaptation de la méthode de l'indice de crue décrite au début de ce chapitre concerne la distribution $V(0, T)$ ajustée sur les échantillons $xmoy(j)$ en chacun des sites étudiés (cf. ci-dessus). De cette façon, la méthode de l'indice de crue présentée au paragraphe 1.4. peut être appliquée.

Tout d'abord, nous définissons pour chaque bassin étudié, l'indice de crue μ comme étant la moyenne des NV valeurs de l'échantillon $xmoy(j)$.

$$\mu = \frac{1}{NV} \sum_{j=1}^{NV} xmoy(j) \quad (5-18)$$

Puis, la méthode des moments pondérés régionaux décrite en début de ce chapitre est appliquée en considérant toutes les distributions expérimentales $xmoy(j)$ normées par l'indice de crue, pour chacun des bassins d'une même région.

3. Présentation des données étudiées

3.1. Les données disponibles

Les bassins ont été étudiés dans le cadre d'une collaboration avec l'INRS-Eau à Québec. Cette collaboration a été financée par le Consulat de France au Québec et visait entre autres à comparer les approches de régionalisation françaises et celles utilisées au Québec.

Au Québec, un précédent projet piloté par l'INRS-Eau a consisté à comparer en Ontario et au Québec les principales méthodes régionales existantes (GREHYS, 1996ab). La base de données ayant servi à cette étude et contenant des chroniques de débits journaliers ainsi que des caractéristiques physiques des bassins nous a donc été fournie par l'INRS-Eau.

La Figure 5-3 indique la localisation des 158 bassins étudiés, en fonction de leur surface. On constate que les grands bassins se situent majoritairement dans les territoires du Nord. La liste de bassins étudiés figure en annexe 6.

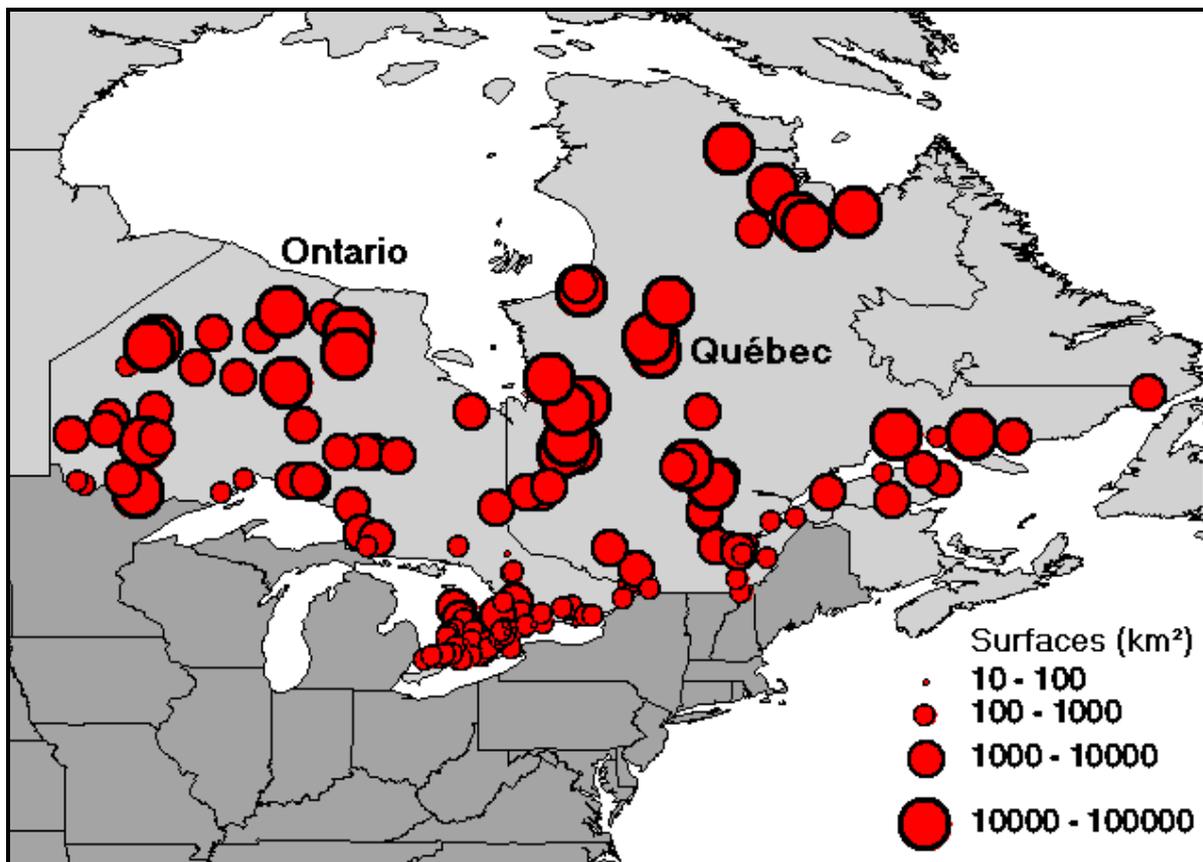


Figure 5-3 : Localisation des bassins étudiés

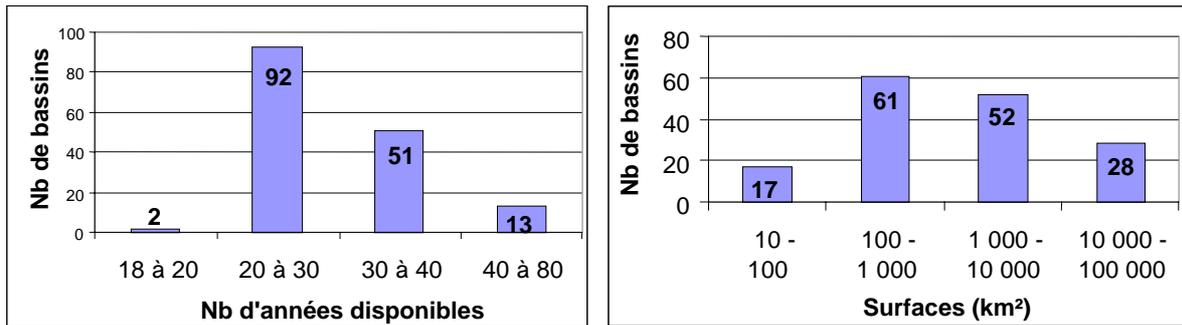


Figure 5-4 : Nombre d'années disponibles et surface des bassins étudiés

La Figure 5-4 donne les histogrammes des tailles des chroniques de débits journaliers disponibles, ainsi que des surfaces des bassins étudiés.

Pour chacun de ces bassins, des caractéristiques physiques ont aussi été fournies par l'INRS-Eau. Celles-ci sont résumées dans le Tableau 5-1.

<i>Québec</i>	<i>Ontario</i>
latitude, longitude	latitude, longitude
surface	surface
surface de lac	pente de la rivière
surface de forêt	longueur de la rivière
altitude moyenne du bassin	surface du bassin en amont des lacs
pente moyenne du bassin	

Tableau 5-1 : Liste des caractéristiques physiques des bassins disponibles

3.2. Etude préliminaire sur la saisonnalité et la dynamique des crues

Les bassins étudiés sont marqués par un régime pluvio-nival. Deux saisons de hautes eaux sont distinguées : l'une au printemps due à la fonte de la neige, l'autre à l'automne, due à une reprise des précipitations, pas encore sous forme de neige. Ce constat avait déjà pu être fait au chapitre 1, traitant de la notion de régime hydrologique, (voir l'exemple de la Rivière Nouvelle au Québec (1140 km²), Figure 1-8).

De ce fait, l'échantillonnage des crues doit impérativement être saisonnalisé afin de ne pas mélanger dans le même échantillon les crues de printemps et d'automne. Leur genèse étant différente, on suppose en effet que les populations dont elles sont issues le sont aussi.

De plus, comme le laisse supposer la grande variabilité des surfaces observées (Figure 5-4), les dynamiques de crues peuvent fortement varier d'un bassin à un autre. A titre d'exemple, la Figure 5-5 présente l'évolution sur une année du débit journalier de deux bassins, l'un de 78 km², l'autre d'environ 50 000 km². On constate effectivement dans le premier cas des crues rapides, avec une première saison au printemps caractérisée par plusieurs événements, probablement dus à la succession de redoux (fonte éventuellement accompagnée de pluie) et

regels. Puis à l'automne, de nouveau quelques crues, moins fortes mais plus rapides, provoquées uniquement par les précipitations. A l'inverse, pour le bassin de 40 000 km², la dynamique est tellement lente, qu'on n'observe qu'une seule crue dans l'année, au printemps, s'étalant jusqu'à l'automne, durant lequel la reprise de la pluviométrie se manifeste par quelques sursauts de la courbe de récession.

Ces deux exemples montrent que :

- les crues les plus importantes ont principalement lieu au printemps ;
- la dynamique de crue peut fortement varier d'un bassin à un autre.

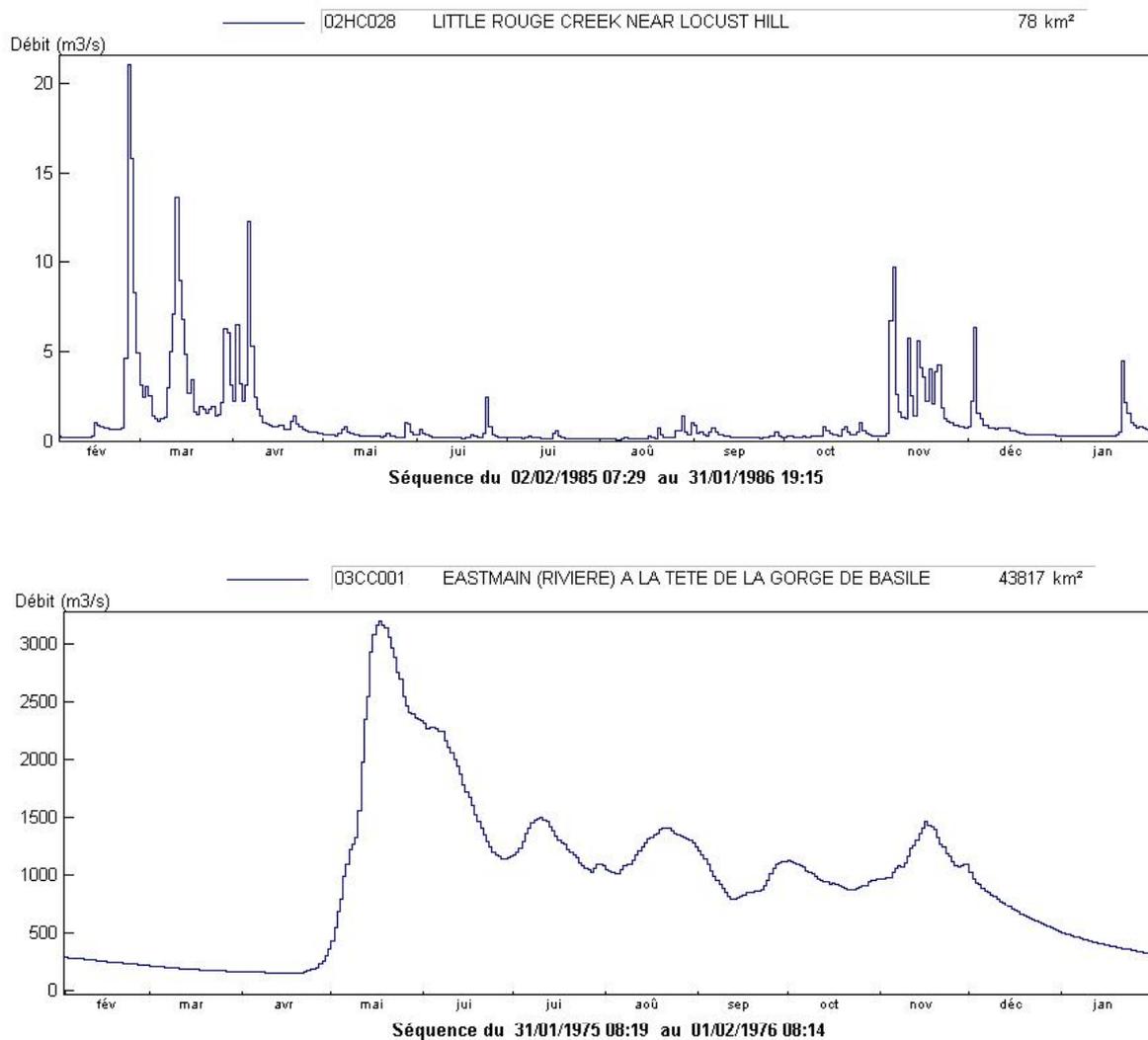


Figure 5-5 : Evolution du débit journalier : comparaison relative à un bassin de 78 km² et de 43 817 km²

Suite à cette étude préliminaire, nous avons décidé d'effectuer un échantillonnage par maximum annuel, en **ne considérant que les valeurs maximales printanières**. En effet, la méthode du maximum annuel est préférable à celle des valeurs supérieures à un seuil conseillée au chapitre 2, puisqu'ici on n'observe parfois qu'une seule pointe dans l'année. La

saisonnalisation de l'échantillonnage est nécessaire pour ne pas mélanger des crues de printemps et d'automne, (Figure 5-5).

Concernant la forte variabilité des durées de crues observées d'un bassin à un autre, nous avons adopté la méthode développée au chapitre 3. **L'échantillonnage multi-durée est réalisé sur des durées n'excédant pas environ 3 fois la durée Δ** . De cette façon, on estime rester dans la durée des crues, comme celles représentées Figure 5-5.

4. Résultats du modèle convergent (local et régional)

4.1. Résultats du modèle QdF local

Le modèle local convergent précédemment décrit a tout d'abord été testé sur les 158 bassins décrits ci-dessus. Un échantillonnage « multi-durée » a permis d'obtenir les quantiles $V_d(T)$ pour quatre durées différentes, allant de 1 jour à une durée d_d voisine de la dynamique des crues observées sur chaque bassin. Cette durée d_d dépend donc du bassin étudié et a été déterminée en l'indexant sur Δ (cf. 3.2.)

Puis le modèle QdF local a été ajusté, permettant d'obtenir les quantiles $V(d,T)$ pour chaque bassin. Afin de représenter les résultats obtenus pour les différents bassins sur un même graphique, on peut remarquer que l'équation (5-9) peut se mettre sous la forme :

$$\frac{V(d,T)}{V(0,T)} = \frac{1}{1 + d/\Delta} \quad (5-19)$$

Ainsi, la Figure 5-6 représente les quantiles observés $V_d(T)$ normés par les quantiles $V(0,T)$ calculés par le modèle QdF, en fonction de d/Δ . De cette façon, si le modèle est validé, les quantiles normés doivent s'ajuster sur l'hyperbole d'équation $y(x)=1/(1+x)$.

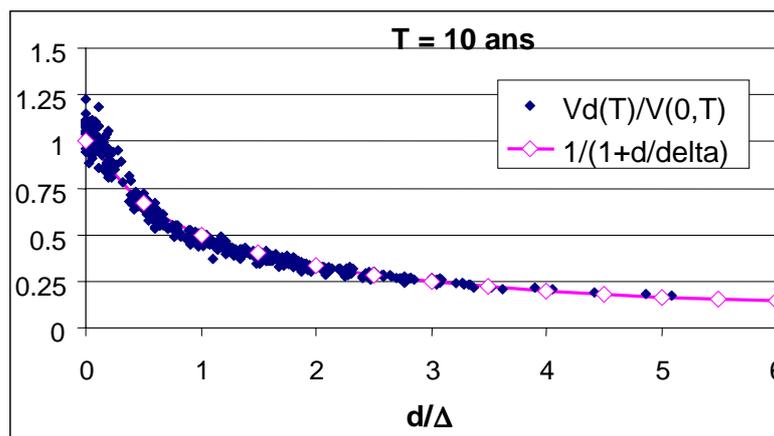


Figure 5-6 : Quantiles observés $V_d(T)$ normés par $V(0,T)$, en fonction de d/Δ

Ces résultats montrent que pour l'exemple de la période de retour 10 ans, le modèle est satisfaisant et en particulier que la forme hyperbolique en $1/(1+d/\Delta)$ modélise bien l'évolution des quantiles $V_d(T)$ en fonction de d .

Afin d'avoir une synthèse de ces résultats pour différentes périodes de retour, l'erreur quadratique moyenne a aussi été calculée, pour les 158 bassins i et les 4 durées d_j étudiées :

$$rRMSE(T) = \left\{ \frac{1}{158} \frac{1}{4} \sum_{i=1}^{158} \sum_{j=1}^4 \left[\frac{V(d_j, T) - V_{d_j}(T)}{V_{d_j}(T)} \right]^2 \right\}_{\text{bassin } i}^{1/2} \quad (5-20)$$

L'intérêt de cette mesure est qu'elle prend à la fois en compte le biais et la dispersion du modèle. La Figure 5-7 indique les résultats obtenus pour les périodes de retour de 1.1, 2, 5, 10 et 100 ans et montre que ceux-ci sont tout à fait corrects.

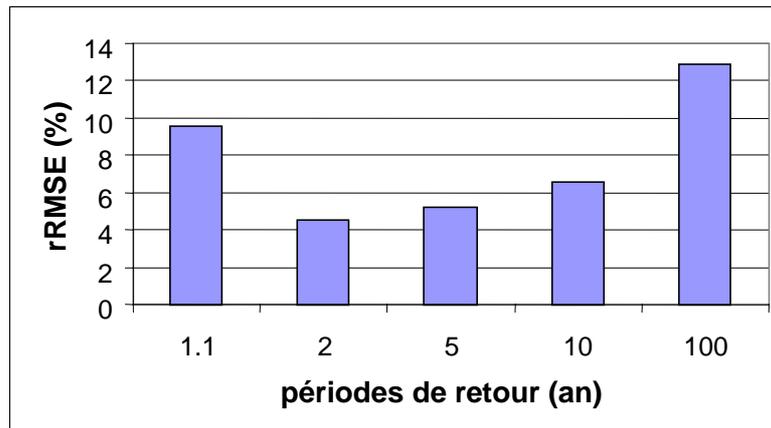


Figure 5-7 : Erreur quadratique moyenne entre les quantiles locaux non consolidés $V_d(T)$ et ceux calculés par le modèle local convergent $V(d,T)$

Le modèle local convergent que nous venons de tester donne donc des résultats satisfaisants. Par rapport au chapitre 3, ce paragraphe apporte les éléments nouveaux suivants :

- D'une part, le modèle est applicable dans un contexte climatique très différent de celui des bassins français jusqu'alors étudiés.
- D'autre part, nous avons appliqué sur un grand nombre de bassin (158) la loi *GEV*, qui n'avait été testée que sur les deux exemples du chapitre 4.
- Enfin, la particularité climatique de la région nous a conduit à effectuer un échantillonnage saisonnalisé, chose qui n'avait pas encore été faite aux chapitres précédents.

4.2. Résultats du modèle QdF régional

4.2.1. Constitution de régions hydrologiques homogènes

Maintenant que le modèle *QdF* a été validé localement, nous allons montrer les résultats obtenus en appliquant le modèle *QdF* régional. Pour cela, sept régions hydrologiques homogènes ont été définies au Québec et cinq en Ontario, dans le cadre du projet GREHYS (1996ab). Il s'agit de régions fixes non contiguës, calculées par Gingras *et al.* (1994) à partir des données physiques de bassin. Pour le Québec, les données utilisées dans l'analyse sont la surface, les pourcentages de forêt et de lac, l'altitude et la pente moyenne du bassin. Pour

l'Ontario, les caractéristiques retenues sont la surface du bassin, la longueur et la pente moyenne de la rivière.

Notre but n'était pas de travailler sur la constitution des régions hydrologiques homogènes, mais dans un premier temps de montrer l'adaptabilité de la méthode de l'indice de crue dans le cas d'une approche débit-durée-fréquence. Il serait tout à fait possible de refaire le travail présenté sur la base d'un autre découpage. Les régions qui ont été utilisées ne le sont qu'à titre d'exemple. Mais, dans l'avenir, il serait bien sûr intéressant de travailler à la constitution de ces régions.

4.2.2. Composition d'un échantillon de calibration et de validation

Afin de valider le modèle régional convergent, nous avons divisé le groupe des 158 en un échantillon de calibration, un autre de validation. Ainsi la forme régionale adimensionnelle du modèle régional convergent a été ajustée pour chacune des régions définies par le GREHYS, en ne considérant que 105 bassins sur les 158 disponibles, les 53 autres servant uniquement à la validation. Une méthode pour éviter de couper en deux notre échantillon aurait été de faire une validation par « Jack Knife » : on ne valide que sur un seul bassin qu'on enlève. Puis on remet cette station dans l'échantillon, et on enlève la suivante. On répète cette procédure pour l'ensemble des bassins.

4.2.3. Résultats

Une fois que la distribution régionale $V_R^*(0,T)$ a été ajustée pour chaque région, selon la méthodologie définie au paragraphe 2.2.2. , il est possible d'obtenir des estimations régionales des quantiles $V(d,T)$ grâce aux relations (5-10) et (5-11), connaissant pour chaque bassin sa région d'appartenance ainsi que ses caractéristiques μ et Δ .

Les résultats présentés distinguent deux cas : le premier est celui d'une estimation « **locale-régionale** » telle que la définit Cunanne (1988). Nous dénormons en chaque site les courbes régionales adimensionnelles par les valeurs de μ et Δ obtenues **localement**. En comparant les résultats obtenus à ceux des quantiles locaux $V_d(T)$ précédemment calculés, cette façon de faire nous permet d'apprécier si les distributions régionales calculées sont bien représentatives des régions sur lesquelles elles ont été ajustées, ou si elle « dégradent » l'information locale.

Ensuite, nous effectuons une estimation « **purement régionale** » toujours au sens de Cunanne (1988). Dans ce cas, les caractéristiques μ et Δ sont déduites de régressions avec les caractéristiques physiques de terrain. Nous avons sélectionné les caractéristiques les plus pertinentes par la méthode Stepwise. Pour le Québec, μ est estimé à partir de la surface, de la longitude et du pourcentage de lac des bassins ; Δ à partir de la surface, de la longitude, et du pourcentage de lac. Pour l'Ontario, μ est estimé à partir de la surface, de la longitude et de la longueur de la rivière ; Δ à partir de la surface de la longitude et du pourcentage de lac. Nous

ne détaillerons pas plus ces régressions, car le chapitre 7 développe la méthode pour des bassins de la Moselle.

La Figure 5-8 montre les résultats obtenus sur l'échantillon de validation qui n'a servi ni au calcul des distributions régionales $V_R^*(0,T)$ ni à celui des régressions évoquées ci-dessus permettant d'estimer μ et Δ .

L'erreur quadratique entre les quantiles régionaux et les quantiles locaux est calculée pour les périodes moyennes de retour suivantes : 1.1, 2, 5, 10 et 100 ans. A chaque fois, le calcul tient compte des 4 durées d étudiées, comme nous l'avons fait précédemment pour le modèle local convergent (5-20). La période de retour de 100 a été étudiée pour rester cohérent avec l'étude du GREHYS (1996ab). Cependant, étant donnée la taille des chroniques, la valeur locale du quantile est la plupart du temps extrapolée, donc peu précise.

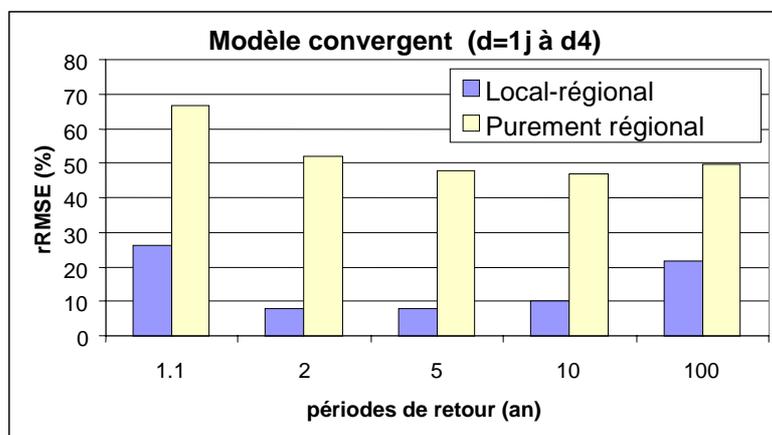


Figure 5-8 : Résultats du modèle régional convergent dénormé par μ et Δ locaux (local-régional) et par μ et Δ estimés à partir des régressions (purement régional)

Pour avoir une idée de la qualité des résultats ci-dessus, nous avons aussi appliqué la méthode de l'indice de crue « classique », c'est à dire uniquement sur la distribution des débits journaliers ($d=1$ jour). Dans ce cas, le modèle local convergent n'est pas utilisé. Nous avons calculé pour chaque région la distribution régionale $V_{d=1j R}^*(T)$, ainsi que des régressions permettant d'obtenir l'indice de crue $\mu_{d=1j}$ relatif aux débits journaliers. Les résultats obtenus, sont présentés Figure 5-9.

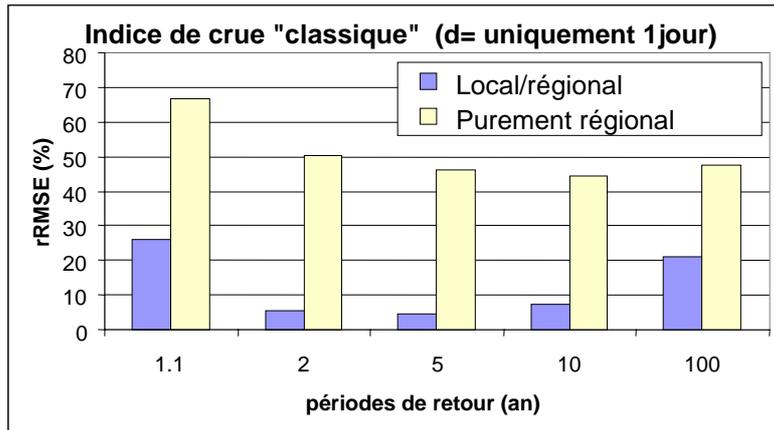


Figure 5-9 : Résultats obtenus avec la méthode de l'indice de crue classique appliquée sur les seuls débits journaliers ($d=1j$)

4.2.4. Commentaires

Les résultats ci-dessus appellent aux commentaires suivants. Tout d'abord l'erreur du modèle régional convergent calculée sur l'ensemble des 4 durées étudiées est comparable à l'erreur de la méthode de l'indice de crue « classique » (c'est à dire pour la seule durée journalière). Ceci est vrai aussi bien pour les estimations « locales-régionales » que « purement régionales ». Le modèle régional convergent n'entraîne donc pas une dégradation de l'erreur d'estimation. L'avantage par rapport à la méthode de l'indice de crue classique limitée aux quantiles de débit journalier, est que le modèle régional convergent permet de calculer tous les quantiles $V(d,T)$.

Enfin, on observe dans les deux cas (modèle régional convergent et méthode de l'indice de crue classique) une forte augmentation de l'erreur lorsque l'on passe des estimations « locales-régionale », aux estimations « purement régionale ». Ceci montre que les courbes régionales adimensionnelles sont bien représentatives des régions sur lesquelles elles ont été ajustées. Le modèle régional convergent parvient à « résumer » en un nombre minimal de paramètres (μ et Δ) toute l'information contenue dans les courbes débit-durée-fréquence d'un bassin. En revanche, l'estimation de ces caractéristiques μ et Δ par des formules de régressions entraîne une erreur importante.

Conclusion du chapitre 5

Ce chapitre nous a donné un point de vue nouveau, aussi bien par le contexte de cette étude, que par les concepts développés, liés à l'hydrologie régionale. Il a été l'occasion de travailler dans un cadre méthodologique différent, en collaboration avec la Chaire d'Hydrologie Statistique de l'INRS-Eau, à Québec.

Tout d'abord, concernant le **modèle local convergent**, la zone d'étude - le Québec et l'Ontario - a permis l'application de ce modèle à des bassins soumis à des conditions climatiques complètement différentes de celles des bassins étudiés aux chapitres 3 et 4. De plus la loi *GEV* a été utilisée ainsi qu'un échantillonnage saisonnalisé, printanier, chose qui n'avait pas encore été réalisée aux chapitres précédents. Cette partie du travail montre donc que le modèle local convergent n'est pas lié à un type de loi ou à une méthode d'échantillonnage, mais peut être adapté dans des contextes spécifiques.

Ensuite, le principal apport de ce chapitre est le développement du **modèle régional convergent**. Pour ce faire, les concepts classiques de l'analyse fréquentielle régionale des crues ont été utilisés. Le modèle obtenu « résume » pour une région hydrologique homogène donnée les courbes débit-durée-fréquence de chaque bassin par seulement deux paramètres : un débit noté μ et la durée caractéristique Δ . Le modèle proposé généralise la méthode de l'indice de crue « classique », sans pour autant en dégrader les résultats. Cette dernière ne permet de calculer les distributions que pour les débits journaliers, alors que le modèle régional convergent permet l'estimation de ces mêmes distributions, pour toute durée d .

Concernant la qualité des résultats, les estimations « purement régionales », c'est à dire en n'utilisant aucune donnée locale amènent à de grosses erreurs. Inversement, les estimations simplement « locales-régionales » c'est à dire pour lesquelles on utilise les valeurs locales de μ et Δ donnent des erreurs acceptables. Ceci montre que la principale source d'erreur des estimations « purement régionales » est due aux régressions. C'est pourquoi une attention toute particulière doit être portée à l'établissement de ces régressions. Ce chapitre n'a pas traité ce problème dans le détail. Il le sera au chapitre 7, pour les sous-bassins de la Moselle.

Comme nous le verrons au chapitre 6, le modèle régional convergent présenté ici possède beaucoup de points communs avec l'approche débit-durée-fréquence développée au Cemagref avant cette thèse (Galéa et Prudhomme, 1997), et dont nous n'avons pas encore parlé. Le chapitre suivant propose de comparer ces deux approches.

Chapitre 6

Modélisation *QdF* « locale-régionale » : comparaison avec l'approche « bassin de référence »

Le chapitre précédent a présenté le modèle régional convergent, s'ajustant sur un ensemble de sites formant une région hydrologique homogène. Sa formulation adimensionnelle est dénormée localement par deux variables : un débit et une durée caractéristique du bassin étudié.

Cette façon de faire comporte des analogies avec l'approche débit-durée-fréquence précédemment développée au Cemagref et dont cette thèse est un prolongement : l'approche « bassin de référence » (Galéa et Prudhomme, 1997). Ce chapitre compare ces deux approches, afin de déterminer leurs points communs et leurs différences. Nous allons considérer :

- d'abord les estimations « **locales-régionales** » au sens de Cunnane (1988), c'est à dire en utilisant une l'information hydrologique locale pour estimer le débit et la durée caractéristique requise par chacune de deux approches
- puis les estimations « **purement régionales** », c'est à dire sans utiliser aucune donnée hydrologique locale.

Ce chapitre concerne le premier point : l'estimation « locale-régionale ». Le cas de l'estimation « purement régionale », plus difficile, est traité dans le dernier chapitre de cette thèse.

L'approche « bassin de référence » est présentée, puis appliquée sur une cinquantaine de bassins français. Ses résultats sont ensuite comparés à ceux du modèle convergent *local*, puis *régional*.

1. Rappels sur l'approche *QdF* « bassin de référence »

L'approche « bassin de référence » (ce nom sera explicité par la suite) est l'approche débit-durée-fréquence développée au Cemagref par Prudhomme (1995) et Galéa et Prudhomme (1997). Ce paragraphe fait de brefs rappels sur le modèle élaboré pour les débits moyens notés $V_d(T)$. Les mêmes développements existent pour les débits seuils $Q_d(T)$, mais nous n'en parlerons pas ici. Les sous paragraphes suivants détaillent les différentes étapes, en prenant bien soin de faire la distinction entre l'analyse **locale** et l'analyse **régionale**, selon les principes définis au paragraphe 1 du chapitre 5.

1.1. Analyse locale

- **Analyse fréquentielle : le modèle local non consolidé**

Le modèle que nous appelons « local non consolidé » est celui dont nous sommes partis, au chapitre 3, pour établir le modèle local convergent (6-1), à la différence près que le premier considère aussi le cas des fréquences rares ($T > T_g$) en faisant une extrapolation par la méthode dite du « gradex esthétique » (6-2) (Michel, 1982) :

$$\text{Pour } 0.5 \leq T(\text{an}) \leq T_g : V_d(T) = [a_q]_d \ln(T) + [x_0]_d \quad (6-1)$$

$$\text{Pour } T > T_g : V_d(T) = V_d(T_g) + [a_p]_d \ln \left(1 + \frac{[a_q]_d}{[a_p]_d} \frac{T - T_g}{T_g} \right) \quad (6-2)$$

avec : $[x_0]_d$ et $[a_q]_d$, respectivement le paramètre de position et d'échelle (ou *gradex*) de la loi exponentielle calée sur les débits observés, pour la durée d fixée, en m^3/s ;

$[a_p]_d$ paramètre d'échelle (ou *gradex*) des pluies, sur la durée d fixée (loi exponentielle), ramené en m^3/s à partir de la surface du bassin et de d ;

T_g : seuil fréquentiel à partir duquel la loi des débits est extrapolée à l'aide de la loi des pluies, généralement fixé à 10 ou 20 ans ;

$\ln(T)$: logarithme népérien de T .

La méthode du gradex esthétique est maintenant formalisée de manière plus générale par le modèle Agrégée (Margoum *et al.*, 1994). Ces deux modèles, gradex esthétique et Agrégée, sont tous deux inspirés de la méthode du gradex développée par Guillot et Duband (1967). Le principe est de rendre la distribution des débits parallèle à celle des pluies, cette dernière étant exprimée en m^3/s à l'aide de la surface du bassin et de la durée d . Par rapport à la méthode du gradex, celle du gradex esthétique opère un raccordement plus progressif entre observations et extrapolations (Figure 6-1).

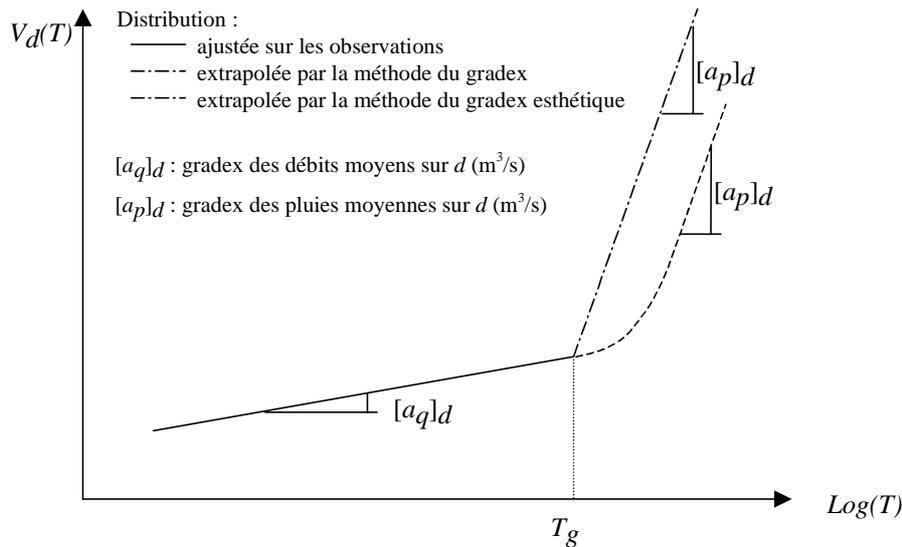


Figure 6-1 : Méthode d'extrapolation du gradex et celle du gradex esthétique

- **Analyse en fonction de la durée d : le modèle local continu**

La seconde étape consiste à étudier la variation des quantiles $V_d(T)$ en fonction de d . Pour cela, les paramètres $[x_0]d$, $[a_q]d$, $[a_p]d$ définis en (6-1) et (6-2) sont remplacés par des fonctions $x_0(d)$, $a_q(d)$ et $a_p(d)$ de la forme suivante :

$$y(d) = \frac{1}{\alpha d + \beta} + \gamma \quad (6-3)$$

Pour chacune des trois fonctions $x_0(d)$, $a_q(d)$ et $a_p(d)$, les paramètres α , β et γ sont ajustés (soit 9 paramètres, pour chaque bassin étudié). A cause de ce grand nombre de paramètres, ce modèle, que nous appelons « modèle local continu » a été très peu utilisé. L'approche régionale de type « bassin de référence », décrite ci-après, lui a été préférée, même dans le cas de bassins observés.

1.2. Analyse régionale : l'approche « bassin de référence »

- **Construction de modèles régionaux relatifs à un bassin de référence**

L'approche type « bassin de référence » consiste à comparer les courbes *QdF* adimensionnalisées de différents bassins, puis à constituer plusieurs familles. Pour chacune d'elles, un bassin a été choisi comme référence, d'où le nom que nous avons donné à cette approche.

Afin de rendre adimensionnelles les courbes *QdF*, deux normes s'avèrent nécessaires : une pour les débits, une autre pour les durées. Après plusieurs essais, le débit de pointe décennal) et une durée caractéristique de crue (D) du bassin étudié ont été choisis. La durée D est définie au sens de la méthode *Socose* (CTGREF, 1980). Cette durée a déjà été présentée au chapitre 1. Nous noterons le débit de pointe décennal $Q(0,10)$. Comme nous l'avons souligné

au chapitre 4 (cf. Figure 4-5), pour la durée d égale zéro, les débits seuils (notés Q) et les débits moyens (notés V) sont équivalents. On a donc égalité entre $Q(0,10)$ et $V(0,10)$.

Si l'on raisonne avec les valeurs adimensionnelles définies ci-dessus, on obtient :

$$d^* = d / D ; V^*(d^*, T) = V(d, T) / Q(0,10) \quad (6-4)$$

$$\text{Pour } 0.5 \leq T(\text{an}) \leq 20: V^*(d^*, T) = a_q^*(d^*) \ln T + x_0^*(d^*) \quad (6-5)$$

$$\text{Pour } T > 20 \text{ ans : } V^*(d^*, T) = V^*(d^*, T_g) + a_p^*(d^*) \ln \left(\frac{a_q^*(d^*) T - T_g}{a_p^*(d^*) T_g} \right) \quad (6-6)$$

$$\text{avec : } a_q^*(d^*) = \frac{1}{x_1 d^* + x_2} + x_3 ; x_0^*(d^*) = \frac{1}{x_4 d^* + x_5} + x_6 ; a_p^*(d^*) = \frac{1}{x_7 d^* + x_8} + x_9 \quad (6-7)$$

Des études effectuées sur de nombreux bassins (français en majorité) ont permis d'identifier trois familles de courbes *QdF* adimensionnelles. Pour chacune d'elles, un bassin a été jugé représentatif, et a été choisi comme référence. Il s'agit de la Mimente à Florac, du Roubion à Soyans, et de la Dragne à Vandenesse, situés respectivement dans les Cévennes, la Drôme et en Bourgogne. Pour ces trois bassins, l'analyse en fonction de d^* précédemment évoquée a été effectuée. Pour chacun d'eux, neuf paramètres x_i ($i=1...9$) ont été ajustés, permettant de déterminer les courbes *QdF* adimensionnelles dites de Florac, de Soyans et de Vandenesse, notées $Q_{\text{Florac}}^*(d^*, T)$, $Q_{\text{Soyans}}^*(d^*, T)$ et $Q_{\text{Vandenesse}}^*(d^*, T)$ (Tableau 6-1). Par abus de langage, on parle des modèles de **Florac**, **Soyans** et **Vandenesse**, mais il s'agit en fait de trois paramétrisations différentes des fonctions définies par les équations (6-4) à (6-7)

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
Vandenesse	2.635	6.190	0.016	1.045	2.385	0.172	1.083	1.750	0.000
Florac	1.120	3.560	0.000	0.950	3.180	0.039	1.560	1.910	0.085
Soyans	0.870	4.600	0.000	1.070	2.500	0.099	0.569	0.690	0.046

Tableau 6-1 : Paramètres des trois modèles *QdF* « bassins de référence » définis pour les débits moyens $V(d, T)$

• Application à un site quelconque

Les trois modèles régionaux de type « bassin de référence » peuvent être utilisés pour estimer les quantiles $V(d, T)$ d'un site quelconque, observé ou non. Les étapes suivantes sont distinguées.

1. Chaque bassin est caractérisé par trois paramètres locaux : la durée caractéristique de crue de *Socose* (D), le débit de pointe décennal $Q(0,10)$ et les valeurs du gradex des pluies $[a_p]_d$ relatives à plusieurs durées d . Les deux premiers paramètres sont estimés directement à partir des mesures de débit lorsqu'elles existent, ou bien à partir de formules empiriques. Celles-ci ont été établies à l'échelle nationale ou régionale et utilisent des variables explicatives facilement estimables, telle que la surface, ainsi que des valeurs

caractéristiques de pluie (CTGREF, 1980). Le gradex des pluies peut être obtenu auprès de Météo-France ou lorsqu'il n'y a pas de données disponibles en consultant des études de synthèse, comme par exemple celle de Bois *et al.* (1995).

2. Choix de l'un des trois modèles type « bassins de référence » (Tableau 6-1): pour différentes valeurs de durées d , le rapport $[a_p]_d/Q(0,10)$ adimensionnel est comparé à deux limites $L_1(d^*)$ et $L_2(d^*)$ (Figure 6-2). Suivant la position de ce rapport (*inférieur à L_1 , compris entre L_1 et L_2 , ou supérieur à L_2*), on choisit le bassin de référence *Vandenesse*, *Florac* ou *Soyans*. Les limites $L_1(d^*)$ et $L_2(d^*)$ sont de la forme $f(x)=1/(a.x+b)$ et ont été déterminées expérimentalement par Prudhomme (1995) :

$$L_1(d^*) = \frac{1}{0.419d^* + 1.580} ; \quad L_2(d^*) = \frac{1}{0.768d^* + 2.332} \quad (6-8)$$

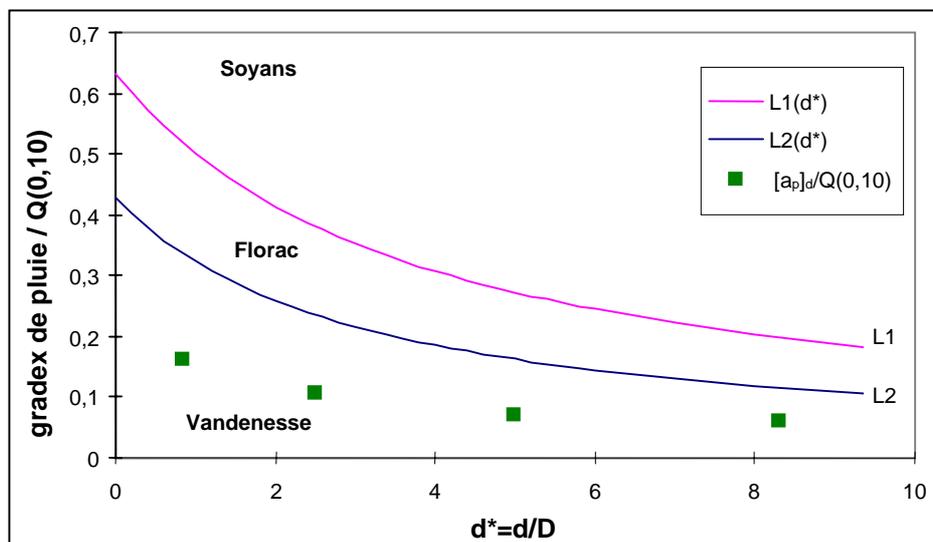


Figure 6-2 : Critère de choix du bassin de référence. Exemple de la Moselle à Rupt-sur-Moselle (153 km²)

3. Calcul des quantiles adimensionnels (formules(6-5), (6-6) et (6-7)) relatifs au bassin de référence choisi (Tableau 6-1) et restitution des quantiles $V(d,T)$ locaux, en dénormant ces quantiles adimensionnels par les paramètres du site cible ($Q(0,10)$ et la durée D) :

$$V(d,T) = V_{ref}^*(d^*,T) \cdot Q(0,10) ; \quad d = d^* \cdot D \quad (6-9)$$

avec : $ref = Florac, Soyans \text{ ou } Vandenesse$

1.3. Remarques sur l'approche « bassin de référence »

- **Lien avec la méthode de l'indice de crue**

Comme le soulignent Ouarda *et al.* (1999), il existe une analogie entre l'approche *QdF* type « bassins de référence » et la méthode de l'indice de crue : dans les deux cas, des quantiles adimensionnels sont dénormés par les caractéristiques du bassin cible. L'approche *QdF* utilise une durée caractéristique et le débit instantané décennal. La méthode de l'indice de crue, qui

ne considère que la distribution des débits instantanés ou journaliers, n'utilise qu'un quantile d'une période de retour d'environ deux ans.

Cependant deux différences subsistent. D'une part, les trois modèles régionaux *QdF* relatifs à un bassin de référence ont chacun été calés, comme leur nom l'indique, sur un seul bassin (*Soyans*, *Florac*, ou *Vandenesse*). A l'inverse, la distribution régionale de la méthode de l'indice de crue est déterminée à partir de tous les bassins jaugés de la région. Dans l'approche *QdF* décrite ci-dessus, le transfert d'information n'est donc pas, à proprement parler, un transfert "régional" mais plutôt un transfert du bassin de référence vers le bassin étudié.

D'autre part, l'approche *QdF*, comme cela est couramment pratiqué en France, tient compte de l'information pluviométrique locale pour extrapoler les distributions de débit aux fréquences rares. Cette information est introduite explicitement pour les trois bassins de référence et implicitement pour un bassin quelconque, grâce au critère de choix. Rappelons que ce critère utilise des gradex de pluie sur différentes durées (Figure 6-2). Pour la méthode de l'indice de crue, l'information pluviométrique peut être utilisée pour constituer les régions hydrologiques homogènes, mais pas pour extrapoler les distributions des débits. L'extrapolation aux fréquences rares est faite uniquement à l'aide de l'information sur les débits, consolidée par l'approche régionale.

- **Lien avec le modèle convergent (local et régional)**

Comme nous l'avons vu, le modèle **local** convergent s'ajuste sur l'ensemble des données disponibles en chaque site observé, alors que l'approche « bassin de référence » effectue un transfert de l'information du bassin de référence vers le bassin étudié. Nous ne nous situons donc pas au même niveau d'analyse : l'un est local, l'autre régional.

Concernant maintenant le modèle **régional** convergent présenté au chapitre 5, les deux analyses sont cette fois régionales. Dans les deux cas, des courbes *QdF* adimensionnelles sont dénormées par une caractéristique de débit, le débit instantané décennal $Q(0,10)$ et une durée caractéristique : la durée D de Socose (approche « bassin de référence ») et la durée Δ (modèle convergent). Cependant, la principale différence réside dans le fait que pour l'approche « bassin de référence », les courbes *QdF* adimensionnelles sont ajustées sur un seul bassin, le bassin de référence, alors que le modèle régional convergent est inspiré de la méthode de l'indice de crue et utilise toute l'information régionale disponible.

Enfin, concernant la propriété de convergence des distributions, bien que celle-ci ne fasse pas partie des hypothèses de l'approche « bassin de référence », elle est implicitement vérifiée, comme le montre la Figure 6-3 présentant les quantiles $V^*(d^*,T)$ calculés à partir des paramètres x_1 à x_6 du Tableau 6-1.

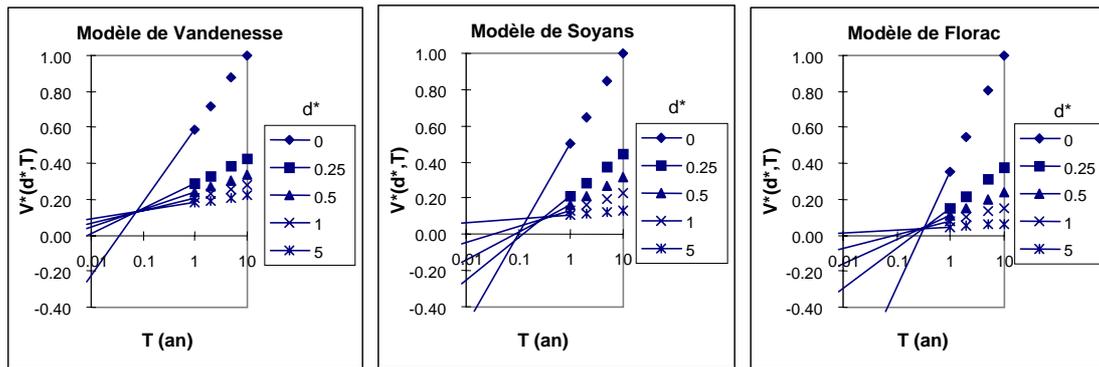


Figure 6-3 : Convergence des distributions observées sur chacun des trois modèles de référence

Dans le paragraphe suivant, nous comparons les résultats de l'approche « bassin de référence » avec ceux du modèle convergent local ou régional.

Cette comparaison est effectuée uniquement pour des périodes de retour inférieures ou égales à 20 ans. Au-delà, les résultats dépendent de la méthode d'extrapolation adoptée.

2. Données étudiées et méthodologie de comparaison

2.1. Données

Les bassins étudiés, de surfaces comprises entre 10 et 500 km² se situent en France et ont été regroupés au sein de deux échantillons. Le premier, que nous appelons « Echantillon France » contient 27 bassins. Ces bassins (plus quelques autres dont les données n'étaient plus disponibles) ont été utilisés par Prudhomme (1995) pour définir les trois typologies et les limites L_1 et L_2 de l'approche bassin de référence (Figure 6-2). Un second échantillon, appelé Echantillon « Moselle », est constitué de 22 bassins situés dans le bassin hydrographique de la Moselle (Figure 6-4). Ce second échantillon a été constitué car les bassins qui le composent n'ont pas été utilisés par Prudhomme (1995). Il sera donc utilisé pour valider objectivement le critère de choix de l'approche « bassin de référence » précédemment évoqué. La liste des bassins étudiés figurent en annexe 7.

D'autre part, en plus des chroniques de débit, nous disposons pour chacun de ces bassins des gradex de pluie, notés $[a_p]_d$, relatifs à différentes durées d .

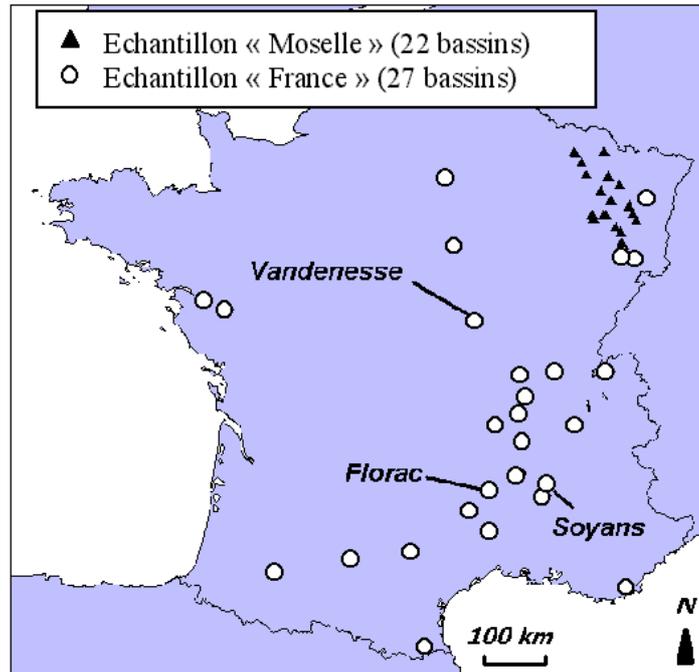


Figure 6-4 : Situation des trois bassins de référence ainsi que des autres bassins étudiés

2.2. Méthodologie

Afin de pouvoir comparer les différentes courbes débit-durée-fréquence testées par la suite, nous avons tout d'abord calculé les quantiles non consolidés $V_d(T)$. La méthode d'échantillonnage et d'ajustement de la loi exponentielle est la même que celle présentée au chapitre 3. Sept durées d_i ont été étudiées. Rappelons que ces durées sont différentes selon les bassins étudiés, et dépendent de la dynamique des crues. Les quantiles $V_d(T)$ ne sont calculés que jusqu'à des périodes de retour de 20 ans. On estime que la loi exponentielle (6-1) n'est plus valide au-delà, et que l'extrapolation qui pourrait être faite (méthode du gradex esthétique, ou autre) n'est pas du champ de notre étude puisqu'elle n'est pas validable.

Pour appliquer les différentes approches précédemment présentées, nous avons besoin de connaître pour chaque bassin étudié :

- le débit instantané décennal $Q(0,10)$ et la durée de Socose D pour l'approche « bassin de référence »,
- un indice de crue μ et la durée Δ pour le modèle régional convergent (*cf.* chap. 5)

Dans la suite de ce chapitre, $Q(0,10)$ est pris égal à $Q_{d=0}(10)$ et D est calculée à partir des hydrogrammes de crue (*cf.* chapitre 1, §3.2.2). Pour le modèle régional convergent, l'indice de crue μ est choisi lui aussi égal à $Q_{d=0}(10)$ afin de pouvoir comparer ces deux approches régionales. La durée Δ est quant à elle issue de l'ajustement du modèle local convergent. (*cf.* chap. 3).

Dans toute la suite de ce chapitre, les différentes caractéristiques $Q(0,10)$, D et Δ sont donc les valeurs locales observées. De ce fait, les quantiles $V(d,T)$ sont des estimations « locales/régionales », au sens de Cunnane (1988) : on utilise à la fois de l'information locale et régionale. Selon lui, l'intérêt est une consolidation de l'information purement locale.

Le chapitre suivant traite du problème de l'estimation purement régionale, c'est à dire sans utiliser l'information hydrologique locale des bassins de validation. Dans ce cas, les caractéristiques $Q(0,10)$, D et Δ sont estimées par régressions multiples.

Chacune des approches testées permet à sa façon de calculer les quantiles $V(d,T)$. Pour comparer objectivement les différents résultats obtenus, le critère suivant est défini :

$$\varepsilon = \frac{1}{N+M} \sum_{i=1}^N \sum_{j=1}^M \frac{|V(d_i, T_j) - V_{d_i}(T_j)|}{V_{d_i}(T_j)} \quad (6-10)$$

avec : N et M : respectivement le nombre de durées et de périodes de retour étudiées.

d_i : les durées étudiées ($N=7$)

T_j : les périodes de retour étudiées ($M=5$) : $T_j = 1 ; 2 ; 5 ; 10$ et 20 ans.

Dans la suite, les quantiles de référence sont les quantiles $V_d(T)$ ajustés localement et non consolidés. Ils sont calculés pour les sept durées étudiées, et pour des périodes de retour observables.

De plus, pour l'approche « bassin de référence » nous avons aussi appliqué le critère de choix décrit au paragraphe 1.2. afin de déterminer, selon ce critère, quel bassin de référence utiliser. Par exemple, la Figure 6-2, relative à la Moselle à Rupt-sur-Moselle, indique qu'il faut choisir le bassin de *Vandenesse*.

3. Résultats de l'approche « bassin de référence » : incidence du choix du bassin de référence

3.1. Comparaison des performances des trois bassins de références

Avant de comparer à proprement parler l'approche « bassin de référence » avec le modèle convergent local (chap. 3) et régional (chap. 5), ce paragraphe présente les résultats de l'approche « bassin de référence », figurant dans Javelle *et al.* (2001). Nous cherchons en particulier à connaître l'incidence du choix du bassin de référence sur le calcul de $V(d,T)$, **pour des périodes de retour inférieures ou égales à 20 ans**, c'est à dire sans utiliser la forme d'extrapolation décrite par les paramètres x_7 à x_9 du Tableau 6-1.

La Figure 6-5 montre les résultats obtenus pour la Moselle à Rupt-sur-Moselle (153 km²). On constate que le choix de *Vandenesse*, effectué par le critère de la Figure 6-2 n'est pas le

meilleur ; cependant il est acceptable. En effet, l'erreur avec le choix de *Soyans* est de 8%, alors qu'avec celui de *Vandenesse* il est de 10%. De manière plus générale, les résultats relatifs aux trois bassins de référence sont assez proches (erreur moyenne de 8, 10 et 12%). Ceci s'explique par le fait que dans tous les cas, on a $Q(d=0,10) = Q_{d=0}(10)$, puisqu'il s'agit de la caractéristique utilisée pour dénormer les courbes QdF adimensionnelles (cf. §1.2.). Le choix du bassin de référence détermine en fait la plus ou moins forte inclinaison du faisceau de courbes autour du « point pivot » relatif au point sur le graphe défini par $V(d=0, 10)$. Cette inclinaison est respectivement faible, moyenne et forte pour *Vandenesse*, *Soyans* et *Florac*. Attention, cette observation n'est valable que pour les quantiles $V(d,T)$ calculés par la relation (6-5), c'est à dire pour des périodes de retour comprises entre 0.5 et 20 ans.

La Figure 6-6 présente les résultats obtenus pour l'ensemble des bassins des échantillons « France » et « Moselle » et compare les performances du modèle avec les trois bassins de référence (performances mesurées par l'écart ε défini à l'équation (6-10)). Pour une meilleure lisibilité, les bassins ont été classés par ordre croissant d'erreur, en considérant d'abord le meilleur des trois modèles « bassins de référence ». D'autre part, le trait plein indique les bassins de référence désignés par le critère de choix (Figure 6-2).

La Figure 6-6 confirme les remarques précédentes (Figure 6-5) : souvent, les résultats relatifs à chaque bassin de référence sont assez proches (une dizaine de %). De plus, le bassin de référence choisi pas le critère n'est pas forcément le meilleur. En effet, il s'avère que pour l'échantillon « France » (qui a donc participé à l'élaboration de ce critère par Prudhomme (1995), ce choix est le meilleur dans 52% des cas. Sur l'échantillon « Moselle » (qui sert donc de validation), cette proportion chute à 23%.

Il s'avère donc que, pour des périodes de retour observables ($0.5 \leq T(\text{an}) \leq 20$ ans), le critère de choix défini par l'approche « bassin de référence » n'est pas performant.

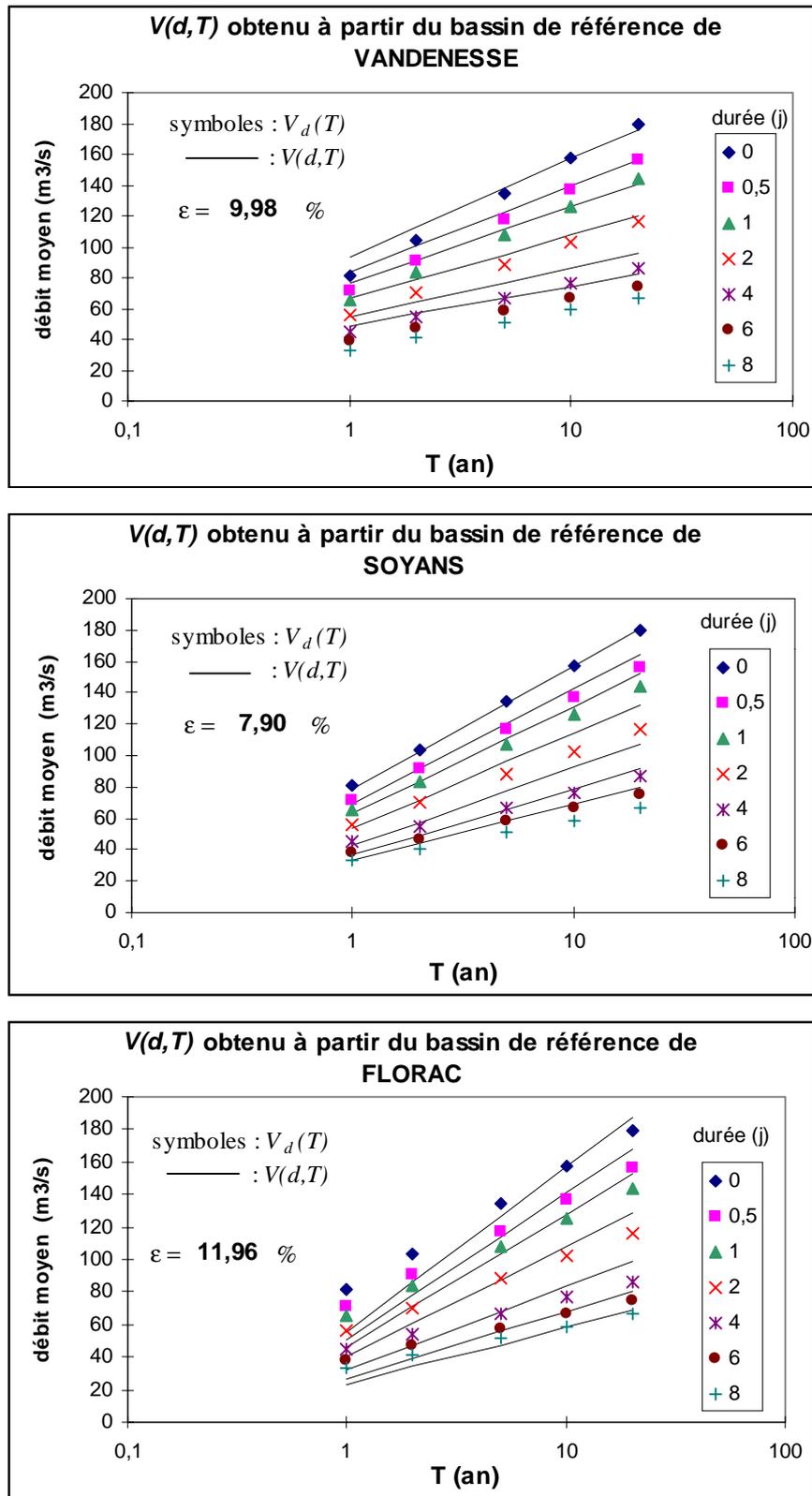


Figure 6-5 : Comparaison des quantiles non consolidés $V_d(T)$, pour $T=1, 2, 5, 10$ et 20 ans et les 7 durées étudiées, avec les quantiles $V(d,T)$ calculés à partir de chacun des trois bassins de référence. Exemple de la Moselle à Rupt-sur-Moselle (153 km^2)

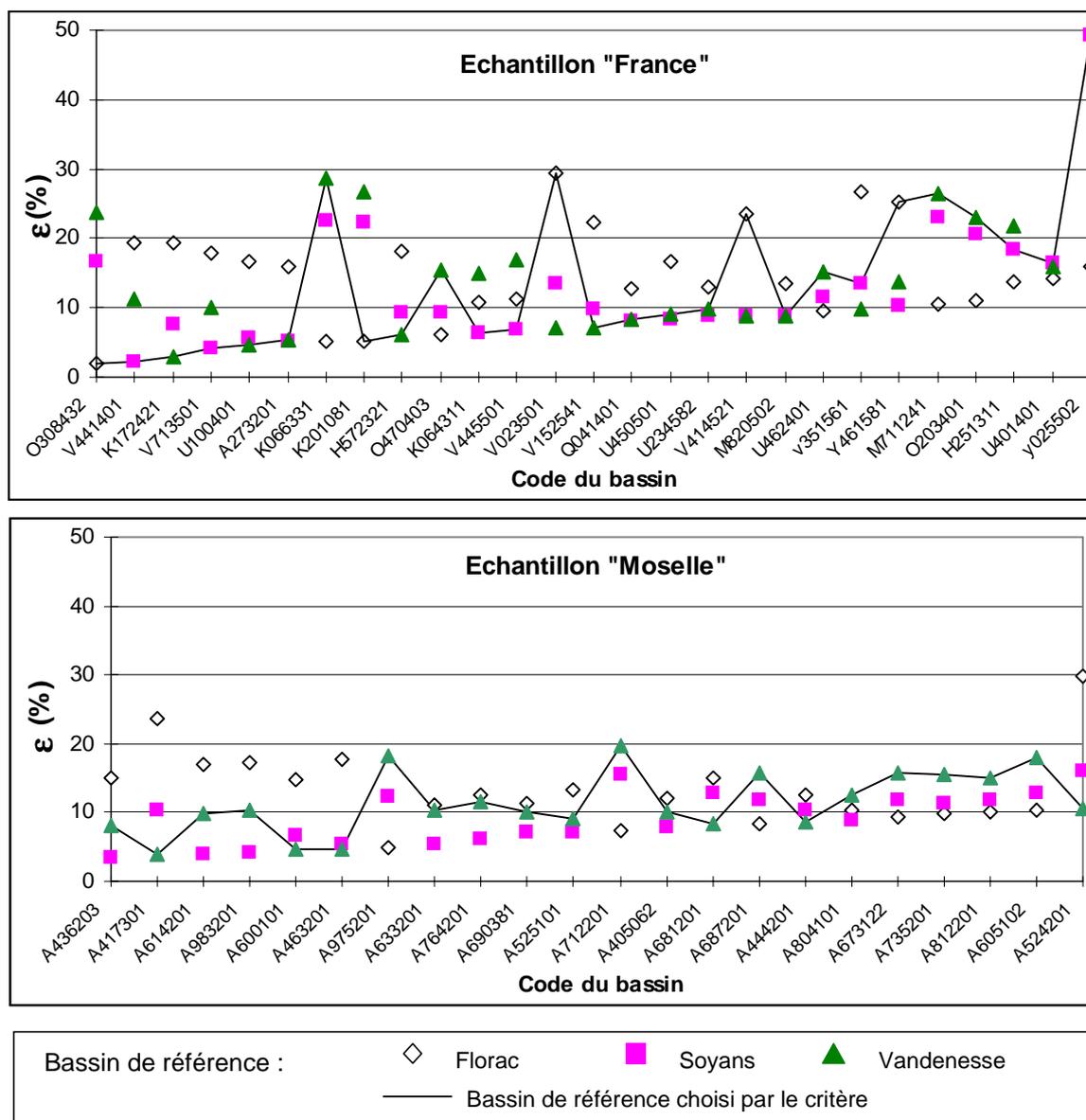


Figure 6-6 : Comparaison des performances des trois bassins de référence et pertinence du critère de choix.

3.2. Comparaison avec les résultats de Prudhomme

Etudions maintenant les résultats obtenus par Prudhomme (1995), et qui ont servi à l'élaboration du critère de choix que nous venons de mettre en défaut. Contrairement à nous, Prudhomme détermine le critère de performance de chacun des bassins de référence en considérant aussi le domaine extrapolé ($T > 20$ ans). Dans ce cas, le critère de choix indique le meilleur des trois bassins de référence pour 83% des 66 bassins qu'elle a testés. L'échantillon « France » de la Figure 6-6 étant composé de 27 bassins issus de l'échantillon de Prudhomme (les autres données n'étaient malheureusement plus disponibles au moment de notre étude), on peut raisonnablement supposer que la proportion de bassins bien classés dans l'échantillon

« France » est, elle aussi, voisine de **83%**. Ce résultat est à comparer aux **52%** que nous avons trouvés Figure 6-6, en ne considérant que le domaine des fréquences observables. Les résultats de Prudhomme montrent donc que si on considère aussi le domaine des fréquences extrapolées, alors le choix du bassin de référence devient plus pertinent.

Afin d'expliquer ce résultat, revenons sur la méthode employée par Prudhomme pour définir les limites $L_1(d^*)$ et $L_2(d^*)$ (6-8) du critère de choix. La Figure 6-7 est similaire à la Figure 6-2, sauf que Prudhomme y a représenté tous les bassins qu'elle a testés : chaque point expérimental correspond au gradex de pluie sur d du bassin étudié, divisé par le débit décennal $Q(0,10)$, en fonction de d/D . (un bassin est donc représenté par 2 à 5 points différents, les gradex de pluie sur d étant disponibles sur 2 à 5 durées d). Comme nous, Prudhomme utilise l'approche « bassin de référence » avec chacun des trois bassins de référence (*Soyans*, *Florac* et *Vandenesse*). On observe effectivement une certaine répartition des points expérimentaux, en haut, au milieu et en bas du graphique, pour les bassins respectivement les mieux modélisés par *Soyans*, *Florac* et *Vandenesse* (Figure 6-7).

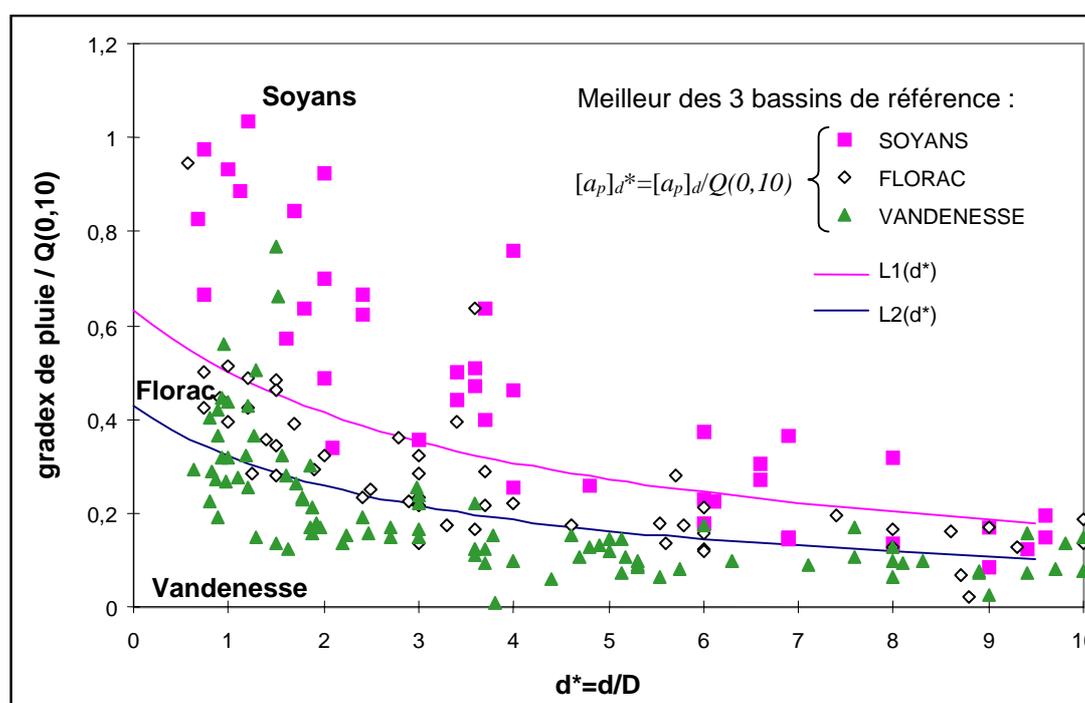


Figure 6-7 : Elaboration des limites $L_1(d^*)$ et $L_2(d^*)$, figure refaite d'après les résultats de Prudhomme (1995)

Prudhomme a ensuite ajusté les deux limites $L_1(d^*)$ et $L_2(d^*)$ définissant les domaines de *Soyans*, *Florac* et *Vandenesse* de façon à maximiser le nombre de bassins bien classés. C'est ainsi qu'elle arrive à 83% de bassins bien classés, ce qui semble très satisfaisant.

L'interprétation de ce critère de choix appelle quelques commentaires. En effet, celui-ci fait intervenir différentes caractéristiques : les gradex de pluie $[a_p]_d$ sur plusieurs durées d , le débit instantané décennal $Q(0,10)$ et la durée D . Pour comprendre sa signification, revenons sur la

fonction $a_p^*(d^*)$ de la relation (6-6), qui, rappelons le, a été ajustée pour chacun des trois bassins de référence (paramètres x_7 à x_9 du Tableau 6-1). Si on représente sur un graphique les fonctions $a_p^*(d^*)$ de chacun des trois modèles de référence, on s'aperçoit, Figure 6-8, que les points expérimentaux se regroupent autour de la courbe $a_p^*(d^*)$ correspondant au bassin de référence donnant les meilleurs résultats.

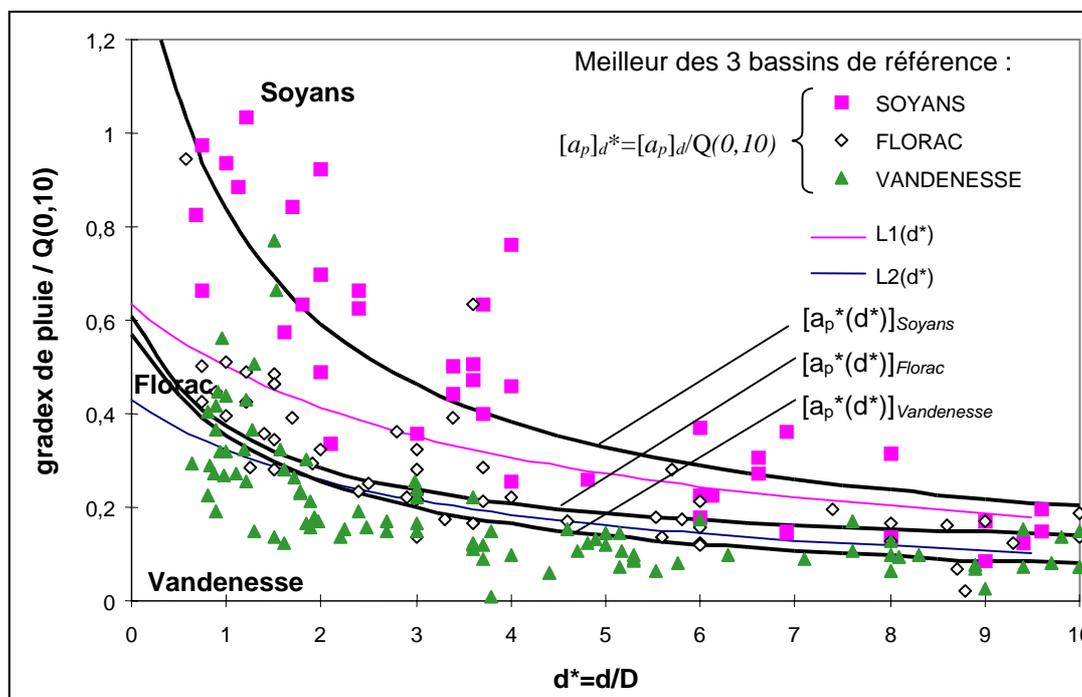


Figure 6-8 : Comparaison des limites $L_1(d^*)$ et $L_2(d^*)$ relatives au critère de choix et des fonctions $a_p^*(d^*)$ relatives aux gradex de pluie modélisés par chacun des bassins de référence par la relation (6-6)

Ce graphe montre que si les points expérimentaux $[a_p]_{d^*}=[a_p]_d/Q(0,10)$ sont proches des courbes $a_p^*(d^*)$ de l'un des trois bassins de référence, alors les quantiles $V(d,T)$ calculés à partir de ce même bassin de référence sont les meilleurs, **à condition de considérer aussi les fréquences extrapolées**. Ceci s'explique par le fait que les gradex de pluie expérimentaux $[a_p]_d$ servent à calculer les extrapolations des quantiles dits « observés » $V_d(T)$, et parallèlement les fonctions $a_p^*(d^*)$, une fois dénormées par $Q(0,10)$, servent à calculer les quantiles cherchés $V(d,T)$.

Le critère de choix défini par Prudhomme est donc une méthode permettant de choisir le bassin de référence qui donne, pour le bassin d'application, les extrapolations les plus proches de celles obtenues par la méthode du gradex esthétique. La typologie définie est donc uniquement liée à cette méthode d'extrapolation.

Si on revient à la « validation » de la méthode présentée Figure 6-7, l'information **locale** pluviométrique est donc utilisée à la fois pour :

- le calcul des quantiles « observés », par la méthode du gradex esthétique, équation. (6-2) ;
- le calcul des courbes *QdF* adimensionnelles, équation (6-6) ;
- le choix du bassin de référence à utiliser (Figure 6-2).

De ce fait, les « bons résultats » obtenus Figure 6-7 étaient prévisibles. Ils ne peuvent donc constituer une **validation** de la typologie établie.

4. Comparaison avec les résultats du modèle convergent

Nous comparons ici les résultats de l'approche « bassin de référence » obtenus au paragraphe précédent, avec ceux du modèle convergent. Les deux niveaux de cette approche sont distingués : le modèle **local** convergent et le modèle **régional** convergent.

4.1. Comparaison avec le modèle local convergent

Ce paragraphe compare maintenant les résultats de l'approche « bassin de référence » avec ceux du modèle local convergent, défini au chapitre 3. Rappelons que ces deux approches se situent à des niveaux d'analyse différents (§1.3.) . La première est régionale, la seconde est locale. Cependant, nous effectuons cette comparaison car l'approche « bassin de référence » est utilisée même lorsque les données locales de débit sont disponibles. En effet, le modèle « local continu » utilisé par Galéa et Prudhomme (1997) (*cf.* deuxième point du §1.1.) comporte 6 paramètres pour $T \leq 20$ ans, 9 pour $T > 20$ ans, et est donc relativement lourd à ajuster.

La Figure 6-9 reprend les résultats de la Figure 6-6, avec en plus, ceux obtenus par le modèle local convergent. On constate une amélioration des résultats dans la presque totalité des cas.

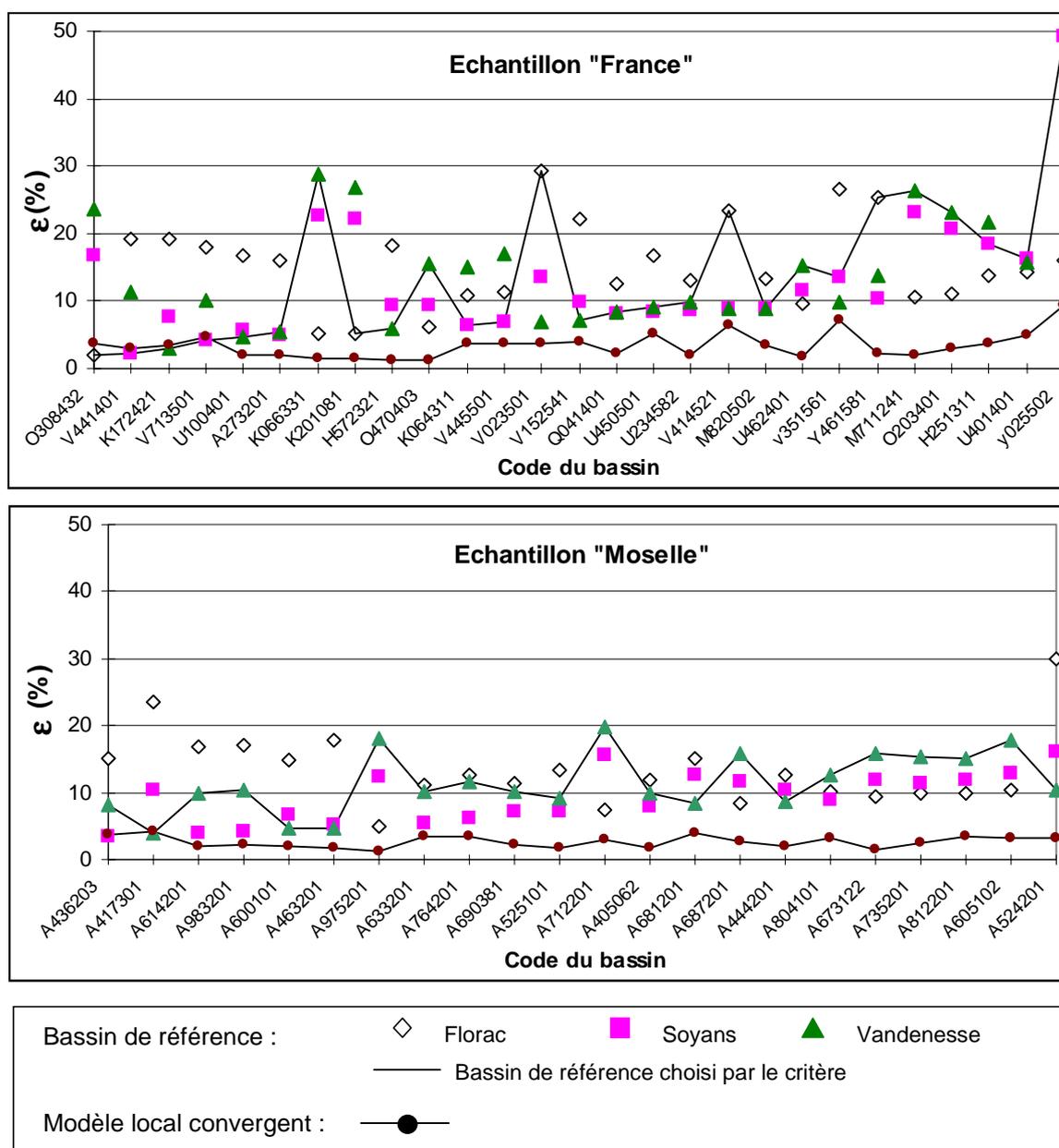


Figure 6-9 : Comparaison de l'approche "bassin de référence" et du modèle local convergent

Signalons le cas particulier des trois premiers bassins de l'échantillon « France », dont les codes sont : O308432, V441401 et K172241, et qui constituent les trois meilleurs résultats de l'approche « bassin de référence ». Il s'agit précisément de la *Mimente* à Florac, du *Roubion* à Soyans, et de la *Dragne* à Vandenesse, les trois bassins de référence ayant servi à caler les courbes *QdF* adimensionnelles. Par conséquent, les résultats obtenus par l'approche « bassin de référence » pour ces trois bassins (et seulement pour ceux là), sont issus de l'ajustement local des 6 paramètres x_1 à x_6 , du Tableau 6-1. Le modèle local convergent ne contenant que 4 paramètres (P , $x_0(0)$, $a_0(0)$ et Δ), il est normal que pour ces trois bassins, l'approche « bassin de référence » donne de meilleurs résultats.

4.2. Comparaison avec le modèle régional convergent

Ce paragraphe compare les résultats de l'approche « bassin de référence » avec ceux du modèle convergent **régional**. Ce modèle comme nous l'avons vu au chapitre précédent, consiste à calculer des courbes *QdF* régionales adimensionnelles, pour une région donnée, puis à dénormer localement ces courbes, pour le bassin étudié, à l'aide de deux caractéristiques : le débit instantané décennal μ et la durée Δ .

4.2.1. Choix de la norme de débit

Le but étant de comparer les résultats obtenus avec ceux de l'approche "bassin de référence" nous adoptons la même norme : le débit instantané décennal $Q(0,10)$. Les différences observées entre les deux approches ne seront donc pas dues au choix de cette norme. Comme nous l'avons vu au chapitre précédent traitant de données du Québec et de l'Ontario, il est aussi possible de choisir un quantile de période de retour plus faible, par exemple la moyenne ou la médiane de l'échantillon des crues. Ce chapitre ne traite pas de l'incidence de ce choix. Dans le cadre d'une comparaison des différentes méthodes régionales existantes à travers le monde, il y aurait lieu de mener cette analyse.

4.2.2. Constitution des régions hydrologiques et test de leur homogénéité

Comme nous l'avons vu au chapitre précédent, différentes méthodes existent pour définir les régions hydrologiques, des plus simples (région géographique) au plus élaborées (région d'influence, ou bien définie par analyse canonique). Nous adoptons ici la méthode la plus simple, consistant à regrouper les bassins selon leur localisation. Le chapitre précédent avait utilisé des régions non contiguës définies sur le Québec et l'Ontario, par analyse en composantes principales (Adamowski, 2000). Les deux régions géographiques que nous avons considérées sont celles définies par nos échantillons : la Moselle et la France sans la Moselle.

L'homogénéité étudiée est donc liée aux paramètres du modèle étudié, ici le modèle local convergent. Nous allons étudier chacun des paramètres du modèle local convergent, normés par le débit instantané décennal :

$$x_0^* = x_0(0) / Q(0,10) ; \quad a^* = a(0) / Q(0,10) \quad \text{et} \quad P^* = P / Q(0,10) \quad (6-11)$$

Rappelons que $x_0(0)$ et $a(0)$ sont respectivement le paramètre de position et le gradex de la distribution des débits $V(0,T)$ et que dans cette étude, P a été fixé au module interannuel \overline{QA} .

Les résultats obtenus pour chaque bassin sont présentés Figure 6-10.

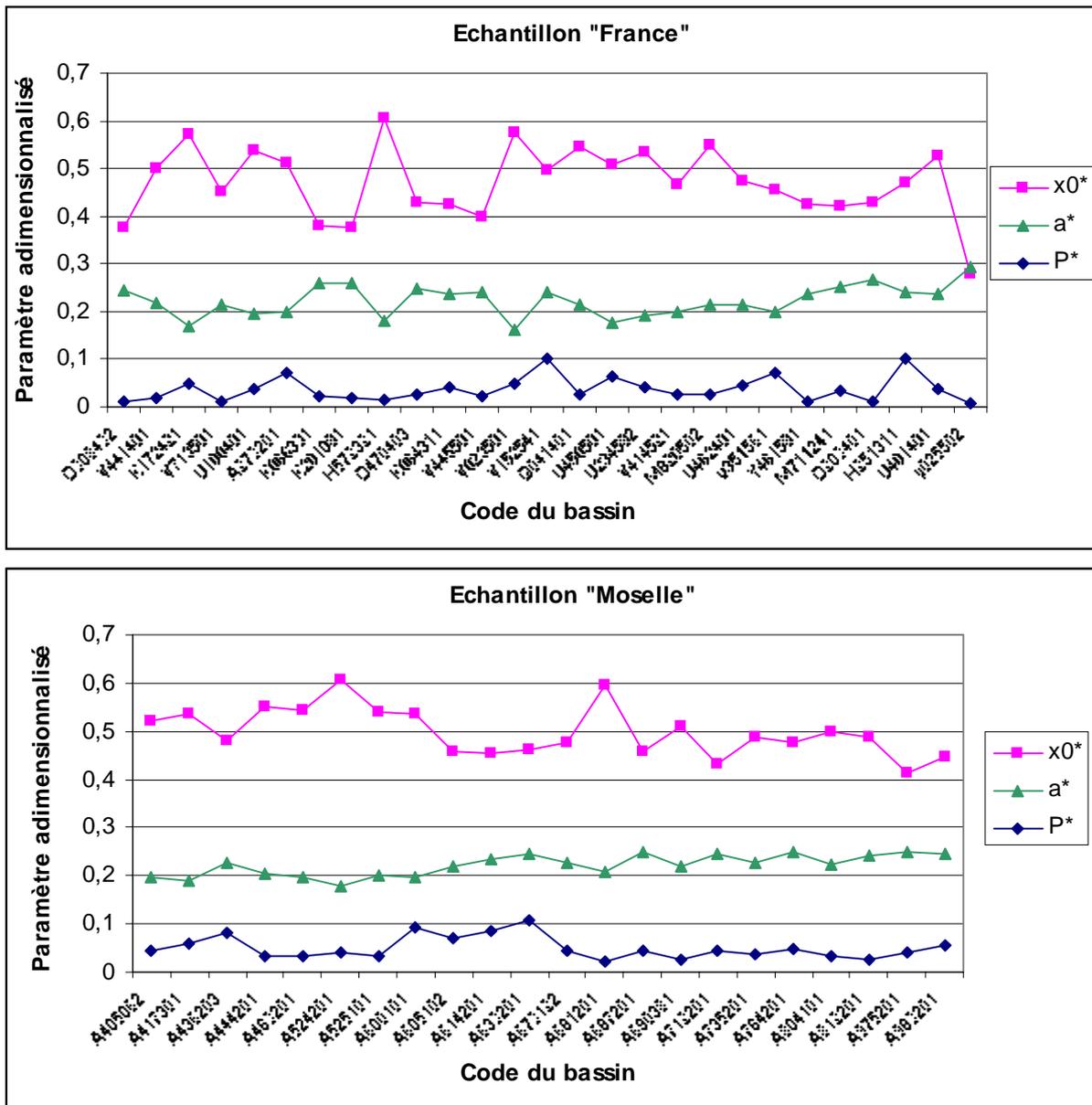


Figure 6-10 : Paramètres du modèle local convergent normés par le débit instantané décennal $Q(0,10)$

La Figure 6-11 donne la moyenne et le coefficient de variation (écart type/moyenne) des paramètres x_0^* , a^* et P^* , pour les deux échantillons (ou “régions”) considérés.

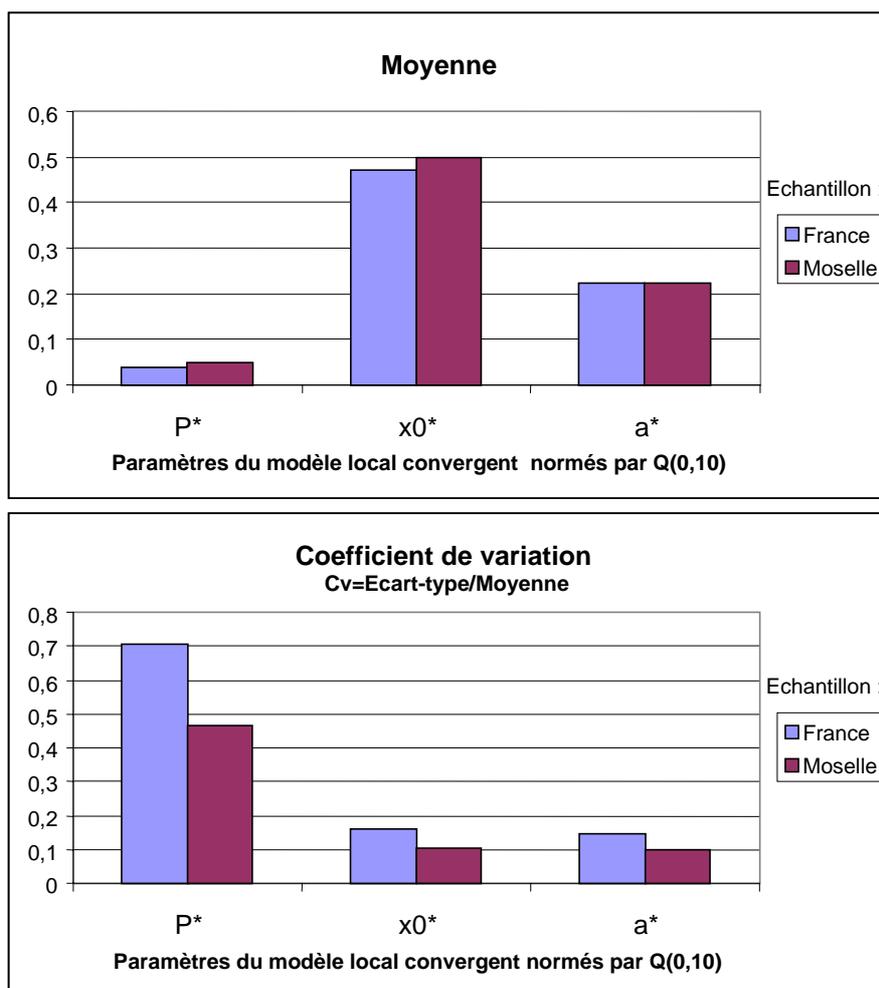


Figure 6-11 : Moyenne et coefficient de variation des paramètres du modèle local convergent normés par le débit instantané décennal $Q(0,10)$

L'étude du coefficient de variation des coefficients locaux au sein de chaque région étudiée permet de juger l'**homogénéité** de ces coefficients (Figure 6-11). En toute rigueur, il faudrait effectuer des tests statistiques pour savoir si l'on peut considérer que les paramètres normés peuvent être considérés comme constants, et donc que la région est homogène. (cf. chapitre 5, §1.2). Nous nous contentons ici de comparer les graphes obtenus. On constate que la moyenne des coefficients (graphique du haut, Figure 6-11), c'est à dire les coefficients régionaux $x_0^*_R$, a^*_R et P^*_R sont relativement voisins, d'un échantillon à l'autre. Par contre, lorsqu'on regarde leur coefficient de variation, ceux de la Moselle sont beaucoup plus faibles. D'un point de vue hydrologique, cela signifie que les bassins de l'échantillon Moselle sont plus homogènes. Ce résultat n'est pas une surprise, puisque les bassins de l'échantillon « France » sont soumis à des climats variés (océanique, alpin, méditerranéen), et ont vraisemblablement des sols et sous-sols de types très différents. A l'inverse, les bassins de la Moselle sont relativement semblables. Nous étudierons d'ailleurs en détail leurs caractéristiques au chapitre 7.

L'étude ci-dessus des coefficients de variation montre que l'hypothèse de **région homogène**, essentielle en hydrologie régionale, est moins bien respectée pour l'échantillon "France" que pour celui de la Moselle. Néanmoins, nous présentons dans la suite l'analyse régionale conduite sur ces deux échantillons, afin d'apprécier l'incidence sur les résultats de cette plus ou moins grande homogénéité.

L'étape suivante consiste à calculer les coefficients régionaux $x_0^*_R$, a^*_R et P^*_R . Comme nous l'avons déjà noté, nous prendrons simplement la moyenne. Néanmoins, pour nous placer dans des conditions de validation, cette moyenne est calculée pour chaque bassin, sans utiliser les données de ce bassin. Ainsi, les seules données locales utilisées dans notre comparaison sont $Q(0,10)$ et la durée Δ . Nous nous mettons donc bien dans les mêmes conditions que l'approche « bassin de référence » qui utilise $Q(0,10)$ et la durée D (Socose).

$$X_i^*_R = \frac{1}{N-1} \left[\sum_{j=1}^{i-1} X_j^* + \sum_{j=i+1}^N X_j^* \right] \quad (6-12)$$

avec X la caractéristique étudiée (respectivement $x_0(0)$, $a(0)$ et P); i l'indice du bassin cible ; j l'indice des autres bassins de la région, N le nombre de bassins de la région.

La Figure 6-12 présente les résultats du calcul des coefficients régionaux $x_0^*_R$, a^*_R et P^*_R selon la relation (6-12) (c'est à dire sans utiliser l'information locale du bassin étudié), pour l'exemple de l'échantillon « Moselle ». Le même calcul a été effectué pour l'échantillon « France ». On constate que les paramètres obtenus sont pratiquement constants, d'un bassin à l'autre. Ceci s'explique par le nombre suffisamment élevé de bassins étudiés et par la bonne homogénéité de cet échantillon.

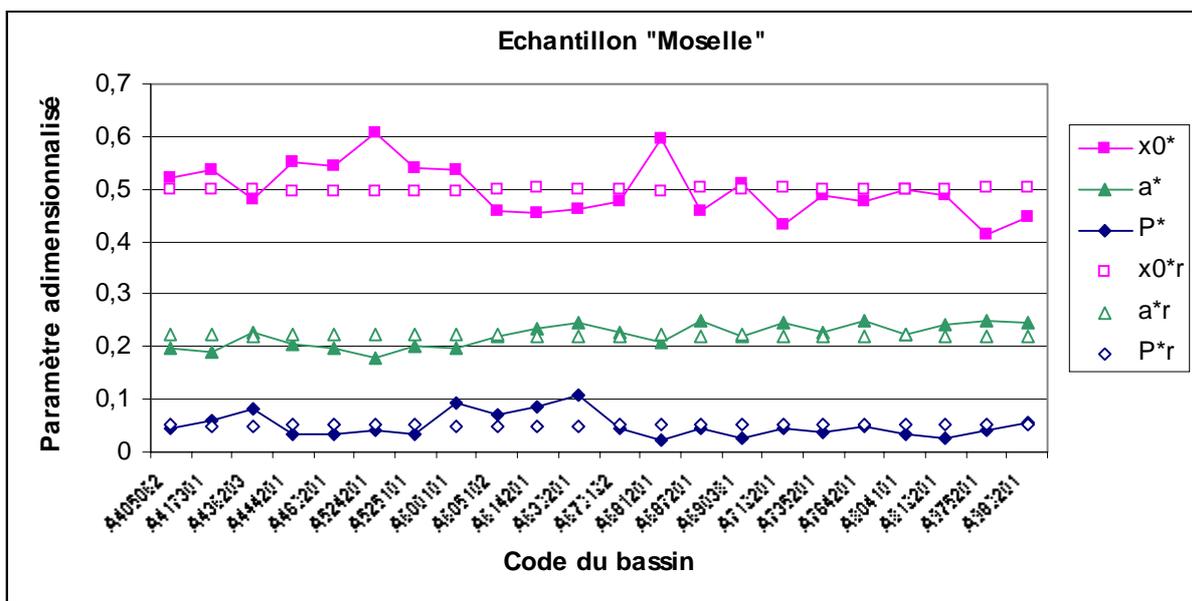


Figure 6-12 : Calcul des paramètres régionaux $x_0^*_R$, a^*_R et P^*_R calculés en chaque site sans utiliser l'information locale, relatifs à l'échantillon "Moselle"

Le modèle régional convergent étant maintenant défini pour chacune des deux régions étudiées (calcul des coefficients $x_0^*_{R}$, a^*_R et P^*_{R}), il est désormais possible d'estimer les quantiles $V(d,T)$ en chaque site en utilisant pour seule information **locale** le débit décennal instantané $Q(0,10)$ et la durée Δ , le reste de l'évolution de $V(d,T)$ en fonction de d et T étant donné par l'information **régionale** (c'est à dire issue de l'ensemble des bassins de la région, sauf celui étudié (*cf.* chapitre 5)).

Contrairement au paragraphe précédent, nous nous situons ici au même niveau d'analyse **régionale** que dans l'approche « bassin de référence ». Dans les deux cas, la seule information **locale** utilisée est liée à un quantile de crue, $Q(0,10)$ et une durée caractéristique Δ ou D , selon l'approche considérée.

La Figure 6-13 présente les résultats finalement obtenus. Ils concernent ceux déjà présentés Figure 6-9, (approche « bassin de référence » et modèle local convergent) avec en plus, ceux du modèle régional convergent.

On constate Figure 6-13 que les résultats du modèle régional convergent sont systématiquement moins bons que ceux du modèle local convergent, ce qui est logique puisque l'information locale utilisée est moins riche. Le modèle régional convergent est meilleur que le bassin de référence désigné par le critère de choix pour 59% des bassins de l'échantillon « France » et pour 91% de ceux de l'échantillon « Moselle ».

Pour y voir plus clair, le paragraphe suivant propose une synthèse de ces résultats, pour permettre notamment de comparer l'échantillon peu homogène « France » et l'échantillon bien homogène « Moselle » (*cf.* début de ce paragraphe).

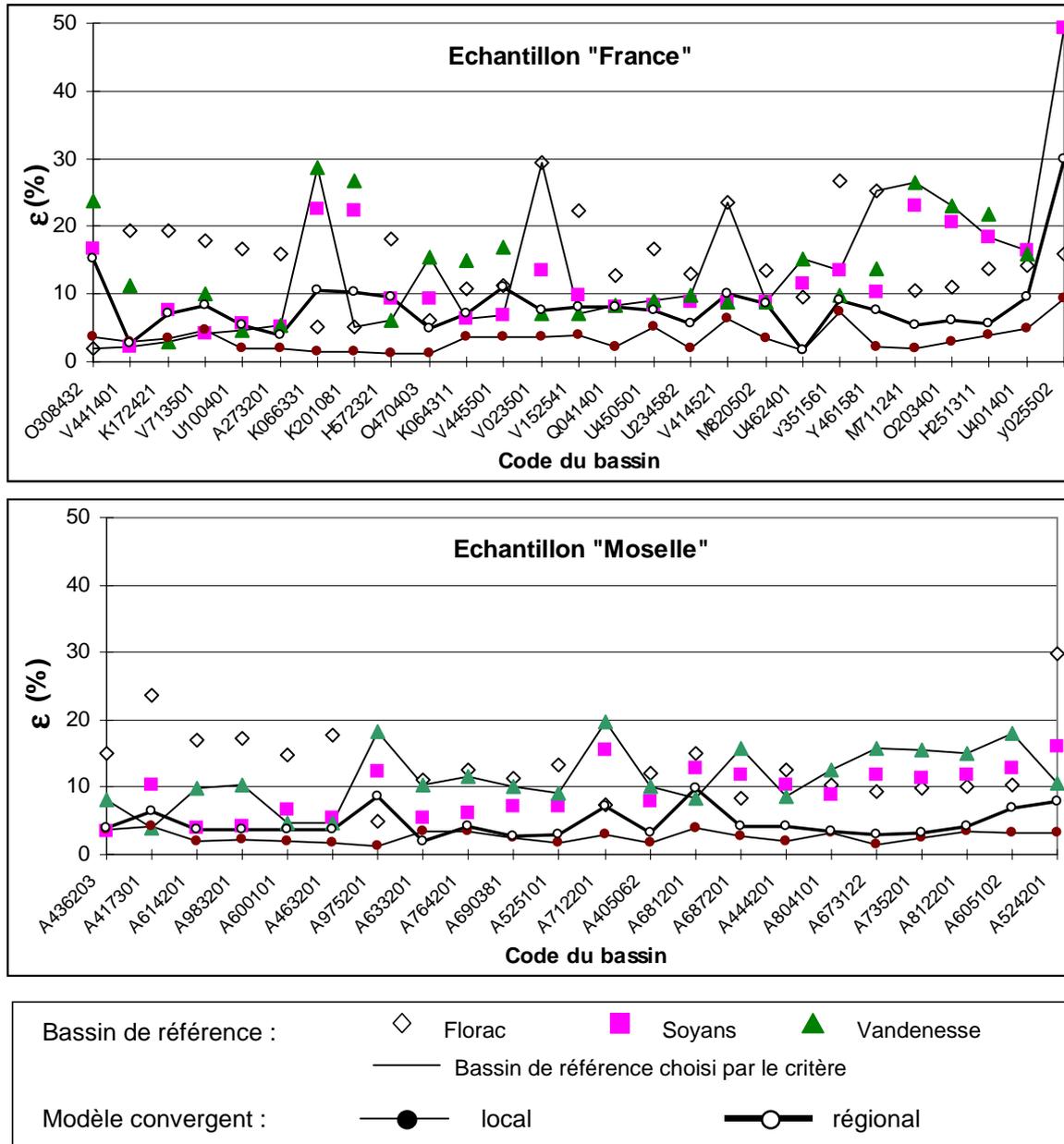


Figure 6-13 : Comparaison de l'approche "bassin de référence" et du modèle convergent local *et* régional

5. Synthèse des résultats

Afin d'effectuer la synthèse des résultats de la Figure 6-13, nous calculons la moyenne ε_{moy} et l'écart type σ_ε du critère ε calculé selon la formule (6-10) pour les deux échantillons, et pour les quatre cas suivants :

- le modèle « bassin de référence » désigné par le critère de choix ;
- le meilleur des trois modèles « bassin de référence », même s'il n'a pas été choisi ;
- le modèle **local** convergent ;
- le modèle **régional** convergent ;

Ces informations sont résumées dans un critère plus général, que nous notons Erreur Quadratique Moyenne (*EQM*) :

$$EQM = \sqrt{\varepsilon_{moy}^2 + \sigma_\varepsilon^2} \quad (6-13)$$

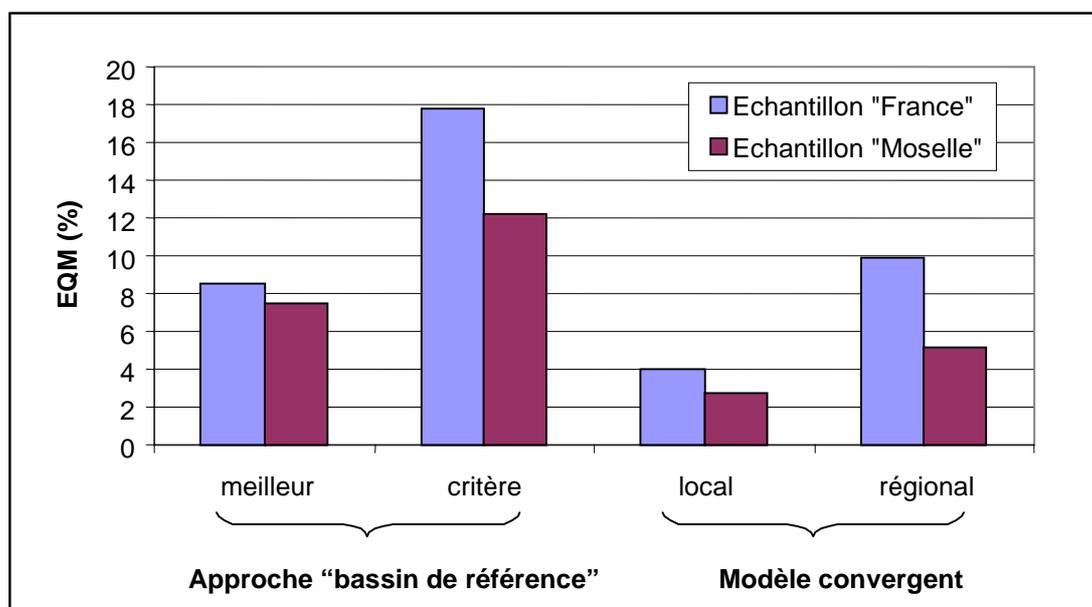


Figure 6-14 : Synthèse des résultats des différentes approches

La Figure 6-14 amène plusieurs remarques. D'une part, on constate que les résultats de l'échantillon Moselle sont systématiquement meilleurs, quelle que soit l'approche testée. Ce groupe de bassins constitue donc bien une région plus homogène que ceux de l'échantillon France, comme nous l'avons signalé dès le début du paragraphe 4.2.

D'autre part, le modèle convergent (local ou régional) est toujours meilleur que l'approche « bassin de référence », sauf lorsqu'on compare le meilleur des trois bassins de l'approche régionale avec le modèle régional convergent sur l'échantillon France. Cependant, le fait de choisir systématiquement le meilleur des trois bassins de référence utilise implicitement la totalité de l'information locale, et non plus uniquement le débit instantané décennal et une

durée caractéristique de crue. Nous ne sommes donc plus dans le cadre d'une l'estimation « locale-régionale ».

Enfin, concernant le critère de choix de l'approche bassin de référence, on voit qu'il est largement dépassé, surtout sur l'échantillon "France", qui pourtant est hétérogène et où, comme nous l'avons dit, le modèle régional convergent est désavantagé par rapport à une approche permettant de choisir entre trois familles de courbes adimensionnelles. Mais comme nous l'avons noté au paragraphe 3.2. ce critère de choix fait explicitement référence à la méthode d'extrapolation du gradex esthétique. On comprend pourquoi dans notre domaine d'étude, celui des fréquences observables, il n'est pas pertinent.

Conclusion du chapitre 6

Ce chapitre a permis de comparer les estimations « locales-régionales » de l'approche « bassin de référence » avec le modèle convergent, local (*cf.* chap. 3) puis régional (*cf.* chap. 5). Les résultats de cette étude réalisée en ne considérant que le domaine des fréquences observables, amènent aux conclusions suivantes.

Concernant l'approche « bassin de référence », la typologie définie n'est pas réellement significative. Le critère de choix ne désigne pas le modèle restituant le mieux les courbes « débit-durée-fréquence », sans pour autant donner de mauvais résultats. Ceci s'explique par le fait que dans le domaine des fréquences observables, les courbes *QdF* des trois bassins de référence sont relativement proches (*cf.* Figure 6-5). En fait, la typologie définie concerne les quantiles de fréquence rare ($T > 20$ ans), obtenus par extrapolation à l'aide des gradex de pluie (méthode du gradex esthétique, Michel, 1982). Elle implique que l'on accepte cette méthode d'extrapolation, inspirée de la théorie du gradex (Guilot et Duband, 1967), et ne trouve donc pas sa justification par une concordance entre valeurs « observées » et « calculées », les valeurs extrapolées n'étant par définition, pas validables.

Concernant la comparaison l'approche « bassin de référence » et du modèle *QdF* convergent (local et régional), celle-ci est en faveur de ce dernier. L'intérêt du modèle proposé est qu'il évite le choix parfois délicat du bassin de référence. En intégrant les concepts plus classiques de l'hydrologie régionale, il propose une démarche moins empirique, tout en intégrant les points forts de l'approche « bassin de référence », comme par exemple le fait de « résumer » les courbes débit-durée-fréquence d'un bassin par seulement deux caractéristiques hydrologiques : un débit et une durée.

Seuls des bassins situés en régime tempéré ont été étudiés dans ce travail. Cependant Meunier (1999, 2001) a aussi testé de façon concluante le modèle convergent (local et régional) en Martinique. Des résultats similaires sont obtenus en Guadeloupe (Galéa et Javelle, 2000). De plus, le chapitre précédent a montré une application satisfaisante sur plus d'une centaine de bassins situés au Québec et en Ontario. Les bons résultats obtenus ne semblent donc pas liés à une zone climatique particulière. Pour s'en assurer, il faudrait sans doute effectuer encore d'autres essais, comme par exemple sur des bassins sahéliens.

Un problème reste à traiter, celui de l'estimation *purement régionale*, c'est à dire pour des bassins où l'on ne dispose d'aucune information hydrologique. Nous l'abordons dans le septième et dernier chapitre de cette thèse.

Chapitre 7

Modélisation QdF « purement régionale »

Le chapitre 6 a comparé les estimations « locales-régionales » de l'approche « bassin de référence » et du modèle régional convergent. Comme nous l'avons souligné, ce type d'approche ne peut pas être appliqué sur des bassins où l'on ne dispose d'aucune donnée de débit. En effet, l'estimation des deux caractéristiques hydrologiques requises (un débit et une durée) est locale.

Le but de ce dernier chapitre est d'effectuer la même comparaison qu'au chapitre 6, mais sans utiliser d'information locale sur les débits. Dans ce cas, les deux caractéristiques hydrologiques des modèles QdF régionaux sont estimées par régressions multiples, à partir des caractéristiques pluviométriques et des caractéristiques « physiques » du bassin (géomorphologie et occupation du sol).

Au chapitre 5, nous avons déjà étudié les estimations « purement régionales » du modèle régional convergent, appliqué au Canada. Ces résultats avaient montré que l'erreur engendrée était très importante, lorsqu'on passait de l'étape « locale-régionale » à celle « purement régionale ». Ce présent chapitre propose de revenir en détail sur ce problème dans le cas des bassins situés sur le bassin hydrographique de la Moselle, en France et en Allemagne. Cette zone a été étudiée dans le cadre d'un projet européen (NOAH : New Opportunities of Altimetry in Hydrology), et de nombreuses données sont disponibles : hydrologiques, pluviométriques, ainsi qu'un modèle numérique de terrain et une carte d'occupation du sol obtenues à partir d'images des satellites SPOT.

Le premier paragraphe de ce chapitre présente les zones et les données étudiées.

Le second explique comment sont sélectionnées les caractéristiques intervenant dans les régressions pour expliquer les différentes grandeurs hydrologiques étudiées.

Le troisième paragraphe teste par la méthode de la régression de crête (*ridge regression*) la robustesse des régressions obtenues.

Enfin, le dernier paragraphe de ce chapitre montre l'application « purement régionale » des modèles QdF testés au chapitre 6, en utilisant cette fois-ci les régressions obtenues dans ce chapitre

1. Présentation du contexte de l'étude et des données

1.1. Objectif

L'objectif de ce chapitre est d'établir par régressions multiples, des relations entre les caractéristiques hydrologiques, les caractéristiques pluviométriques, et les caractéristiques physiques de bassin. Ces caractéristiques sont schématisées Figure 7-1.

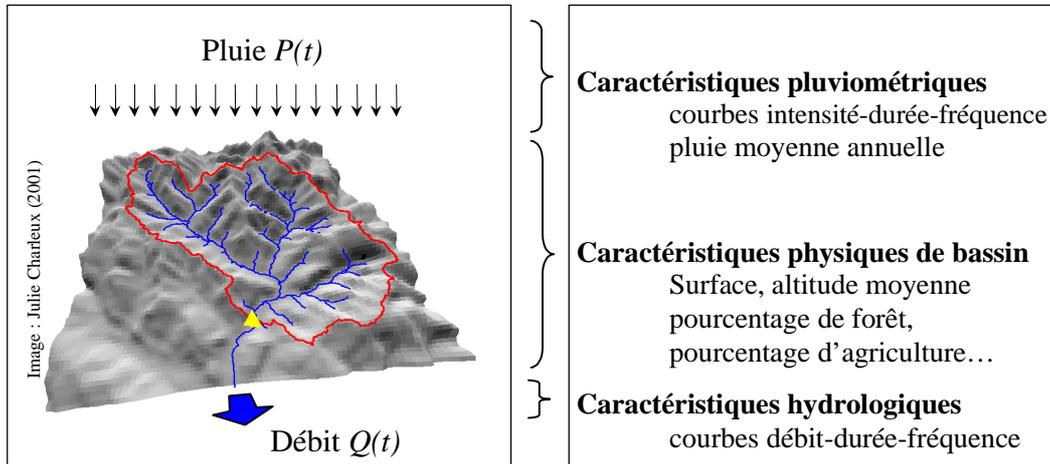


Figure 7-1 : Schématisation des différentes caractéristiques étudiées

L'intérêt de ces formules de régression est de pouvoir être appliquées sur des bassins où l'on ne dispose d'aucune information sur les débits. Les caractéristiques hydrologiques étudiées sont soit directement des quantiles de crue (*cf.* chap. 5, §1.3), soit des caractéristiques hydrologiques plus « intégratrices », permettant de dénormer localement des modèles régionaux adimensionnels. Il s'agit par exemple de la méthode de l'indice de crue (*cf.* chap. 5, §1.4), ou bien de la généralisation que nous en avons faite, le modèle régional convergent (*cf.* chap. 5, §2.), ou encore de l'approche « bassin de référence » (*cf.* chap. 6, §1.).

De nombreuses formules de régressions existent un peu partout dans le monde. On peut citer les relations établies aux Etats-Unis par Thomas et Benson (1970), en Grande Bretagne par le Flood Study Report (NERC, 1975), ou bien en France pour le débit instantané décennal et la durée de Socose (CTGREF *et al.*, 1980). Par exemple, pour le débit instantané décennal ($Q(0,10)$ selon nos notations, $QIXA10$ selon Oberlin (1992)), la relation établie pour la France à partir de 630 bassins (méthode Crupedix) fait intervenir la surface (S), la pluie décennale journalière ($PJXA10$), et un coefficient régional (R), permettant de corriger localement le résultat de la régression :

$$\hat{Q}(0,10) = QIXA10 = S^{0,8} \left(\frac{PJXA10}{80} \right)^2 R \quad (7-1)$$

De même, une formule a été établie pour la France pour estimer la durée D de Socose, requise par l'approche bassin de référence présentée au paragraphe précédent :

$$\ln \hat{D} = -0,69 + 0,32 \ln S + 2,2 \sqrt{\frac{PA}{PJXA10} \cdot \frac{1}{t_a}} \quad (7-2)$$

Avec S , la surface, PA la pluie moyenne annuelle, $PJXA10$ le pluie journalière décennale et t_a la température moyenne annuelle.

Ce type de formules ne donne en général qu'un ordre de grandeur de la « vraie » valeur. Faute de mieux, celles-ci sont très utilisées en ingénierie, lorsqu'on ne dispose pas d'information sur les débits. A titre indicatif, avec la formule (7-1), on a 70% de chance que la vraie valeur $Q(0,10)$ se trouve dans l'intervalle $[\hat{Q}(0,10)/2; 2\hat{Q}(0,10)]$, ce qui est considérable. L'incertitude est la même pour l'estimation de la durée D .

1.2. Le projet NOAH et sa zone d'étude

L'étude suivante a été réalisée dans le cadre du projet Européen NOAH (New Opportunities of Altimetry in Hydrology). Le but de ce projet était de voir comment les données issues de l'Observation de la Terre pouvaient être utilisées en hydrologie et en hydraulique. Pour plus de détails, on peut se référer au rapport final du projet (Noah, 2000), ou plus simplement, sur le site internet suivant <http://www.egeriesa-noah.lu>).

La zone étudiée est le bassin hydrographique de la Moselle, couvrant une superficie d'environ 30 000 km², au Luxembourg, en Allemagne et en France. Cette zone a été choisie pour plusieurs raisons. Tout d'abord, il s'agit d'un bassin transfrontalier. Ensuite, cette zone relativement restreinte regroupe une assez grande variabilité de paysages, et donc, de conditions d'écoulement. Par exemple, l'amont du bassin se situe dans le massif des Vosges, recouvert de forêts. Plus en aval, la Moselle traverse un paysage très plat voué à l'agriculture. En son nord, le bassin est de nouveau bordé par une petite chaîne de montagne, dans sa partie luxembourgeoise et allemande. Enfin, cette région est à la fois économiquement très développée et soumise à des inondations fréquentes. Au cours des cent dernières années, les inondations les plus dévastatrices se sont produites par ordre d'importance en 1947 (une douzaine de victimes et plusieurs milliards de francs de dégâts, puis en 1919 et enfin en 1983 (source : http://www.republicain-lorrain.fr/XX_siecle/4_saison/2panno.htm).

1.3. Les données disponibles

1.3.1. Les données hydrologiques

Au total, les chroniques de débit mesurées en 54 stations de jaugeages ont été étudiées. Les surfaces des bassins versants sont comprises entre 16 et 1250 km². Nous avons volontairement fait le choix de ne pas étudier les bassins plus grands pour trois raisons :

- afin de limiter au maximum le nombre de bassins emboîtés ;
- afin de travailler sur des échelles de surface relativement comparables

- afin de n'étudier que des stations « pas trop influencées ». Pour des surfaces supérieures, les stations sont situées sur des parties de la Moselle et de la Meuse fortement anthropisées (voie navigable, pompage, rejets urbains, dérivation de canaux, etc...).

N'ayant pu avoir accès aux données luxembourgeoises, les stations traitées proviennent de la partie allemande et française du bassin de la Moselle. Pour la partie française (47 stations), les données ont été achetées à la base nationale HYDRO. Pour 39 de ces stations, il s'agit de chroniques de *débit instantané* au pas de temps variable (*cf.* chap. 1 pour la définition de ces chroniques). Pour 8 stations, ce type de chroniques n'était pas disponible, et nous avons utilisé des chroniques de *débit moyen journalier*. Pour la partie allemande (7 stations), les chroniques étudiées sont constituées de *débites journaliers*, et ont été fournies gratuitement par le Landesamt für Umweltschutz (LfU) et le Landesamt für Wasserwirtschaft, (LfW). La liste des stations étudiées figure en annexe 8.

1.3.2. Les données pluviométriques

Pour la France, des chroniques de *pluie locale journalière* provenant de 192 stations ont été mises à disposition par Météo-France. Les données allemandes sont des *pluies journalières spatialisées*, relatives à chaque bassin étudié. Nous verrons dans la suite comment ces différents types de données (locales et spatiales) ont été analysés.

La Figure 7-2 localise les différents stations hydrographiques et pluviographiques étudiées.

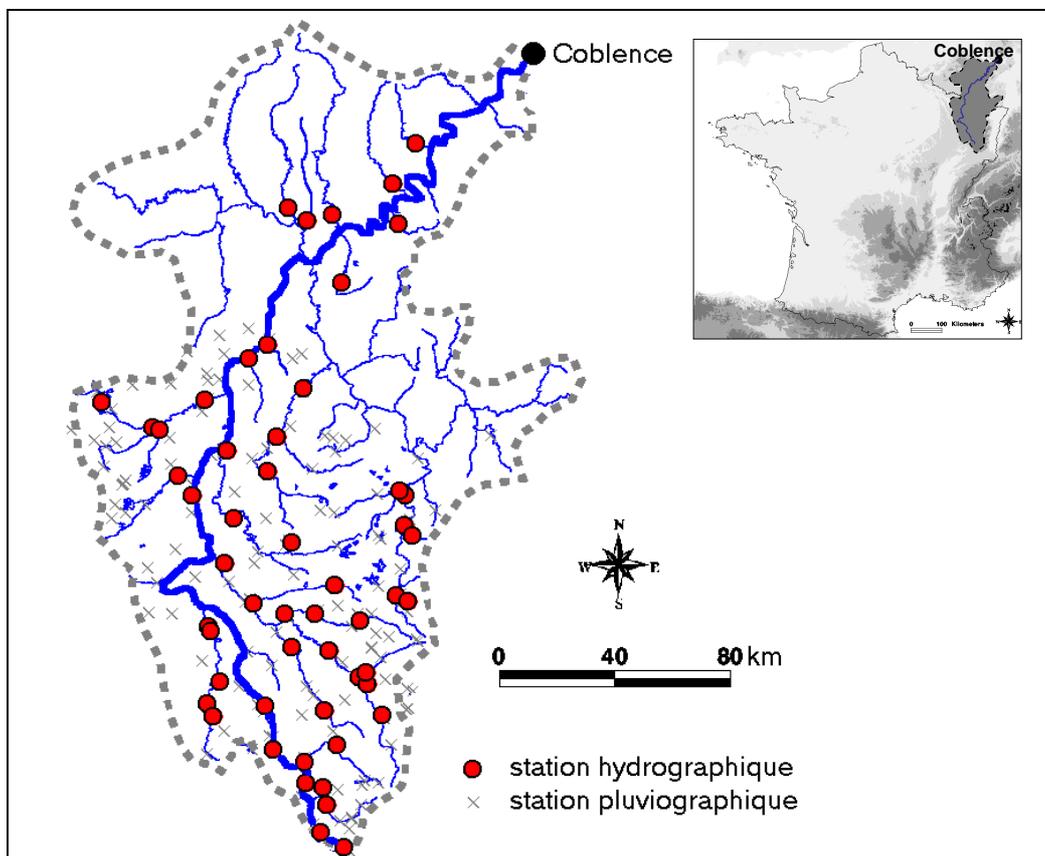


Figure 7-2 : Localisation des stations hydrographiques et pluviographiques étudiées

1.3.3. Les données issues de la télédétection

Les données de télédétection analysées dans le cadre du projet NOAH proviennent de satellites SPOT. 21 couples stéréoscopiques en mode panchromatique (10 mètres de résolution), ainsi que 21 scènes en mode multispectral (20 mètres de résolution) ont été fournies par SPOT Image. Chacune des ces scènes couvre une surface de 60x60 kilomètres. La Figure 7-3 indique les différentes prises de vue effectuées. Sur l'exemple d'image présentée, on distingue nettement les méandres que fait la Moselle, peu avant de se jeter dans le Rhin.

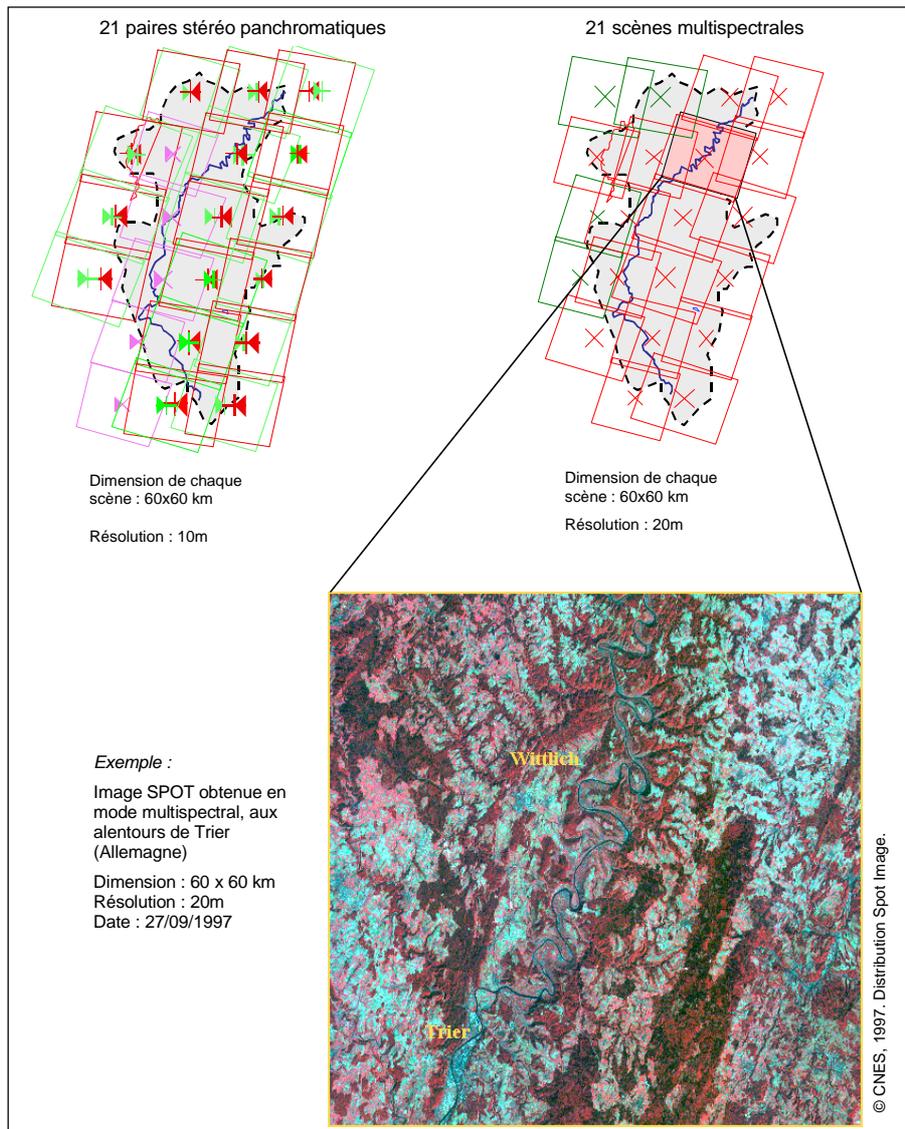


Figure 7-3 : Couverture du bassin de la Moselle par les différentes prises de vue SPOT

Ces couples stéréoscopiques et ces images multispectrales ont respectivement servi à calculer un modèle numérique de terrain et une carte d'occupation du sol, pour tout le bassin de la

Moselle. Dans les deux cas, la résolution finale est de 20 mètres. On estime la précision verticale du MNT à 10 mètres. Ces traitements ont été effectués par la société ISTAR.

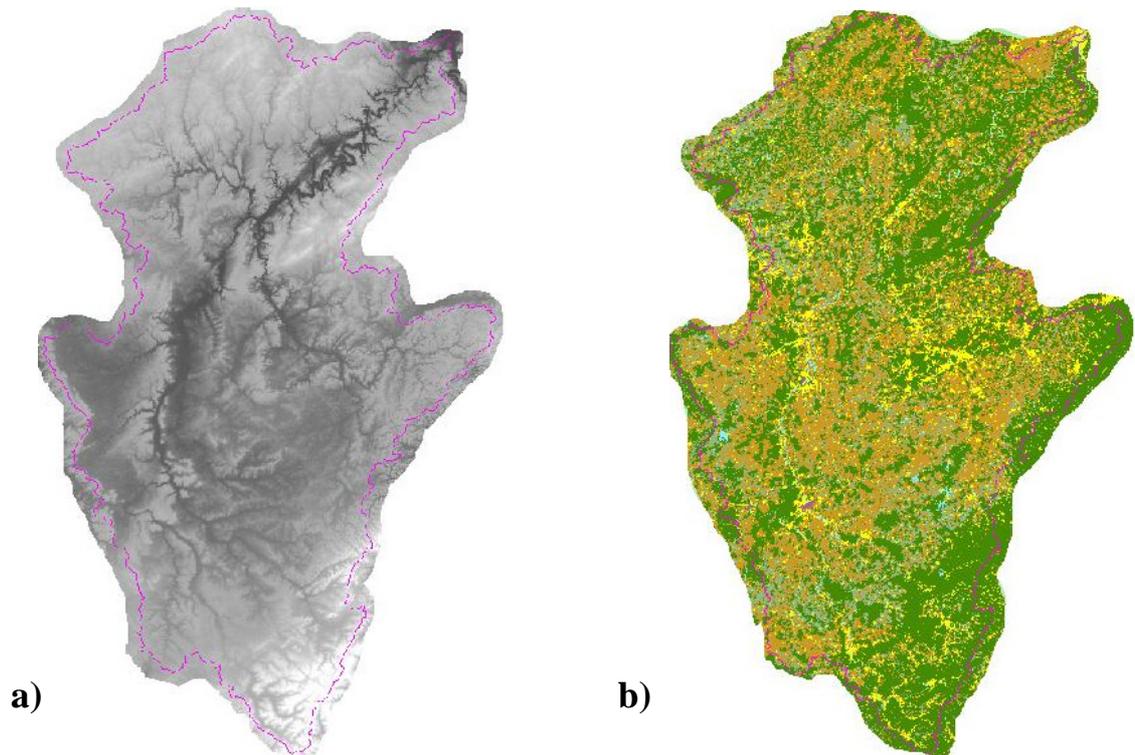


Figure 7-4 : Modèle numérique de terrain (a) et carte d'occupation du sol (b) calculés sur l'ensemble du bassin de la Moselle (30 000 km²) à une résolution de 20 mètres à partir d'images SPOT

1.4. Les caractéristiques calculées à partir des données disponibles

1.4.1. Les caractéristiques hydrologiques

Les *caractéristiques hydrologiques* étudiées par la suite sont :

- des quantiles de crue $V_d(T)$;
- la durée Δ , relative au modèle local convergent (Chap. 3)

Les quantiles $V_d(T)$, sont relatifs aux durées d de 0 (débits instantanés), 1, 6 et 15 jours, et pour **une période de retour de 10 ans**. Le cas de la durée 0 concerne les débits instantanés et n'a pas pu être étudié pour les 15 chroniques de débit journalier. La période de retour de 10 ans a été choisie pour les raisons suivantes :

- Cette période de retour avait déjà été étudiée en France par la méthode Crupedix (CTGREF, 1980), et est requise par l'approche bassin de référence (débit instantané décennal)
- Cette période de retour n'étant pas trop élevée, les quantiles peuvent être estimés de façon relativement précise à partir des seules données observées. Il n'est pas nécessaire d'extrapoler les distributions fréquentielles de débit.

Dans l'analyse par régression multiple réalisée au paragraphe suivant, ces *caractéristiques hydrologiques* constituent les *variables à expliquer*. Pour simplifier les notations, elles seront notées :

- $Q10_0$: débit décennal instantané, précédemment noté $Q_{d=0}(T=10)$, en m^3/s
- $Q10_1$: débit décennal moyen sur 1 jour, $V_1(10)$, en m^3/s
- $Q10_6$: débit décennal moyen sur 6 jours, $V_6(10)$, en m^3/s
- $Q10_{15}$: débit décennal moyen sur 15 jours, $V_{15}(10)$, en m^3/s
- $DELTA$: la durée caractéristique de crue du modèle local convergent, en jour.

Remarque : Contrairement à l'approche développée pour le modèle convergent (chap. 3), les durées d étudiées sont ici les mêmes quel que soit le bassin étudié. En conséquence, il peut arriver que pour les bassins rapides, la durée de 15 jours dépasse la durée des crues réellement observées. Cependant, cette durée est suffisamment courte pour que, dans la majorité des cas, on reste à peu près dans la dynamique de la crue. Par contre, la durée $DELTA$ a été ajustée selon la méthodologie du chapitre 3, c'est à dire avec des durées différentes selon le bassin considéré.

1.4.2. Les caractéristiques pluviométriques

Concernant les *caractéristiques pluviométriques*, il faut considérer séparément les données françaises ponctuelles, et allemandes déjà spatialisées. Plusieurs méthodes existent pour calculer une pluie de bassin, comme par exemple celle des polygones de Thiessen. Après avoir consulté différents avis sur la question (comité de thèse du 07/04/1999), une approche simple et pragmatique a été adoptée. Elle consiste, pour chaque bassin, à calculer chaque jour la moyenne des pluies journalières de tous les postes inclus dans le bassin, et fonctionnant ce jour là. Cette méthode correspond à celle des polygones de Thiessen simplifiée, sans pondération en fonction de la surface des polygones. Cette simplification a été adoptée pour deux raisons :

- Le grand nombre de bassins à traiter rend compliqué un calcul manuel des polygones de Thiessen, d'autant plus que pour un même bassin, ces polygones sont à réajuster chaque fois qu'un des postes pluviométriques comporte des lacunes.
- La précision sur la pluie de bassin obtenue est jugée suffisante, compte tenu de notre problématique.

Une fois la chronique journalière de pluie de bassin déterminée, celle ci est traitée selon l'approche intensité-durée-fréquence. Exactement de la même façon que pour l'analyse débit-durée-fréquence, on obtient des quantiles d'intensité moyenne de pluie sur une durée donnée d , pour une période de retour T , notés $I_d(T)$. Pour être cohérent avec les durées et les périodes de retour calculées pour des débits $V_d(T)$, la période de retour 10 ans, ainsi que les durées de 1, 3, 5, 10, 15 et 20 jours ont aussi été étudiées. Ces caractéristiques pluviométriques sont notées par la suite :

- PA : pluie (de bassin) moyenne annuelle (mm)

- *I10_1* : intensité (de bassin) décennale moyenne sur 1 jour (mm/j)
- *I10_3* : " " " " sur 3 jours "
- *I10_6* : " " " " sur 6 jours "
- *I10_10* : " " " " sur 10 jours "
- *I10_15* : " " " " sur 15 jours "
- *I10_20* : " " " " sur 20 jours "

1.4.3. Les caractéristiques physiques de bassin

L'extraction des *caractéristiques physiques de bassin* est réalisée à partir du Modèle Numérique de Terrain et de la carte d'occupation du sol précédemment présentés. Ces traitements ont été réalisés par Julie Charleux dans le cadre d'une thèse sur l'utilisation des MNT en hydrologie (Charleux, 2001). Celle-ci traite en particulier des problèmes liés à la résolution des MNT, et à l'extraction automatique des caractéristiques présentées ci-après. Ces caractéristiques dépendent de la résolution utilisée, ici, elles ont toutes été obtenues à partir d'une résolution de 20 mètres :

- *AREA* : la surface du bassin, en km² ;
- *ORIENT* : l'angle définissant l'orientation (voir Figure 7-5), en degrés ;
- *SHAPE* : l'indice d'élongation du bassin (0=un segment de droite, 1=un cercle, voir Figure 7-5) ;
- *ZM* : l'altitude moyenne du bassin, en m ;
- *SLOPE* : la moyenne des pentes locales calculées en chaque maille du bassin, en %
- *TOPO50* : la valeur médiane de la distribution de l'indice topographique défini par Beven et Kirkby (1979). Cet indice se calcule en chaque maille du bassin de la façon suivante :

$$I = \ln\left(\frac{A}{\tan \beta}\right) \quad (7-3)$$

Avec *A* surface drainée amont, et β , la pente locale.

- *DTE_DRAI* : la densité de drainage du bassin, définie à partir de la longueur des drains (*L*, en km) et de la surface du bassin (en km²) :

$$DTE_DRAI = \frac{L}{AREA} \quad (7-4)$$

- *FOR* : Pourcentage de forêt, entre 0 et 1
- *AGRI* : Pourcentage d'agriculture, entre 0 et 1
- *URB* : Pourcentage de zone urbaine, entre 0 et 1

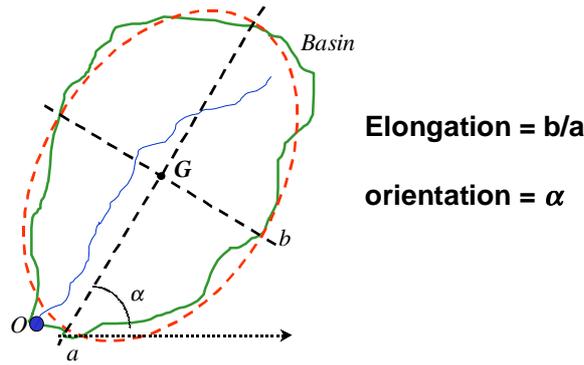


Figure 7-5 : Définition de l'élongation et de l'orientation

2. Estimation des caractéristiques hydrologiques par régressions multiples

2.1. Présentation des outils statistiques utilisés

Le but est d'estimer par régressions multiples les caractéristiques hydrologiques (§1.4.1.) à partir des caractéristiques pluviométriques (§1.4.2.) et des caractéristiques physiques de bassin (1.4.3.). Ce paragraphe présente brièvement les outils utilisés par la suite. Pour plus de détails, on peut par exemple se reporter à Duband (1992).

2.1.1. La régression linéaire multiple

Objectif

Le but est d'estimer une variable Y , appelée *variable à expliquer*, grâce à un ensemble de p *variables explicatives*, notées X_1 à X_p . L'estimation de Y , notée \hat{Y} est donnée par la relation linéaire suivante :

$$\hat{Y}_i = \alpha_0 + \alpha_1 X_{1i} + \dots + \alpha_p X_{pi} = Y_i - \varepsilon_i \quad (7-5)$$

$\alpha_1, \dots, \alpha_p$ sont les coefficients de la régression, α_0 la constante et ε le résidu de la régression.

La formulation ci-dessus est relative à la population d'où sont issues les variables Y et X_1 à X_p . Dans la pratique, on n'obtient qu'une estimation des coefficients de cette régression à partir des échantillons disponibles. La relation obtenue est notée :

$$\hat{Y}_i = a_0 + a_1 X_{1i} + \dots + a_p X_{pi} = Y_i - e_i \quad (7-6)$$

a_j est l'estimation de α_j

Estimation des coefficients de régression a_j

L'estimation des coefficients de la régression s'effectue en minimisant l'erreur quadratique entre la variable à expliquer et son estimation : $\sum_{i=1}^n (Y_i - \hat{Y}_i)^2$; n étant le nombre total

d'observations. Les valeurs a_0, \dots, a_p minimisant l'erreur quadratique ci-dessus sont déterminées en résolvant un système à $p+1$ équations et $p+1$ inconnues. Ce système est obtenu en écrivant que la dérivée partielle de l'erreur quadratique par rapport à a_i ($i=0, 1, \dots, p$) est nulle. En adoptant la notation matricielle suivante :

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \quad X = \begin{bmatrix} 1 & X_{11} & X_{21} & \dots & X_{p1} \\ 1 & X_{12} & X_{22} & & X_{p2} \\ \vdots & & & & \vdots \\ 1 & X_{1n} & X_{2n} & \dots & X_{pn} \end{bmatrix} \quad \text{et} \quad A = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_p \end{bmatrix} ;$$

le système d'équations à résoudre s'écrit (Duband, 1992) :

$$(X'X)A = X'Y \quad (7-7)$$

où X' est la transposée de X . En notant $(X'X)^{-1}$ la matrice inverse de $(X'X)$, la solution de ce système est donnée par :

$$A = (X'X)^{-1} X'Y \quad (7-8)$$

Coefficients de détermination

R^2 donne le pourcentage de la variance de Y expliquée par le modèle. Un R^2 nul signifie qu'il n'y a aucune liaison linéaire entre Y et son estimation \hat{Y} . Inversement, un R^2 de 1 signifie que l'estimation de Y est parfaite.

$$R^2 = 1 - \frac{\sum_{i=1}^n e_i^2}{\sum_{i=1}^n Y_i^2} \quad \text{ou encore} \quad R^2 = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (7-9)$$

2.1.2. Test de la pertinence des variables explicatives

Variance de l'estimation des paramètres de la régression

Les paramètres de la régression a_0, \dots, a_p étant des estimations des « vraies » valeurs $\alpha_0, \dots, \alpha_p$, il est possible de quantifier la fluctuation de ces estimations en calculant leur variance. Pour ce faire, on repart de la matrice $(X'X)^{-1}$ précédemment calculée, que l'on note :

$$(X'X)^{-1} = \begin{bmatrix} b_{00} & b_{10} & b_{20} & \dots & b_{p0} \\ b_{01} & b_{11} & b_{21} & & b_{p1} \\ \vdots & & & & \vdots \\ b_{0p} & b_{1p} & b_{2p} & \dots & b_{pp} \end{bmatrix} \quad (7-10)$$

La variance des paramètres a_0, \dots, a_p est alors donnée par :

$$\hat{v}ar(a_j) = b_{jj} \cdot \hat{v}ar(e) = b_{jj} \frac{\sum_{i=1}^n e_i^2}{n-p-1} \quad (7-11)$$

$\hat{v}ar(e)$ étant l'estimation de la variance de l'erreur résiduelle.

Test de signification de Student

Afin de savoir si la *variable explicative* X_j est significative, c'est à dire qu'elle permet réellement d'*expliquer* la variable Y , il est possible d'effectuer un test « de Student » sur le paramètre de régression a_i . Pour ce faire, on teste l'hypothèse suivante :

Hypothèse : « *La variable explicative X_j n'est pas significative* »

Ce qui revient à dire que le coefficient de régression a_i est suffisamment faible pour être négligé, ou bien qu'il est de signe opposé à α_j , le « vrai » coefficient.

Pour ce faire, on calcule le rapport suivant :

$$t = \frac{a_j}{\hat{\sigma}(a_j)} = \frac{a_j}{\sqrt{\hat{v}ar(a_j)}} \quad (7-12)$$

que l'on compare aux fractiles de la loi de Student $t_{1-p/2}(v)$, v étant le nombre de degrés de liberté ($v=n-p-1$) et p la **probabilité de se tromper en rejetant l'hypothèse testée**.

En résumé, plus la variable X_j est significative, plus la valeur absolue de t est forte, et plus p , le risque de rejeter à tort l'hypothèse que la variable n'est pas significative, est faible.

Remarque

Afin de déterminer la pertinence de la variable X_j , il est aussi possible d'effectuer un test de Student sur le coefficient de régression partielle. Ce coefficient mesure l'intensité de la relation entre Y et X_j , lorsque l'on élimine l'influence des autres variables explicatives sur ces deux variables (Duband, 1992). Cette démarche est équivalente à celle que nous utilisons.

2.1.3. Sélection des variables les plus pertinentes : la méthode Stepwise

Une méthode utilisée pour déterminer quelles sont les *variables explicatives* les plus pertinentes est celle de la sélection par Stepwise (ou méthode par adjonction). Cette méthode consiste à chercher d'abord la variable la plus corrélée avec la *variable à expliquer*, puis à rajouter pas à pas les variables provoquant le plus fort accroissement de R^2 . A chaque nouvelle variable, on teste si cet accroissement est significatif, afin de ne pas introduire de variable non pertinente, et l'on élimine les variables déjà introduites, mais qui deviennent éventuellement moins intéressantes, suite aux nouveaux ajouts.

Ces résultats ont été obtenus à l'aide du logiciel STATISTICA. Les échantillons de calibration dépendent de la variable à expliquer. Ils contiennent 27 bassins pour *Q10_0*, et 36 bassins pour les autres variables à expliquer (cf. Tableaux 7-2 et 7-3).

Nb de var. explicatives	Variables explicatives	a_j	Ecart type de a_j	t	p (%)	R^2 non biaisé
1	constante	-0,24	0,48	-0,50	61,89	0,77
	AREA	0,82	0,09	9,41	0,00	
2	constante.	-3,14	0,95	-3,29	0,00	0,84
	AREA	0,83	0,07	11,35	0,00	
	I10_10	1,02	0,30	3,34	0,00	
3	Constante	-4,92	0,87	-5,64	0,00	0,90
	AREA	0,81	0,06	13,99	0,00	
	I10_10	2,20	0,38	5,79	0,00	
	SLOPE	-0,73	0,18	-4,00	0,00	
4	Constante	-4,15	0,76	-5,49	0,00	0,93
	AREA	0,81	0,05	16,94	0,00	
	I10_10	1,95	0,32	6,01	0,00	
	SLOPE	-0,60	0,15	-3,89	0,00	
	SHAPE	0,48	0,14	3,39	0,00	

Tableau 7-4 : Méthode Stepwise appliquée à la variable à expliquer *Q10_0*

Nb de var. explicatives	variables explicatives	a_j	Ecart type de a_j	t	p (%)	R^2 non biaisé
1	Constante	-0,86	0,36	-2,37	0,02	0,84
	AREA	0,89	0,07	13,53	0,00	
2	Constante	-3,49	0,70	-5,02	0,00	0,89
	AREA	0,91	0,05	16,76	0,00	
	I10_10	0,93	0,22	4,18	0,00	
3	Constante	-5,61	0,83	-6,80	0,00	0,92
	AREA	0,91	0,05	19,74	0,00	
	I10_10	1,54	0,25	6,12	0,00	
	AGRI	1,31	0,36	3,68	0,00	
4	Constante	-5,06	0,85	-5,93	0,00	0,93
	AREA	0,89	0,05	19,67	0,00	
	I10_10	-0,64	1,24	-0,52	0,61	
	AGRI	1,58	0,38	4,21	0,00	
	I10_15	2,14	1,19	1,79	0,08	

Tableau 7-5 : Méthode Stepwise appliquée à la variable à expliquer *Q10_1*

Nb de var. explicatives	variables explicatives	a_j	Ecart type de a_j	t	p (%)	R^2 non biaisé
1	Constante	-1,90	0,29	-6,62	0,00	0,91
	AREA	0,98	0,05	18,76	0,00	
2	Constante	-4,37	0,33	-13,05	0,00	0,97
	AREA	0,99	0,03	33,21	0,00	
	I10_15	0,96	0,11	8,45	0,00	
3	Constante	0,03	1,79	0,02	98,68	0,97
	AREA	1,02	0,03	34,52	0,00	
	I10_15	1,76	0,34	5,16	0,00	
	PA	-0,95	0,38	-2,49	1,82	
4	Constante	2,45	2,10	1,17	0,25	0,98
	AREA	1,01	0,03	35,95	0,00	
	I10_15	2,84	0,63	4,52	0,00	
	PA	-1,19	0,38	-3,11	0,00	
	I10_3	-0,99	0,50	-2,00	0,05	

Tableau 7-6 : Méthode Stepwise appliquée à la variable à expliquer *Q10_6*

Nb de var. explicatives	variables explicatives	a_j	Ecart type de a_j	t	p (%)	R^2 non biaisé
1	Constante	-2,44	0,31	-7,86	0,00	0,90
	AREA	1,01	0,06	17,86	0,00	
2	Constante	-5,26	0,31	-17,02	0,00	0,98
	AREA	1,02	0,03	37,01	0,00	
	I10_15	1,09	0,10	10,47	0,00	
3	Constante	-5,46	0,33	-16,42	0,00	0,98
	AREA	1,02	0,03	37,63	0,00	
	I10_15	1,22	0,14	8,99	0,00	
	FOR	-0,33	0,23	-1,45	0,16	
4	Constante	-7,14	0,69	-10,31	0,00	0,98
	AREA	1,02	0,02	41,21	0,00	
	I10_15	1,01	0,15	6,86	0,00	
	FOR	-0,86	0,29	-3,00	0,01	
	ZM	0,41	0,15	2,70	0,01	

Tableau 7-7 : Méthode Stepwise appliquée à la variable à expliquer *Q10_15*

Nb de var. explicatives	variables explicatives	a_j	Ecart type de a_j	t	p (%)	R^2 non biaisé
1	Constante	2,69	0,22	11,99	0,00	0,31
	AGRI	-2,38	0,58	-4,12	0,00	
2	Constante	0,86	0,47	1,83	0,08	0,54
	AGRI	-2,46	0,47	-5,22	0,00	
	AREA	0,34	0,08	4,27	0,00	
3	Constante	3,99	1,20	3,31	0,00	0,62
	AGRI	-3,61	0,60	-6,05	0,00	
	AREA	0,34	0,07	4,61	0,00	
	I10_15	-1,08	0,39	-2,78	0,01	
4	Constante	8,41	2,68	3,14	0,00	0,65
	AGRI	-8,80	2,89	-3,04	0,00	
	AREA	0,35	0,07	4,91	0,00	
	I10_15	-1,27	0,39	-3,28	0,00	
	FOR	-5,53	3,02	-1,83	0,08	

Tableau 7-8 : Méthode Stepwise appliquée à la variable à expliquer *DELTA*

Pour chacune des régressions présentées dans les tableaux ci-dessus, nous avons comparé les valeurs « observées^{*} » à celles calculées, pour l'échantillon de calibration, mais aussi de validation. A titre d'exemple, les figures suivantes illustrent plus en détail les résultats de l'estimation du débit instantané décennal *Q10_0*, présenté au Tableau 7-4. Ces graphiques étant sur une échelle logarithmique, nous avons aussi représenté l'erreur relative $(obs-cal)/cal$ en fonction des valeurs observées. La Figure 7-6 concerne l'échantillon de calibration. La Figure 7-7 présente les résultats sur l'échantillon de validation.

* Il s'agit là d'un abus de langage. Les valeurs « observées » sont en fait des estimations locales.

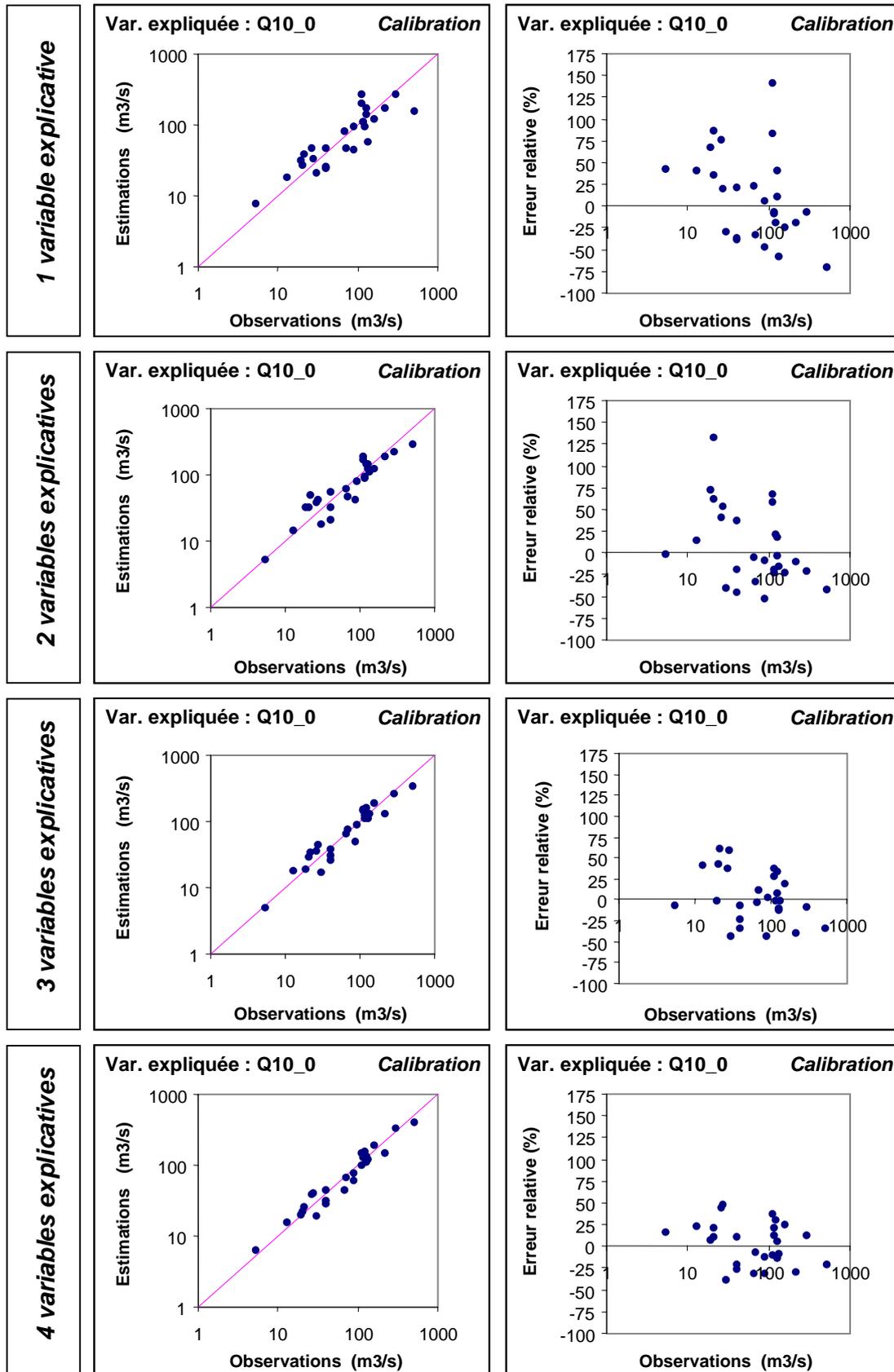


Figure 7-6 : Comparaison des estimations et des « observations » pour le débit instantané décennal *Q10_0* en fonction du nombre de variables explicatives : échantillon de calibration

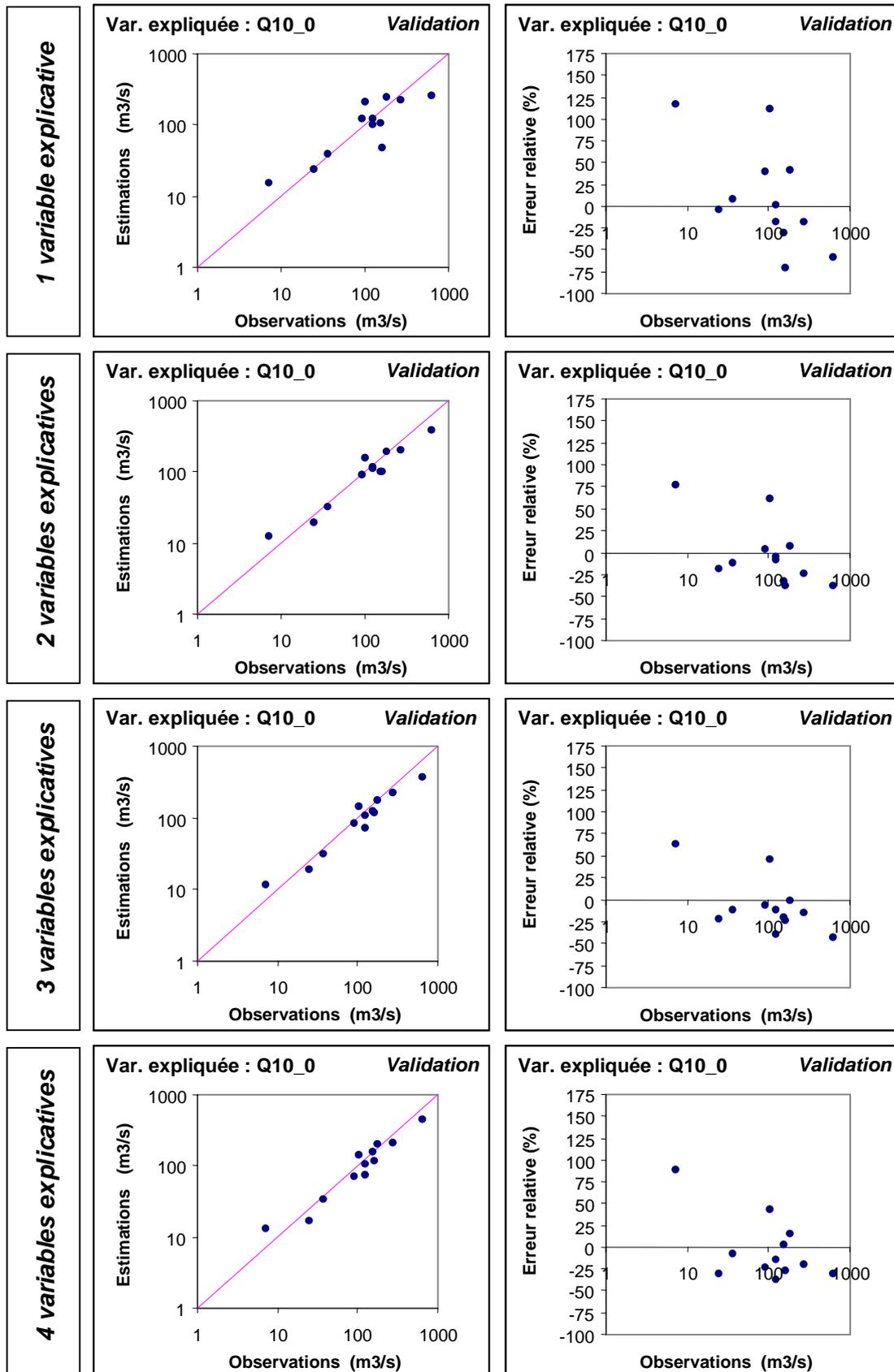


Figure 7-7 : Comparaison des estimations et des « observations » pour le débit instantané décennal *Q10_0* en fonction du nombre de variables explicatives : échantillon de validation

2.4. Premiers commentaires

Afin d’avoir une première synthèse des résultats présentés ci-dessus, nous avons représenté l’évolution de la racine carrée de l’erreur relative quadratique moyenne (*rRMSE*), en fonction du nombre de variables explicatives introduites (Figure 7-7). Ce critère a été préféré au R^2 , car comme les débits sont exprimés en m³/s, et en plus transformés en valeurs logarithmiques, on obtient toujours de très fortes valeurs de R^2 , cachant parfois des résultats en fait moyens. Le critère *rRMSE* a déjà été utilisé au chapitre précédent, nous le rappelons ici :

$$rRMSE = \left[\sum_{i=1}^n \left(\frac{cal_i - obs_i}{obs_i} \right)^2 \right]^{1/2} \tag{7-13}$$

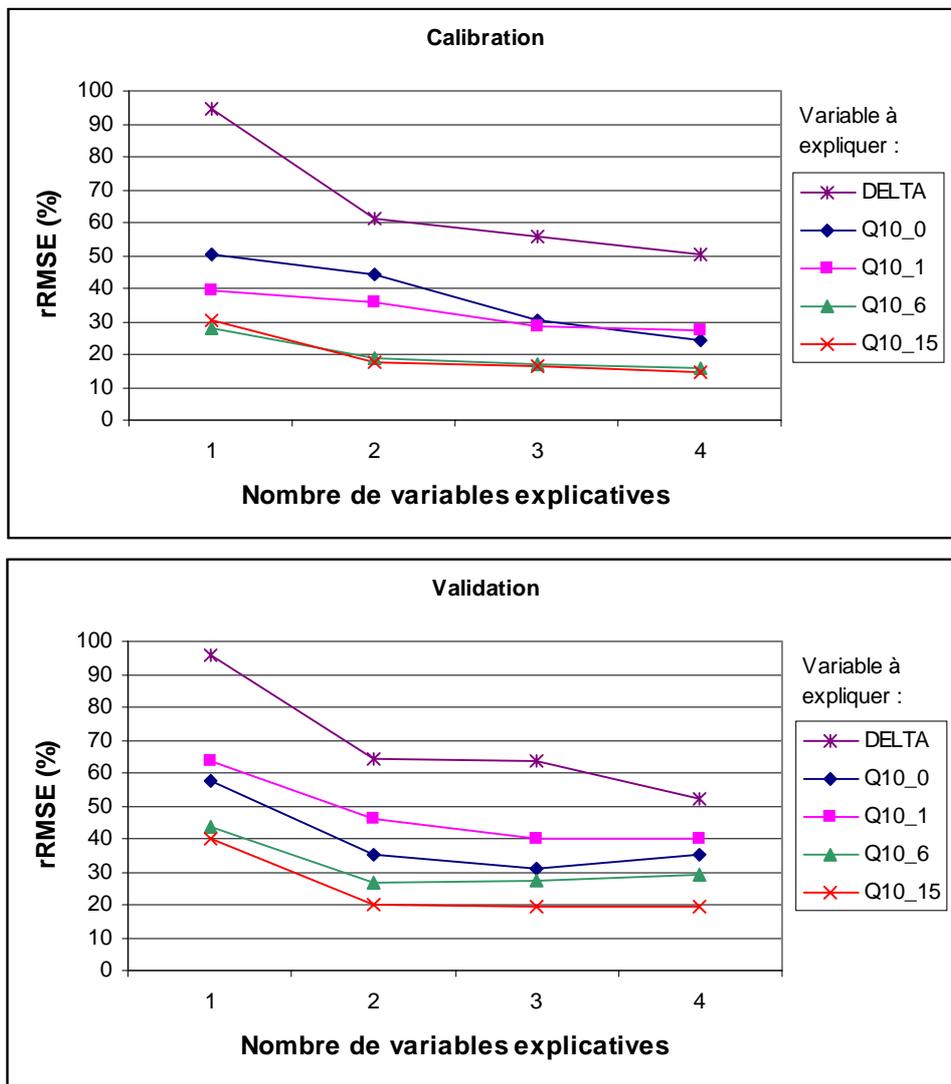


Figure 7-8 : Evolution de l’erreur quadratique moyenne (*rRMSE*) en fonction du nombre de variables explicatives, pour les échantillons de calibration et de validation

Les résultats de la **calibration** (graphique du haut, Figure 7-8) montrent une diminution assez sensible de l’erreur quadratique avec le nombre de variables explicatives introduites. Ceci était déjà visible sur les tableaux précédents, où l’on peut constater une augmentation du R^2 ,

ainsi qu'un fort pouvoir explicatif de toutes les variables introduites (valeurs de p , la probabilité de se tromper presque toujours inférieures à 1%).

Cependant, ces résultats sont à nuancer. En effet, lorsqu'on regarde maintenant les résultats de la **validation** (graphique du bas, Figure 7-7), le gain associé à l'introduction d'une nouvelle variable est beaucoup moins net. Au-delà de la troisième variable introduite, le gain est nul ($Q10_3$ et $Q10_15$), voire même négatif ($Q10_0$ et $Q10_6$). Seule *DELTA* semble profiter de l'introduction d'une quatrième variable.

Avant de discuter du rôle des différentes caractéristiques introduites pour expliquer les débits, il nous faut être sûr que les variables explicatives introduites sont réellement pertinentes. Pour cela, une méthode consiste à utiliser la « ridge regression » pour tester la robustesse des régressions obtenues. C'est ce que propose le paragraphe suivant.

3. Test à l'aide de la "ridge regression" de la robustesse des régressions obtenues

3.1. Présentation de la ridge regression (régression de crête)

Comme nous l'avons souligné au paragraphe 2.2. les variables explicatives sélectionnées sont très corrélées les unes avec les autres. Ainsi, même si le test de Student s'avère significatif (§2.1.2.), les coefficients de la régression sont peu stables à cause du problème de colinéarité. Dans ces conditions Champelly (1999) conseille trois méthodes de régression :

- la régression sur composante principale,
- la régression *PLS* (Partial Least Square) et
- la *ridge regression*.

Ce paragraphe utilise cette dernière méthode, introduite par Hoerl (1962), Hoerl et Kennard (1970). Le principe consiste à introduire une "perturbation" dans la matrice de corrélation et à étudier l'évolution des coefficients de régression en fonction de cette perturbation.

Afin de ne pas dépendre des différentes unités employées, Y , la variable à expliquer et X_j , les p variables explicatives ($j=1, \dots, p$) sont centrées réduites :

$$Y_i^* = \frac{Y_i - \bar{Y}}{\sigma(Y)} \quad ; \quad X_{ij}^* = \frac{X_{ij} - \bar{X}_j}{\sigma(X_j)} \quad (7-14)$$

L'équation (7-5) devient alors :

$$\hat{Y}_i^* = a_1^* X_{1i}^* + \dots + a_p^* X_{pi}^* = Y_i^* - e_i^* \quad (7-15)$$

Les coefficients a_j^* sont les coefficients de régression entre variables centrées réduites. Par définition, la constante $a_0^*=0$. La relation entre a_j^* et a_j ($j=1$ à p) est la suivante :

$$a_j^* = a_j \frac{\sigma(X_j)}{\sigma(Y)} \quad (7-16)$$

En notant A^* le vecteur des coefficients a_j^* , la relation (7-8) devient alors :

$$A^* = (X^{*t} X^*)^{-1} X^{*t} Y^* \quad (7-17)$$

Les variables étant centrées réduites, la matrice $X^{*t}X^*$ correspond à la matrice de corrélation. Si on note r_{kl} le coefficient de corrélation entre X_k et X_l , alors $X^{*t}X^*$ s'écrit :

$$(X^{*t} X^*) = \begin{bmatrix} 1 & r_{12} & \dots & r_{1p} \\ r_{12} & 1 & \dots & r_{2p} \\ \vdots & \vdots & & \vdots \\ r_{1p} & r_{2p} & \dots & 1 \end{bmatrix} \quad (7-18)$$

Dans le cadre de la ridge regression, un terme λ est ajouté sur la diagonale de la matrice de corrélation. En notant I la matrice unité, on a :

$$(X^{*t} X^*) + \lambda \cdot I = \begin{bmatrix} 1+\lambda & r_{12} & \dots & r_{1p} \\ r_{12} & 1+\lambda & \dots & r_{2p} \\ \vdots & \vdots & & \vdots \\ r_{1p} & r_{2p} & \dots & 1+\lambda \end{bmatrix} \quad (7-19)$$

Il est ensuite possible, pour la valeur de λ testée, de calculer une nouvelle estimation des coefficients de régressions :

$$A^*(\lambda) = (X^{*t} X^* + \lambda \cdot I)^{-1} X^{*t} Y^* \quad (7-20)$$

On a pour habitude de représenter sur un même graphique l'évolution des différents coefficients $a_j^*(\lambda)$ ($j=1, \dots, p$; p , le nombre de variables explicatives), pour des valeurs de λ généralement comprises entre 0 et 1. Ainsi, pour chaque variable explicative étudiée, on obtient sa "trace" en fonction de λ (Attention, le terme de « trace » n'est pas ici au sens de la trace d'une matrice. Il s'agit simplement de l'évolution de a_j^* en fonction de λ)

Hoerl (1962) cherche la valeur de λ (différente de zéro) pour laquelle l'estimation donnée par le vecteur $A^*(\lambda)$ est meilleure que celle donnée par le vecteur $A^*(0)$. Dans la suite de ce paragraphe, nous utilisons la ridge regression dans une optique légèrement différente. Celle-ci est décrite par Duband (1975). Il ne s'agit pas de trouver la valeur de λ optimale, mais d'éliminer les variables explicatives non pertinentes. En effet, la trace des différents coefficients $a_i^*(\lambda)$ permet d'apprécier leur stabilité, et par conséquent, la pertinence de la variable explicative X_i correspondante. Selon Duband (1975), une "trace instable" présente les symptômes suivants :

- variation non monotone ;
- changement de signe ;

- taux de variation très fort pour $0 < \lambda < 0,02$;
- changement de position relative par rapport aux traces des autres variables.

Une fois les variables non pertinentes éliminées, les coefficients de régression sont de nouveau calculés, cette fois-ci de façon classique, c'est à dire pour $\lambda=0$.

3.2. Application de la ridge regression

Malgré l'expression de prime abord un peu compliquée de $a_j^*(\lambda)$ donnée par la relation (7-20) les calculs suivants ont simplement été effectués sous un tableur EXCEL, à partir de la matrice des corrélations présentée aux Tableaux 7-2 et 7-3. En effet, Duband (1975) montre que les coefficients a_j^* peuvent s'exprimer de façon plus simple :

$$a_j^* = -\frac{d_{0j}}{d_{00}} \quad (7-21)$$

d_{ij} étant les composantes de la matrice inverse de la matrice de corrélation r_{ij} modifiée selon (7-19). Les indices i et j sont compris en 0 et p , le nombre de variables explicatives (la première ligne et la première colonne de cette matrice font référence à la variable à expliquer Y , les autres, aux variables explicatives X_j).

Nos calculs ont pu être validés grâce un à exemple fourni par Duband (Duband, 1975). Les résultats obtenus sont exactement identiques. Cette comparaison a aussi été effectuée avec le logiciel STATISTICA et est parfaitement cohérente.

Dans un premier temps, nous avons calculé les traces relatives à chacune des variables à expliquer étudiées au paragraphe 2.3. : *Q10_0*, *Q10_1*, *Q10_6*, *Q10_15* et *DELTA* expliqués avec quatre variables explicatives. Ces résultats sont présentés par les graphiques de gauche de la Figure 7-9. On constate que, sauf pour *Q10_0*, tous ces graphiques présentent une ou plusieurs traces « instables », c'est à dire pour lesquelles on observe un changement de signe ou un très fort taux de variation (cf. fin du §3.1.).

Puis, dans un second temps, nous avons enlevé la ou les variables explicatives désignées « instables » sur les graphiques de gauches. Ces résultats sont présentés par les graphiques de droite de la Figure 7-9. Les variables supprimées sont les suivantes :

- Pour *Q10_0*, aucune variable n'a été enlevée, car elles sont toutes robustes.
- Pour *Q10_1*, *I10_10* a été supprimée.
- Pour *Q10_6*, *PA* et *I10_3* ont été éliminées.
- Pour *Q10_15*, *FOR*, mais aussi *ZM*, ont été retirées. En effet, cette dernière n'a pas été déclarée instable sur le graphique de gauche, mais, lorsqu'on enlève *FOR*, son fort pouvoir explicatif relevé au Tableau 7-7 ($p=0,01\%$) chute à 42%. *ZM* a donc aussi été supprimée.

Pour *DELTA*, lorsqu'on élimine *FOR*, *I10_15* change à son tour de signe. Ces variables ont donc toutes les deux été retirées.

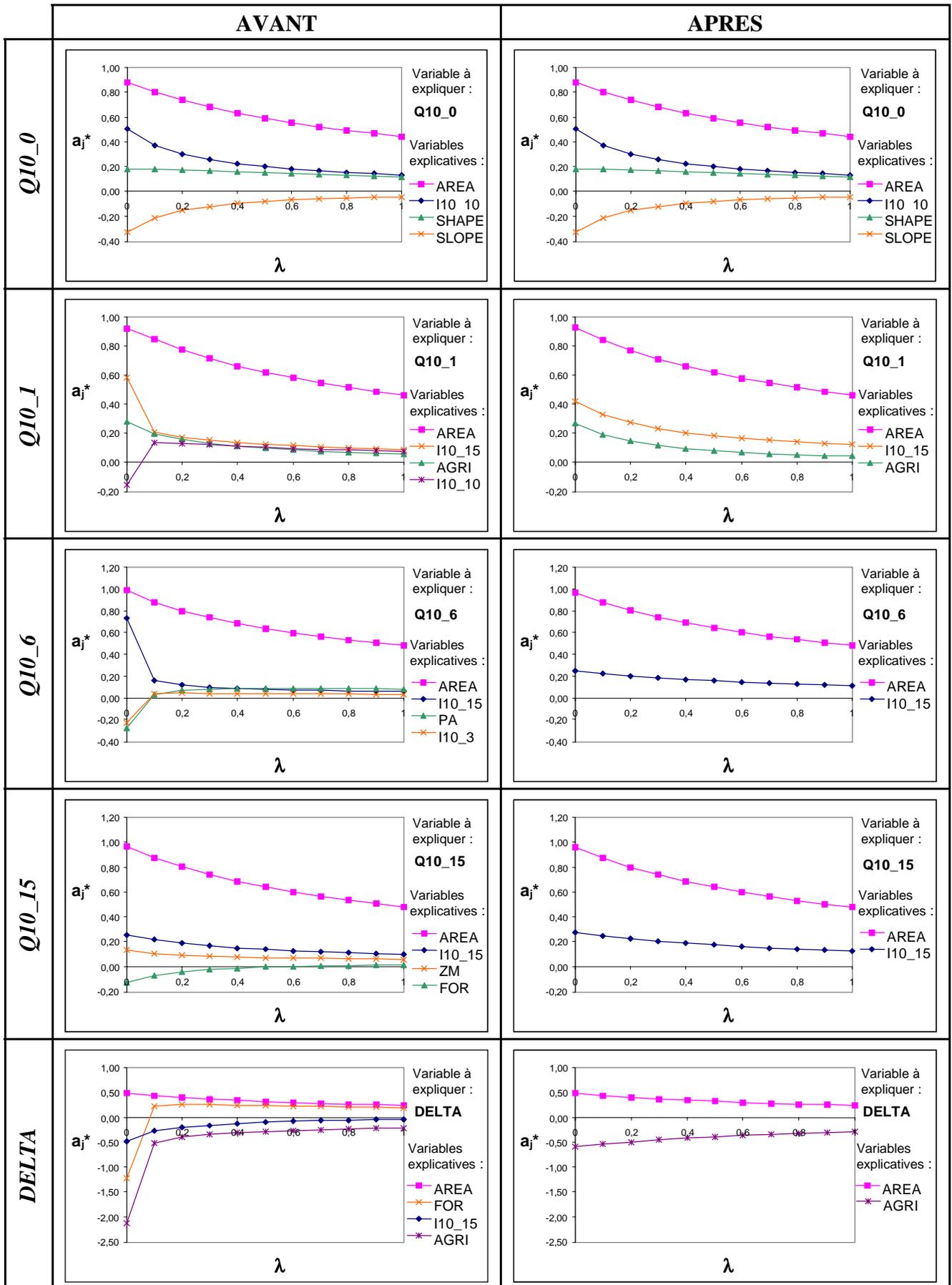


Figure 7-9 : Traces obtenues AVANT et APRES avoir enlevé les variables « instables »

La ridge regression permet donc de confirmer les doutes déjà formulés au paragraphe 2.4. sur la pertinence de certaines variables explicatives, pourtant établie par le test de Student

Cette mise à défaut du test de Student s'explique par la très forte corrélation entre certaines variables explicatives, constatée au paragraphe 2.2. , mais aussi par la faible taille de l'échantillon de calibration : 27 bassins pour *Q10_0* et 36 bassins pour les autres variables à expliquer.

Comme nous l'avons montré, l'intérêt de la ridge regression, est d'identifier les variables explicatives réellement pertinentes. Cependant, comme toute méthode statistique, cette approche n'est pas infaillible. Certaines variables non pertinentes peuvent subsister, ou inversement. A titre d'exemple, les résultats sur la variable à expliquer *DELTA* peuvent paraître un peu sévères. Le paragraphe 2.4. montrait en effet un réel gain lorsqu'on passait au modèle à quatre paramètres, aussi bien en calibration qu'en validation. L'application de la ridge regression penche clairement pour un modèle à 2 paramètres. Il faut cependant noter que le cas de *DELTA* est un peu particulier : c'est pour cette variable que l'on observe les plus mauvais résultats : $R^2=65\%$, même en calibration, et avec le modèle à 4 paramètres. Pour rester cohérent avec toute la démarche, nous nous sommes rangés à l'avis de la ridge regression, dont le principal intérêt reste son objectivité.

3.3. Résultats finaux et interprétation

3.3.1. Présentation des régressions finalement obtenues

Les graphiques suivants présentent les régressions obtenues avec les variables explicatives finalement retenues. Nous distinguons les résultats de la calibration et ceux de la validation. De plus, nous avons représenté les valeurs estimées par la régression en fonction des valeurs observées, ainsi que l'erreur relative (*cal-obs*)/*obs* en fonction des valeurs observées.

Ces régressions sont résumées dans le Tableau 7-9. Outre les variables explicatives et leurs coefficients de régression, figurent le paramètre *t* utilisé lors du test de Student, ainsi que l'erreur quadratique moyenne *rRMSE*, calculée selon (7-13). On peut voir la hiérarchie des variables explicatives, ainsi que la qualité des estimations, en calibration et en validation.

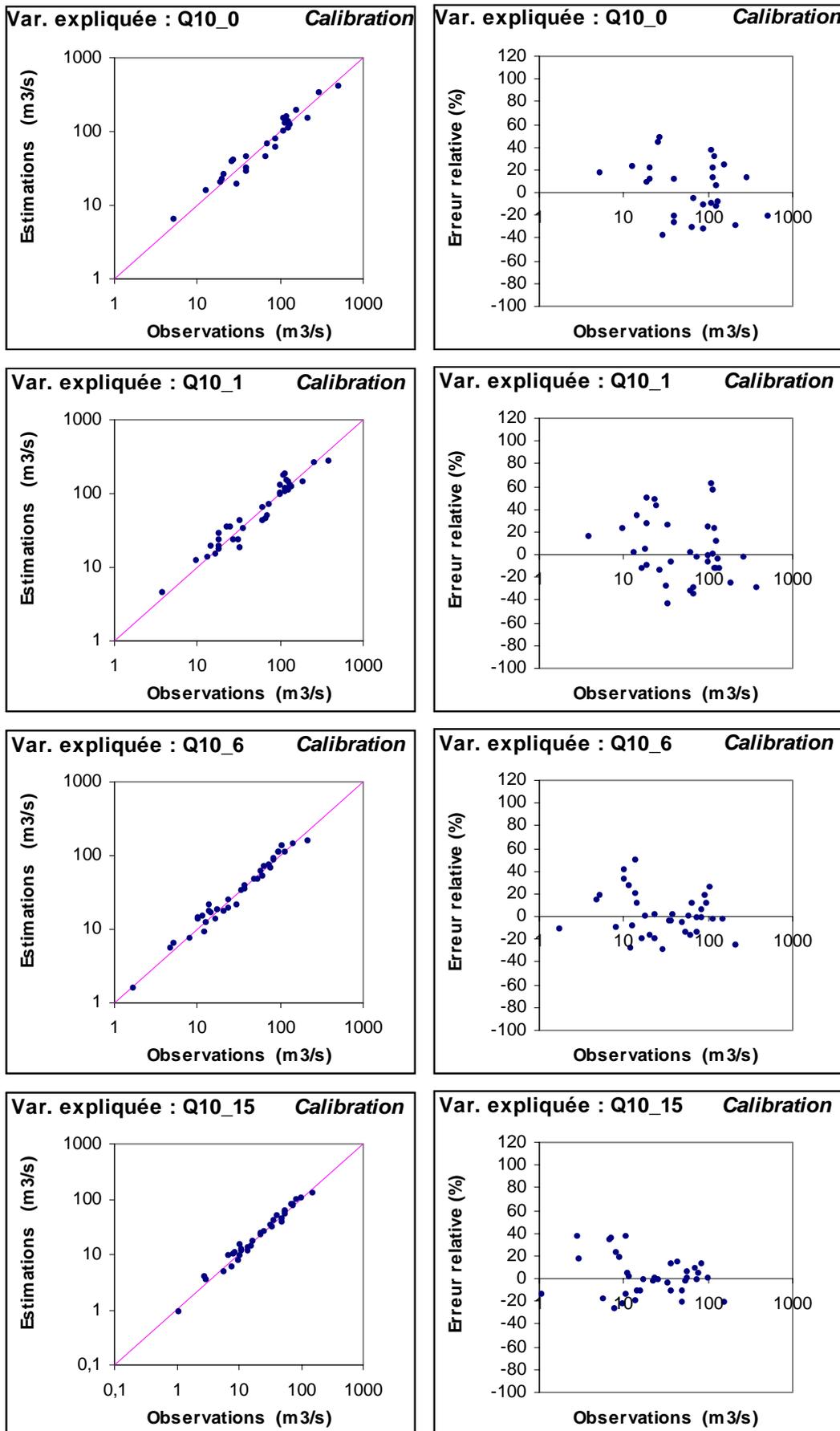


Figure 7-10 : Estimation de $Q10_0$, $Q10_1$, $Q10_6$ et $Q10_{15}$: CALIBRATION

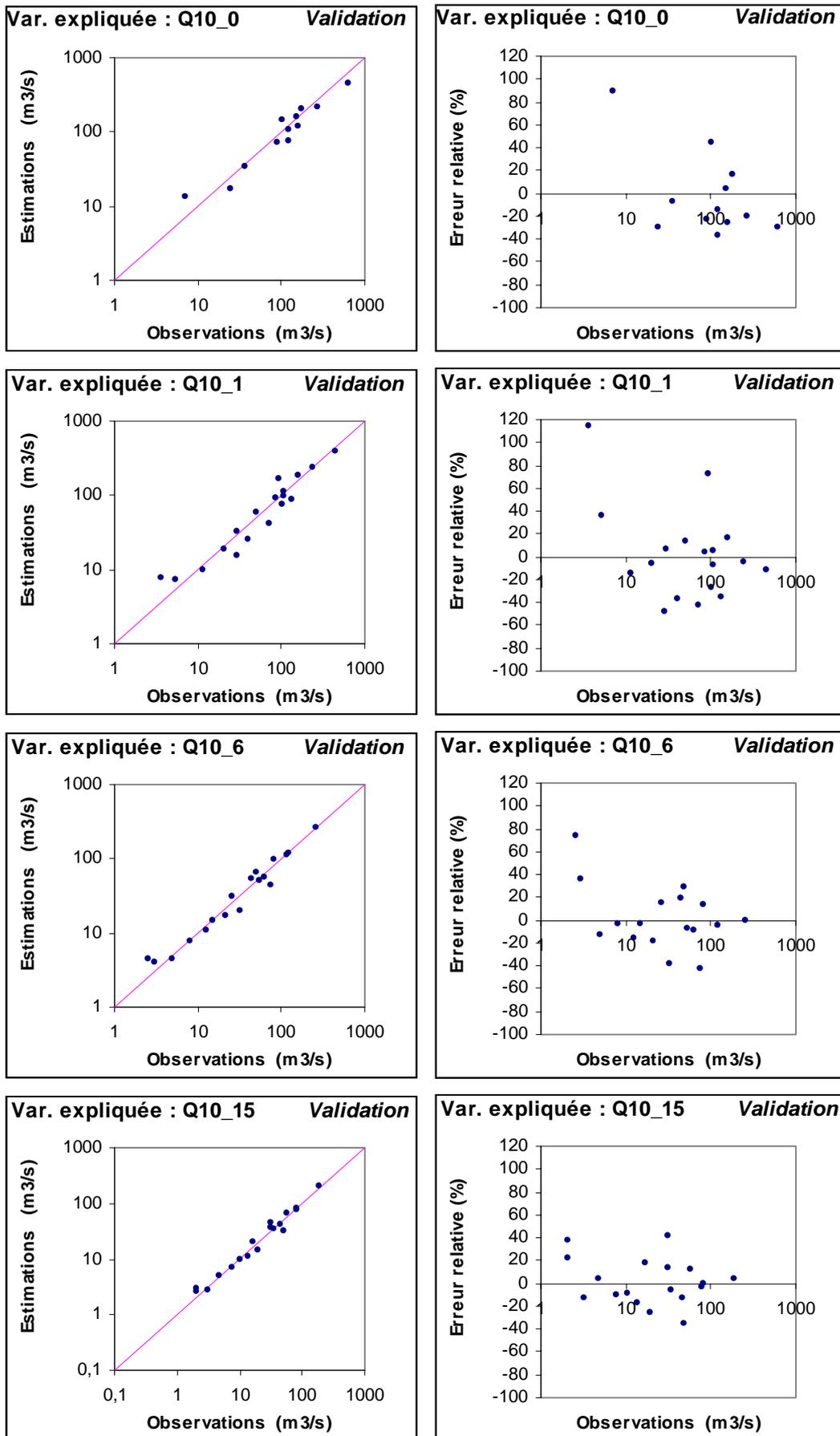
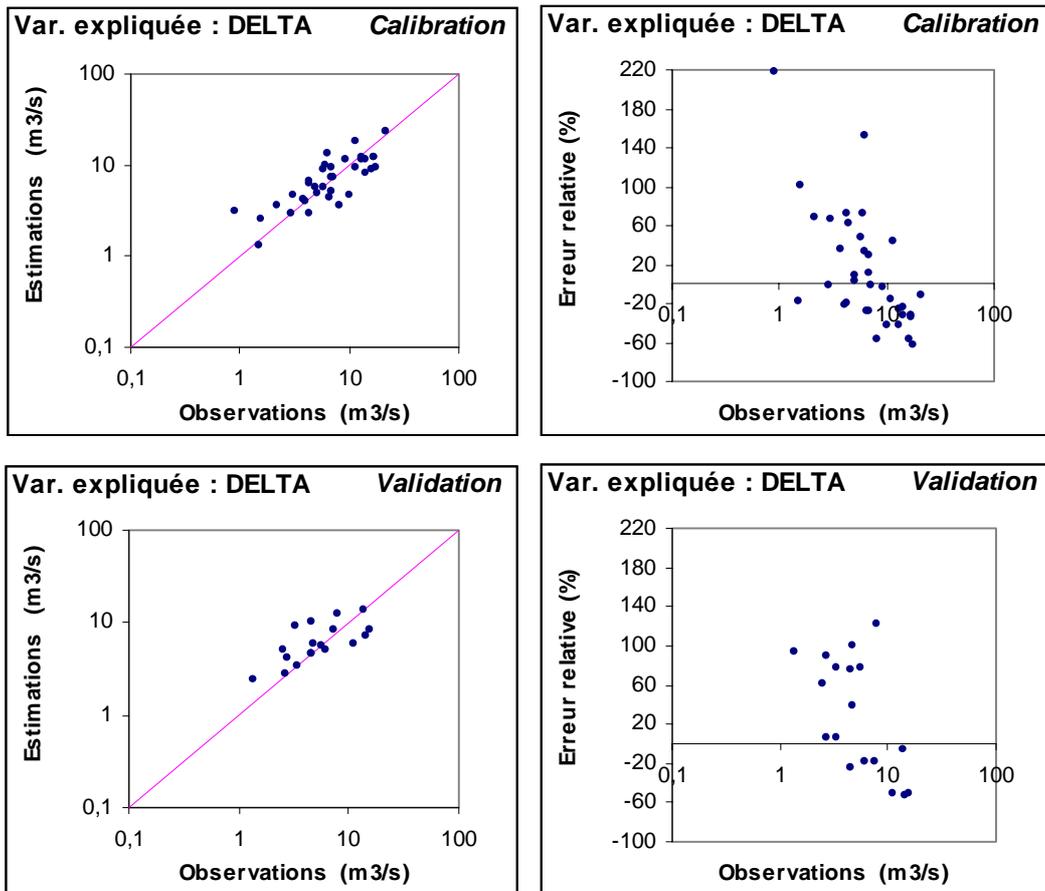


Figure 7-11 : Estimation de Q_{10_0} , Q_{10_1} , Q_{10_6} et Q_{10_15} : VALIDATION


 Figure 7-12 : Estimation de *DELTA* : CALIBRATION et VALIDATION

Régressions	<i>rRMSE</i> (%)	
	Calibration	Validation
$Q_{10_0} = 1,57 \cdot 10^{-2} \cdot AREA^{0,81} \cdot I_{10_10}^{1,95} \cdot SLOPE^{-0,60} \cdot SHAPE^{0,48}$ (t=16,94) (t=6,01) (t=-3,89) (t=3,39)	24,3%	35,4%
$Q_{10_1} = 5,03 \cdot 10^{-3} \cdot AREA^{0,90} \cdot I_{10_15}^{1,54} \cdot (1 + AGRI)^{1,52}$ (t=20,43) (t=6,64) (t=4,28)	27,1%	39,4%
$Q_{10_6} = 1,26 \cdot 10^{-2} \cdot AREA^{0,99} \cdot I_{10_15}^{0,96}$ (t=33,21) (t=8,45)	18,7%	26,6%
$Q_{10_15} = 5,18 \cdot 10^{-3} \cdot AREA^{1,02} \cdot I_{10_15}^{1,09}$ (t=37,01) (t=10,47)	17,3%	19,8%
$DELTA = 54,1 \cdot AREA^{0,34} \cdot (1 + AGRI)^{-2,46}$ (t=4,27) (t=-5,22)	56,0%	63,4%

Tableau 7-9 : Régressions finalement retenues

3.3.2. Première remarque : évolution des résultats en fonction de la variable à expliquer

Le Tableau 7-9 montre que la variable la moins bien estimée est la durée *DELTA*, avec une erreur quadratique de 56% en calibration et 63% en validation.

Concernant les quantiles décennaux, on constate une amélioration des résultats lorsque *d* augmente. On passe d'une erreur de 23% en calibration et 35% en validation pour *Q10_0* à 17% et 20% pour *Q10_15*. Cependant, la validation de *Q10_1* vient contredire cette affirmation. Mais, comme nous l'avons vu, les échantillons de validation et de calibration utilisés pour *Q10_0* sont plus petits que ceux utilisés pour les autres variables. En particulier, ils ne contiennent pas les données allemandes, pour lesquelles seules les chroniques de débit journalier étaient disponibles. Ainsi, certains de ces bassins peuvent présenter des singularités. Par exemple, Figure 7-11, on constate que pour quelques bassins (non présents dans l'échantillon de *Q10_0*), l'estimation de *Q10_1* est particulièrement mauvaise. En particulier, un bassin de validation a une erreur de 115%. Les causes de ce mauvais résultat n'ont pu être déterminées. Cependant, ne faisant pas partie de l'échantillon de calibration, ce bassin n'est pas venu influencer sur la détermination des variables explicatives ni sur les coefficients de régression.

3.3.3. Deuxième remarque : hiérarchie des différentes variables explicatives

La variable de Student *t*, Tableau 7-9, indique le pouvoir explicatif de chacune des différentes variables des régressions. On note la même hiérarchie, quelle que soit la variable à expliquer considérée, sauf pour *DELTA*. Cette hiérarchie est la suivante.

Première caractéristique : la surface

En premier lieu, on constate le poids prépondérant de la **surface** (*AREA*) dans les régressions. Ceci avait déjà été souligné lors de l'étude des corrélations simples entre les différentes variables. Ce résultat était donc attendu, mais on peut tout de même noter que l'exposant de *AREA* intervenant dans la régression est respectivement de 0,81, 0,91, 0,99, 1,02 pour respectivement *Q10_0*, *Q10_1*, *Q10_6* et *Q10_15*. Remarquons que l'exposant de 0,81 relatif aux quantiles instantanés est très proche de celui de 0,8 déterminé par la formule de Crupedix (7-1) à l'échelle de la France.

Ainsi, lorsque la durée augmente, l'exposant de *AREA* tend vers 1. Comme le montre les tableaux 7-4 à 7-7 du paragraphe 2.3. ce résultat ne dépend pas du nombre de variables explicatives introduites. Ceci montre qu'*AREA* est très peu liée aux autres variables explicatives, comme on l'avait déjà vu sur la matrice des corrélations simples (§2.2.).

Concernant *DELTA*, même si *AREA* n'est pas la variable la plus explicative, son rôle est également déterminant.

Seconde caractéristique : la pluie

La seconde variable la plus explicative est liée à la **pluie**. Parmi les caractéristiques pluviométriques testées, à savoir la pluie annuelle (*PA*) et les intensités décennales moyennes sur 1, 3, 6, 10, 15 et 20 jours (*I10_1*, *I10_3*, *I10_6*, *I10_10*, *I10_15*, *I10_20*), les deux variables sélectionnées sont *I10_10* (pour *Q10_0*) et *I10_15* (pour *Q10_1*, *Q10_6*, *Q10_15*). Ainsi les quantiles de pluie moyenne sur 10 à 15 jours sont les plus explicatifs, même lorsqu'on étudie des débits calculés sur des pas de temps beaucoup plus fins. Ceci peut éventuellement s'expliquer par le fait que ces durées de 10 à 15 jours entrent mieux dans la dynamique des bassins étudiés. Il faut cependant noter que ces quantiles de pluie sont corrélés avec l'altitude moyenne et que cette liaison augmente avec le pas de temps considéré. Par exemple, le coefficient de corrélation entre *ZM* et *I10_d* passe de 52% à 78% lorsque *d* varie de 1 à 20 jours (Tableau 7-3). Les variables *I10_10* et *I10_15* ont donc peut être aussi été sélectionnées parce qu'elles contiennent indirectement l'altitude.

Le cas de *DELTA* est un peu particulier. La pluie *I10_10* avait bien été sélectionnée par la méthode Stepwise. Mais suite à l'analyse de la ridge regression, nous l'avons retirée.

Troisième (et quatrième) caractéristique : les caractéristiques physiques de bassin

Après les deux premiers types de variables décrites ci-dessus (surface et pluie), les plus explicatives sont les caractéristiques physiques de bassin (excepté la surface). Celles-ci dépendent de la variable explicative considérée :

- Pour *Q10_0*, il s'agit de la pente *SLOPE* et du coefficient de forme *SHAPE* ;
- pour *Q10_1* et *DELTA*, du pourcentage d'agriculture *AGRI* ;
- enfin, pour *Q10_6* et *Q10_15*, aucune caractéristique physique de bassin (excepté *AREA*) n'a finalement été retenue.

3.3.4. Interprétation des résultats et rôle de la fonction de transfert

Revenons tout d'abord sur l'amélioration des résultats en fonction de la caractéristique hydrologique considérée. Pour un même volume précipité, la valeur du débit de pointe instantané dépend fortement du transfert au sein du bassin étudié : fort débit pour un bassin rapide, faible débit pour un bassin lent. Si on étudie le volume écoulé sur une durée de plus en plus longue, le rôle du transfert devient de moins en moins important. **Les plus mauvais résultats sont donc observés là où le transfert joue un rôle déterminant, c'est à dire pour le débit instantané *Q10_0* et la durée *DELTA*.** En effet, même si *DELTA* n'est pas à proprement parler un temps de transfert, nous avons montré que cette variable est fortement liée à la forme des crues observées (voir chapitre 3). **A l'inverse, les résultats sont les meilleurs là où le transfert intervient le moins**, c'est à dire pour *Q10_15*, relatif au volume maximal écoulé sur 15 jours.

Un autre élément peut aussi venir expliquer ce point. A cause de l'extrapolation de la courbe de tarage, la valeur du débit de pointe instantané est probablement moins précise que celle du

débit moyenné sur 15 jours. Cette réduction des erreurs de mesure pour les grandes durées contribue à améliorer les résultats des régressions.

Étudions maintenant le rôle des caractéristiques physiques de bassin intervenant dans l'estimation des différentes caractéristiques hydrologiques.

Concernant *Q10_0*, les coefficients de régression indiquent que ce débit augmente avec le coefficient de forme *SHAPE*. Rappelons que *SHAPE*=0 correspond à un bassin très allongé et *SHAPE*=1 à un bassin rond. Ceci peut s'expliquer par le fait que dans ce dernier cas les temps de transfert sont moins longs, et donc le débit de pointe plus fort. Concernant la variable *SLOPE*, on note un coefficient négatif, ce qui signifie que *Q10_0* diminue lorsque la pente augmente. Si l'on fait le même raisonnement que précédemment, l'augmentation de la pente entraînerait plutôt une diminution des temps de transfert, et donc une augmentation du débit de pointe. Cette apparente contradiction s'explique par le fait que la pente est positivement corrélée avec le pourcentage de forêt *FOR* ($R=88\%$, cf. Tableau 7-2). Le signe négatif du coefficient de *SLOPE* indique donc plutôt que le débit de pointe diminue lorsque le pourcentage de forêt augmente, ce qui paraît plausible (transfert plus lent).

Pour *Q10_1*, la caractéristique de terrain sélectionnée est le pourcentage d'agriculture *AGRI*, avec un coefficient de régression positif. Les surfaces agricoles tendraient donc à augmenter le débit sur une courte durée. Comme nous l'avons déjà noté, de nombreuses variables étant inter-corrélées, les résultats doivent être interprétés prudemment. Ici, l'agriculture est très fortement inversement corrélée à la forêt ($R=-0.99$). Le résultat obtenu tient donc aussi du fait que la forêt a tendance à ralentir les écoulements.

Concernant *Q10_6* et *Q10_15*, aucune des caractéristiques physiques de bassin testées ne parvient à améliorer significativement les résultats. Comme nous l'avons déjà souligné, c'est aussi pour ces caractéristiques que les résultats sont les meilleurs. Cela signifie que les caractéristiques physiques de bassin sont inutiles tant que le transfert n'intervient pas. Lorsque la durée est courte (*Q10_0* ou *Q10_1*), ces caractéristiques physiques deviennent utiles, mais elle ne sont pas fortement explicatives.

Les régressions relatives à la durée *DELTA* sont les plus mauvaises. Les signes des coefficients de régression indiquent que *DELTA* augmente lorsque *AGRI* diminue et lorsque *AREA* augmente. Si l'on considère que *DELTA* dépend des vitesses de transfert sur le bassin étudié (*DELTA* fort, transfert lent), les variations de *DELTA* avec *AGRI* et *AREA* sont plausibles. Notons de plus que *AGRI* est fortement inversement corrélé avec la forêt ($R=-0.99$, cf. Tableau 7-3): il semble que la forêt ralentit le transfert et que les zones agricoles l'accélèrent.

3.4. Conclusion sur les régressions

Nous restons prudents dans l'interprétation du rôle physique des différentes variables explicatives. En effet, la taille des échantillons est faible, les variables explicatives étudiées sont pratiquement toutes liées à l'altitude.

Les variables explicatives sont par ordre d'importance, la surface, la pluie (pluie décennale sur 10 ou 15 jours), puis les caractéristiques physiques de bassin (excepté la surface). Ces dernières ne jouent un rôle que pour les caractéristiques très liées à la fonction de transfert ($DELTA$, $Q10_0$ et $Q10_1$). C'est aussi pour ces caractéristiques que l'on observe les moins bons résultats.

Les caractéristiques physiques de bassins aident donc à mieux prendre en compte le transfert au sein du bassin. Malheureusement, leur pouvoir explicatif reste faible.

4. Application « purement régionale » de l'approche débit-durée-fréquence

4.1. Objectif et intérêt de la démarche

L'analyse des régressions multiples présentées en début de ce chapitre permet d'estimer des caractéristiques hydrologiques sur des bassins où l'on ne dispose pas d'information sur les débits. On pourrait se contenter de ces résultats, et ainsi estimer simplement les quantiles $V_d(T)$ étudiés ($T=10$ ans, $d=0, 1, 6$ et 15 jours), respectivement notés $Q10_0$, $Q10_1$, $Q10_6$ et $Q10_{15}$ dans les régressions du Tableau 7-9. L'inconvénient de cette approche (appelée régression directe au chapitre 5, §1.3), est que pour estimer des quantiles relatifs à une autre période de retour ou à une autre durée, il faut établir de nouvelles régressions. Tout le travail des paragraphes 2. et 3. est donc à refaire !

L'utilisation de modèles régionaux adimensionnels présentés aux chapitres 5 et 6 permet de résoudre ce problème. L'estimation par régressions de deux caractéristiques hydrologiques (un quantile de crue et une durée) donne accès à la totalité de l'information résumée par ces modèles adimensionnels. Les deux approches débit-durée-fréquence possibles sont :

- l'approche « bassin de référence » de Galéa et Prudhomme (1997), pour laquelle il faut :
 - i) connaître le débit instantané décennal $Q(0,10)$ et la durée de Socose D du bassin étudié,
 - ii) choisir parmi les trois possibles le bassin de référence à utiliser (cf. chap. 6) ;
- le modèle **régional** convergent (cf. chap. 5) pour lequel il faut connaître $Q(0,10)$ et la durée de crue Δ relative au modèle **local** convergent (cf. chap. 3).

Le chapitre 6 a comparé les deux approches ci-dessus, mais seulement dans le cadre d'une analyse « locale-régionale » : les données locales ont été utilisées pour déterminer les caractéristiques $Q(0,10)$, D et Δ , selon l'approche appliquée.

La suite de ce chapitre étudie les estimations cette fois-ci « purement régionales » de ces modèles : les caractéristiques $Q(0,10)$ et Δ sont déterminées à partir de régressions précédemment présentées. La durée D sera estimée au prochain paragraphe.

4.2. Rappel de la méthodologie

- **Echantillonnage multi-durée**

L'échantillonnage multi-durée est présenté au chapitre 3. Six durées d ont été étudiées, comprises entre 0 et d_6 , puis le modèle local convergent ajusté sur les échantillons obtenus. La durée d_6 est déterminée après plusieurs itérations, jusqu'à ce qu'elle soit de l'ordre $1,5\Delta$. Les durées étudiées dépendent donc de chaque bassin étudié, contrairement aux régressions multiples précédemment présentées, où nous avons fait le choix de n'étudier que les durées de 0, 1, 6 et 15 jours.

- **Le modèle régional convergent**

Nous rappelons ici les principales étapes de l'application du modèle régional convergent :

- Pour chaque bassin i de l'échantillon de calibration, ajustement du modèle local convergent, sous sa formulation simplifiée, c'est à dire avec P fixé à zéro. En effet, comme nous l'avons vu, ce paramètre n'améliore pas significativement les résultats.

$$V(d,T)_i = \frac{x_0(0)_i + a(0)_i \ln(T)}{1 + d / \Delta_i} \quad (7-22)$$

- Normalisation des paramètres de positions $x_0(0)$ et d'échelle $a(0)$ par le débit instantané décennal local :

$$x_{0i}^* = \frac{x_0(0)_i}{Q(0,10)_i} \quad ; \quad a_i^* = \frac{a(0)_i}{Q(0,10)_i} \quad (7-23)$$

- Calcul des coefficients régionaux x_{0R} et a_R en faisant la moyenne des coefficients locaux, obtenus sur les bassins de l'échantillon de calibration :

$$x_{0R}^* = \frac{1}{N} \sum_{j=1}^N x_{0j}^* \quad ; \quad a_R^* = \frac{1}{N} \sum_{j=1}^N a_j^* \quad (7-24)$$

Puis deux estimations sont possibles :

- Estimation « locale-régionale » en chaque bassin en dénormant le modèle régional convergent par le débit instantané décennal $Q(0,10)$ local et la durée Δ locale :

$$V(d,T)_{\text{régional}}^{\text{local}} = \frac{x_{0R}^* + a_R^* \ln(T)}{1 + d / \Delta_{\text{local}}} \cdot Q(0,10)_{\text{local}} \quad (7-25)$$

- Estimation « purement régionale » en chaque bassin, cette fois ci en dénormant le modèle régional convergent par le débit instantané décennal $Q(0,10)$ et la durée Δ obtenue par régressions à partir des caractéristiques pluviométriques et physiques du bassin :

$$V(d, T)_{\substack{\text{purement} \\ \text{régional}}} = \frac{x_0^* + a^* \ln(T)}{1 + d / \Delta_{\text{régional}}} \cdot Q(0, 10)_{\text{régional}} \quad (7-26)$$

Pour les estimations régionales de $Q(0, 10)$ et Δ , nous avons utilisé les régressions résumées aux Tableau 7-9.

Remarque : Contrairement au chapitre précédent, nous testons ici le modèle convergent dans sa forme simplifiée ($P=0$), ce qui nous permettra de voir l'incidence de ce choix.

- **L'approche « bassin de référence »**

De la même façon, l'approche « bassin de référence » décrite au chapitre 5 a été appliquée, selon une démarche « locale-régionale », puis « purement régionale ». Ici, les deux caractéristiques à connaître sont le débit décennal instantané $Q(0, 10)$ et la durée de crue D dite « de Socose » (cf. chap. 5, §1.). Comme nous l'avons vu, $Q(0, 10)$ peut être estimé grâce à la formule de régression de $Q10_0$ Tableau 7-9. Par contre, la durée D n'a pas été précédemment étudiée. Nous avons donc aussi établi une formule de régression pour cette caractéristique, ajustée sur les bassins de calibration. Les variables explicatives ont été sélectionnées par la méthode Stepwise :

$$D_{\substack{\text{purement} \\ \text{régional}}} = 13,16 \cdot AREA^{0,44} \cdot (1+AGRI)^{-3,26} \cdot I10_15^{-1,33} \quad (7-27)$$

(t=6,14) (t=-4,27) (t=-2,86)

La Figure 7-13 présente les résultats obtenus, sur les 27 bassins de calibration, et les 12 de validation. L'erreur quadratique $rRMSE$ est respectivement de 55% et 53%. Ainsi, tout comme pour le durée Δ étudiée plus en détail précédemment, les résultats ne sont pas très bons.

Comme nous l'avons vu au chapitre 5, le choix du modèle de référence, entre *Soyans*, *Florac*, et *Vandenesse* pose problème. Le critère de choix proposé n'est pas pertinent pour les périodes de retour que nous étudions (périodes observables). De plus, celui-ci fait intervenir les caractéristiques $Q(0, 10)$ et D supposées inconnues dans le cadre d'une estimation purement régionale et il y a peu de chance pour que le critère s'améliore lorsqu'on l'utilise avec les estimations purement régionales de ces caractéristiques. Pour résoudre ce problème, nous avons appliqué le critère de choix sur chaque bassin de l'échantillon de calibration en utilisant les données locales. Puis nous avons observé que le bassin de *Vandenesse* était systématiquement choisi. Suite à cette observation faite sur l'échantillon de calibration, nous avons donc choisi d'appliquer systématiquement le bassin de *Vandenesse*. Nous sortons donc ici un peu de la méthodologie énoncée par Galéa et Prudhomme (1997), mais nous n'utilisons aucune des données hydrologiques locales des bassins de validation.

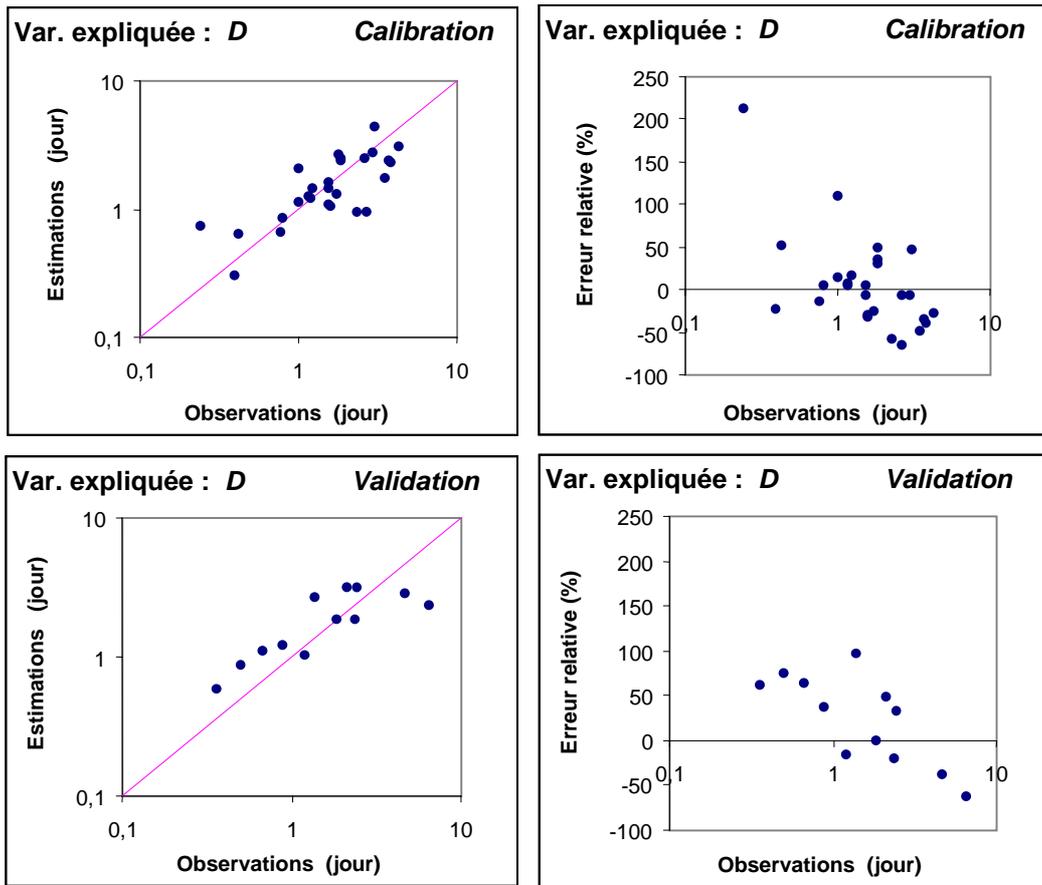


Figure 7-13 : Estimations purement régionales de la durée de crue *D* de Socose

- **Critères de performance**

Afin d’apprécier les performances de chacune des approches, nous avons calculé les erreurs suivantes :

$$rRMSE(T) = \left[\frac{1}{6} \frac{1}{N} \sum_{k=1}^6 \sum_{i=1}^N \left(\frac{V(d_k, T)_i - V_{d_k}(T)_i}{V_{d_k}(T)_i} \right)^2 \right]^{1/2} \quad (7-28)$$

$V_d(T)$ est l’estimation locale au bassin i , pour la durée d_k et la période de retour T . $V(d, T)$ est l’estimation « locale-régionale » ou « purement régionale », de l’approche « bassin de référence » ou du modèle régional convergent. Enfin N est le nombre de bassins étudié : 27 en calibration, 12 en validation.

4.3. Résultats et discussion

- **Estimations « locales-régionales »**

La Figure 7-14 présente l’erreur quadratique $rRMSE$ (7-28) obtenue pour les deux modélisations étudiées, l’approche « bassin de référence » et le modèle régional convergent,

et dans le cas d'une estimation « locale-régionale ». Les périodes de retour étudiées sont 1, 2, 5, 10 et 20 ans.

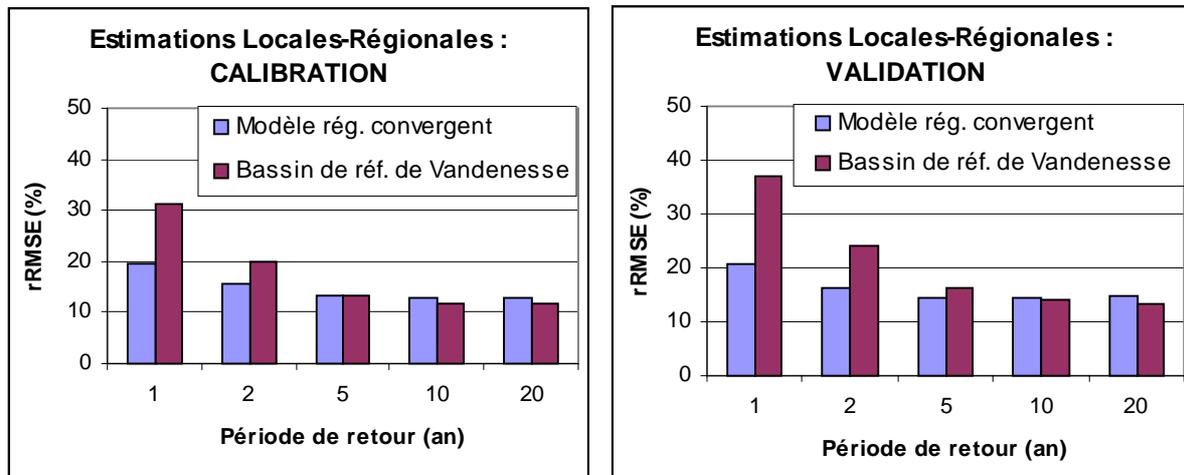


Figure 7-14 : Estimations « locales-régionales » obtenues avec le bassin de référence de *Vandenesse* et avec le modèle régional convergent

Tout d'abord, concernant les estimations « locales-régionales » (graphiques du haut, Figure 7-14), dans tous les cas l'erreur diminue lorsque la période de retour augmente. Ceci s'explique par le fait que le modèle régional convergent et l'approche « bassin de référence » utilisent tous les deux le débit instantané décénal local $V_0(10)$. Il est donc logique d'observer de meilleurs résultats autour de cette période de retour. Si l'on avait considéré des périodes de retour supérieures, il est très probable que les erreurs auraient de nouveau augmenté.

Pour ce qui est des différences observées entre l'approche « bassin de référence » et le modèle régional convergent, on constate une dérive plus forte des erreurs, lorsqu'on s'éloigne de $T=10$ ans pour l'approche « bassin de référence ». Ceci est cohérent avec le chapitre 6. Les résultats avaient été présentés différemment, en calculant un critère de performance par bassin. On observait aussi un avantage pour le modèle régional convergent.

Concernant les différences entre les échantillons de calibration et de validation, celles-ci sont insignifiantes pour l'approche « bassin de référence ». En effet, quel que soit l'échantillon, cette approche utilise ici les mêmes données : le modèle régional adimensionnel a été ajusté sur le bassin de *Vandenesse* et les caractéristiques $Q(0,10)$ et D sont locales. Pour le modèle régional convergent, on constate aussi de très faibles différences. Ceci indique que le modèle régional convergent adimensionnel calé sur l'échantillon de calibration décrit bien l'échantillon de validation. Ce résultat s'explique par la bonne homogénéité des bassins de la Moselle, déjà soulignée au chapitre 6.

- **Estimations « purement régionales »**

La Figure 7-15 présente les résultats obtenus pour les estimations « purement régionales ».

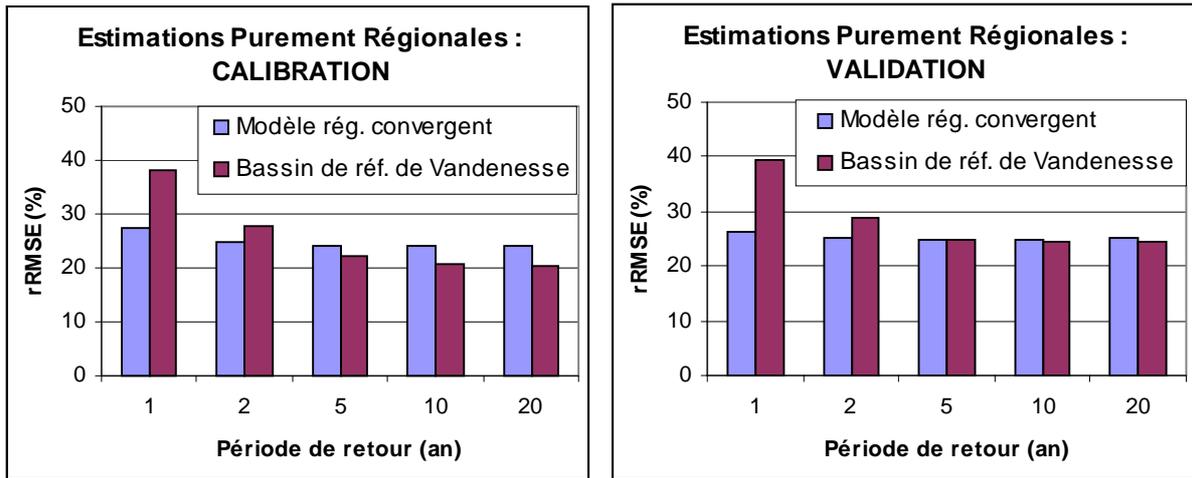


Figure 7-15 : Estimations « purement régionales » obtenues avec le bassin de référence de *Vandenesse* et avec le modèle régional convergent

Par rapport à la Figure 7-14, on constate une dégradation des résultats lorsqu'on passe aux estimations « purement régionales ». Ceci s'explique évidemment par le fait que les caractéristiques utilisées pour dénormer les modèles régionaux adimensionnels sont maintenant issues de régressions. Cependant, si on compare ces résultats à ceux obtenus au chapitre 5 pour le Québec et l'Ontario, les erreurs étaient multipliées par cinq, lorsqu'on passait aux estimations purement régionales (Figure 5-7). Sur les résultats de la Figure 7-15, l'erreur n'est que doublée. Ce réel progrès s'explique par la relative bonne qualité des régressions utilisées. D'une part le choix des variables explicatives s'avère pertinent (*cf.* paragraphe 2. et 3.), d'autre part les bassins de la Moselle présentent une assez bonne homogénéité dans leur comportement. Si nous avions utilisés les résultats de régressions calées au niveau national (par exemple les relations (7-1) et (7-2) donnant respectivement le débit instantané décennal et la durée de Socose) il est probable que nous aurions obtenu de plus mauvais résultats.

Concernant les différences entre l'approche « bassin de référence » et le modèle régional convergent, on constate ici aussi une plus grande dérive des résultats lorsqu'on s'éloigne de $T=10$ ans pour l'approche bassin de référence. En particulier l'erreur pour la période de retour de 1 an devient considérable, proche de 40%.

En validation, les résultats restent assez satisfaisants, et proches de ceux de la calibration. Les erreurs sont de l'ordre de 25% pour le modèle régional convergent, et comprises entre 25% et 40% pour l'approche bassin de référence. Ceci montre donc que les modèles régionaux testés sont robustes et donnent un relativement bon ordre de grandeur des différents quantiles $V(d,T)$. Rappelons qu'à aucun moment de l'analyse, nous n'avons utilisé les données hydrologiques de ces bassins de validation, sauf, bien sûr pour le calcul de l'erreur d'estimation ! Cependant, la faible taille de l'échantillon de validation (12 bassins) nous oblige à rester prudents sur ces résultats néanmoins encourageants.

Conclusion du chapitre 7

En traitant du problème des estimations « purement régionales » ce chapitre vient clore le travail présenté dans cette thèse.

La démarche suivie est la suivante :

- i. calcul de nombreuses caractéristiques « physiques » de bassin, relatives à leur géomorphologie et à leur occupation du sol, calculées grâce à un modèle numérique de terrain et une carte d'occupation du sol, obtenues à partir des satellites SPOT.
- ii. utilisation de la *ridge regression* pour déterminer objectivement les caractéristiques réellement explicatives
- iii. utilisation des régressions par les modèles *QdF* régionaux afin d'obtenir des estimations « purement régionales » du régime de crue.

Concernant les régressions (points i. et ii.), les résultats montrent que pour les différents quantiles de crue étudiés la variable la plus explicative est nettement la surface du bassin ; vient ensuite la pluie décennale sur 10 ou 15 jours ; puis enfin, les caractéristiques physiques telles que le pourcentage de zone agricole ou de forêt, l'indice de forme du bassin ou la pente moyenne.

Les meilleurs résultats sont obtenus pour les débits décennaux calculés sur de longues durées (6 et 15 jours, notés $Q10_6$ et $Q10_{15}$). Malgré le grand nombre de variables testées, aucune caractéristique physique n'a été retenue dans ces régressions. A l'inverse, les moins bons résultats sont observés pour les courtes durées (débit instantané $Q10_0$ et débit moyen sur 1 jour $Q10_1$), et pour les durées caractéristiques de crue Δ et D . Contrairement au cas précédent, les caractéristiques physiques viennent améliorer significativement les résultats.

Du fait de la forte inter-corrélation entre les variables explicatives sélectionnées, (presque toutes liées à l'altitude), l'interprétation du rôle physique de chacune d'elles est difficile. Notons simplement que les caractéristiques physiques sont pertinentes lorsque la variable hydrologique étudiée est liée à la fonction de transfert au sein du bassin. Ces caractéristiques physiques permettent donc d'un peu mieux tenir compte de ce processus, malheureusement, leur pouvoir explicatif reste faible.

Concernant le point iii. ci-dessus, les deux approches débit-durée-fréquence régionales déjà étudiées aux chapitres précédents ont été comparées : l'approche « bassin de référence » de Galéa et Prudhomme (1997) et le modèle régional convergent, élaboré au cours de cette thèse. On note que cette nouvelle approche donne en général de meilleurs résultats. Cependant, dans les deux cas, l'erreur d'estimation est principalement liée à celles des régressions. C'est donc dans ce domaine qu'il faut à l'avenir porter les efforts de recherches.

Conclusion générale et perspectives



Illustration 4 : Crue du 3 novembre 1859, Jardin de Ville, Grenoble
Source : Pilot J.-J.-A., Grenoble inondé, Maisonville & Jourdain, 1859. Dessin de D. Raoult

Conclusion générale

L'objectif principal de ce travail a été de développer un modèle fréquentiel capable de rendre compte du régime des crues d'un bassin. Comme nous l'avons souligné en introduction, ce type d'information est par exemple utile dans le cadre de la prévention des risques liés aux inondations (connaissance de l'aléa).

La première partie de la thèse a été consacrée à l'analyse **locale** du régime, par opposition à l'analyse **régionale**, traitée dans la seconde partie.

Analyse locale du régime des crues

Le régime des crues est décrit par une analyse en débit, durée et fréquence. Nous avons cherché à modéliser les courbes débit-durée-fréquence obtenues, à l'aide d'un minimum de paramètres. Des propriétés d'invariance d'échelle sur les différentes distributions observées permettent la formulation simple et facilement ajustable d'un modèle que nous avons appelé « **modèle *QdF* local convergent** ». Une tentative d'interprétation, assez simpliste, a été faite. En supposant que les hydrogrammes de crue se déduisent les uns des autres par affinité, la propriété observée sur les distributions est expliquée.

Parmi les paramètres du modèle proposé se trouve la durée Δ . Cette durée est ajustée au même titre que les autres paramètres. Nous avons montré qu'elle avait un sens physique bien précis : elle informe sur la plus ou moins grande durée de l'ensemble des crues observées et est ajustée de façon automatique sur l'ensemble de la chronique étudiée. Nous avons ainsi défini un paramètre décrivant la dynamique des crues, et calculable de façon entièrement objective, contrairement à certaines durées caractéristiques difficiles à évaluer.

Le modèle local convergent a été testé sur une centaine de bassins. Ses bons résultats nous ont permis d'envisager dans la seconde partie de cette thèse son application dans le cadre d'une analyse régionale.

Analyse régionale du régime des crues

Un modèle *QdF* régional a été développé en généralisant la méthode de l'indice de crue, classiquement utilisée en régionalisation. Ce modèle, appelé « **modèle *QdF* régional convergent** », définit pour une région donnée des courbes *QdF* adimensionnelles, qu'il faut dénormer à l'aide de deux caractéristiques hydrologiques du bassin étudié : un débit de crue μ et la durée Δ , évoquée ci-dessus.

Afin de tester ce modèle, différents niveaux d'analyse ont été distingués. Pour cela, nous avons adopté le vocabulaire de Cunnane (1988), parlant d'estimations :

- « locales-régionales » lorsqu'une distribution régionale est combinée à des paramètres du régime des crues estimés localement ;
- « purement régionales » lorsqu'elles ne font appel qu'aux données régionales. Ces estimations peuvent donc être faites sur des bassins où l'on ne dispose d'aucune information sur les débits, et sont déterminées à l'aide de régressions faisant appel aux caractéristiques physiques et climatiques des bassins.

Le modèle régional convergent donne des résultats plutôt meilleurs que l'approche utilisée jusqu'à présent : l'approche « bassin de référence » (Galéa et Prudhomme, 1997). Ceci s'explique par le fait que le modèle proposé utilise la totalité de l'information régionale, alors que l'approche antérieure n'utilise que l'information relative à trois bassins de référence. Pour ce qui concerne les estimations « purement régionales », quelle que soit l'approche considérée, l'erreur observée est en grande partie due à la faible qualité des régressions, malgré tout le soin que nous y avons apporté (test d'un grand nombre de variables explicatives et utilisation de la *ridge regression*).

Le modèle convergent (local et régional) s'est inspiré de l'approche « bassin de référence », et a profité de l'expérience acquise ces dernières années au Cemagref en matière d'analyse débit-durée-fréquence.

- Au niveau de l'analyse locale, il simplifie l'ancienne formulation ;
- Au niveau de l'analyse régionale, il intègre les concepts classiques de l'hydrologie fréquentielle régionale, en généralisant la méthode de l'indice de crue.

L'approche proposée doit donc être vue comme une évolution significative de l'approche « bassin de référence », vers des concepts plus classiques de l'analyse régionale.

Perspectives

Les résultats du modèle *QdF* (local et régional) convergent autorisent des perspectives intéressantes.

Tout d'abord, d'un point de vue purement conceptuel, l'interprétation de la propriété de convergence que nous avons faite (affinité des hydrogrammes de crue) doit être approfondie. Cette explication est schématique, cependant il est clair que les propriétés observées doivent trouver leur explication dans la forme des crues observées.

D'un point de vue plus opérationnel les efforts pour rendre ce travail directement utilisable doivent porter dans différentes directions :

- Le problème des extrapolations aux fréquences rares doit être traité. Il faudra faire un choix, quant à la méthode d'extrapolation, et l'intégrer dans le modèle.
- La définition des régions homogènes n'a pas été abordée, nous nous sommes contentés soit de reprendre les régions déjà existantes du Québec et de l'Ontario, soit d'étudier une seule région géographique : les sous-bassins de la Moselle. Il serait utile d'intégrer les paramètres de la modélisation débit-durée-fréquence dans l'établissement de ces régions, notamment pour des régions de type voisinage.
- Le choix de la norme de débit n'a pas été testé lui non plus. Pour le Québec-Ontario, nous avons utilisé un quantile instantané de période de retour d'environ 2 ans, pour les données françaises, un quantile de période de retour 10 ans. Il faudra à l'avenir étudier l'influence de ce choix.
- L'estimation des caractéristiques hydrologiques doit être améliorée. Les méthodes de régressions sont très utiles, mais montrent cependant leurs limites. Les travaux de Sauquet (2000), permettant une interpolation des caractéristiques hydrologiques, basée sur la structure du réseau ouvrent de nouvelles perspectives qu'il faudrait approfondir.
- Enfin, la diffusion de cette nouvelle méthode passe nécessairement par l'élaboration d'outils. Le meilleur exemple à ce jour est fourni par le Flood estimation Handbook (Institute of Hydrology, 1999). Cet outil destiné aux bureaux d'études et services opérationnels regroupe dans un Système d'Information Géographique les données et les méthodes nécessaires pour effectuer une analyse régionale de façon simple et conviviale.

Le développement de chacun des ces points permettrait alors d'obtenir un outil réellement utile dans le cadre de la prévention contre les inondations.

Références bibliographiques

- Adamson, P. T., Metcalfe, A. V. et Parme, t., B. (1999). "Bivariate extreme value distribution: an application of the Gibbs sampler to the analysis of floods." Water Resources Bulletin, **35(9)**: 2825-2835.
- Ambroise, B. (1998). "Genèse des débits dans les petits bassins versants ruraux en milieu tempéré : 1- Processus et facteurs." Revue des Sciences de l'Eau, **11/4**: 471-495.
- Ambroise, B. (1999). "Genèse des débits dans les petits bassins versants ruraux en milieu tempéré : 2- Modélisation systémique et dynamique." Revue des Sciences de l'Eau, **12/1**: 123-153.
- Arnaud, P. et Lavabre, J. (1999). "Using a stochastic model for generating hourly hyetographs to study extreme rainfalls." Hydrological Sciences Journal, **44**: 433-446.
- Arnell, N. W., Krasovskaïa, I. et Gottshalk, L. (1993). "River flow regimes in Europe." Flow Regimes From International Experimental an Network Data (FRIEND), Vol 1 : Hydrological Studies (Ed. by A. Gustard), Institut of Hydrology, Wallingford, Oxfordshire, UK: 112-121.
- Ashkar, F. (1980). Partial duration series models for flood analysis, Ecole polytechnique de Montréal, Montréal, Canada.
- Balocki, J. B. et Burges, S. J. (1994). "Relationships between n-day flood volumes for infrequent large floods." Journal of Water Resource Planning and Management, **120(6)**: 794-818.
- Bargaoui, Z. K., Fortin, V., Bobée, B. et Duckstein, L. (1998). "Une approche floue pour la détermination de la région d'influence de la station hydrométrique." Revue des Sciences de l'Eau, **11**: 255-282.
- Beckinsale, R. P. (1969). "River regimes." Water, Earth, and Man. M. C. Ltd. Londres. **1**: 455-471.
- Bendjoudi, H., Hubert, P., Schertzer, D. et Lovejoy, S. (1997). "Interprétation multifractale des courbes intensité-durée-fréquence des précipitations." Comptes Rendus de l'Académie des Sciences, **325**(série IIA): 323-326.
- Benzecri, J. P. (1973). "L'analyse des données, T.2 : L'analyse des correspondances." Dunod, Paris: 619p.
- Bernier, J. (1967). Sur la théorie du modèle de renouvellement et son application en hydrologie, Rapport EDF, HE 40 81-09: 22p.
- Black, A. R. et Werritty, A. (1997). "Seasonality of flooding: a case study of North Britain." Journal of Hydrology, **195**: 1-25.
- Blazkova, S. et Beven, K. (1997). "Flood frequency prediction for data limited catchment in the Czech Republic using a stochastic rainfall model and TOPMODEL." Journal of Hydrology, **195**: 256-278.
- Bobée, B. (1999). "Estimation des événements extrêmes de crue par analyse fréquentielle : une revue critique." La Houille Blanche, **7/8**: 100-105.
- Bobée, B. et Abi-Zeid, I. (1999). "La modélisation stochastique des étiages : une revue bibliographique." Revue des Sciences de l'Eau, **12(3)**: 459-484.

- Bobée, B., Cavadias, G. S., Ashkar, F., Bernier, J. et Rasmussen, P. (1993). "Towards a systematic approach to comparing distribution used in flood frequency analysis." Journal of Hydrology, **142**: 121-136.
- Bois, P. (1976). Contribution à l'analyse et à la prévision de variables hydrométéorologiques. Applications à la prévision des débits du Niger et des avalanches à Davos., Thèse de Doctorat, Institut national polytechnique de Grenoble: 218p.
- Bois, P. (1986). "Contrôle des séries chronologiques corrélées par étude du cumul des résidus de la corrélation." 2e journées hydrologiques de l'Orstom, 16-17 sept., 89-100
- Bois, P., Mailloux, H., Obled, C. et De Saintignon, F. (1995). Atlas expérimental des risques de pluie intense dans la région Cévenne Vivarais, Pôle Grenoblois des Risques Naturels.
- Borgman, L. E. (1963). "Risk criteria." Proceedings of A.S.C.E., août 1963, 1-35
- Burn, D. H. (1990). "An appraisal of the "region of influence" approach to flood frequency analysis." Hydrological Sciences Journal, **35(2)**: 149-165.
- Burn, D. H. (1990). "Evaluation of regional flood frequency analysis with a region of influence approach." Water Resources Research, **26(10)**: 2257-2256.
- Burn, D. H. (1997). "Catchment similarity for regional flood frequency analysis using seasonality measures." Journal of Hydrology, **202**: 212-230.
- Calver, A., Lamb, R. et Crewett, J. (2000). "Generalised Flood frequency estimation using continuous simulation." European Conference on Advances in Flood Research, Postdam, Postdam Institute for Climate Impact Research, 1-3 Novembre 2000, 412-421
- Cameron, D. S., Beven, K., Tawn, J., Blazkova, S. et Naden, P. (1999). "Flood frequency estimation by continuous simulation for gauged upland catchment (with uncertainty)." Journal of Hydrology, **219**: 169-187.
- Cavadias, G. S. (1990). "The canonical correlation approach to regional flood estimation." Regionalisation in Hydrology. Proc. of the Ljubljana Symposium, IAHS, Publ. No 191: 171-177.
- Champelly, S. (1999). "Le modèle linéaire généralisé.", Cours d'analyse de données, 28/06/99, Université Lyon 1.
- Chaput, N. (1999). Modèle débit-durée-fréquence des régimes d'étiages, concept et usage opérationnel, Mémoire de fin d'étude, ENSA de Renne, Cemagref-Lyon: 78p+Annexes.
- Charleux, J. (2001). Qualité des MNT pour l'hydrologie (thèse en cours), Thèse de Doctorat, Université de Marne-la-Vallée.
- CTGREF, S., DIAME, SH (1980). "Synthèse nationale sur les crues des petits bassins, fascicule 2 : la méthode SOCOSE ; fascicule 3 : la méthode CRUPEDIX.",: 39p.
- 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. (1979). "A note on the Poisson assumption in partial duration series models." Water Resources Research, **15 (2)**: 489-494.
- Cunnane, C. (1988). "Method and merits of regional flood frequency analysis." Journal of Hydrology, **100**: 269-290.

- Dalrymple, T. (1960). "Flood frequency analysis." U.S. Geol. Surv. Water Supply Pap., **1543A**.
- Dauge, Y. (1999). Cadre de débat et de concertation dans le domaine des inondations, Rapport de mission pour le Premier Ministre sur la prévention du risque d'inondation, Assemblée Nationale: 55p.
- de Marsily, G. (1994). "Quelques réflexions sur l'utilisation des modèles en hydrologie." Revue des Sciences de l'Eau, **7/3**: 219-234.
- Duband, D. (1975). Ridge Regression, Note interne EDF-DTG: 14p.
- Duband, D. (1992). "Cours d'Hydrologie Approfondie", Ecole Nationale Supérieure d'Hydraulique et de Mécanique de Grenoble.
- Dupuis, L. et Rasmussen, P. (1993). Evaluation de méthodes "indice de crue" pour l'estimation d'une distribution régionale, Rapport interne N°I-125, INRS-Eau: 26p+Annexes.
- Dussaux, V. (2000). Modèles Débit-Durée-Fréquence des régimes d'étiages, modèles de conceptualisation adaptée aux distributions fréquentielles les plus utilisées, Rapport de Stage ENTP, Cemagref-Lyon: 48p+annexes.
- Fortin, V., Bobée, B., Duckstein, L. et Bargaoui, Z. K. (1995). "Détermination floue des zones hydrologiques homogènes." Modeling and Management of Sustainable Basin-scale Water Resource Systeme, **IASH publ. 231**: 367-375.
- Galéa, G. et Javelle, P. (2000a). Modèles débit-durée-fréquence de crue en Guadeloupe, Rapport d'étude, protocole Cemagref-Lyon, DIREN Guadeloupe et Météo-France, Cemagref-Lyon: 25p+Annexes.
- Galéa, G., Javelle, P. et Chaput, N. (2000b). "Un modèle débit-durée-fréquence pour caractériser le régime d'étiage d'un bassin versant." Revue des Sciences de l'Eau, **13/4** : 411-430.
- Galéa, G. et 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.
- Gilard, O. (1998). "Les bases techniques de la méthode Inondabilité.", Cemagref Editions. 158
- Gingras, D., Adamowski, K. et Pilon, P. J. (1994). "Regional flood equation for the province of Ontario and Quebec." Water Resources Bulletin, **30(1)**: 55-67.
- Greenwood, J. A., Landwehr, J. M. et Matalas, N. C. (1979). "Probability weighted moments : Definition and relation to parameter of several distributions expressible in inverse form." Water Resources Research, **15(5)**: 1049-1054.
- GREHYS, G. d. r. e. h. s. (1996). "Inter-comparaison of regional flood frequency procedures for Canadian rivers." Journal of Hydrology, **186**: 85-103.
- GREHYS, G. d. r. e. h. s. (1996). "Presentation and review of some methods for regional flood frequency analysis." Journal of Hydrology, **186**: 63-84.
- Grelot, F. (1999). Prévention des inondations : Eléments de réflexion pour une intégration économique et opérationnelle de la perception du risque, Rapport de DEA, ENGREF-ENS Cachan: 59p.

- Grésillon, J. M. (1999). "Dimensions temporelles des régimes." Note de travail dans le cadre du projet "régime" du Programme National de Recherche en Hydrologie (n°97-PNRH-23): 11p.
- Guillot, P. et Duband, D. (1967). "La méthode du GRADEX pour le calcul de la probabilité des crues à partir des pluies." Journée de la SHF: question 1, rapport 7, Paris, Septembre 1967.
- Hoerl, A. E. (1962). "Application of ridge analysis to regression problems." Chemical Engineering Progress, **58**: 54-59.
- Hoerl, A. E. et Kennard, R. W. (1970). "Ridge Regression: Biased Estimation for Nonorthogonal Problems." Technometrics, **12(1)**.
- Hosking, J. R., Wallis, J. R. et Wood, E. F. (1985). "Estimation of the generalized extreme-value distribution by the method of probability-weighted moment." Technometrics, **27(3)**: 251-261.
- Hosking, J. R. M. et Wallis, J. R. (1993). "Some statistics useful in regional frequency analysis." Water Resources Research, **24(4)**: 588-600.
- Hubert, P., Javelle, P., Bendjoudi, H., Lang, M., Schertzer, D. et Tchiguirinskaïa, I. (2000). "Deux modélisations invariantes d'échelle des courbes QdF." Colloque PNRH 2000, Toulouse, 16-17 mai 2000,
- Hydrology, I. o. (1999). "Flood Estimation Handbook (5 volumes).", Institute of Hydrology, Wallingford.
- Jambu, M. (1976). "Sur l'interprétation mutuelle d'une classification hiérarchique et d'une analyse des correspondances." Revue des Statistiques Appl., **24(2)**: 45-73.
- Jambu, M. (1978). "Classification automatique pour l'analyse des données, 1. Méthodes et algorithmes.", Dunod, Paris, 310p.
- Javelle, P., Galéa, G. et Grésillon, J. M. (1998). "A new formulation of flow-duration-frequency models." FRIEND-AMHY annual meeting, Istanbul October, IHP series, Unesco Paris.
- Javelle, P., Galéa, G. et Grésillon, J. M. (2000). "L'approche débit-durée-fréquence : historique et avancées." Revue des Sciences de l'Eau, **13/3** : 303-321.
- Javelle, P., Grésillon, J. M. et Galéa, G. (1999). "Modélisation des courbes débit-durée-fréquence en crue et invariance d'échelle." Comptes Rendus de l'Académie des Sciences, Sciences de la terre et des planètes, **329**: 39-44.
- Kieffer-Weisse, A. (1998). Etude des précipitations de pas de temps court en relief accidenté (Alpes Françaises) - Méthode de cartographie des précipitations extrêmes., Thèse de l'Institut Polytechnique de Grenoble, Grenoble: 314p.
- Krasovskaia, I. (1997). "Entropy-based grouping of flow regimes." Journal of Hydrology, **202**: 173-191.
- Krasovskaia, I. et Gottschalk, L. (1992). "Stability of flow regimes." Nordic Hydrology, **23**: 137-154.
- Lamb, R. (1999). "Calibration of a conceptual rainfall-runoff model for flood frequency estimation by continuous simulation." Water Resources Research, **35(10)**: 3103-3114.
- Lang, M. (1995). Les chroniques en hydrologie : Modélisation comparée par un système de gestion de base de données relationnel et orienté objet ; Traitements de bases et

- intervalles de confiance des quantiles de crues ; Techniques d'échantillonnage par la méthode du renouvellement, Université Joseph Fourier: 296p.
- Lang, M. (1999). "Theoretical discussion and Monte-Carlo simulations for a Negative Binomial process paradox." Stochastic environmental research and risk assessment, **13**: 183-200.
- Lang, M. (2000). Analyse probabiliste des crues pour la prévention du risque d'inondation. Cemagref-Lyon, Unité de Recherche en Hydrologie-Hydraulique, Rapport d'habilitation à diriger des recherches. Institut National Polytechnique de Grenoble: 57p.
- Lang, M., Ouarda, T. B. M. J. et Bobée, B. (1999). "Towards operational guidelines for over-threshold modeling." Journal of Hydrology, **225**: 103-117.
- Lang, M., Rasmussen, P., Oberlin, G. et Bobée, B. (1997). "Echantillonnage par valeurs supérieures à un seuil : modélisation des occurrences par la méthode de renouvellement / Over-threshold sampling : Modelling of occurrences by renewal processes (en français et en anglais)." Revue des Sciences de l'Eau, **10/3**: 279-299.
- Lecce, S. A. (2000). "Spatial variations in the timing of annual floods in the southeastern United States." Journal of Hydrology, **235**: 151-169.
- Lefrou, C., Martin, X., Labarthe, J.-P., Varret, J., Mazière, B., Tordjeman, R. et Feunteun, R. (2000). Les crues des 12,13 et 14 novembre 1999 dans les départements de l'Aude, de l'Hérault, des Pyrénées-Orientales et du Tarn, Conseil Général des Ponts et Chaussées - Inspection Générale de l'Environnement: 140p.
- Magilligan, F. J. et Graber, B. E. (1996). "Hydroclimatological and geomorphic controls on the timing and spatial variability of floods in New England, USA." Journal of Hydrology, **178**: 159-180.
- Mahalanobis, P. C. (1936). "On the generalized distance in statistics." Proc. Natl. Inst. Sci. India, **12**: 49-55.
- Maione, U., Mignosa, P. et Tomirotti, M. (2000). "Sul calcolo del volume degli invasi di laminazione." L'Aqua, **5**: 19-28.
- Mardia, K. V. (1972). "Statistics of directional data.", Academic, New York.
- Margoum, M., Oberlin, G., Lang, M. et Weingartner, R. (1994). "Estimation des crues rares et extrêmes : principes du modèle Agréee." Hydrologie Continentale, **9(1)**: 85-100.
- Meunier, M. (1999). Les torrents du Nord-Ouest de la Martinique - Hydrologie des crues, érosion, hydraulique et dynamique torrentielle, Cemagref Editions: 403p.
- Meunier, M. (2001). "Regional flow-duration-frequency model for tropical island of Martinique." Journal of Hydrology, **accepté**.
- Michel, C. (1982). "Extrapolation par la méthode du GRADEX." Note interne KG 03.05.82, Cemagref-Antony, Div. Hydrologie.
- Moallemi, A. (1988). Mise au point d'une méthode d'estimation de la crue décennale sur un petit bassin non jaugé, Thèse de Doctorat, Université Paris-Sud Orsay: 424p.
- NERC (1975). "Estimation of flood volumes over different durations." Flood Studie Report, Vol. 1, chap. 5: 243-264.

- NOAH (2000). New Apportunities for Altimetry in hydrology - Final Report, European Commission DG XII, Environment and Climat, Theme 3, Space and Techniques, CEO, ref NOAH ENV4-CT96-0371: 318p.
- Oberlin, G. (1992). "Normalisation des variables dans les modèles hydrologiques descriptifs." Informations Techniques du Cemagref, n°85, note 4: 8p.
- Oberlin, G. et Hubert, P. (1999). "Refondation du concept de régime hydrologique." XXIIe Assemblée Générale de l'Union Géodésique et Géophysique Internationale, Birmingham, 18-30 juillet 1999, 269-277
- Ouarda, T. B. M. J., Boucher, G., Rasmussen, P. F. et Bobée, B. (1997). "Regionalization of floods by canonical analysis. Operational Water Management." E. J.C. Refsgaard and E.A. Karalis, Balkema. Rotterdam, The Netherlands: 297-302.
- Ouarda, T. B. M. J., Girard, C., Cavadias, G. S. et Bobée, B. (2000). "Regional flood frequency estimation with canonical correlation analysis." Journal of Hydrology, soumis.
- Ouarda, T. B. M. J., Haché, M., Bruneau, P. et Bobée, B. (2000). "Regional flood peak and volume estimation in a northern Canadian Basin." Journal of Cold Regions Engineering, American Society of Civil Engineering (ASCE), 14(4): 176-191.
- Ouarda, T. B. M. J., Lang, M., Bobée, B., Bernier, J. et Bois, P. (1999). "Synthèse de modèles régionaux d'estimation de crue utilisée en France et au Québec." Revue des Sciences de l'Eau, 12(1): 155-182.
- Pandey, G. R. et Nguyen, V. T. V. (1999). "A comparative study of regression based methods in regional flood frequency analysis." Journal of Hydrology, 225: 92-101.
- Pardé, M. (1955). "Fleuves et rivières." Paris, Armand Colin. 224p
- Prudhomme, C. (1995). Modèles synthétique des connaissances en hydrologie, Application à la régionalisation des crues en Europe Alpine et Méditerranéenne, Thèse de Doctorat, Université Montpellier II: 397p.
- Rasmussen, P., Bobée, B. et Bernier, J. (1994). "Comparaison de modèles d'estimation régionale de crue : méthodologie générale." Revue des Sciences de l'Eau, 7(1): 23-41.
- Rasmussen, P., Mathier, L., Perron, H., Ouarda, T. B. M. J., Ashkar, F., Bobée, B. et Bernier, J. (1995). Résultats d'une intercomparaison de modèles d'estimation régionale de crue - Rapport du projet de recherche CRSNG Stratégique (STR0118482) : Méthodologie d'estimation régionale des débits de crue : application à la région Québec-Ontario, INRS-Eau: 30p + annexes.
- Ribeiro-Corréa, J., Cavadias, G. S., Clément, B. et Rouselle, J. (1995). "Identification of hydrological neighborhoods using canonical correlation analysis." Journal of Hydrology, 173: 71-89.
- Ribot-Bruno, J. et Breil, P. (1998). Suivi hydrométrique de la SEMINE, Cemagref.
- Roche, M. (1963). "Hydrologie de surface." Gauthier-Villars, Paris. 430p
- Rosbjerg, D. (1985). "Estimation in partial duration with independent and dependent peak values." Journal of Hydrology, 76: 183-195.
- Sauquet, E. (2000). Une cartographie des écoulements annuels et mensuels d'un grand bassin versant structurée par la topologie du réseau hydrographique, Thèse de Doctorat, Institut National Polytechnique de Grenoble: 334p+Annexes.

- Sauquet, E., Gottschalk, L. et Leblois, E. (2000a). "Mapping average annual runoff: A hierarchical approach applying a stochastic interpolation scheme." Hydrological Sciences Journal, **Accepté**.
- Sauquet, E., Krasovskaïa, I. et Leblois, E. (2000b). "Mapping mean monthly runoff pattern using EOF analysis." Hydrology and Earth System Sciences, **4(1)**: 79-93.
- Schober, A. (1996). Analyse de la variabilité des paramètres caractéristiques de l'hydrologie d'un bassin versant et modélisation des crues en présence de données hydrologiques succinctes, Thèse de Doctorat. Université Joseph Fourier, Grenoble: 300p.
- Shane, R. M. et Lynn, W. R. (1964). "Mathematical model for flood risk evaluation." Journal of the hydraulics division. Proc. of the Amer. soc. of Civ. eng., **HY6**: 1-20.
- Sherman, L. K. (1932). "Streamflow from rainfall by unit-graph method." Engineering News Record, **108**: 505-510.
- Sherwood, J. M. (1994). "Estimation of volume-duration-frequency relations of ungauged small urban streams in Ohio." Water Resources Bulletin, **30(2)**: 261-269.
- Thomas, D. M. et Benson, M. A. (1970). "Generalization of streamflow characteristics from drainage-basin characteristics." USGS Water Supply paper 1975.
- Tollan (1975). "Hydrologiske regioner i Norden." Vannet i Norden, N°1.
- Weingartner, R. et Aschwanden, H. (1992). Régimes d'écoulement comme base pour l'estimation des valeurs moyennes des débits, in Atlas hydrologique de la Suisse, planche 5.2. Berne, Suisse, Service hydrologique et géologique national.
- Wiltshire, S. E. (1986). "Identification of homogeneous regions for flood frequency analysis." Journal of Hydrology, **84**: 287-302.
- Yue, S., Ouarda, T. B. M. J., Bobée, B., Legendre, P. et Bruneau, P. (1999). "The Gumbel mixed model for flood frequency analysis." Journal of Hydrology, **226**: 88-100.

ANNEXES

ANNEXE 1

RAPPELS SUR LES LOIS DE DISTRIBUTIONS ET COMPARAISON ENTRE VALEURS SUPERIEURES A UN SEUIL ET MAXIMALES ANNUELLES

L'ajustement d'une distribution théorique sur un échantillon de valeurs maximales (supérieures à un seuil ou maximales annuelles) consiste à considérer l'échantillon comme un sous ensemble d'une population infinie. Le but recherché est d'estimer le plus précisément possible les caractéristiques de cette population. Selon les notations adoptées, nous noterons les fonctions relatives aux valeurs supérieures à un seuil avec un indice s , et celles relatives aux maxima-annuelles avec un indice x . Lorsqu'il n'y a pas d'indice, la formulation peut s'appliquer indifféremment aux deux cas.

Définitions

Sur cette population, une **fonction de répartition** $F(x)$ est définie. Cette fonction exprime la probabilité de tirer au hasard une valeur X inférieure ou égale à x :

$$F(x) = Prob(X \leq x) \quad (\text{A-1})$$

La fonction de répartition $F(x)$ de la population peut être estimée par la **fonction de répartition expérimentale** $F_{\text{exp}}(x)$ (ou encore fréquence expérimentale), calculée à partir de l'échantillon. Si les N valeurs x_i sont classées par ordre croissant, la fonction de répartition empirique est donnée par :

$$F_{\text{exp}}(x_i) = \frac{i - a}{N + b} \quad (\text{A-2})$$

où a et b dépendent des lois de probabilité ajustées. De très nombreuses valeurs sont conseillées dans la littérature. En général, a est compris entre 0 et 0.5, et b entre 0 et 1.

La période de retour $T(x)$ représente la durée moyenne entre deux événements successifs, supérieures ou égaux à la valeur x . **La notion « d'événements successifs » n'est pas la même selon que l'on considère des valeurs maximales annuelles, ou des valeurs supérieures à un seuil.** Aussi, il faut bien faire la différence entre les périodes de retour $T_x(x)$, relative aux valeurs maximales annuelles, et $T_s(x)$, relatives aux valeurs supérieures à un seuil :

$T_s(x)$ est la durée moyenne entre des crues dépassant le débit x .

$T_x(x)$ est la durée moyenne entre les années contenant une ou plusieurs crues dépassant le débit x .

Dans le cas des valeurs maximales annuelles, la période de retour $T_x(x)$ est reliée à la fonction de répartition $F_x(x)$ par la relation suivante :

$$F_x(x) = 1 - 1/T_x(x) \quad (\text{A-3})$$

Dans le cas des valeurs supérieures à un seuil, la relation entre la période de retour $T_s(x)$ et la fonction de répartition $G_s(x)$ dépend du processus d'occurrence des événements. Si le processus d'occurrence est décrit par une loi de Poisson, on a alors la relation suivante (Rosberg, 1985). Lang (1999) donne d'autres relations, lorsque le processus d'occurrence est décrit par d'autres lois (par exemple, la loi binomiale négative).

$$G_s(x) = 1 - 1/(\mu T_s(x)) \quad (\text{A-4})$$

avec μ le nombre moyen d'événements par an.

De nombreuses lois sont proposées dans la littérature (Bobée et Ashkar, 1993). Le Tableau A-1 donne des exemples de lois souvent utilisées dans la littérature, pour la description des valeurs maximales annuelles ou supérieures à un seuil. Il en existent bien d'autres, comme par exemple la loi Log Pearson III officiellement conseillée aux Etats Unis, ou bien la loi Logistique Généralisée, récemment recommandée en Angleterre par le Flood Estimation Handbook (1999).

Equivalence entre les valeurs supérieures à un seuil et maximales annuelles

Lorsque le processus des valeurs supérieures à un seuil suit une loi de Poisson il existe relation entre les distributions des valeurs maximales annuelles et des valeurs supérieures à un seuil. Cette relation est donnée par :

$$F_x(x) = \exp\{-\mu[1 - G_s(x)]\} \quad (\text{A5})$$

ce qui, en terme de période de retour, donne :

$$1/T_x = 1 - \exp(-1/T_s) \quad (\text{A-6})$$

On trouvera dans Lang (1997) ces relations développées pour d'autres processus.

Pour les lois présentées dans le Tableau A-1 on peut alors montrer qu'une distribution de valeurs supérieures à un seuil décrite par une loi exponentielle est équivalente à une distribution de valeurs maximales annuelles décrite par une loi de Gumbel. De la même façon, il y a équivalence entre la loi de Pareto généralisée et la loi *GEV*.

Type d'Echantillonnage	Type de loi	Fonction de répartition	Quantile
Maxi Annuel	Gumbel	$F_x(x) = \exp\left[-\exp\left(-\frac{x-x_0}{a}\right)\right]$	$x(T_x) = x_0 - a \ln(-\ln(1-1/T_x))$
	Generalised Extreme Values (GEV)	$F_x(x) = \exp\left[-\left(1-k\frac{x-x_0}{a}\right)^{1/k}\right]$ ($k \neq 0$)	$x(T_x) = x_0 + \frac{a}{k} \left[1 - (-\ln(1-1/T_x))^k\right]$
Sup-seuil	Exponentielle	$G_s(x) = 1 - \exp\left[-\frac{x-x_0}{a}\right]$	$x(T_s) = (x_0 + a \ln \mu) + a \ln T_s$
	Pareto généralisé	$G_s(x) = 1 - \left(1-k\frac{x-x_0}{a}\right)^{1/k}$ ($k \neq 0$)	$x(T_s) = x_0 + \frac{a}{k} \left[1 - \frac{1}{(\mu T_s)^k}\right]$ avec μ : nb moyen de crues par an

Tableau A-1: Lois à 2 et 3 paramètres utilisées dans l'étude des valeurs maximales annuelles et des valeurs supérieures à un seuil

Les paramètres x_0 , a et k sont souvent appelés paramètres de position, d'échelle et de forme.

Les lois de Gumbel (2 paramètres) et GEV (3 paramètres) sont utilisées pour décrire des valeurs maximales annuelles. De même, la loi exponentielle et la loi de Pareto Généralisée sont souvent utilisées pour décrire des valeurs supérieures à un seuil, pour lesquelles le processus d'occurrence est supposé suivre une loi de Poisson.

Dans la pratique, on a l'habitude de représenter les quantiles $x(T)$ issus de ces lois sur une échelle appropriée, en fonction de la période de retour. Pour la loi de Gumbel, on travaillera en fonction de la variable réduite $u = -\ln(-\ln(1-1/T_x))$ de façon à ce que les quantiles théoriques soient représentés par une droite. Pour la loi exponentielle, on tracera les quantiles en fonction de $\ln(T_s)$, pour la même raison.

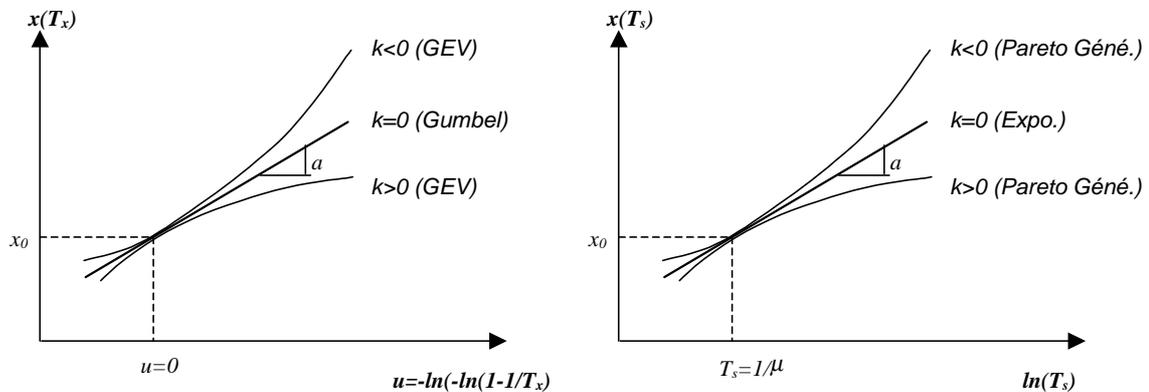


Figure A-1 : Représentation des quantiles $x(T)$ en fonction de leur période de retour

La loi de Gumbel est un cas limite de la loi *GEV*, lorsque k tend vers 0. De même, la loi exponentielle est un cas limite de la loi de Pareto généralisée lorsque k tend vers 0. En France, pour la loi de Gumbel et la loi exponentielle, le paramètre d'échelle a est souvent appelé gradex (gradient des valeurs exponentielles).

A titre d'exemple, la Figure A-2 présente pour l'exemple de la Nied Allemande traité au chapitre 1 une comparaison entre l'échantillon des valeurs supérieures à un seuil obtenu avec les paramètres suivants : ($\mu=2$, $COEF1=1$ et $COEF2=10j$), et celui des valeurs maximales annuelles. Pour les valeurs maximales annuelles, le début de l'année hydrologique est fixé en juillet. Les deux échantillons sont représentés en fonction de la période de retour empirique T_s relative aux valeurs supérieures à un seuil. Pour se faire, les périodes de retour empiriques T_x de l'échantillon des valeurs maximales annuelles ont été converties en T_s d'après la relation (A-6). Pour l'exemple présenté, on constate une bonne cohérence entre ces deux modes d'échantillonnage.

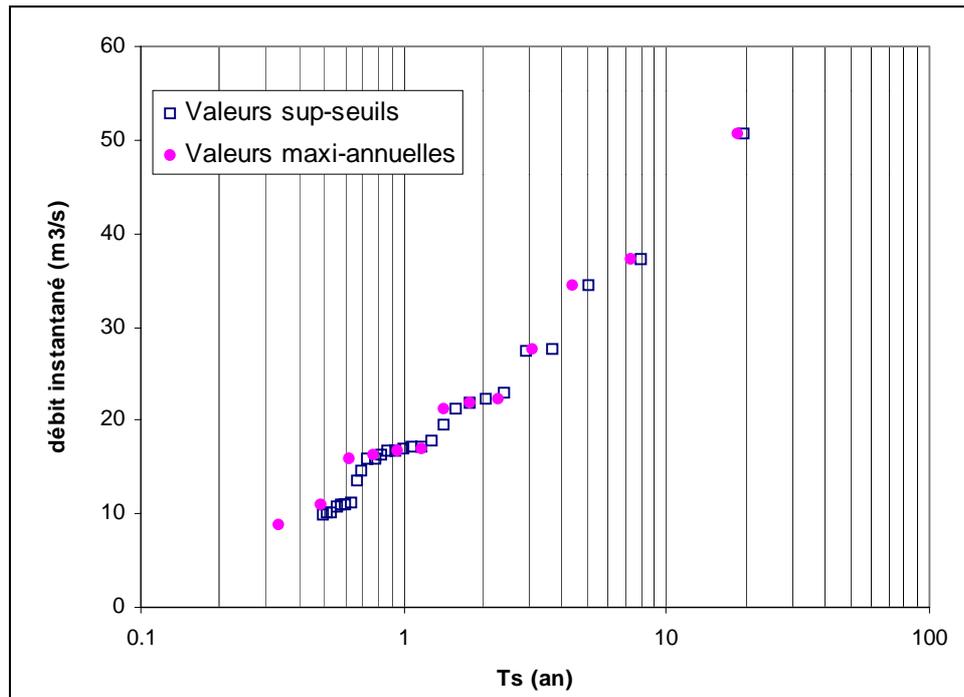


Figure A-2 : Comparaison d'un échantillon obtenu par valeurs supérieures à un seuil et par valeurs maximales annuelles. Exemple de la Nied Allemande à Faulquemont ($S=187 \text{ km}^2$)

Ajustement des paramètres

Considérons une loi de probabilité dont la densité $f(x)$ est définie par p paramètres a_1, \dots, a_p . Le but est de déterminer les paramètres a_1, \dots, a_p de façon à ce que les caractéristiques statistiques de la population décrite par $f(x)$ se rapprochent le plus possible des caractéristiques statistiques observées sur l'échantillon. Deux principales méthodes existent pour ajuster une loi de probabilité sur un échantillon : la méthode des moments et celle du maximum de vraisemblance. A cela, il faut rajouter une variante de la méthode des moments : la méthode des moments pondérés.

Méthode des moments

On écrit que les p premiers moments de la loi $f(x)$ sont égaux aux p premiers moments empiriques calculés sur l'échantillon. On obtient ainsi un système de p équations à p inconnues. Les p inconnues sont les p paramètres a_k que l'on peut ainsi déterminer.

Les moments relatifs à la population s'écrivent :

Moment non centré d'ordre 1 (ou espérance mathématique) :

$$\mu_1 = E(X) = \int_{-\infty}^{+\infty} x f(x) dx$$

Moment centré d'ordre k ($k=2, 3, \dots, p$) :

$$\mu_k = E[(X - \mu_0)^k] = \int_{-\infty}^{+\infty} (x - \mu_0)^k f(x) dx \quad (\text{A-7})$$

Les moments empiriques (non biaisés) estimés à partir de l'échantillon s'écrivent :

Moment non centré d'ordre 1 (moyenne) :

$$m_1 = \frac{1}{N} \sum_{i=1}^N x_i$$

Moment centré d'ordre 2 (variance) :

$$m_2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2 \quad (\text{A-8})$$

Moment centré d'ordre 3 :

$$m_k = \frac{N}{(N-1)(N-2)} \sum_{i=1}^N (x_i - m_1)^k$$

Finalement, les p paramètres de la loi $f(x)$ sont déterminés en écrivant l'égalité entre les moments théoriques μ_k et les moment expérimentaux m_k ($k=1, \dots, p$).

Méthode des moments pondérés

Une variante de la méthode des moments consiste à raisonner non plus sur les moments, mais sur les moments pondérés définis par Greenwood (1979). Cette méthode a été mise au point par Hosking (1985).

De manière générale, les moments pondérés relatifs à la population s'écrivent :

$$M(i, j, k) = E[X^i F^j (1-F)^k] = \int_0^1 x(F)^i F^j (1-F)^k dF \quad (\text{A-9})$$

Le cas où $j=k=0$ correspond à la définition des moments traditionnels.

Dans la méthode, on fixe $i=1$ et $k=0$, les moments pondérés sont alors plus simplement définis par :

$$\beta_j = \int_0^1 x(F) F^j dF \quad (\text{A-10})$$

Les moments pondéré empiriques débiaisés, se calculent de la façon suivante (Landwehr *et al.*, 1979) :

$$b_j = \frac{1}{N} \sum_{i=1}^N \frac{(i-1)(i-2)\dots(i-j)}{(N-1)(N-2)\dots(N-j)} x_i \quad (\text{A-11})$$

De la même façon que précédemment, on cherchera à égaliser les moments pondérés théoriques β_j et expérimentaux b_j .

Méthode du maximum de vraisemblance :

Le principe de la méthode du maximum de vraisemblance est de maximiser la probabilité pour que si l'on tire N variables dans la population, on obtienne l'échantillon observé. Si les variables aléatoires sont indépendantes, la probabilité pour que les N tirages donnent l'échantillon des valeurs x_i est donnée par :

$$V = \prod_{i=1}^N f(x = x_i) dx \quad (\text{A-12})$$

Cette probabilité est fonction de p paramètres a_k que nous cherchons à déterminer. Elle est maximisée en écrivant que sa dérivée par rapport à chacun des p paramètres a_k est égale à zéro. On obtient donc p équations à p , dont les p inconnues sont les paramètres a_k .

ANNEXE 2

DESCRIPTIF DES BASSINS AU CHAPITRE 3

Code HYDRO	Rivière	Localisation	S (km ²)	Département
A143201	LA THUR	WILLER-SUR-THUR	159	68
A273201	LA BRUCHE	WISCHES	230	67
A405062	LA MOSELLE	RUPT-SUR-MOSELLE	153	88
A417301	LA CLEURIE	CLEURIE	68	88
A420063	LA MOSELLE	ST-NABORD (NOIR GUEUX)	621	88
A436203	LA VOLOGNE	CHENIMENIL (2)	355	88
A444201	LE DURBION	VAXONCOURT	145	88
A463201	L' AVIERE	FRIZON (BASSE FRIZON)	106	88
A511061	LA MOSELLE	TONNOY	1990	54
A524201	LE GITE	VELOTTE-ET-TATIGNECOURT (VELOTTE)	114	88
A525101	LE MADON	MIRECOURT	382	88
A543101	LE MADON	PULLIGNY	940	54
A573061	LA MOSELLE	TOUL	3350	54
A600101	LA MEURTHE	FRAIZE	69	88
A605102	LA MEURTHE	ST-DIE	374	88
A614201	LE RABODEAU	MOYENMOUTIER	153	88
A627101	LA MEURTHE	DAMELEVIERES	2280	54
A633201	LA PLAINE	RAON-L' TAPE (LA TROUCHE)	116	88
A656111	LA VEZOUZE	LUNEVILLE	559	54
A673122	LA MORTAGNE	GERBEVILLER	493	54
A681201	LE SANON	LAGARDE	73	57
A687201	LE SANON	DOMBASLE-SUR-MEURTHE	284	54
A690381	LE PETIT RHONE	ST-NICOLAS-DE-PORT	21,4	54
A692101	LA MEURTHE	LANEUVEVILLE-DEVANT-NANCY	2780	54
A695301	L' AMEZULE	LAY-ST-CRISTOPHE	84,8	54
A712201	L' ESCH	JEZAINVILLE	231	54
A735201	LE RUPT DE MAD	ONVILLE	358	54
A764201	LA PETITE SEILLE	CHATEAU-SALINS	143	57
A782101	LA SEILLE	NOMENY	925	54
A793061	LA MOSELLE	HAUCONCOURT	9387	57
A804101	L' ORNE	BONCOURT	412	54
A812201	L' YRON	JARNY	383	54
A975201	LA NIED FRANCAISE	CONDE-NORTHEN (PONTIGNY)	499	57
A983201	LA NIED ALLEMANDE	FAULQUEMONT	187	57
A994201	LA NIED	BOUZONVILLE	1160	57
H310201	L' OUANNE	TOUCY	153	89
H572301	L' ORGEVAL	BOISSY-LE-CHATEL (LE THEIL, ULTRA-SONS)	104	77
H572302	LES AVENELLES	BOISSY-LE-CHATEL (LES AVENELLES, ULTRA-SONS)	45,7	77
H572303	LES AVENELLES	ST-GERMAIN-SOUS-DOUE (LA GOUGE)	24,7	77
H572321	LE FOSSE ROGNON	DOUE (MELARCHEZ, ULTRA-SONS)	7,1	77
K064311	LA MARE	SA-MARCELLIN-EN-FOREZ (VERINES)	95,2	42
K066331	LA COISE	LARAJASSE (LE NEZEL)	61	69
K172421	LA DRAGNE	VANDENESSE	115	58
K201081	L' ALLIER	LAVEYRUNE (ROGLETON 1)	48,8	7
K212301	LE CHAPEAUROUX	CHATEAUNEUF-DE-RANDON (HERMET)	109	48
L621692	LE BRIGNON	PAULMY (2)	75	37
L811401	LA VIETTE	BEAULIEU-SOUS-PARTHENAY (PONT DE LA BORDE)	33	79
M711241	LA SEVRE NANTAISE	TIFFAUGES (LA MOULINETTE)	814	85
M820502	L' OGNON	LES SORINIERES (VILLENEUVE)	147	44
N302401	LE LOUING	CHANTONNAY (ST-PHILBERT DU PONT CHARRAULT)	120	85

Code HYDRO	Rivière	Localisation	S (km ²)	Département
N311301	LE PETIT LAY	ST-HILAIRE-LE-VOUHIS	262	85
N330821	LA DOULAYE	MOUTIERS-SUR-LE-LAY	38	85
N340301	L' YON	DOMPIERRE-SUR-YON	42	85
O062401	LE VOLP	MONTBERAUD (STE-CROIX-VOLVESTRE)	91	31
O203401	L' AUSSONNELLE	SEILH	192	31
O308432	LA MIMENTE	FLORAC	125	48
O470403	LE DADOU	PAULUNET (ST-JEAN-DE-JEANNE)	72	81
P218431	LA TOURMENTE	ST-DENIS-LES-MARTEL	202	46
Q041401	LE LOUET	SOMBRUN	84,3	65
U100401	L' OGNON	SERVANCE (FOURGUENONS)	73,5	70
U102501	LE RAHIN	PLANCHER-BAS	33	70
U131401	L' OUCHE	LA BUSSIERE-SUR-OUCHÉ	304	21
U131402	L' OUCHE	STE-MARIE-SUR-OUCHÉ (PONT DE PANY)	442	21
U132401	L' OUCHE	PLOMBIERES-LES-DIJON	655	21
U234582	LA ROSEMontoise	CHAUX	25,4	90
U401401	LA REYSSOUZE	MONTAGNAT	84,4	1
U450501	L' ARDIERES	BEAUJEU	54,5	69
U462401	L' AZERGUES	CHATILLON	336	69
V023501	LA MENOGE	VETRAZ-MONTHOUX	160	74
V101581	LA SEMINE	CHATILLON-EN-MICHAILLE (COZ)	183	1
V123561	L' EAU MORTE	DOUSSARD	92,5	74
V152541	L' AINAN	ST-GEOIRE-EN-VALDAINE (1)	41	38
V271201	L' AIN	PONT-D'AIN	2760	1
V292401	L' ALBARINE	ST-RAMBERT-EN-BUGEY	232	1
V292402	L' ALBARINE	CHALEY	135	1
V294201	L' AIN	CHAZEY-SUR-AIN	3630	1
V301501	L' YZERON	CRAPONNE	48	69
V351561	LE RUISSEAU DES PREAUX	BOURG-ARGENTAL	22,1	42
V361401	LA GALAURE	ST-UZE	232	26
V414521	LA GLUEYRE	GLUIRAS (TISONECHE)	71	7
V421401	LA DROME	LUC-EN-DIOIS	194	26
V422501	LE BEZ	CHATILLON-EN-DIOIS	227	26
V427501	LA GERVANNE	BEAUFORT-SUR-GERVANNE	108	26
V441401	LE ROUBION	SOYANS	186	26
V445501	LE JABRON	SOUSPIERRE	85	26
V713501	LE GARDON DE ST-JEAN	CORBES (ROC COURBE)	263	30
W110501	LE GELON	LA CROIX-DE-LA-ROCHETTE	65,4	73
W353402	L' HERBASSE	CLERIEUX (PONT DE L'HERBASSE)	187	26
X121632	LA RAVINE ELEMENTAIRE	DRAIX (LA ROUBINE)	13	4
X200202	LE VERDON	ALLOS (LA FOUX)	10,1	4
Y022401	LE TECH	LE TECH (PONT DE LA VIERGE)	132	66
Y024403	LE TECH	AMELIE-LES-BAINS-PALALDA (AVAL 2)	376	66
Y024521	LE MONDONY	AMELIE-LES-BAINS-PALALDA	32,3	66
Y025404	LE TECH (PARTIEL)	REYNES (ST-PAUL SUR TECH)	473	66
Y028405	LE TECH	PALAU-DEL-VIDRE (ELNE)	726	66
Y043642	LE CADY	VILLEFRANCHE-DE-CONFLENT	60,2	66
Y046403	LA TET (PARTIELLE)	RODES	974	66
Y047403	LA TET (PARTIELLE)	PERPIGNAN	1300	66
Y131401	LE FRESQUEL	VILLEPINTE	216	11
Y200201	L' HERAULT	VALLERAUGUE	46,2	30
Y402201	L' ARC	MEYREUIL (BAYEUX)	303	13
Y412201	L' ARC	AIX-EN-PROVENCE (ST-PONS)	605	13
Y461581	LE VALESCURE	COLLOBRIERES (AVAL)	9,22	83
Y543621	LA VERNE	LA MOLE (LES CABRIS)	38	83
Y551404	LA SIAGNE	CALLIAN (AJUSTADOUX)	182	83

ANNEXE 3

PARAMETRES DU MODELE QdF LOCAL CONVERGENT AJUSTES AVEC $P=0$

$$V(d,T) = \frac{x_0(0) + a(0)\ln(T)}{1 + d/\Delta}$$

station	Nb d'années d'observation	P (m3/s)	$x_0(0)$ (m3/s)	a(0) (m3/s)	Delta (j)
A143201	21,38	0	47,34	20,06	7,22
A273201	20,75	0	40,51	16,36	10,58
A405062	17,67	0	75,70	28,88	5,97
A417301	26,68	0	20,70	7,05	5,03
A420063	17,71	0	244,39	88,20	6,30
A436203	17,74	0	54,69	24,57	14,99
A444201	18,02	0	34,24	11,74	2,97
A463201	26,60	0	19,46	6,72	3,16
A511061	11,63	0	351,82	139,23	8,96
A524201	17,74	0	23,14	6,18	3,40
A525101	32,17	0	76,02	27,83	2,89
A543101	16,75	0	157,83	55,22	4,81
A573061	17,78	0	474,91	204,05	10,69
A600101	14,41	0	11,05	3,47	16,50
A605102	10,91	0	50,67	22,97	13,89
A614201	14,01	0	16,72	8,09	16,22
A627101	16,88	0	224,99	100,92	13,48
A633201	14,72	0	9,68	4,63	16,79
A656111	23,51	0	70,98	26,30	5,75
A673122	16,71	0	59,61	25,96	4,80
A681201	12,71	0	25,32	8,38	2,16
A687201	9,79	0	31,04	16,54	7,02
A690381	20,01	0	2,79	1,17	2,11
A692101	11,76	0	215,68	95,51	16,02
A695301	24,18	0	13,15	5,12	2,90
A712201	26,79	0	13,49	7,13	7,85
A735201	26,76	0	41,43	19,70	4,99
A764201	28,65	0	12,17	6,38	8,04
A782101	26,66	0	46,39	24,81	17,85
A793061	14,91	0	813,92	365,38	14,99
A804101	16,72	0	57,36	25,40	5,20
A812201	17,87	0	72,39	34,83	3,54
A975201	23,17	0	36,77	21,95	11,13
A983201	13,60	0	16,55	8,63	15,11
A994201	32,87	0	75,28	39,68	15,66
H310201	22,67	0	15,47	6,82	2,73
H572301	30,51	0	11,39	4,77	3,12
H572302	23,66	0	5,97	2,15	2,88
H572303	19,31	0	3,88	1,40	3,20
H572321	30,83	0	1,73	0,50	1,73
K064311	19,60	0	8,32	4,44	4,25
K066331	18,79	0	10,48	6,68	2,97
K172421	22,53	0	20,64	6,14	5,29
k201081	18,56	0	32,63	22,28	4,35
K212301	18,71	0	17,94	9,06	4,21
L621692	20,69	0	7,12	3,19	2,03
l811401	14,69	0	8,53	1,81	3,15
M711241	20,36	0	109,55	61,60	5,27
M820502	23,39	0	22,78	8,22	4,06

station	Nb d'années d'observation	P (m3/s)	x ₀ (0) (m3/s)	a(0) (m3/s)	Delta (j)
N302401	20,64	0	13,82	6,39	5,48
N311301	20,58	0	45,16	18,27	3,87
n330821	24,19	0	7,88	1,96	2,47
N340301	15,94	0	11,78	5,50	1,91
O062401	24,48	0	47,26	22,85	1,00
O203401	22,62	0	26,43	15,89	2,13
O308432	13,78	0	123,51	83,35	1,69
O470403	22,73	0	25,15	14,82	3,36
P218431	19,00	0	22,18	7,05	7,27
Q041401	18,07	0	23,39	8,91	2,10
U100401	21,32	0	46,21	16,02	3,53
U102501	18,83	0	23,59	6,08	3,95
U131401	10,20	0	30,98	12,99	6,61
U131402	9,32	0	36,81	15,42	8,56
u132401	14,25	0	51,63	22,77	11,29
U234582	16,27	0	12,38	4,14	5,56
U401401	25,02	0	7,70	3,20	3,70
U450501	22,92	0	6,65	2,43	8,80
U462401	21,12	0	45,77	19,87	7,93
V023501	9,78	0	43,36	11,33	2,69
V101581	10,52	0	128,60	45,37	3,31
V123561	22,89	0	25,69	7,87	4,98
V152541	14,35	0	5,32	2,60	15,19
V271201	10,29	0	813,55	319,30	5,44
V292401	22,76	0	77,66	24,05	4,10
V292402	13,68	0	57,50	15,01	3,88
V294201	15,01	0	801,79	321,66	7,92
V301501	27,04	0	4,03	1,97	5,51
V351561	13,89	0	2,44	1,04	7,72
V361401	12,47	0	33,06	15,71	1,92
V414521	30,97	0	37,70	16,05	2,98
V421401	14,78	0	27,70	15,43	5,06
V422501	15,01	0	37,63	16,05	6,89
V427501	22,01	0	21,58	9,42	2,46
V441401	15,85	0	46,15	19,88	0,96
V445501	27,04	0	13,82	7,80	1,31
V713501	19,03	0	254,85	118,50	1,42
W110501	13,26	0	8,56	2,60	10,13
w353402	18,93	0	43,94	24,81	1,07
x121632	9,19	0	12,73	5,94	0,01
X200202	10,64	0	4,85	1,75	3,62
Y022401	18,95	0	26,37	18,71	3,91
Y024403	9,26	0	100,29	73,91	1,79
Y024521	24,14	0	23,32	16,30	1,78
Y025404	20,81	0	201,45	173,89	1,84
Y028405	11,26	0	369,21	336,50	1,33
Y043642	24,50	0	19,38	15,03	2,02
Y046403	18,78	0	105,86	60,14	4,72
Y047403	21,00	0	198,76	143,71	2,41
Y131401	19,69	0	18,40	12,95	2,89
Y200201	13,91	0	85,96	45,62	1,36
Y402201	23,33	0	39,41	30,55	1,19
Y412201	18,71	0	48,98	32,05	2,14
Y461581	24,18	0	4,86	2,73	1,94
Y543621	25,19	0	23,95	15,58	1,09
Y551404	10,02	0	48,15	14,44	12,23

ANNEXE 4

PARAMETRES DU MODELE QdF LOCAL CONVERGENT
AJUSTES AVEC $P = \overline{QA}$

$$V(d, T) = \frac{x_0(0) + a(0) \ln(T) - P}{1 + d/\Delta} + P$$

station	Nb d'années d'observation	P (m3/s)	x(0) (m3/s)	a(0) (m3/s)	Delta (m3/s)
A143201	21,38	5,08	49,81	20,92	5,04
A273201	20,75	6,13	43,02	16,85	6,58
A405062	17,67	7,08	79,90	30,27	4,36
A417301	26,68	2,37	22,11	7,07	3,19
A420063	17,71	25,14	253,74	95,84	4,45
A436203	17,74	9,92	60,84	27,80	7,50
A444201	18,02	1,94	34,71	12,18	2,50
A463201	26,60	1,22	19,88	7,10	2,56
A511061	11,63	46,31	377,20	148,83	5,76
A524201	17,74	1,51	23,76	6,44	2,73
A525101	32,17	4,79	77,80	29,25	2,38
A543101	16,75	11,49	160,84	55,23	3,95
A573061	17,78	66,06	512,75	210,22	6,29
A600101	14,41	2,19	12,27	4,44	7,23
A605102	10,91	8,84	55,20	26,09	7,35
A614201	14,01	3,47	18,78	9,63	7,48
A627101	16,88	37,25	251,56	108,17	7,15
A633201	14,72	2,26	10,63	4,58	8,63
A656111	23,51	6,59	71,75	28,12	4,63
A673122	16,71	6,06	61,65	29,76	3,44
A681201	12,71	0,95	25,34	8,57	2,01
A687201	9,79	2,99	31,99	17,35	5,43
A690381	20,01	0,14	2,81	1,21	1,87
A692101	11,76	37,65	229,21	95,67	9,68
A695301	24,18	0,72	13,41	5,03	2,50
A712201	26,79	1,40	14,00	7,10	6,13
A735201	26,76	3,40	43,53	18,60	3,92
A764201	28,65	1,27	12,58	6,60	6,18
A782101	26,66	7,50	47,72	23,85	14,88
A793061	14,91	138,57	851,97	379,86	9,72
A804101	16,72	3,74	58,46	25,41	4,42
A812201	17,87	4,07	73,21	36,17	3,06
A975201	23,17	3,87	39,16	23,87	7,93
A983201	13,60	2,10	17,48	9,00	10,72
A994201	32,87	11,06	78,83	40,59	11,44
H310201	22,67	0,99	16,15	7,03	2,18
H572301	30,51	0,65	11,47	4,86	2,77
H572302	23,66	0,30	6,01	2,20	2,57
H572303	19,31	0,21	3,94	1,42	2,79
H572321	30,83	0,05	1,74	0,51	1,63
K064311	19,60	0,87	8,69	4,83	2,88
K066331	18,79	0,69	11,63	7,91	1,91
K172421	22,53	1,96	21,66	6,87	3,58
k201081	18,56	1,66	32,81	23,20	3,88
K212301	18,71	1,69	19,78	9,97	2,33
L621692	20,69	0,31	7,26	3,31	1,79
l811401	14,69	0,45	8,67	1,72	2,74
M711241	20,36	9,61	115,73	67,03	3,86
M820502	23,39	1,09	22,94	8,46	3,60

station	Nb d'années d'observation	P (m3/s)	x ₀ (0) (m3/s)	a(0) (m3/s)	Delta (j)
N302401	20,64	1,11	14,48	6,39	4,35
N311301	20,58	2,49	46,02	18,23	3,31
n330821	24,19	0,34	7,96	1,93	2,20
N340301	15,94	0,39	11,82	5,56	1,81
O062401	24,48	1,09	47,35	23,16	0,95
O203401	22,62	0,77	26,55	16,18	1,99
O308432	13,78	3,57	124,74	86,05	1,54
O470403	22,73	1,59	25,69	15,20	2,94
P218431	19,00	2,06	23,17	7,13	5,32
Q041401	18,07	1,13	23,62	9,38	1,82
U100401	21,32	3,40	47,38	16,43	2,76
U102501	18,83	1,83	24,01	6,35	3,13
U131401	10,20	2,70	31,82	13,47	5,08
U131402	9,32	3,64	40,48	16,69	6,21
u132401	14,25	6,78	54,33	23,24	8,39
U234582	16,27	1,00	12,67	4,32	4,43
U401401	25,02	0,50	7,78	3,31	3,16
U450501	22,92	0,92	7,02	2,47	5,70
U462401	21,12	4,66	47,04	21,66	5,89
V023501	9,78	4,31	46,59	13,77	1,64
V101581	10,52	9,31	133,77	48,94	2,54
V123561	22,89	2,84	27,37	8,04	3,12
V152541	14,35	1,06	5,88	2,47	8,21
V271201	10,29	110,82	863,67	355,21	3,59
V292401	22,76	6,27	79,02	24,91	3,32
V292402	13,68	4,46	58,52	16,78	3,03
V294201	15,01	127,23	862,55	384,13	4,73
V301501	27,04	0,37	4,19	2,08	4,14
V351561	13,89	0,42	2,74	1,11	3,81
V361401	12,47	2,10	33,66	16,85	1,58
V414521	30,97	2,18	38,17	16,58	2,51
V421401	14,78	2,45	28,66	16,88	3,80
V422501	15,01	4,20	39,75	18,55	4,26
V427501	22,01	1,19	22,02	9,84	2,13
V441401	15,85	1,81	46,90	20,63	0,82
V445501	27,04	0,78	14,47	8,27	0,96
V713501	19,03	7,11	256,67	122,53	1,30
W110501	13,26	1,51	9,36	2,42	5,57
w353402	18,93	1,85	44,46	25,81	0,95
x121632	9,19	0,01	12,72	5,94	0,01
X200202	10,64	0,39	5,38	1,91	2,19
Y022401	18,95	1,43	26,86	19,62	3,27
Y024403	9,26	6,37	101,55	77,33	1,49
Y024521	24,14	0,55	23,40	16,69	1,67
Y025404	20,81	7,91	208,21	180,86	1,55
Y028405	11,26	9,64	371,88	345,39	1,24
Y043642	24,50	1,02	19,66	17,60	1,61
Y046403	18,78	10,06	110,49	68,57	3,26
Y047403	21,00	10,49	204,58	143,63	2,07
Y131401	19,69	1,08	19,18	13,32	2,39
Y200201	13,91	2,59	86,49	47,05	1,24
Y402201	23,33	1,23	39,88	31,10	1,08
Y412201	18,71	3,22	50,85	33,86	1,70
Y461581	24,18	0,13	4,95	2,78	1,78
Y543621	25,19	0,45	24,31	16,07	1,01
Y551404	10,02	5,59	49,88	14,90	8,61

ANNEXE 5

APPLICATION DU MODELE QdF CONVERGENT AUX ETIAGES

Cette annexe concerne la suite de l'application du modèle *QdF* convergent aux étiages, présentée au chapitre 4, §3. La lecture de ce paragraphe est un préalable à celle de cette annexe.

1. Consolidation des observations en fonction de U_F

L'hypothèse de base du modèle local convergent est la suivante :

Toute distribution de durée d_1 peut se déduire d'une distribution de durée d_2 par une affinité ayant pour axe la droite d'équation $y(x)=P$

L'exemple de la Loire à Montjean semble confirmer cette hypothèse, en fixant simplement P à zéro. En effet, si on norme chaque distribution par le quantile de fréquence 0.5, c'est dire $U_F=0$, on note que toutes les distributions se superposent.

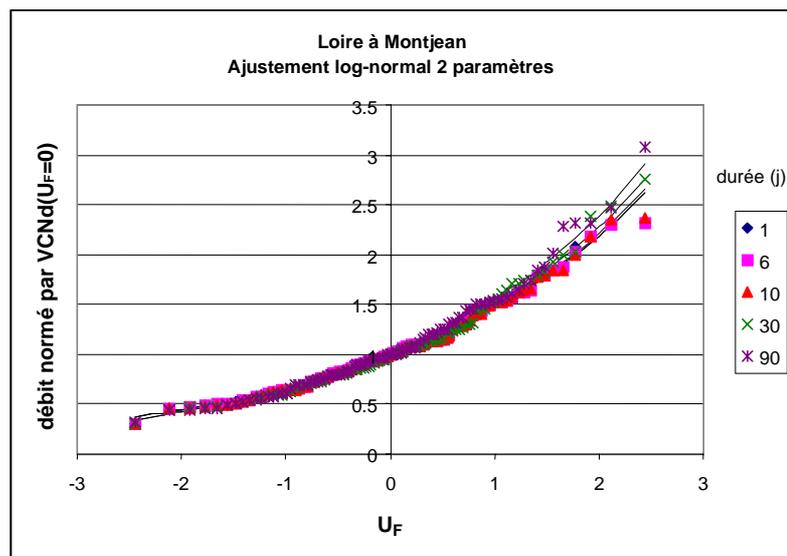


Figure A-3 : Vérification de l'affinité sur les débits moyens minimaux *VCNd*

Il est pratique de représenter les débits en échelle logarithmique afin que les quantiles théoriques de la loi log-normal s'alignent sur une droite. Dans ce cas, la propriété d'affinité vérifiée par les débits en valeur naturelle se traduit en valeurs logarithmiques par un parallélisme des distributions (Figure A-4).

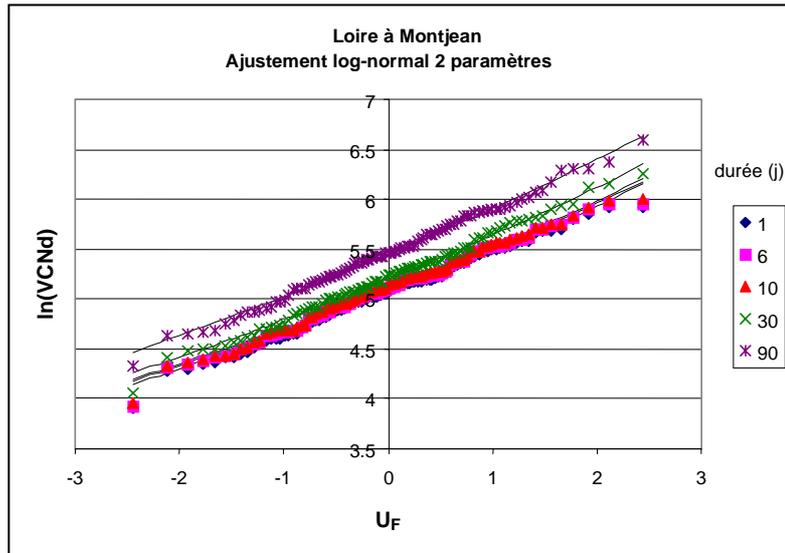


Figure A-4 : Parallélisme observé des distributions des débits moyens minimaux $VCNd$ en échelle logarithmique

Ce parallélisme implique que les paramètres s_d de la loi ajustée sur chaque échantillon (relation 4-33) ne dépendent pas de d . Afin de déterminer le paramètre noté s^c consolidé sur l'ensemble des distributions, chaque distribution expérimentale vcn_d est translatée de $-m_d$ de façon à ce que toutes les distributions obtenues se confondent et passent par zéro. En effet, d'après (4-33), m_d correspond à la valeur des distributions pour $U_F=0$.

$$\ln[vcn_d(j)] = \ln[vcn_d(j)] - m_d \tag{A-13}$$

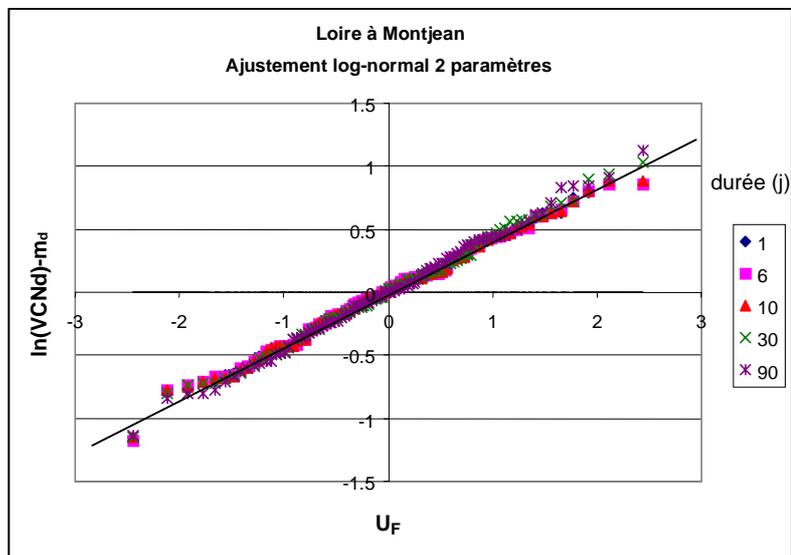


Figure A-5 : Distributions expérimentales des débits moyens minimaux $VCNd$ en échelle logarithmique, translatées de $-m_d$, et ajustement des distributions “consolidées”

Le paramètre consolidé s^c est égal à la pente définie par l'ensemble de ces distributions confondues, et le modèle « consolidé » est donné par l'expression :

$$\ln[VCN_d^c(u_F)] = s^c \cdot u_F + m_d \tag{A-14}$$

Cette équation est identique à (4-33), à la différence près que s^c ne dépend pas de d .

Si maintenant on revient en valeurs naturelles, par construction, les distributions consolidées convergent parfaitement vers un point de coordonnée $(-\infty, 0)$, de la même façon que pour les crues.

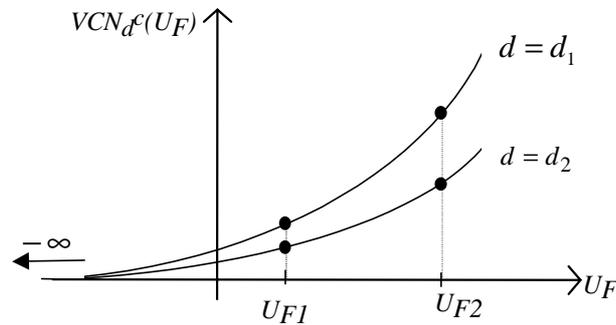


Figure A-6 : Propriété d'affinité des distributions appliquée aux étiages

De façon analogue aux crues, la propriété d'affinité Figure A-6 permet d'établir la relation suivante, indépendante de U_F .

$$\frac{VCN_{d_2}^c(u_{F1})}{VCN_{d_1}^c(u_{F1})} = \frac{VCN_{d_2}^c(u_{F2})}{VCN_{d_1}^c(u_{F2})} \quad (\text{A-15})$$

2. Modélisation en fonction de d

L'étape suivante consiste à définir les quantiles pour toute durée d , c'est à dire comme étant une fonction de d . Ces quantiles, notés $VCN(U_F, d)$ seront donnés par :

$$\ln(VCN(d, u_F)) = s^c \cdot u_F + m(d) \quad (\text{A-16})$$

s^c ayant déjà été défini au paragraphe précédent (ajustement consolidé), il ne reste plus qu'à déterminer la fonction $m(d)$.

Pour ce faire, les hypothèses du modèle sont les suivantes :

- La première a été énoncée au début du paragraphe précédent et est relative à l'affinité des distributions (Figure A-6).
- La deuxième hypothèse concerne l'évolution des quantiles, pour u_F fixé. On suppose que $VCN(d, U_F)$ varie **linéairement** en fonction de d . Cette hypothèse est fondée sur de nombreuses observations. A titre d'exemple, la Figure A-7 donne l'évolution observée pour les quantiles $VCNd(U_F)$.

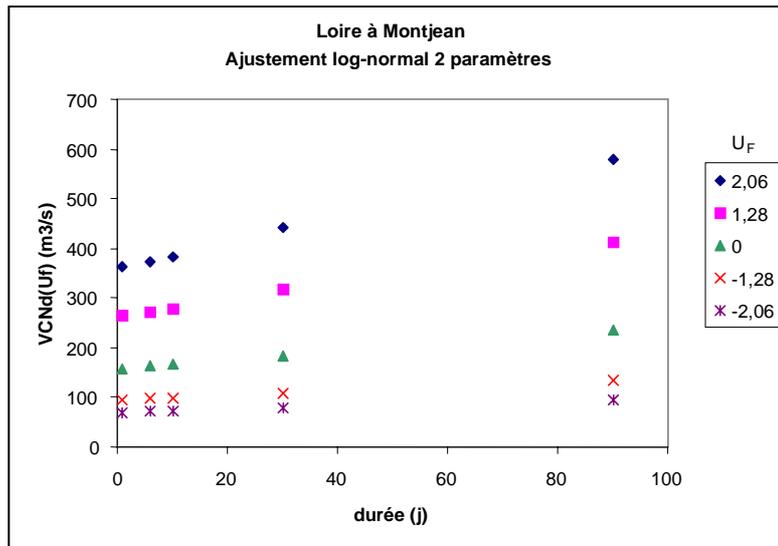


Figure A-7 : Evolution observée des quantiles $VCNd(U_F)$, en fonction de d pour U_F fixé

La relation (A-15) ne dépendant pas de U_F , nous allons appliquer la transformation suivante aux différentes valeurs étudiées (quantile VCN et valeur observée de l'échantillon vcn):

$$X^* = \frac{X}{VCN_{d=1j}^c(U_F)} \quad (\text{A-17})$$

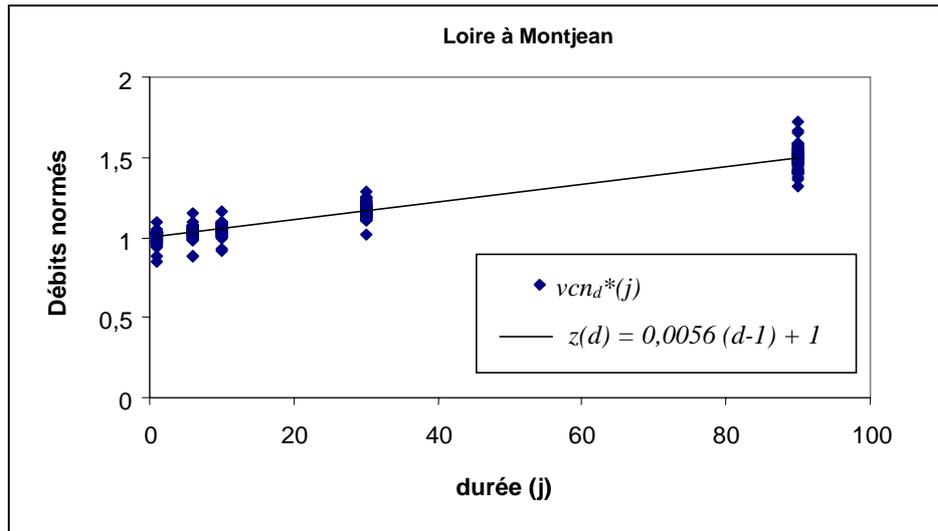
En appliquant la transformation ci-dessus aux quantiles $VCN(d, U_F)$ on obtient des variables ne dépendant que de d (hypothèse 1) :

$$z(d) = VCN^*(d, U_F) = \frac{VCN(d, U_F)}{VCN_{d=1j}^c(U_F)} \quad (\text{A-18})$$

De plus, l'hypothèse 2 indique que $z(d)$ est de la forme $z(d) = ad + b$. En imposant que $VCN(d = 1, U_F) = VCN_{d=1j}^c(U_F)$, $z(d)$ ne dépend alors plus que du paramètre a , $z(1)$ étant alors égal à 1 :

$$z(d) = VCN^*(d, U_F) = \frac{VCN(d, U_F)}{VCN_{d=1j}^c(U_F)} = a(d - 1) + 1 \quad (\text{A-19})$$

La Figure A-8 montre le résultat de la transformation (A-17) ci-dessus appliquée aux valeurs échantillonnées $vcn_d(r)$, r étant le rang dans l'échantillon servant à calculer U_F . Le coefficient a de l'équation (A-19) est alors le coefficient directeur de la droite ajustée dans le nuage de point. Cette façon d'ajuster est équivalente à celle présentée Figure 3-9 (chapitre 3) concernant la durée Δ relative au modèle des crues. Nous l'avons adoptée car elle tient compte de l'ensemble des valeurs échantillonnées $vcn_d(j)$, toutes durées confondues.

Figure A-8 : Ajustement en fonction de d

En particulier, on peut écrire pour $u_F = 0$:

$$\frac{VCN(d, u_F = 0)}{VCN(d = 1j, u_F = 0)} = \frac{e^{m(d)}}{e^{m(1j)}} = a(d - 1) + 1 \quad (\text{A-20})$$

Finalement, on obtient le modèle suivant :

$$VCN(d, u_F) = e^{s^c \cdot u_F} \cdot m(1j) \cdot \left[\frac{d-1}{\Delta e} + 1 \right] \quad (\text{A-21})$$

où le coefficient a a été remplacé par $1/\Delta e$, de façon à avoir la dimension d'un temps (j). La durée Δe est l'équivalent pour les étiages de la durée Δe que nous avons définie pour les crues (cf. chapitre 3).

Remarques :

La méthode d'ajustement proposée est inspirée de la première méthode présentée pour les crues, dite « des ajustements consolidés », au chapitre 3, §2.2.2. De même, il est possible d'adapter la seconde méthode que nous avons présentée, dite « simplifiée » (cf. chapitre 3, §2.2.3).

3. Application du modèle sur 36 sous-bassins de la Moselle

La validation du modèle proposé a été effectuée sur une partie des 105 bassins étudiés au chapitre 3. Ces bassins sont au nombre de 36 et sont tous situés sur la partie française du bassin de la Moselle (Figure A-9). Leurs surfaces sont comprises entre 16 et 10 000 km².

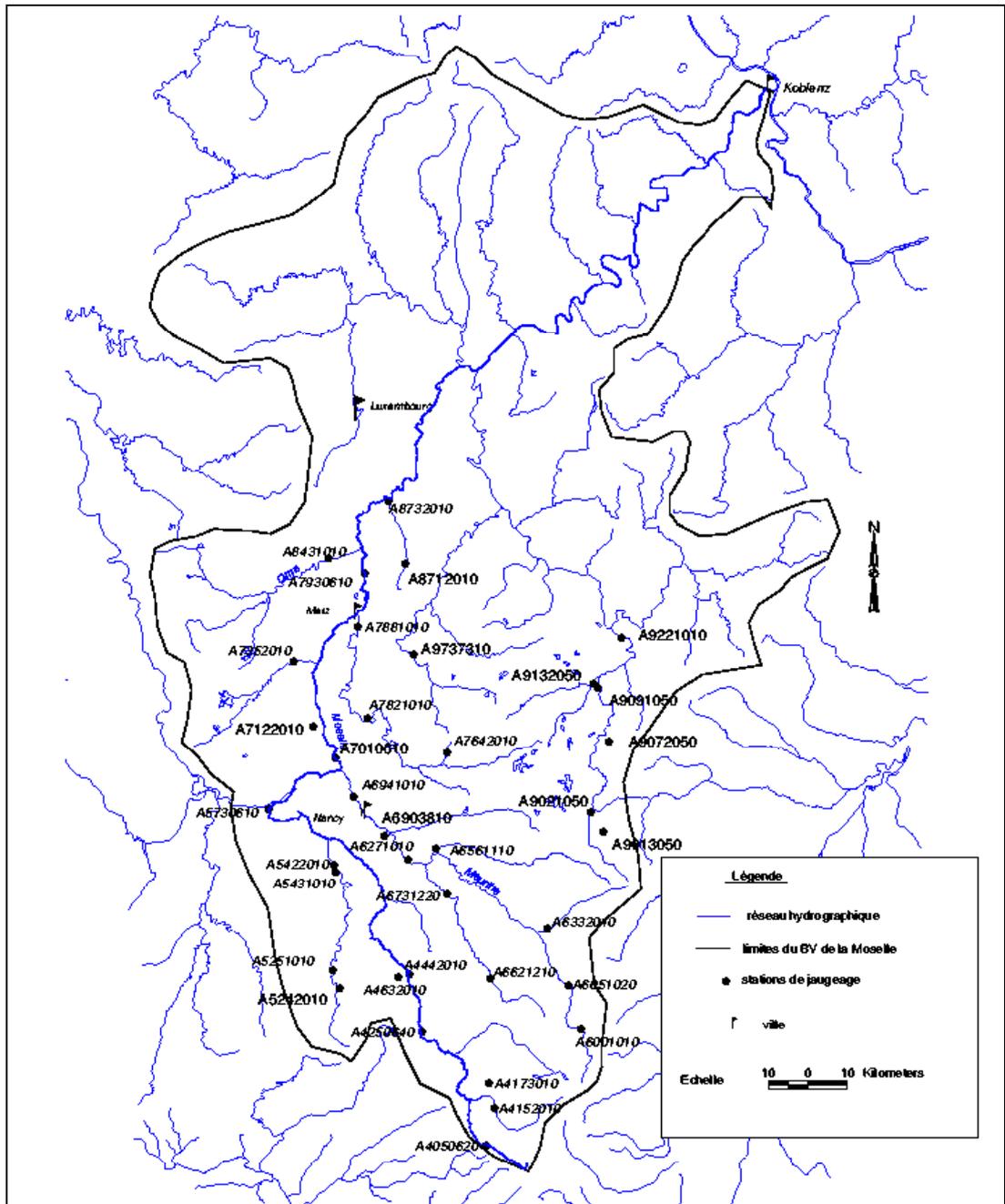


Figure A-9 : Localisation des stations étudiées (Chaput, 1999)

Afin de comparer sur un même graphique les quantiles issus des observations $VCNd(U_F)$ et de la modélisation $VCN(d, U_F)$ des 36 sous-bassins, nous procédons de façon analogue aux crues. Les quantiles sont transformés selon la relation (A-17). Cette transformation ramène la comparaison des quantiles $VCNd(U_F)$ et $VCN(d, U_F)$ à celle des valeurs $VCNd^*(U_F)$ et $VCN^*(d, U_F)$. L'intérêt est que, d'après (A-19), les valeurs théoriques $VCN^*(d, U_F)$ sont égales à $\frac{(d-1)}{\Delta e} + 1$ (rappelons que $a=1/\Delta e$). Il devient alors possible de représenter sur un même graphe (Figure A-10) les résultats des 36 sous-bassins de la Moselle. Cette représentation permet d'estimer directement, toute durée confondue, l'erreur relative maximale due à la modélisation, par exemple 12% pour T=2ans et environ 16% pour T=5ans (humide) (cf. chap.4 §3.).

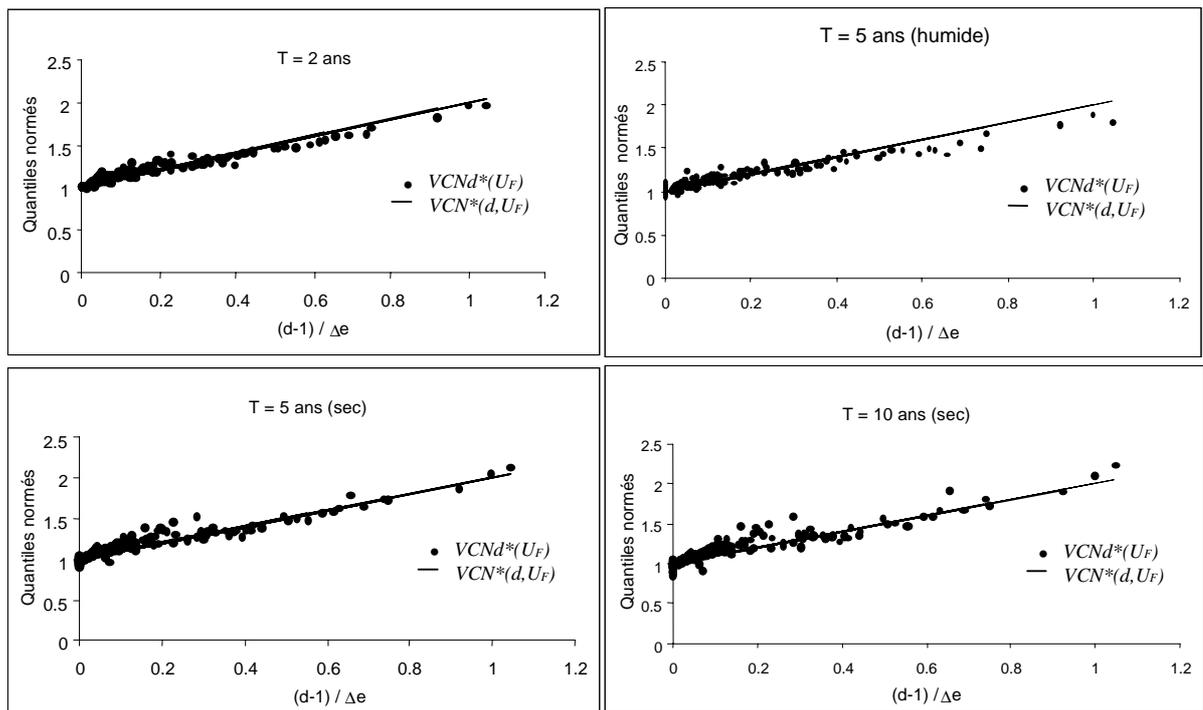


Figure A-10 : Validation du modèle (Chaput, 1999)

4. Conclusion

Nous avons montré que les concepts développés pour l'analyse débit-durée-fréquence des crues pouvaient être adaptés aux étiages. Nous n'avons présenté ici que les résultats appliqués aux débits moyens $VCNd$, mais l'analogie peut être poussée plus loin. Ainsi Galéa *et al.*, (2001) donnent, comme pour les crues, une relation entre les débits moyens $VCNd$ et les débits seuils $QCNd$. D'autre part, le choix de la loi étant laissé libre, Dussaux (2000) présente ces développements adaptés à d'autres types de lois (loi de Weibull à 2 et 3 paramètres, loi-log-normale à 3 paramètres).

ANNEXE 6

LISTE DES BASSINS ETUDIÉS AU CHAPITRE 5

Code	Nom	Surface (km ²)	Nb d'année d'observation	region
01BF001	NOUVELLE (RIVIERE) AU PONT	1140	30	QUE1
01BH001	DARTMOUTH (RIVIERE) PRES DE CORTEREAL	748	22	QUE2
01BH002	YORK (RIVIERE) A SUNNY BANK	1010	34	QUE2
02AB008	NEEBING RIVER NEAR THUNDER BAY	187	40	ONT4
02AC001	WOLF RIVER AT HIGHWAY NO. 17	736	23	ONT2
02BA002	STEEL RIVER NEAR TERRACE BAY	1190	23	ONT5
02BB002	BLACK RIVER NEAR MARATHON	1980	21	ONT5
02BB003	PIC RIVER NEAR MARATHON	4270	24	ONT5
02BD003	MAGPIE RIVER NEAR MICHIPICOTEN	1930	50	ONT1
02BF001	BATCHAWANA RIVER NEAR BATCHAWANA	1190	26	ONT1
02BF002	GOULAIS RIVER NEAR SEARCHMONT	1160	26	ONT1
02CA002	ROOT RIVER AT SAULT STE. MARIE	108	23	ONT4
02CF007	WHITSON RIVER AT CHELMSFORD	272	33	ONT5
02DD008	DUCHESNAY RIVER NEAR NORTH BAY	90	25	ONT4
02EA005	NORTH MAGNETAWAN RIVER NEAR BURK'S FALLS	321	78	ONT3
02EA010	NORTH MAGNETAWAN RIVER ABOVE PICKEREL LAKE	149	26	ONT1
02EC002	BLACK RIVER NEAR WASHAGO	1445	79	ONT2
02EC009	HOLLAND RIVER AT HOLLAND LANDING	181	28	ONT3
02EC010	SCHOMBERG RIVER NEAR SCHOMBERG	42,9	27	ONT4
02EC011	BEAVERTON RIVER NEAR BEAVERTON	282	25	ONT5
02ED003	NOTTAWASAGA RIVER NEAR BAXTER	1180	46	ONT1
02ED007	COLDWATER RIVER AT COLDWATER	177	28	ONT3
02FA001	SAUBLE RIVER AT SAUBLE FALLS	927	37	ONT2
02FB007	SYDENHAM RIVER NEAR OWEN SOUND	181	60	ONT1
02FC001	SAUGEEEN RIVER NEAR PORT ELGIN	3960	80	ONT1
02FC002	SAUGEEEN RIVER NEAR WALKERTON	2150	80	ONT1
02FC011	CARRICK CREEK NEAR CARLSRUHE	163	39	ONT3
02FE008	MIDDLE MAITLAND RIVER NEAR BELGRAVE	648	26	ONT2
02FE009	SOUTH MAITLAND RIVER AT SUMMERHILL	376	26	ONT1
02FF002	AUSABLE RIVER NEAR SPRINGBANK	865	48	ONT2
02FF007	BAYFIELD RIVER NEAR VARNA	466	27	ONT1
02GA010	NITH RIVER NEAR CANNING	1008	58	ONT3
02GA017	CONESTOGO RIVER AT DRAYTON	326	21	ONT3
02GA018	NITH RIVER AT NEW HAMBURG	552	43	ONT2
02GB007	FAIRCHILD CREEK NEAR BRANTFORD	360	29	ONT1
02GB009	KENNY CREEK NEAR BURFORD	92	28	ONT1
02GC002	KETTLE CREEK AT ST. THOMAS	329	40	ONT3
02GC010	BIG OTTER CREEK AT TILLSONBURG	342	33	ONT3
02GC018	CATFISH CREEK NEAR SPARTA	287	29	ONT3
02GD010	FISH CREEK NEAR PROSPECT HILL	150	48	ONT3
02GD019	TROUT CREEK NEAR FAIRVIEW	36	28	ONT1
02GD020	WAUBUNO CREEK NEAR DORCHESTER	96	29	ONT1
02GE005	DINGMAN CREEK BELOW LAMBETH	146	28	ONT3
02GG002	SYDENHAM RIVER NEAR ALVINSTON	730	46	ONT2
02GG004	BEAR CREEK ABOVE WILKESPORT	609	18	ONT2
02GG005	SYDENHAM RIVER AT STRATHROY	172	25	ONT5
02GG006	BEAR CREEK NEAR PETROLIA	267	27	ONT3
02HA006	TWENTY MILE CREEK AT BALLS FALLS	293	37	ONT3
02HB004	EAST OAKVILLE CREEK NEAR OMAGH	199	37	ONT3
02HB012	GRINDSTONE CREEK NEAR ALDERSHOT	83	28	ONT1
02HC009	EAST HUMBER RIVER NEAR PINE GROVE	197	40	ONT4
02HC013	HIGHLAND CREEK NEAR WEST HILL	88	35	ONT4
02HC018	LYNDE CREEK NEAR WHITBY	106	32	ONT4
02HC019	DUFFINS CREEK ABOVE PICKERING	94	33	ONT4
02HC023	COLD CREEK NEAR BOLTON	62	29	ONT4

Code	Nom	Surface (km ²)	Nb d'année d'observation	region
02HC025	HUMBER RIVER AT ELDER MILLS	303	31	ONT3
02HC027	BLACK CREEK NEAR WESTON	58	27	ONT4
02HC028	LITTLE ROUGE CREEK NEAR LOCUST HILL	78	30	ONT4
02HC029	LITTLE DON RIVER AT DON MILLS	130	29	ONT4
02HC030	ETOBICOKE CREEK BELOW QUEEN ELIZABETH HIGHWAY	204	27	ONT3
02HC031	WEST HUMBER RIVER AT HIGHWAY NO. 7	148	24	ONT4
02HC032	EAST HUMBER RIVER AT KING CREEK	95	26	ONT1
02HC033	MIMICO CREEK AT ISLINGTON	71	27	ONT4
02HD002	GANARASKA RIVER NEAR DALE	232	23	ONT4
02HD006	BOWMANVILLE CREEK AT BOWMANVILLE	82,9	35	ONT4
02HD007	SOPER CREEK AT BOWMANVILLE	78	26	ONT4
02HD008	OSHAWA CREEK AT OSHAWA	125	34	ONT4
02HD009	WILMOT CREEK NEAR NEWCASTLE	83	29	ONT1
02HE001	BLOOMFIELD CREEK AT BLOOMFIELD	13,9	21	ONT4
02HJ001	JACKSONS CREEK AT PETERBOROUGH	110	32	ONT4
02HL004	SKOOTAMATTA RIVER NEAR ACTINOLITE	671	38	ONT2
02HL005	MOIRA RIVER NEAR DELORO	308	28	ONT1
02HM004	WILTON CREEK NEAR NAPANEE	112	28	ONT1
02HM005	COLLINS CREEK NEAR KINGSTON	155	24	ONT5
02JB003	KINOJEVIS (RIVIERE) EN AVAL DE LA RIVIERE VILLEMONTÉL	1676	29	QUE1
02JB004	KINOJEVIS (RIVIERE) EN AVAL DU LAC PREISSAC	1026	32	QUE3
02JB013	KINOJEVIS (RIVIERE) A CLERICY	2609	27	QUE1
02JC008	BLANCHE RIVER ABOVE ENGLEHART	1780	26	ONT5
02LB006	CASTOR RIVER AT RUSSELL	427	46	ONT3
02LB007	SOUTH NATION RIVER AT SPENCERVILLE	246	46	ONT5
02LB008	BEAR BROOK NEAR BOURGET	440	38	ONT3
02LD005	PETITE NATION (RIVIERE DE LA) PRES DE RIPON	1330	25	QUE1
02LH002	DESERT (RIVIERE) EN AMONT DE LA RIVIERE DE L'AIGLE	1780	39	QUE1
02MC001	RAISIN RIVER NEAR WILLIAMSTOWN	358	33	ONT2
02NE007	CROCHE (RIVIERE) A LA CROCHE	1847	27	QUE1
02NE011	CROCHE (RIVIERE) EN AVAL DU RUISSEAU CHANGY	1579	29	QUE1
02OE018	HALL (RIVIERE) PRES D'EAST HEREFORD	218	44	QUE2
02OE027	EATON (RIVIERE) PRES DE LA RIVIERE SAINT-FRANCOIS-3	642	41	QUE2
02PA007	BATISCAN (RIVIERE) EN AVAL DE LA RIVIERE DES ENVIES	4400	27	QUE1
02PC009	PORTNEUF (RIVIERE) PRES DE PORTNEUF	339	27	QUE4
02PE009	GOUFFRE (RIVIERE DU) A BAIE-SAINT-PAUL	862	26	QUE2
02PG004	LOUP (RIVIERE DU) A LA ROUTE NO. 289	500	22	QUE2
02PJ007	BEAURIVAGE (RIVIERE) A SAINTE-ETIENNE	709	68	QUE4
02PJ030	FAMINE (RIVIERE) A SAINT-GEORGES	682	28	QUE4
02PL001	BECANCOUR (RIVIERE) A LYSER	1410	46	QUE1
02PL005	BECANCOUR (RIVIERE) EN AMONT DE LA RIVIERE PALMER	933	26	QUE2
02QA002	RIMOUSKI (RIVIERE)-1	1610	31	QUE1
02QB011	CAP CHAT (RIVIERE) A CAP-CHAT	725	27	QUE2
02QC001	MADELEINE (RIVIERE) A RIVIERE-LA-MADELEINE	1200	40	QUE1
02RD002	MISTASSIBI (RIVIERE)	9350	39	QUE4
02RD003	MISTASSINI (RIVIERE) EN AMON DE LA RIVIERE MISTASSIBI	9713	38	QUE4
02RF001	CHAMOUCOUANE (RIVIERE) A LA CHUTE A MICHEL	15333	40	QUE4
02RF002	CHAMOUCOUANE (RIVIERE) EN AVAL DE LA RIVIERE DU CHEF	11163	31	QUE5
02RF006	CHAMOUCOUANE (RIVIERE)	4429	29	QUE5
02UC002	MOISIE (RIVIERE)	19019	29	QUE5
02VA001	TONNERRE (RIVIERE AU)	684	36	QUE2
02VC001	ROMAINE (RIVIERE)	13028	37	QUE5
02WA001	NABISIPI (RIVIERE)	2080	25	QUE1
02XC001	SAINT-PAUL (RIVIERE) EN AVAL DE LA RIVIERE BUJEAULT	6710	25	QUE7
03AB002	WASWANAPI (RIVIERE) A LA CHUTE ROUGE	31611	26	QUE6
03AC004	BELL (RIVIERE) EN AMONT DU LAC MATAGAMI	22060	32	QUE6
03AD001	NOTTAWAY (RIVIERE) A LA TETE DU LAC SOSCUMICA	57500	21	QUE6

Code	Nom	Surface (km ²)	Nb d'année d'observation	region
03BA003	TEMISCAMIE (RIVIERE) PRES DE LAC ALBANEL	7280	24	QUE4
03BC002	RUPERT (RIVIERE DE) EN AVAL DU LAC NEMISCAU	40964	30	QUE6
03BE001	BROADBACK (RIVIERE) EN AVAL DE LA RIVIERE OUSOUAGAMI	17431	21	QUE7
03CC001	EASTMAIN (RIVIERE) A LA TETE DE LA GORGE DE BASILE	43817	20	QUE6
03DC002	GRANDE RIVIERE (LA) EN AMONT DE LA RIVIERE DE PONTOIS	37577	20	QUE6
03DD002	DE PONTOIS (RIVIERE) EN AMONT DE LA RIVIERE SAKAMI	13200	30	QUE6
03DD003	DE PONTOIS (RIVIERE) PRES DE LA GRANDE RIVIERE	19100	19	QUE6
03EA001	BALEINE (GRANDE RIVIERE DE LA) A LA SORTIE DU LAC BIENVILLE	21000	29	QUE6
03EC001	DENYS (RIVIERE) PRES DE LA GRANDE RIVIERE DE LA BALEINE	4660	31	QUE5
03ED001	BALEINE (GRANDE RIVIERE DE LA) EN AMONT DE LA RIVIERE DENYS-	35564	31	QUE7
03ED004	COATS (RIVIERE) PRES DE LA GRANDE RIVIERE DE LA BALEINE	7298	24	QUE7
03HA001	ARNAUD (PAYNE)(RIVIERE) EN AMONT DE LA RIVIERE HAMELIN-1	26900	20	QUE7
03JB001	FEUILLES (RIVIERE AUX) EN AVAL DE LA RIVIERE PELADEAU	41700	24	QUE7
03KA001	MELEZES (RIVIERE AUX) EN AMONT DE LA RIVIERE DU GUE	8831	29	QUE7
03KC004	MELEZES (RIVIERE AUX) PRES DE LA RIVIERE KOKSOAK	42538	30	QUE7
03LF002	CANIAPISCAU (RIVIERE) A LA CHUTE DE LA PYRITE	86929	31	QUE7
03MB002	BALEINE (RIVIERE A LA) PRES DE L'EMBOUCHURE-1	29800	31	QUE6
04CA002	SEVERN RIVER AT OUTLET OF MUSKRAT DAM LAKE	36500	27	ONT5
04CA003	ROSEBERRY RIVER ABOVE ROSEBERRY LAKES	619	23	ONT2
04CB001	WINDIGO RIVER ABOVE MUSKRAT DAM LAKE	10800	25	ONT5
04CE002	FAWN RIVER BELOW BIG TROUT LAKE	4350	22	ONT5
04DA001	PIPESTONE RIVER AT KARL LAKE	5960	27	ONT5
04DB001	ASHEWEIG RIVER AT STRAIGHT LAKE	7950	26	ONT5
04DC001	WINISK RIVER BELOW ASHEWEIG RIVER TRIBUTARY	50000	27	ONT5
04DC002	SHAMATTAWA RIVER AT OUTLET OF SHAMATTAWA LAKE	4710	27	ONT5
04EA001	EKWAN RIVER BELOW NORTH WASHAGAMI RIVER	10400	25	ONT5
04FA003	PINEIMUTA RIVER AT EYES LAKE	4900	26	ONT5
04FB001	ATTAWAPISKAT RIVER BELOW ATTAWAPISKAT LAKE	24200	27	ONT5
04FC001	ATTAWAPISKAT RIVER BELOW MUKETEI RIVER	36000	26	ONT5
04GA002	CAT RIVER BELOW WESLEYAN LAKE	5390	24	ONT5
04JA002	KABINAKAGAMI RIVER AT HIGHWAY NO. 11	3780	36	ONT5
04JC003	SHEKAK RIVER AT HIGHWAY NO. 11	3290	36	ONT5
04JD005	PAGWACHUAN RIVER AT HIGHWAY NO. 11	2020	26	ONT5
04JF001	LITTLE CURRENT RIVER AT PERCY LAKE	5360	26	ONT5
04LJ001	MISSINAIBI RIVER AT MATTICE	8940	74	ONT5
04MF001	NORTH FRENCH RIVER NEAR THE MOUTH	6680	27	ONT5
04NA001	HARRICANA (RIVIERE) A AMOS	3703	61	QUE6
05PA006	NAMAKAN RIVER AT OUTLET OF LAC LA CROIX	13400	72	ONT5
05PB014	TURTLE RIVER NEAR MINE CENTRE	4870	78	ONT5
05PC010	STURGEON RIVER NEAR BARWICK	168	33	ONT3
05PC011	PINEWOOD RIVER NEAR PINEWOOD	461	42	ONT3
05QA001	ENGLISH RIVER NEAR SIOUX LOOKOUT	13900	58	ONT5
05QA004	STURGEON RIVER AT MCDOUGALL MILLS	4450	32	ONT5
05QC003	TROUTLAKE RIVER BELOW BIG FALLS	2370	24	ONT5
05QE008	CEDAR RIVER BELOW WABASKANG LAKE	1690	24	ONT2
05QE009	STURGEON RIVER AT OUTLET OF SALVESEN LAKE	1530	34	ONT2

ANNEXE 7

LISTE DES BASSINS ETUDIES AU CHAPITRE 6

Echantillon « Moselle »

Code	Rivière	Localisation	S (km ²)	Dépt.
A405062	LA MOSELLE	RUPT-SUR-MOSELLE	153	88
A417301	LA CLEURIE	CLEURIE	68	88
A436203	LA VOLOGNE	CHENIMENIL (2)	355	88
A444201	LE DURBION	VAXONCOURT	145	88
A463201	L' AVIERE	FRIZON (BASSE FRIZON)	106	88
A524201	LE GITE	VELOTTE-ET-TATIGNECOURT (VELOTTE)	114	88
A525101	LE MADON	MIRECOURT	382	88
A600101	LA MEURTHE	FRAIZE	69	88
A605102	LA MEURTHE	ST-DIE	374	88
A614201	LE RABODEAU	MOYENMOUTIER	153	88
A633201	LA PLAINE	RAON-L' TAPE (LA TROUCHE)	116	88
A673122	LA MORTAGNE	GERBEVILLER	493	54
A681201	LE SANON	LAGARDE	73	57
A687201	LE SANON	DOMBASLE-SUR-MEURTHE	284	54
A690381	LE PETIT RHONE	ST-NICOLAS-DE-PORT	21,4	54
A712201	L' ESCH	JEZAINVILLE	231	54
A735201	LE RUPT DE MAD	ONVILLE	358	54
A764201	LA PETITE SEILLE	CHATEAU-SALINS	143	57
A804101	L' ORNE	BONCOURT	412	54
A812201	L' YRON	JARNY	383	54
A975201	LA NIED FRANCAISE	CONDE-NORTHEN (PONTIGNY)	499	57
A983201	LA NIED ALLEMANDE	FAULQUEMONT	187	57

Echantillon « France »

Code	Rivière	Localisation	S (km ²)	Dépt.
A273201	LA BRUCHE	WISCHES	230	67
H251311	LE THOLON	CHAMPVALLON	131	89
H572321	LE FOSSE ROGNON	DOUE (MELARCHEZ, ULTRA-SONS)	7,1	77
K064311	LA MARE	SA-MARCELLIN-EN-FOREZ (VERINES)	95,2	42
K066331	LA COISE	LARAJASSE (LE NEZEL)	61	69
K172421	LA DRAGNE	VANDENESSE	115	58
K201081	L' ALLIER	LAVEYRUNE (ROGLETON 1)	48,8	7
M711241	LA SEVRE NANTAISE	TIFFAUGES (LA MOULINETTE)	814	85
M820502	L' OGNON	LES SORINIÈRES (VILLENEUVE)	147	44
O203401	L' AUSSONNELLE	SEILH	192	31
O308432	LA MIMENTE	FLORAC	125	48
O470403	LE DADOU	PAULUNET (ST-JEAN-DE-JEANNE)	72	81
Q041401	LE LOUET	SOMBRUN	84,3	65
U100401	L' OGNON	SERVANCE (FOURGUENONS)	73,5	70
U234582	LA ROSEMONTAISE	CHAUX	25,4	90
U401401	LA REYSSOUZE	MONTAGNAT	84,4	1
U450501	L' ARDIÈRES	BEAUJEU	54,5	69
U462401	L' AZERGUES	CHATILLON	336	69
V023501	LA MENOGE	VETRAZ-MONTHOUX	160	74
V152541	L' AINAN	ST-GEOIRE-EN-VALDAINE (1)	41	38
V351561	LE RUISSEAU DES PREAUX	BOURG-ARGENTAL	22,1	42
V414521	LA GLUEYRE	GLUIRAS (TISONÈCHE)	71	7
V441401	LE ROUBION	SOYANS	186	26
V445501	LE JABRON	SOUSPIERRE	85	26
V713501	LE GARDON DE ST-JEAN	CORBES (ROC COURBE)	263	30
Y025502	L' AMPLE	REYNES (LE VILA)	47,8	66
Y461581	LE VALESCURE	COLLOBRIÈRES (AVAL)	9,22	83

ANNEXE 8

LISTE DES BASSINS ETUDIES AU CHAPITRE 7

Code	Rivière	Location	Surface (km²)	Dépt.	type de chronique
A402061	LA MOSELLE	FRESSE-SUR-MOSELLE	72	88	QJ
a405062	LA MOSELLE	RUPT-SUR-MOSELLE	153	88	QV
a415201	LA MOSELOTTE	VAGNEY (ZAINVILLERS)	187	88	QV
a417301	LA CLEURIE	CLEURIE	68	88	QV
a420063	LA MOSELLE	ST-NABORD (NOIR GUEUX)	621	88	QV
a425064	LA MOSELLE	EPINAL	1220	88	QV
a433301	LE NEUNE	LAVELINE-DEVANT-BRUYERES	94	88	QV
a436203	LA VOLOGNE	CHENIMENIL (2)	355	88	QV
a444201	LE DURBION	VAXONCOURT	145	88	QV
a524201	LE GITE	VELOTTTE-ET-TATIGNECOURT	114	88	QV
a525101	LE MADON	MIRECOURT	382	88	QV
a530301	LE COLON	XARONVAL	62,1	88	QV
A542201	LE BRENON	AUTREY	139	54	QJ
a543101	LE MADON	PULLIGNY	940	54	QV
a605102	LA MEURTHE	ST-DIE	374	88	QV
a614201	LE RABODEAU	MOYENMOUTIER	153	88	QV
a620103	LA MEURTHE	RAON-L' TAPE	727	88	QV
A622101	LA MEURTHE	AZERAILLES	960	54	QJ
a633201	LA PLAINE	RAON-L' TAPE (LA TROUCHE)	116	88	QV
A644301	LE VACON	BARBAS	35,4	54	QJ
A654301	LE RUISSEAU DES AMIS	MARAINVILLER	35,6	54	QJ
a656111	LA VEZOUZE	LUNEVILLE	559	54	QV
A662121	LA MORTAGNE	STE-HELENE	100	88	QJ
a673122	LA MORTAGNE	GERBEVILLER	493	54	QV
a681201	LE SANON	LAGARDE	73	57	QV
a687201	LE SANON	DOMBASLE-SUR-MEURTHE	284	54	QV
A695301	L' AMEZULE	LAY-ST-CHRISTOPHE	84,8	54	QJ
a721301	LE RUISSEAU DU TREY	VANDIERES	38,7	54	QV
a735201	LE RUPT DE MAD	ONVILLE	358	54	QV
a764201	LA PETITE SEILLE	CHATEAU-SALINS	143	57	QV
a782101	LA SEILLE	NOMENY	925	54	QV
a788101	LA SEILLE	METZ	1280	57	QV
a800621	LE RUISSEAU DE VAUX	MORGEMOULIN	42,3	55	QV
a804101	L' ORNE	BONCOURT	412	54	QV
a812201	L' YRON	JARNY	383	54	QV
a843101	L' ORNE	ROSSELANGE	1226	57	QV
A873201	LA CANNER	KOENIGSMACKER	110	57	QJ
a885301	LE RUISSEAU DE MONTENACH	SIERCK-LES-BAINS	47	57	QV
a900105	LA SARRE BLANCHE	LANEUVEVILLE-LES-LORQUIN	64	57	QV
a901305	LA SARRE ROUGE	VASPERVILLER	90	57	QV
a907205	L' ISCH	POSTROFF	145	57	QV
a909105	LA SARRE	KESKASTEL	878	67	QV
a909106	LA SARRE	DIEDENDORF	721	67	QV
a913205	L' ALBE	SARRALBE (RECH)	407	57	QV
a973731	L' ELVON	BAZONCOURT	16,2	57	QV
a975201	LA NIED FRANCAISE	CONDE-NORTHEN (PONTIGNY)	499	57	QV
a994201	LA NIED	BOUZONVILLE	1160	57	QV
alsd	?	?	263	Allemagne	QJ
hent	?	?	103	Allemagne	QJ
kord	?	?	817	Allemagne	QJ
papi	?	?	170	Allemagne	QJ
pelt	?	?	173	Allemagne	QJ
plat	?	?	380	Allemagne	QJ
przu	?	?	577	Allemagne	QJ

