

Higher National School of Hydraulic

The Library

Digital Repository of ENSH



المدرسة الوطنية العليا للري

المكتبة

المستودع الرقمي للمدرسة العليا للري



The title (العنوان):

Construction locale et régionale des courbes intensité durée-  
fréquence basée sur l'invariance d'échelle..

The paper document Shelf mark (الشفرة) : 8-0003-20

APA Citation ( APA توثيق ):

Boukhelifa, Mustapha. (2020). Construction locale et régionale des courbes intensité  
durée-fréquence basée sur l'invariance d'échelle. [Thèse de doctorat, ENSH].

The digital repository of the Higher National School for Hydraulics "Digital Repository of ENSH" is a platform for valuing the scientific production of the school's teachers and researchers.

Digital Repository of ENSH aims to limit scientific production, whether published or unpublished (theses, pedagogical publications, periodical articles, books...) and broadcasting it online.

Digital Repository of ENSH is built on the open DSpace software platform and is managed by the Library of the National Higher School for Hydraulics. <http://dspace.ensh.dz/jspui/>

المستودع الرقمي للمدرسة الوطنية العليا للري هو منصة خاصة بنمّين الإنتاج العلمي لأساتذة و  
باحثي المدرسة.

يهدف المستودع الرقمي للمدرسة إلى حصر الإنتاج العلمي سواء كان منشورا أو غير منشور  
(أطروحات، مطبوعات، مباحثات، مقالات البوريات، كتب....) و بثه على الخط.

المستودع الرقمي للمدرسة مبني على المنصة المفتوحة DSpace و يتم إدارته من طرف مديرية  
المكتبة للمدرسة العليا للري.

كل الحقوق محفوظة للمدرسة الوطنية العليا للري.



**ENSH-(n° d'ordre)**

**THESE DE DOCTORAT  
DE L'ECOLE NATIONALE SUPERIEURE D'HYDRAULIQUE**

Présentée par

Monsieur **BOUKHELIFA Mustapha**

**Pour obtenir le grade de  
DOCTEUR DE L'ECOLE NATIONALE SUPERIEURE D'HYDRAULIQUE**

Sujet de la thèse :

**CONSTRUCTION LOCALE ET RÉGIONALE DES COURBES INTENSITÉ-  
DURÉE-FRÉQUENCE BASÉE SUR L'INVARIANCE D'ÉCHELLE :  
APPLICATION À LA RÉGION CENTRE DU NORD ALGÉRIEN**

Thèse présentée et soutenue à Blida le 27/10/2020 devant le jury composé de :

M. SALAH Boualem	Professeur, ENSH Blida	Président
M. MEDDI Mohamed	Professeur, ENSH Blida	Directeur de thèse
M. GAUME Éric	HDR, Univ. G. Eiffel, Nantes, France	Co-directeur de thèse
M. AMMARI Abdelhadi	MCA, ENSH Blida	Examineur
M. BERMAD Abdelmalek	Professeur, ENP El Harrach	Examineur
M. HACHAMA Mohamed	Professeur, Univ. Khemis-Miliana	Examineur
M <sup>me</sup> . BENMAMAR Saadia	Professeur, ENP El Harrach	Invitée



## Remerciements

Je tiens à remercier chaleureusement et à exprimer toute ma gratitude à mes deux directeurs de thèse de doctorat, messieurs MEDDI Mohamed et Eric GAUME. J'ai toujours trouvé dans votre encadrement de la bienveillance, de l'enthousiasme, du dynamisme, des conseils et orientations avisés et plus que tout un accompagnement scientifique (à mes yeux) sans nul autre pareil. A quelques jours de la finalisation de la rédaction de ce manuscrit de thèse, un ancien directeur de l'Ecole Nationale Supérieure d'Hydraulique de Blida me disait ceci : « presque dix années ont passé entre ta première inscription en thèse de doctorat et son aboutissement, car tu as manqué de chance, c'est tout, la vie est ainsi ». En travaillant sous votre direction, cette "malchance" au début de ma carrière d'enseignant-chercheur, Chikh Meddi et Eric vous m'avez aidé, chacun de vous d'une manière bien différente, à la surmonter. Vous avez mon éternelle reconnaissance. Encore une fois, merci à vous du fond de mon être. Aussi, un remerciement affectueux à mon épouse Asma TAIBI pour sa relecture assidue du manuscrit de thèse et ses encouragements incessants.

J'adresse également mes sincères remerciements aux membres du jury qui me font l'honneur d'évaluer ces travaux : au Professeur Salah BOUALEM en sa qualité de président (que je considérerai toujours comme mon honorable enseignant), à mon cher confrère de l'ENSH Docteur Abdelhadi AMMARI, aux professeurs BENMAMAR Saadia et BERMAD Abdelmalek de l'Ecole Nationale Polytechnique et au professeur HACHAMA Mohamed de l'université Djilali Bounaama Khemis-Miliana. Aussi, un grand merci au Professeur MANSOURI Rachid de l'université de Guelma pour sa précieuse aide.

Mes remerciements vont aussi à toute l'équipe du laboratoire Eau et Environnement du département GERS de l'IFSTTAR (actuellement rebaptisé en université Gustave Eiffel) de Nantes de m'avoir accueilli pendant une année dans le cadre du programme national exceptionnel de bourses de recherches Algériennes. Particulièrement, Olivier Payrastre et François Bourgin pour leurs critiques constructives, lors des présentations des résultats préliminaires, qui ont permis d'améliorer ce travail. Sans oublier tous les doctorants et stagiaires avec qui j'ai échangé de tous et de rien pendant une centaine d'heures de transport en tramway/bus de Nantes à Bouguenais et vice versa...

J'adresse par ailleurs mes remerciements à l'Agence Nationale des Ressources hydrauliques (ANRH) pour avoir mis à disposition les données pluviographiques utilisées dans cette thèse, particulièrement Amine KHELFI et Hamid ALI. Mes remerciements vont également à feu Djamel BOUCHERF (رحمه الله) de l'ONM pour son aide.

Enfin, mes vifs remerciements aux collègues de l'ENSH qui m'ont encouragé et soutenu, surtout le fameux « kemelt wella mazal » (Je pense à toi Dr Ayoub Zeroual). Aussi à tout membre du personnel de soutien de l'ENSH qui a contribué de près ou de loin à l'encadrement administratif de cette thèse, particulièrement à la directrice du laboratoire Génie de l'eau et de l'environnement (GEE) Docteur MEDDI Hind qui a toujours su faciliter les choses.

## Dédicaces

Je dédie cette infime contribution à la recherche scientifique, à mon épouse Asma et mes chers enfants Maya, Dania, Edris et petit ange qui ont énormément donné de leur temps (et ce n'est rien de le dire : vacances, weekend, fêtes...) au profit de mes interminables heures de simulation et de rédaction. Je ne te remercierai jamais assez Asma pour ton soutien indéfectible, tes encouragements répétés, ton aide permanente, ton accompagnement et les sacrifices que tu as consenti. Je te suis à jamais reconnaissant.

Je dédie également ce travail à mes chers parents Hamid et Fatiha qui m'ont soutenu dans tous les sens du terme au cours de mon parcours scolaire et universitaire. Merci de vous être préoccupés de mes études de la sixième, du BEF, du BAC, de l'ingénieur d'état, du Magister et enfin du doctorat es science (celui la fut plus long - même lent- que la norme) avec les mêmes égards et attentions bien que j'eusse dépassé les 25 ans au cours des deux derniers diplômes. Vous avez mon éternelle gratitude.

Enfin, et comment oublier tous les membres de ma famille, je dédie ce travail à mon frère Sami et sa petite tribu qui s'enquiert de nous de l'autre côté de la Méditerranée et de leur aide pendant mon année de bourse, à mes grands-parents Ma el Hadja, Abi et Mazola qui se réjouissent toujours de savoir que j'ai encore obtenu un diplôme, et tous les autres que je ne cite pas, qu'ils ne m'en tiennent pas rigueur.

En dernier mais non des moindres, à mes beaux-parents Mourad et Zouleikha pour leur soutien permanent, ainsi que l'ensemble de ma belle-famille, Maçaab, Khouloud, Mouad, Lynda, Mohamed, Yasmine, et particulièrement Meriem, Assia et Sarya qui prennent soin de mes enfants au moment même où je rédige cette dédicace : merci pour toutes les autres fois aussi.

## الملخص

تمثل منحنيات الشدة - المدة - التردد (IDF) أحد النماذج الإحصائية حول الأمطار الأكثر استخدامًا في الهندسة الهيدرولوجية وفي مشاريع تحليل المخاطر. ومع ذلك، نادرًا ما تم تقييم الإرتيابات المتعلقة بتطور منحنيات الشدة - المدة - التردد في الماضي. يمكن أن تكون هذه الإرتيابات كبيرة للغاية، لا سيما في مناطق العالم حيث لا تزال سلاسل تسجيلات الأمطار الموثوقة محدودة العدد وغالبًا ما تكون قصيرة جدًا كذلك. للاختيار المناسب يمكن، سواء من خلال صياغة منحنيات الشدة - المدة - التردد أو طرق الاستدلال الإحصائي، أن يقلل بشكل كبير من هذه الشكوك كما هو موضح في هذه الأطروحة.

في هذا العمل، سيتم تسليط الضوء على العلاقة بين صياغة منحنيات الشدة - المدة - التردد وبعض الخصائص، التي لوحظت في كثير من الأحيان، من سلاسل قياس الأمطار، وهي الثبات السلمي البسيط والبنى متعددة الكسور. إذا افترضنا أن هذه الخصائص صحيحة: إذن، فإن صياغة منحنيات الشدة - المدة - التردد هي نتاج دالة توزيع احتمالية بلا أبعاد لتراكمات/شدة الحد الأقصى من الأمطار السنوية وعامل المقياس المرتبط بالمدة. وبالتالي، يمكن تقييم معلمته بطريقة متكاملة (أي على أساس عينة مشتركة واحدة من سلاسل تراكمات/شدة الحد الأقصى من الأمطار السنوية على مدى فترات تتراوح من 15 دقيقة إلى 24 ساعة).

يوفر استخدام طريقة الاستدلال الإحصائي، استنادًا إلى الاحتمالية ونهج Bayesian Markov Chain Monte Carlo (MCMC)، في تقييم الإرتيابات تقديرًا ثابتًا لهذه الأخيرة حول جميع معلمات العلاقة الشدة - المدة - التردد، وكذلك كميات الأمطار المقابلة. تم اختبار هذه المنهجية على مجموعة بيانات كبيرة إلى حد ما من 48 محطة مطر، موزعة على الجزء المركزي الأوسط من الجزائر (25000 كيلومتر مربع)، في ظل أنظمة مناخية مختلفة. النهج المتكامل متسق بلا شك مع التقديرات الناتجة من تعديلات الحد الأقصى من الأمطار السنوية لكل فترة على حدة. بالإضافة إلى ذلك، يتم تقليل فاصل الثقة بشكل كبير.

يتوضح كذلك أن هذا النهج المتكامل قوي جدًا: بمعنى أنه، على عكس الطريقة التقليدية المستندة إلى التقديرات لكل مدة على حدة، فإنه يوفر تقديرات عقلانية للكميات، حتى عندما تكون سلاسل القياس المتوفرة قصيرة جدًا (10 سنوات). هذا يمثل ميزة معتبرة ووفرة نوعية للتطبيقات الهندسية. في خطوة ثانية، هذا التقدير البايزي القوي والمتكامل لمنحنيات الشدة - المدة - التردد المقترحة هنا يخضع للتمديد من أجل تحديد جهوي لكميات الأمطار القصوى. يجمع هذا النهج الأخير بين طرق الاندماج والتحليلات المحلية والجهوية للحد الأقصى من الأمطار السنوية. بعد ذلك، يتم تقييم تأثير هذا الإجراء المقترح على الإرتيابات من كميات الأمطار التي تم مكانيا استيفاءها بواسطة Krigeage.

يتم تعيين الحد الأعلى لمجال المصدقية بنسبة 70٪ لكميات المطر المختلفة (أي على مدار فترات مختلفة) الممثل من خلال الجمع بين الإرتيابات في التقدير المحلي والاستيفاء المكاني. نظرًا لانخفاض الإرتيابات في تقديرات هطول الأمطار المحلية والاتساق المكاني الكبير للقيم المقدرة، يبدو أن النهج المقترح يتجاوز بشكل كبير النهج التقليدي حتى بالنسبة للتقديرات في المواقع الغير-مقاسة.

## Résumé

Les courbes Intensité-Durée-Fréquence (IDF) représentent un des modèles statistiques, sur les pluies, parmi les plus couramment utilisés en ingénierie hydrologie et dans les projets d'analyse du risque. Néanmoins, les incertitudes reliées à l'élaboration de ces courbes IDF, ont été rarement évaluées dans le passé. Ces incertitudes peuvent être très conséquentes, particulièrement dans les régions du monde où les séries des enregistrements pluviographiques fiables sont encore limitées en nombre et bien souvent assez courtes. La sélection appropriée, aussi bien de la formulation des courbes IDF que des méthodes d'inférence statistique, peut significativement réduire ces incertitudes tel qu'illustré ici. Dans ce travail, il sera mis en avant le lien existant entre la formulation des courbes IDF et quelques propriétés, souvent observées, des séries pluviométriques à savoir l'invariance d'échelle simple et les structures multifractales. Si l'on assume que ces propriétés sont valides : alors, la formulation des courbes IDF est le produit d'une fonction de distribution de probabilités adimensionnelle des cumuls/intensités des pluies maximales annuelles et d'un facteur d'échelle dépendant de la durée. Ainsi, ses paramètres peuvent être évalués de manière intégrée (i.e. sur la base d'un échantillon conjoint unique des séries des cumuls/intensités de pluies maximales annuelles sur une gamme de durées allant de 15 minutes à 24 heures). L'utilisation d'une méthode d'inférence statistique basée sur la vraisemblance et une approche Bayésienne Markov Chain Monte Carlo (MCMC), fournit une estimation consistante des incertitudes sur tous les paramètres de la relation IDF, ainsi que pour les quantiles de pluies correspondants. Cette méthodologie a été testée sur un jeu de données assez large de 48 postes pluviographiques, répartis sur la partie centrale Nord de l'Algérie (25000 km<sup>2</sup>), sous différents régimes climatiques. L'approche intégrée est indubitablement cohérente avec les estimations issues d'ajustements des pluies maximales annuelles pour chacune des durées séparément. En outre, les intervalles de crédibilité sont significativement réduits. De plus, cette approche intégrée apparaît comme étant très robuste : dans le sens où contrairement à la méthode traditionnelle, basée sur des estimations pour chaque durée séparément, elle fournit des estimations raisonnables des quantiles, même quand les séries d'observation disponibles sont assez courtes (10 années). Ceci représente un avantage significatif pour des applications d'ingénierie. Dans une deuxième étape, cette robuste estimation Bayésienne intégrée des courbes IDF proposée ici est étendue à une détermination régionale des quantiles sur les extrêmes pluviométriques. Cette dernière approche combine entre des méthodes de fusion et d'analyses locales et régionales sur les pluies maximales annuelles. Par la suite, l'impact de cette procédure proposée, sur les incertitudes des quantiles de pluies spatialement interpolés par Krigeage, est évalué. La borne supérieure de l'intervalle de crédibilité à 70% pour différents quantiles de pluie (i.e. sur diverses durées) est cartographiée en fusionnant les incertitudes d'estimation locale et d'interpolation spatiale. En raison des faibles incertitudes sur les estimations locales des pluies et de la grande cohérence spatiale des valeurs estimées, l'approche proposée paraît significativement surpasser l'approche traditionnelle même pour des estimations sur les sites non-jaugés.

## Abstract

Intensity-duration-frequency (IDF) curves are one of the most common rainfall statistical models used in hydrologic design and risk analysis projects. The uncertainties related to the elaboration of these IDF curves have nevertheless seldom been evaluated in the past. These uncertainties may be large, especially in the regions of the world where reliable recorded pluviographic series are still limited in number and often short. An appropriate selection of the IDF formulation as well as of the statistical inference methods may reduce significantly these uncertainties as will be illustrated herein. This work will recall the existing link between the IDF formulation and some properties of the rainfall series that are often observed such as simple scaling and multifractal structure. A simple formulation for the IDF curves can be derived if these properties are assumed to be valid: the IDF formulation is then the product of a dimensionless probability distribution function for the annual maximum rainfall intensities/depths and a duration-dependent scaling factor. Its parameters can be evaluated in an integrated way (i.e. based on a unique pooled sample of rainfall annual maximum depths/intensities over a range of durations: from 15 minutes to 24 hours). The use of likelihood-based Bayesian Markov Chain Monte-Carlo (MCMC) statistical inference methods for this evaluation provides consistent uncertainties for all the parameters of the IDF relation and for the corresponding rainfall quantiles. This methodology has been tested on a large dataset of 48 rain-gauge records, spread over the North Central part of Algeria (25000 km<sup>2</sup>), under various climatic regimes. The integrated approach is undoubtedly consistent with estimates from annual maximum rainfall fitted to single durations. Furthermore, credibility intervals are significantly reduced. Also, this integrated approach appears to be robust: unlike the traditional method based single durations, it generally provides reasonable quantile estimates, even when short observed series (10 years) are available. This is a significant advantage for engineering applications. In a second step, the robust integrated Bayesian estimation of IDF curves proposed herein is extended to regional extreme rainfall return level estimation. This latter approach combines local and regional merging and analysis methods for rainfall annual maximum series. Then, the impact of the proposed procedure on the uncertainties of spatially interpolated rainfall quantiles based on kriging is evaluated. Upper bounds of the 70% credibility intervals for various rain-rate quantiles are mapped, merging local estimation and spatial interpolation uncertainties. Due to lower point rainfall estimation uncertainties and higher spatial consistency of the estimated values, the proposed approach appears to significantly outperform the traditional approach also if implemented to provide estimates at ungauged sites.



# Tables des matières

Remerciements	iii
Dédicaces	iv
الملخص	v
Résumé	vi
Abstract	vii
Table des matières	viii
Table des figures	xi
Liste des tableaux	xiv

---

<b>1. Chapitre 1. Introduction générale</b>	<b>1</b>
<b>1.1 Contexte local, problématique et motivation</b>	<b>2</b>
<b>1.2 Objectifs de la thèse</b>	<b>5</b>
<b>1.3 Vue d'ensemble</b>	<b>5</b>
<b>1.4 Organisation du mémoire</b>	<b>10</b>
<b>2. Chapitre 2 Les extrêmes pluviométriques : compréhension, méthodes d'analyse et utilisation</b>	<b>12</b>
<b>2.1 Introduction</b>	<b>13</b>
<b>2.2 Qu'est-ce que les précipitations extrêmes ?</b>	<b>13</b>
<b>2.3 Procédures d'échantillonnage des extrêmes pluviométriques</b>	<b>14</b>
2.3.1 La mesure des précipitations	14
2.3.2 Les séries des maxima annuels (SMA)	22
2.3.3 Les séries de durées partielles (SDP)	23
2.3.4 SMA vs SDP : Quelle approche choisir ?	23
2.3.5 Conclusion	25
<b>2.4 Modélisation statistique des valeurs extrêmes des précipitations</b>	<b>25</b>
2.4.1 Approche par analyse fréquentielle classique	26
2.4.2 Perspective Bayésienne et procédure Markov Chain Monte Carlo (MCMC)	32
2.4.3 Approches alternatives	36
2.4.4 Vers l'hégémonie de l'inférence fréquentielle Bayésienne	41
2.4.5 La distribution GEV : Le choix adéquat	41
2.4.6 Contrôle des ajustements : tests d'adéquation	45
2.4.7 Evaluation des incertitudes d'estimation	46
2.4.8 Analyse fréquentielle régionale des précipitations extrêmes	50

2.4.9	Stationnarité vs non-stationnarité .....	55
<b>2.5</b>	<b>Invariances d'échelle simple et propriétés multifractales des extrêmes pluviométriques .....</b>	<b>56</b>
2.5.1	Qu'est-ce que l'invariance d'échelle simple (Simple Scaling) .....	57
2.5.2	Caractère multifractal des séries des maxima annuels sur diverses durées.....	58
2.5.3	Invariance d'échelle simple et multifractalité dans la modélisation des précipitations extrêmes 59	
2.5.4	Invariance d'échelle simple vs invariance d'échelle multiple .....	60
<b>2.6</b>	<b>Cartographie des précipitations extrêmes .....</b>	<b>60</b>
2.6.1	Interpolation spatiale par Krigeage.....	61
2.6.2	Analyse fréquentielle locale vs régionale .....	63
2.6.3	Méthode par estimation régionale des quantiles des pluies .....	64
2.6.4	Méthode par interpolation des quantiles de pluie au site .....	65
2.6.5	Analyses multivariées dans la spatialisation des précipitations .....	65
2.6.6	Quelle approche retenir dans la cartographie des pluies extrêmes ?.....	66
<b>2.7</b>	<b>Caractérisation du risque : période de retour, probabilité de défaillance et précision d'estimation. ....</b>	<b>67</b>
2.7.1	Période de retour vs. Probabilité de défaillance.....	67
2.7.2	Précision et incertitude sur les estimations de quantiles .....	68
<b>2.8</b>	<b>Courbes IDF : Outil de synthèse des extrêmes pluviométriques .....</b>	<b>70</b>
2.8.1	Utilisation en analyse hydrologique : Conception des systèmes de gestions des eaux de pluie.	71
2.8.2	Formulation empirique des courbes IDF .....	72
2.8.3	Formulation semi-empirique des courbes IDF .....	73
2.8.4	Formulation mathématique des courbes IDF .....	74
<b>2.9</b>	<b>Conclusion du chapitre 2.....</b>	<b>77</b>
<b>3.</b>	<b>Chapitre 3 Description de la zone d'étude et données utilisées .....</b>	<b>79</b>
<b>3.1</b>	<b>Introduction.....</b>	<b>80</b>
<b>3.2</b>	<b>Situation géographique de la zone d'étude .....</b>	<b>80</b>
<b>3.3</b>	<b>Importance démographique et socio-économique.....</b>	<b>81</b>
<b>3.4</b>	<b>Synthèse hydrologique.....</b>	<b>81</b>
<b>3.5</b>	<b>Données utilisées : enregistrements pluviographiques.....</b>	<b>82</b>
3.5.1	Sélection des séries d'observations.....	83
3.5.2	Extraction et validation des séries des maxima annuels de pluies .....	84
<b>3.6</b>	<b>Climatologie des régions centrales du nord de l'Algérie .....</b>	<b>85</b>
3.6.1	Précipitations moyennes et maximales annuelles .....	85
3.6.2	Effet de la topographie sur les précipitations maximales annuelles .....	88
<b>3.7</b>	<b>Quelques caractéristiques statistiques.....</b>	<b>91</b>
<b>3.8</b>	<b>Conclusion chapitre 3 .....</b>	<b>95</b>
<b>4.</b>	<b>Chapitre 4 Etude des structures probabilistes des séries des maxima annuels .....</b>	<b>97</b>
<b>4.1</b>	<b>Introduction.....</b>	<b>98</b>
<b>4.2</b>	<b>Vérification locale de l'invariance d'échelle simple et des structures fractales.....</b>	<b>98</b>
4.2.1	Analyse d'échelles des moments .....	98
4.2.2	Validation de la relation d'exposant d'échelle.....	101
4.2.3	Vérification supplémentaire sur l'applicabilité de l'invariance d'échelle simple .....	103
<b>4.3</b>	<b>Vérification de l'homogénéité régionale des séries des maxima annuels de pluies.....</b>	<b>104</b>
4.3.1	Délimitation initiale des régions homogènes .....	105
4.3.2	Tests de discordance et d'homogénéité HW (intégration spatiale).....	106
4.3.3	Tests de discordance et d'homogénéité HW (intégration spatio-temporelle).....	107

4.3.4	Analyse supplémentaire sur l'invariance d'échelle spatio-temporelle.....	110
<b>4.4</b>	<b>Conclusion du chapitre 4.....</b>	<b>111</b>
<b>5.</b>	<b>Chapitre 5 Etude des Estimations Bayésiennes intégrées des courbes Intensités-Durées-Fréquences (IDF).....</b>	<b>113</b>
<b>5.1</b>	<b>Introduction.....</b>	<b>114</b>
<b>5.2</b>	<b>Formulation Bayésienne intégrée des courbes IDF.....</b>	<b>115</b>
<b>5.3</b>	<b>Méthodologie d'évaluation des performances.....</b>	<b>117</b>
<b>5.4</b>	<b>Calage local du modèle IDF Bayésien Intégré.....</b>	<b>118</b>
<b>5.5</b>	<b>Robustesse envers les courtes séries d'observations.....</b>	<b>125</b>
<b>5.6</b>	<b>Sensibilité des intervalles de crédibilité à la dépendance intra-site des séries temporelles.....</b>	<b>127</b>
<b>5.7</b>	<b>Quelques discussions et conclusion.....</b>	<b>128</b>
<b>6.</b>	<b>Chapitre 6 Cartographie des courbes IDF : de l'approche locale vers l'analyse régionale.....</b>	<b>131</b>
<b>6.1</b>	<b>Introduction.....</b>	<b>132</b>
<b>6.2</b>	<b>Cartographie des quantiles de pluies à partir d'estimation locale.....</b>	<b>133</b>
6.2.1	Analyse des variogrammes expérimentaux vs. ajustés (M0 et M1).....	133
6.2.2	Interpolation spatiale des quantiles aux sites : Modèle M0 vs. Modèle M1.....	134
6.2.3	Intégration des incertitudes sur les quantiles et sur l'interpolation dans la cartographie des courbes IDF : modèle M0 vs modèle M1.....	137
<b>6.3</b>	<b>Cartographie des quantiles de pluies à partir d'estimation régionale.....</b>	<b>138</b>
6.3.1	Modèle M2 : Estimation Bayésienne régionale des courbes IDF.....	139
6.3.2	Modèle M3 : Estimation Bayésienne intégrée régionale des courbes IDF.....	140
6.3.3	Intégration des incertitudes sur les quantiles et sur l'interpolation dans la cartographie des courbes IDF : modèle M2 vs modèle M3.....	141
<b>6.4</b>	<b>Evaluation et validation des modèles de cartographie des courbes IDF.....</b>	<b>142</b>
<b>6.5</b>	<b>Conclusion du chapitre 6.....</b>	<b>144</b>
<b>7.</b>	<b>Conclusion générale.....</b>	<b>146</b>
<b>7.1</b>	<b>Résultats obtenus.....</b>	<b>148</b>
<b>7.2</b>	<b>Quelques perspectives.....</b>	<b>149</b>
<b>8.</b>	<b>Liste des références bibliographiques.....</b>	<b>153</b>
<b>9.</b>	<b>Annexes.....</b>	<b>170</b>

## Tables des Figures

Figure 1.1 Schématisation de la fractalité des pluies (invariance d'échelle au sens large) en fonction de l'exposant d'échelle $k(q)$ et de l'ordre $q$ des moments statistiques.....	7
Figure 1.2 Principe de la cartographie des précipitations extrêmes proposée dans ce travail. ....	9
Figure 2.1 Reconstitution du hyétogramme du 09/11/2001 au 10/11/2001 à Bouzaréah (Alger) (ISL-BRGM, 2005).....	18
Figure 2.2 Boxplot de l'erreur relative sur l'estimation des quantiles 100 ans pour trois méthodes standard d'inférence. Un total de 1000 séries tirées dans une distribution parente GEV (10, 10, -0.2). A gauche $N=30$ ans et à droite $N=100$ ans. ....	33
Figure 2.3 Les 6 premiers niveaux de : (a) l'ensemble de Cantor (1984) et du triangle de Sierpinski (1915) (Panthou, 2013). ....	58
Figure 2.4 Schématisation de l'analyse d'échelle des moments : (à gauche) Log-log linéarité entre les moments et les durées, (à droite) Relation d'exposant d'échelle (Panthou et al., 2014) .....	59
Figure 2.5 Exemple d'un modèle sphérique ajusté à semivariogramme expérimental.....	62
Figure 2.6 Distribution des fréquences de dépassement de quantiles centennaux de 2000 échantillons de 30 valeurs tirés dans une distribution GEV de référence (station 020506 de Bouzeréah) .....	68
Figure 2.7 Illustration de la distribution prédictive (en rouge) avec la distribution GEV de référence ajustée par Bayesian MCMC (en noir) et par maximum de vraisemblance classique (bleu). Les intervalles de crédibilité à 70% (en gris) sont déduits de la distribution à posteriori de Bayesian MCMC. ....	69
Figure 2.8 Courbes IDF modélisées par formulation semi-empirique ( Boukhelifa et al., 2011).....	70
Figure 3.1 Zone d'étude, topographie et postes pluviographiques (PG) .....	80
Figure 3.2 Extrait de la carte du réseau d'observations hydrologiques de l'ANRH.....	81
Figure 3.3 Période de fonctionnement des stations pluviographiques de la région centre Nord de l'Algérie. ....	84
Figure 3.4 Totaux pluviométriques annuels du centre Nord de l'Algérie .....	86
Figure 3.5 Totaux pluviométriques pour des durées de : (a) 15 minutes, (b) 60 minutes et 1440 minutes pour 49 stations PG dans le centre nord de l'Algérie.....	87
Figure 3.6 Diagramme de dispersion de quelques caractéristiques géographiques et statistiques des stations pluviographiques du Centre Nord de l'Algérie.....	89
Figure 3.7 Diagramme de dispersion des cumuls maximum annuels moyen sur 4 durées d'agrégation (15, 60, 180 et 1440 minutes) pour 49 stations PG du centre Nord de l'Algérie. ....	90
Figure 3.8 Graphique des cumuls de pluies maximum vs. moyens pour les 49 SMA sur 8 durées d'agrégation .....	91
Figure 3.9 Graphique du L-coefficient d'asymétrie vs. L-coefficient de variation sur 8 durées d'agrégation pour 49 stations PG dans le centre Nord de l'Algérie. ....	93
Figure 3.10 Diagramme des L-moments pour les SMA de 49 PG du centre Nord de l'Algérie. ....	94
Figure 4.1 (à droite) : Régression log-log linéaire entre les moments d'ordre $q$ et les durées d'agrégation $d$ . (à gauche) Ajustement d'une régression linéaire à la fonction d'exposant d'échelle $k(q)$ et les ordres $q$ des moments. (a) et (b) pour la station 33. (c) et (d) pour la station 30. ....	99
Figure 4.2 Valeurs des coefficients de détermination $R^2$ entre $\log EY_d$ et $\log d$ , pour 48 stations pluviographiques. Avec $q=(0.5, 1, 1.5, 2, 2.5, 3)$ . (à gauche) : Cas A (15 min – 24h). (à droite) : Cas B (1h – 24h).....	101

Figure 4.3 Evolution des médianes de cumuls de pluies avec les durées (sur une échelle log-log). Points noirs : Valeurs empiriques des médianes pour chacune des séries de durées d. Trait en gras : Médianes théoriques calculées par le modèle IDF Bayésien Intégré. Traits fins : Intervalles de crédibilités à 80% sur les médianes.....	102
Figure 4.4 Délimitation initiale des Régions pluviométriques homogènes .....	106
Figure 4.5 Carte des régions et sous-région homogènes du centre Nord de l'Algérie.....	109
Figure 4.6 MSA en considérant le modèle M3A (intégration spatio-temporelle de 15minutes à 24heures). (a) Sous-région A et (b) sous-région D .....	110
Figure 5.1 Quantiles 100 ans des cumuls de pluie : Model M0 vs. Model M1 .....	121
Figure 5.2 Les Boxplots représentent la distribution Postérieure (Bayesian MCMC) sur l'exposant d'échelle b estimé pour les cas A et B. Carrés Bleu (Estimation par Maximum de vraisemblance Bayesian MCMC). Carrés rouges (estimation par MSA).....	122
Figure 5.3 Séries des maxima annuels de cumuls de pluie ajustés à une distribution GEV. (à gauche : a, b, c, d) : Pour le cas A (15 min – 24h), (à droite : e, f, g, h) : Pour le cas B (60 min – 24 h). Les croix pour les observations. Traits fins : modèle M0. Trait gras : Modèle M1. Trait en tiret : intervalles de crédibilité. ....	124
Figure 5.4 Séries des cumuls de pluies maximales annuelles ajustées avec une distribution GEV. Les croix pour les observations. Traits fins : modèle M0. Trait gras : Modèle M1. Trait en tiret: intervalles de crédibilité.....	125
Figure 5.5 Boxplots de l'erreur relative entre quantiles de références (T=10ans) estimés par les deux modèles M0 et M1 et les quantiles basés sur des séries de 10 années de maxima annuels de pluies tirées aléatoirement. (à gauche) : méthode M0, (au milieu) : méthode M1 cas 2, (à droite) : méthode M1 cas 2. ....	126
Figure 5.6 Séries des cumuls de pluies maximales annuelles ajustées avec une distribution GEV. Les croix pour les observations. Traits fins : modèle M0. Trait gras : Modèle M1. Trait en tiret : intervalles de crédibilité à 90%. Trait gras en gris montre les résultats obtenus avec la méthode M1 calée sur seulement 3 durées (15min, 60 min. et 24h).....	128
Figure 6.1 Boxplot représentent la distribution postérieure (Bayesian MCMC) sur l'exposant d'échelle b estimé pour le cas B. Carrés Bleu (Estimation par Maximum de vraisemblance Bayesian MCMC). Carrés rouges (estimation par MSA). Carrés Verts (estimation empirique ANRH – classique en trois étapes).....	132
Figure 6.2 Graphiques des variogrammes expérimentaux vs ajustés .....	134
Figure 6.3 Cartographie des quantiles 10 ans sur d=1440 minutes estimés par la méthode M0. ....	135
Figure 6.4 Cartographie des quantiles 10 ans sur d=1440 minutes estimés par la méthode M1. ....	135
Figure 6.5 Variances d'interpolation sur les quantiles 10 ans pour d=1440 minutes estimés par la méthode M0. ....	136
Figure 6.6 Variances d'interpolation sur les quantiles 10 ans pour d=1440 minutes estimés par la méthode M1. ....	136
Figure 6.7 Cartographie des bornes supérieures des ICs à 70% combinées aux incertitudes d'interpolation. Période de retour 10ans pour d=1440 minutes estimés par la méthode M0.....	137
Figure 6.8 Cartographie des bornes supérieures des ICs à 70% combinées aux incertitudes d'interpolation. Période de retour 10ans pour d=1440 minutes estimés par la méthode M1.....	137
Figure 6.9 Cartographie des quantiles 10 ans sur d=1440 minutes estimés par la méthode M2. ....	139

Figure 6.10	Variances d'interpolation sur les quantiles 10 ans pour d=1440 minutes estimés par la méthode M2.....	140
Figure 6.11	Cartographie des quantiles 10 ans sur d=1440 minutes estimés par la méthode M3. Sous-région A (région 2). .....	140
Figure 6.12	Incertitudes d'interpolation sur les quantiles 10 ans pour d=1440 minutes estimés par la méthode M3. Sous-région A (région 2).....	141
Figure 6.13	Cartographie de la borne supérieure des ICs à 70% , combinant incertitudes d'inférence statistique et d'interpolation spatiale, pour T = 10 ans sur d=1440 minutes estimés par la méthode M3. Sous-région A (région 2). .....	141
Figure 6.14	Ajustements des SMA pour d=60 min. par le modèle M1 (station 33 : en noir), M2 (région 2, en rouge) et M3 (sous-région D en bleu).....	143
Figure 6.15	Ajustements des SMA pour d=60 min. par le modèle M1 (station 12 : en bleu), et M3 (sous-région C en rouge) .....	143

## Liste des tableaux

Tableau 1.1: Exemple de crues et inondations en Algérie.....	3
Tableau 1.2 : Exemple des densités des réseaux PV et PG en région Méditerranéenne.....	4
Tableau 3.1 Coordonnées des stations pluviographiques du Centre Nord de l'Algérie .....	82
Tableau 3.2 Résultats de la régression linéaire multivariée entre caractéristiques géographiques et statistiques des précipitations extrêmes du centre Nord de l'Algérie. ....	91
Tableau 4.1 Valeurs du coefficient de détermination $R^2$ entre $k(q)$ et $q$ pour les cas A (15 min- 24 h) et le cas B (1h – 24h) .....	100
Tableau 4.2 Résultats des tests d'homogénéité de Hosking et Wallis, Statistique $H1$ pour les deux cas A (15 min- 24 h) et le cas B (1h – 24h).....	103
Tableau 4.3 Résultats des tests de Discordance $D_i$ et d'homogénéité $H1$ de HW pour une régionalisation spatialisée sur différentes durées d'agrégation.....	106
Tableau 4.4 Résultats des tests de Discordance $D_i$ et d'homogénéité $H1$ de HW pour une régionalisation avec intégration spatio-temporelle. ....	107
Tableau 4.5 Délimitation et description des régions et sous-régions homogènes du centre Nord de l'Algérie. Nota : les stations en gras n'ont pas été considérées pour former les régions homogènes.....	108
Tableau 5.1 Statistique AD, cohérence et critère d'évaluation des performances .....	119
Tableau 5.2 Coefficient de corrélation moyen entre les séries de cumuls de pluies mesurés sur diverses <i>durées pour 48 stations pluviographiques</i> .....	127





## **Chapitre 1**

### **Introduction générale**



## 1.1 Contexte local, problématique et motivation

Depuis plusieurs années, et de façon récurrente, plusieurs régions Algériennes subissent des inondations et des crues de sévérités variables. Quelles que soit les conditions climatiques (méditerranéenne, aride, semi-aride, etc.) de ces lieux, le constat est toujours similaire : inondation du milieu urbain et péri-urbain, cours d'eau en crue provoquant des débordements sur les réseaux routiers et stagnation des eaux sur des terres agricoles, même l'habitat rural n'est pas épargné (voir tableau 1.1). Bien évidemment, cet état de fait alarmant peut être imputé à différentes causes (ou leurs conjonctures) telles que les changements climatiques, les défauts de conception et le manque de maintenance des structures de gestions des eaux de pluies, et une urbanisation toujours croissante. Cependant, il y a un fait qui ne peut être éludé, à savoir le manque de données hydro-météorologiques en termes de longueurs des séries d'observations et de densité des réseaux de mesures. Ainsi, les mesures hydrologiques sont de première importance dans les études de conception d'ouvrages hydrotechniques, d'analyse du risque et de prise de décision (Chow et al., 1988). De plus, l'utilisation des mesures de pluies, comme base de l'ingénierie hydrologique, trouve encore plus son importance dans la conception des infrastructures de gestion des eaux de pluies en milieu urbain, où l'environnement naturel a été modifié et les mesures hydrométriques historiques, quand elles existent, ne sont pas représentatives du futur système modifié (Koutsoyiannis & Langousis, 2011).

Les enregistrements pluviométriques, et particulièrement pluviographiques (mesure des cumuls des pluies en fonction du temps) fournissant les mesures de pluies à des échelles temporelles très fines (de l'ordre de la minute), sont le point de départ de toute analyse d'ingénierie hydraulique pratique en milieu urbain, étant donné que les phénomènes d'inondations et de crues s'y déroulent bien souvent à une échelle infra-journalière (Willems, 2012). Ainsi, leur utilisation en analyse statistique sur les précipitations est fortement liée à leur disponibilité et à la qualité des observations (Grimaldi et al., 2011). Aussi, la fiabilité des estimations des événements extrêmes repose essentiellement sur la longueur (en année) des séries de mesure (Meylan & Musy, 1999). En effet, les manuels des bonnes pratiques hydrologiques soulignent tous l'importance de la longueur des séries de mesures. A titre d'exemple au Royaume-Uni l'Institute of Hydrology (1999) recommande d'appliquer la règle d'une longueur d'enregistrement égale à 5 fois  $T$  ( $T$  étant la période de retour) pour déterminer la période de retour maximale qu'il est possible de calculer. De même, le CERTU (2003) en France recommande d'avoir une longueur de série de mesure de 5 à 7 fois plus grande que la période de retour avec une longueur de 3 fois  $T$  minimale. Aux Etats-Unis, la pratique usuelle est que la période de retour doit être inférieure ou égale à 2 fois la longueur des séries de mesures (Stedinger et al., 1993). Au-delà des définitions différentes sur la longueur minimale des séries de mesure à retenir, tous ces référentiels nationaux ont comme point commun de relier fiabilité d'estimation avec la longueur des séries de mesures. En effet, l'implémentation des méthodes d'analyse fréquentielle, sur des échantillons de tailles limitées, tend à affecter les estimations avec de grandes incertitudes (Gaume, 2018).

Tableau 1.1: Exemple de crues et inondations en Algérie

**Skikda : Inondation urbaine du 01-09-2019 avec 140mm de pluie en 24h.**



(AF Rédaction, 2019)

**Chiffa (Blida) : Crue de l'oued Chiffa le 15-04-2018 avec 60mm de pluie**



(ObservAlgerie.com, 2018)

**Merdja (Laghouat) : Crue de l'oued M'zi le 30-09-2016.**



(Arbani, 2016)

Au regard de ce qui a été dit au paragraphe précédent, il est assez difficile de concilier des estimations suffisamment fiables et une analyse statistique classique basée sur de courtes séries d'observation. Pour appliquer les procédures et méthodologies qui seront développées dans ce travail, le cadre d'étude retenu ne compte que 48 stations pluviographiques avec en moyenne 22 années d'enregistrement, avec un maximum de 38 années et un minimum de 8 années. L'application de procédures d'analyse fréquentielle classique pour l'estimation des événements pluvieux extrêmes sur la base des séries de mesures courtes aboutit au mieux à des estimations de faible précision (grandes incertitudes), au pire à des aberrations (Boukhelifa et al., 2018). En plus de la faible durée des enregistrements, les réseaux de mesures pluviographiques (PG) et pluviométriques (PV) algériens sont caractérisés par un nombre réduit de postes. En effet, certaines agglomérations urbaines importantes (Alger Centre, Tizi Ouzou, Blida, Tipasa, etc.) sont carrément dépourvues de postes pluviographiques. Toutefois, comparativement à des régions en bordure de Méditerranée (voir tableau 1.2), subissant plus ou moins les mêmes aléas pluviométriques, la densité du réseau de mesure Algérien est bien meilleure qu'en Tunisie et en Espagne, équivalente à certaines régions d'Italie, mais aussi deux fois moins importante qu'en France et jusqu'à 5 fois plus faible que dans les régions très développées du Nord de l'Italie.

Tableau 1.2 : Exemple des densités des réseaux PV et PG en région Méditerranéenne

Auteurs	Zones	Superficie (km <sup>2</sup> )	Pluviomètres		Pluviographes	
			Nbr PV	Densité (Nbr. / 1000km <sup>2</sup> )	Nbr. PG	Densité (Nbr. / 1000km <sup>2</sup> )
ABH-AHS (2002)	Bassin Algérois (Côtiers, Sebaou et Isser)	12500	123	9.8	28	2.2
Boukhelifa et al. (2018)	Centre Nord de l'Algérie	25000	-	-	48	1.9
Blanchet et al. (2016)	Centre Sud de la France	38000	840	22	213	5.6
Roche et al. (2012)	France	643801	3500	5.4	1500	2.3
Di Baldassarre, Brath, et al. (2006); Di Baldassarre, Castellarin, et al. (2006)	Régions d'Emilie-Romagne et de Marches (Italie)	35800	394	11	131	3.7
Libertino et al. (2018)	Région du Piémont (Italie)	30000	-	-	400	13
Ghanmi et al. (2016)	Région de Tunis (PV)/ Nord de la Tunisie (PG)	2697/ 23570	24	8.9	15	0.6
Salas & Fernández (2007)	Espagne	505990	-	-	63	0.1

## 1.2 Objectifs de la thèse

Ce travail de thèse se focalise sur un seul des aspects, cités en début de la section précédente, responsables des inondations en Algérie à savoir : comment dépasser les limitations, dues à une faible densité du réseau de mesure pluviographique et à des séries de mesures relativement courtes, pour l'application d'une analyse fréquentielle des événements pluviométriques extrêmes. Plus clairement, l'objectif à atteindre est de fournir des outils d'ingénierie permettant des conceptions d'ouvrages hydrauliques et des analyses de risques suffisamment fiables et précises, même dans les zones faiblement voire non-jaugées.

Traditionnellement, les estimations des distributions fréquentielles pour diverses durées d'agrégation sont synthétisées sous forme de courbes Intensité-Durée-Fréquence (IDF). Toutefois, les incertitudes associées à l'élaboration de ces courbes IDF, n'ont été que très rarement évaluées dans le passé. De même, la plupart des modèles de courbes IDF, ne tirent pas profit des structures probabilistes des pluies pour établir une formulation des courbes IDF théoriquement cohérente et hydrologiquement robuste. Ainsi, ce travail vise à proposer une méthodologie et des procédures permettant de faire : (a) la meilleure utilisation possible des données disponibles et de leurs propriétés dans le calage des courbes IDF, (b) une évaluation efficiente des incertitudes sur les paramètres du modèle IDF et les quantiles correspondants, (c) une régionalisation des courbes IDF sur toute la zone d'étude, en combinant les incertitudes issues des estimations des distributions fréquentielles et celles issues des procédures d'interpolation, dans le but d'établir une cartographie des extrêmes pluviométriques limitant les risques de défaillances des ouvrages hydrauliques.

## 1.3 Vue d'ensemble

Il est assez difficile de donner une définition unique à un événement pluviométrique extrême. En effet, définir un événement extrême comme étant un événement remarquable reste assez subjectif. Ainsi, il est tout à fait possible qu'une pluie de période de retour décennale de 50 mm provoque une inondation en un lieu A, et passe inaperçue dans un lieu B. Par contre une pluie centennale de 150 mm provoquera très certainement des troubles dans les deux lieux. Et en rajoutant la notion de durée sur laquelle la quantité de pluie est survenue, il est encore plus compliqué de définir l'évènement extrême. Il devient, donc clair, qu'un événement pluvieux extrêmes est spécifique à une zone et est intimement lié à sa distribution fréquentielle établie sur une durée bien définie. Etant donné que le dimensionnement d'ouvrages hydrauliques est basé sur des grandeurs, définies comme étant les plus fortes valeurs probables au cours de la durée de vie de l'ouvrage (10, 50 voire 100 années), il n'est pas envisageable d'en faire une estimation déterministe. Alors, seule l'approche probabiliste reste significative et efficace, et doit être basée sur les observations historiques des pluies (Koutsoyiannis & Langousis, 2011).

L'analyse fréquentielle des événements extrêmes est une approche nécessitant un traitement spécifique avant et après sa mise en œuvre. Tout d'abord, il s'agit de recueillir les données à utiliser : observations au sol ou à distance (radar et satellite), ou encore des

mesures réelles ou simulées. Une fois le recueil fait, il s'agira d'extraire les données dites extrêmes à travers une procédure d'échantillonnage de type séries des maxima annuels (SMA) ou séries de durées partielles (SDP), puis s'assurer de leur fiabilité. Il n'a pas été établi une prévalence absolue d'une procédure par rapport à l'autre, juste des critères spécifiques à assurer et un contexte bien défini avant l'application de l'une ou de l'autre procédure d'échantillonnage (Lang et al., 1999). Une fois l'échantillon à étudier défini et la loi statistique à utiliser choisie (pour des applications aux évènements pluvieux extrêmes le choix se rapporte à 2 ou 3 lois, cet aspect est développé en détail au §2.4.5), l'analyse fréquentielle à proprement dit peut-être mise en œuvre. Néanmoins, une multitude de processus statistiques inférentiels (Moments, L-moments, maximum de vraisemblance, etc.) est à disposition de l'analyste sans compter sur les nouvelles techniques émergentes telles les super-statistiques et les metastatistiques. Seulement, ces derniers aspects cités peuvent être plus facilement cernés au regard du contexte de ce travail qui nécessite les estimations les plus précises et robustes possible, tout en évaluant les incertitudes qui y sont reliées. La justification sur le choix des techniques et procédures à utiliser dans l'analyse fréquentielle est amplement développée dans le chapitre 2 avec un argumentaire assez pragmatique.

En ingénierie d'hydrologie urbaine, l'analyse fréquentielle des évènements extrêmes sert de base de calcul aux intensités de pluies sur diverses durées résumées souvent sous forme de courbes IDF. La genèse, le développement et l'usage de ces dernières est aussi discuté en fin de chapitre 2. Aussi, il est utile de signaler que la méthodologie d'élaboration du modèle de courbes IDF développée ici, est basée sur une hypothèse de stationnarité. Sa justification est également traitée au même chapitre.

La partie centrale du Nord de l'Algérie a été retenue comme cas d'étude pour la mise en œuvre des méthodologies, proposées et testées dans ce travail, permettant d'atteindre les objectifs précédemment mentionnés. Le centre nord de l'Algérie s'étend sur plus de 25000 km<sup>2</sup> et couvre une dizaine de wilayas parmi les plus importantes du pays d'un point de vue socio-économique. En effet, cette zone compte plusieurs bassins industriels et deux des plus grandes plaines agricoles de l'Algérie (Cheliff et Mitidja). Aussi, cette zone dispose d'un des meilleurs réseaux pluviographiques du pays avec 48 pluviographes opérationnels. Ce qui favorise la mise en place de tests empiriques extensifs et la consolidation des approches retenues. Une plus large revue de la région d'étude et des données est présentée en chapitre trois.

Dans une profonde réflexion sur les notions d'invariance d'échelle et des structures multifractales dans la modélisation des précipitations, Veneziano et al. (2006) attestent que ces dernières sont arrivées à maturité et qu'une théorie cohérente des extrêmes appliquant ce principe a émergé. Il est donc temps de la mettre en œuvre, de façon simple, dans le cadre d'analyse de robustesse, de validation et de régionalisation. Aussi, il est nécessaire d'identifier les modèles de pluie les plus performants en équilibrant entre la précision, la simplicité et la robustesse. Ce travail de thèse s'inscrit pleinement dans cette déclaration.

La notion d'invariance d'échelle, et particulièrement l'invariance d'échelle simple (*simple scaling* en anglais), est souvent considérée (Durrans, 2013) comme un costume taillé sur mesure pour atteindre la cohérence théorique dans la formulation des courbes IDF. En

effet, l'invariance d'échelle simple (au sens strict : voir §2.5) signifie que les fonctions de distribution de probabilités sur diverses échelles (ici la durée d'agrégation) sont toutes identiques à un facteur d'échelle près. Cette définition est analogue à la notion fondamentale d'indice de crue (ou pluie) en analyse fréquentielle régionale (Hosking & Wallis, 1997). D'un autre côté, les propriétés multifractales aussi appelées structure probabiliste des pluies, peuvent être mises en avant à travers une analyse d'échelle des moments de pluies (Burlando & Rosso, 1996; Sivakumar, 2001). Une analyse d'échelle des moments (MSA : *Moment Scaling Analysis* en anglais) de pluies consiste à vérifier l'existence d'une fractalité et de son type (Gupta & Waymire, 1990). D'un côté, la fractalité existe si les relations entre les logarithmes des moments des cumuls/intensités des pluies d'ordre  $q$  et les durées d'agrégations sont représentées par des droites de pentes  $k(q)$ . Et d'un autre côté, la nature de la relation entre la fonction d'exposant d'échelle  $k(q)$  et l'ordre  $q$  des moments, définit le type de la fractalité. En effet, la fractalité des pluies maximales annuelles sur diverses durées d'agrégation peut être définie comme une invariance d'échelle au sens large. Elle est dite : multifractale si la relation entre les exposants d'échelle  $k(q)$  et l'ordre  $q$  des moments est non linéaire, monofractal si cette relation est une droite ne passant pas par l'origine, autosimilaire s'il y a log-log linéarité (c'est-à-dire une droite passant par l'origine) (García-Marín et al., 2013) (figure 1.1). L'hypothèse d'invariance d'échelle doit obligatoirement être concordante avec les données - au risque d'introduire des biais d'estimation- ce qui est vérifiable à travers de multiples tests, tels que MSA, test d'homogénéité, etc. (voir chapitre 4 pour plus de détails sur les tests appliqués).

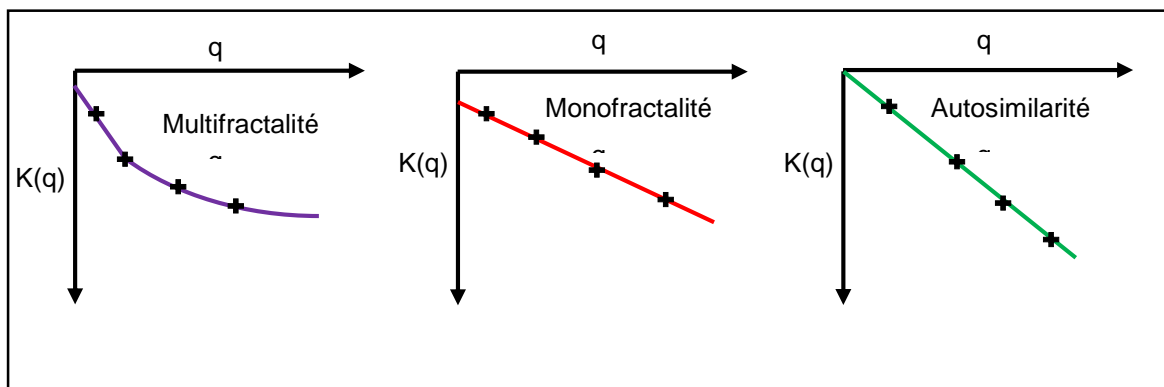


Figure 1.1 Schématisation de la fractalité des pluies (invariance d'échelle au sens large) en fonction de l'exposant d'échelle  $k(q)$  et de l'ordre  $q$  des moments statistiques

L'utilisation de l'invariance d'échelle simple et les avantages en découlant présentent beaucoup plus de subtilité qu'il n'y paraît, particulièrement dans la définition du lien entre invariance d'échelle, fractalité et formulation des courbes IDF. En effet, il a été montré que seule une définition autosimilaire de la fonction d'exposant d'échelle est théoriquement cohérente (voir démonstration en §2.8.4) avec le principe d'invariance d'échelle simple au sens strict (Boukhelifa et al., 2018; Menabde et al., 1999).

Bien que présentant une incohérence théorique, la méthodologie classique d'élaboration des courbes IDF est bien maîtrisée et totalement satisfaisante particulièrement dans les régions où les enregistrements de pluies sont de grandes tailles avec une faible variabilité d'échantillonnage (Koutsoyiannis et al., 1998). Cette incohérence théorique réside dans le



fait que les quantiles de pluies sur une courbe de même fréquence proviennent d'ajustements différents, d'une même loi statistique, chacun correspondant aux diverses durées d'agrégation. L'invariance d'échelle simple avec une définition linéaire de la fonction d'exposant d'échelle permettent une formulation théoriquement consistante du modèle de courbes IDF. Ainsi, le modèle IDF est le produit d'une fonction de distribution de probabilités adimensionnelle des cumuls/intensités de pluies maximales annuelles et un facteur d'échelle dépendant de la durée. Ce qui permet au final de mélanger toutes les séries de diverses durées en une unique série ou un échantillon conjoint (*pooled* en anglais). Ensuite, il est possible d'inférer tous les paramètres du modèle IDF en une seule étape en maximisant le logarithme de la vraisemblance du modèle IDF à travers un algorithme de calcul Bayésien Markov Chain Monte Carlo (MCMC). Ce processus d'inférence statistique, susmentionné, est considéré comme étant le mieux adapté pour un calcul efficient des incertitudes d'estimation (Coles & Pericchi, 2003; Reis & Stedinger, 2005). Cette méthodologie d'élaboration des courbes IDF est décrite de façon détaillée (cf. chapitre 2) et évaluée de manière approfondie (cf. chapitre 5) afin de mettre en avant tous les avantages qu'apporte une estimation Bayésienne intégrée des courbes IDF.

De nombreux endroits de la région d'étude ne disposent pas de mesures pluviographiques, ce qui pose un problème de taille aux ingénieurs concepteurs dans leurs études. En effet, les données et estimations, issues de stations de mesures existantes, ne peuvent être directement transposées vers les lieux non-jaugés. Surtout que les stations de mesures ont des régions d'influence assez limitées, particulièrement quand il s'agit d'observations horaires, et à plus forte raison quand le relief est hétérogène. Ce qui est typiquement le cas de notre région d'étude (Meddi & Toumi, 2015). Bien évidemment, il existe des solutions, établies depuis plusieurs années maintenant, permettant de dépasser cette problématique. Ainsi, la géostatistique a permis de mettre en avant différentes techniques d'interpolation spatiale telles que la pondération inverse à la distance, le Krigeage, etc. (Diggle & Ribeiro, 2007; Kitanidis, 1997). Ces dernières techniques sont assez limitées pour des applications directes en spatialisation des extrêmes pluviométriques particulièrement quand la densité du réseau de mesure est assez faible. En effet, la supériorité de la cartographie des variables hydrologiques en combinant analyse fréquentielle régionale et procédure géostatistique est bien établie (Faulkner & Prudhomme, 1998; Prudhomme, 1999; Reed et al., 1999; Stewart et al., 1999). Ce qui consiste en une inférence statistique sur un échantillon unique, composé de données provenant de différents sites, sous réserves d'homogénéité de la région, puis d'interpoler, par Krigeage ou autre, le paramètre de position (typiquement la moyenne ou la médiane) de la série de données de chaque stations, et enfin, déduire le quantile de pluie – en multipliant le paramètre de position par le quantile régional adimensionnel – en chaque point de la grille ainsi formée et établir une cartographie des évènements pluvieux extrêmes (figure 1.2). Cette méthodologie est mise en œuvre au chapitre 6 en combinant différents schémas d'analyse statistique fréquentielle sur les pluies extrêmes avec une procédure d'interpolation par Krigeage ordinaire afin d'identifier l'approche la plus adéquate.

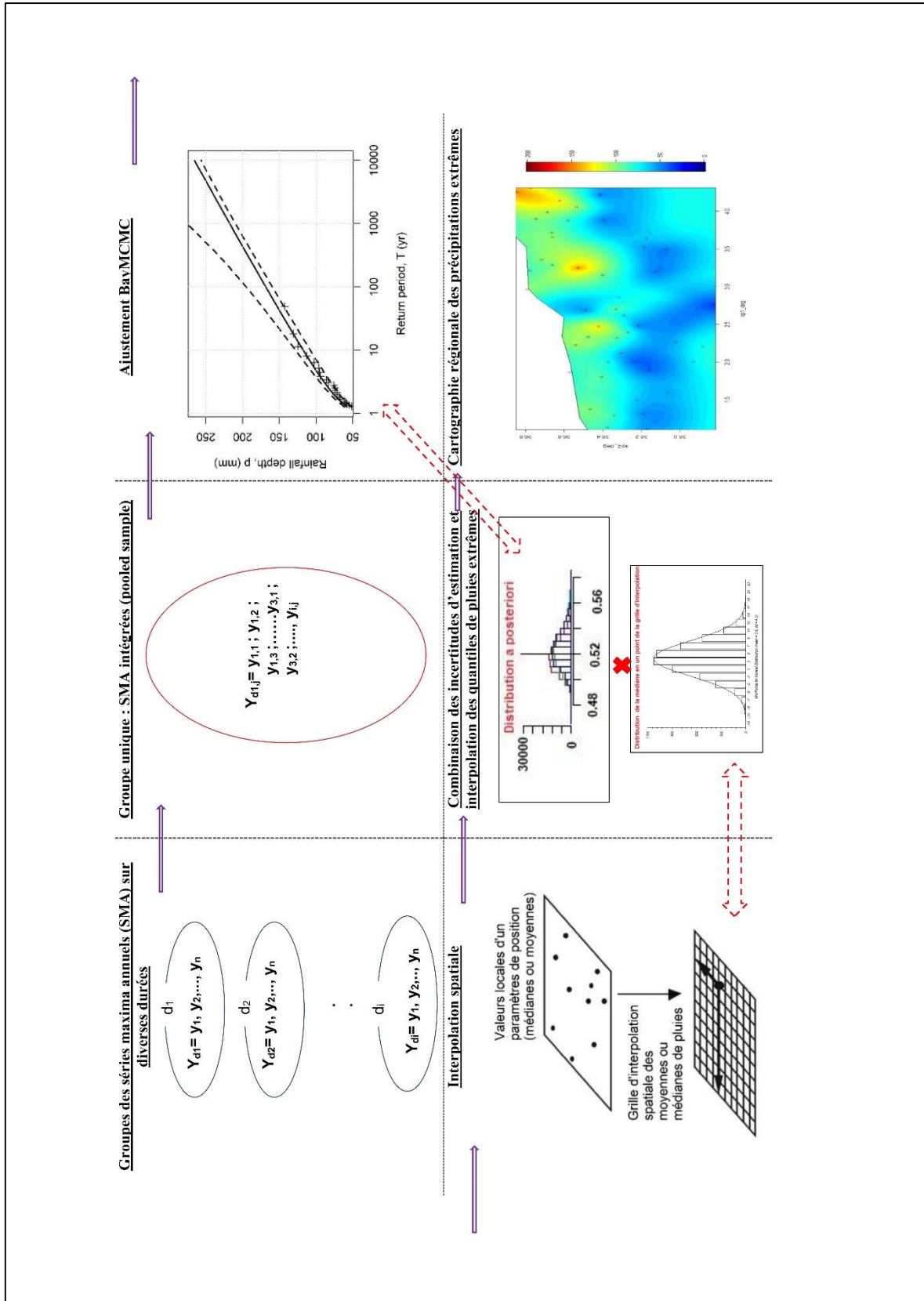


Figure 1.2 Principe de la cartographie des précipitations extrêmes proposée dans ce travail.

## 1.4 Organisation du mémoire

Ce travail est structuré autour de 6 chapitres. Un premier chapitre introductif, suivi d'un second chapitre qui fait une synthèse bibliographique relativement exhaustive sur l'analyse fréquentielle des événements pluviométriques extrêmes et toutes les notions reliées. Le troisième chapitre fait une large revue de la zone d'étude et des données utilisées dans ce travail. Le chapitre quatre concerne l'étude probabiliste des structures temporelles et spatiales des précipitations maximales annuelles sur diverses durées d'agrégations dans la zone d'étude. Le cinquième chapitre est quant à lui consacré à la validation du modèle de courbes IDF reposant sur les notions d'invariance d'échelle simple et de multifractalité, puis à l'évaluation de ses performances et enfin à la mise en exergue des avantages d'une estimation Bayésienne intégrée des courbes IDF. Au final, le chapitre six évalue la possibilité d'une implémentation régionale du modèle IDF proposé.



## **Chapitre 2 :**

### **Les extrêmes pluviométriques : compréhension, méthodes d'analyse et utilisation**

## 2.1 Introduction

Les extrêmes pluviométriques jouent un rôle très important dans les pratiques d'ingénierie de conception et de gestion des ressources en eau. En plus des différentes approches d'inférences statistiques pour sa mise en œuvre, une analyse fréquentielle, sur les précipitations extrêmes, est reliée à un certain nombre de procédures préparatoires (échantillonnage, choix de la loi statistique, etc.) et de post-traitements (adéquation, évaluation des performances, etc.). Au fil des années, de nombreuses techniques et approches, pour l'implémentation d'une même procédure, ont été proposées et parfois il est difficile de connaître celles qui sont les mieux adaptées à chaque cas de figure. Dans ce qui suit, un certain nombre de ces techniques, parmi les plus pertinentes, sont décrites. Leurs avantages et inconvénients, ainsi que leurs limitations, sont discutés dans ce chapitre. De même qu'un large aperçu des différents usages (local et régional) et formulation des courbes IDF.

## 2.2 Qu'est-ce que les précipitations extrêmes ?

Les précipitations représentent la variable hydrologique clé reliant (et jouant un rôle dominant) entre les processus atmosphériques et ceux de la surface terrestre (Teegavarapu et al., 2019). Le plus souvent les précipitations sont mesurées sur une échelle de temps journalière à partir de laquelle sont constitués les totaux pluviaux mensuels, saisonniers et annuels, etc. Cependant, les précipitations sont aussi mesurées sur des échelles de temps plus fines (heures, minutes) permettant, entre autres, de distinguer ce qui est appelé : précipitation extrême, sur diverses durées d'agrégations. En effet, les extrêmes pluviométriques peuvent être définis comme toutes quantités de pluies précipitées sur une échelle de temps continue pouvant aller de quelques minutes à vingt-quatre heures (étendue dans certains cas jusqu'à 48 ou 72 heures) et présentant des cumuls élevés (voire très élevés) comparativement aux mesures courantes. L'étude des distributions fréquentielles de ces pluies extrêmes trouve toute son importance dans l'estimation des événements pluvieux remarquables et les probabilités d'occurrence leurs correspondant (El Adlouni & Ouarda, 2010). Plus spécifiquement, c'est la partie supérieure de ces distributions fréquentielles, usuellement appelée la queue de la distribution, qui régit la grandeur et la fréquence des événements extrêmes. Les queues de toutes les fonctions de distribution de probabilités peuvent être catégorisées en deux familles : les distributions à queue à décroissance lente (i.e. sous-exponentielle – relativement à une loi exponentielle – ou en anglais *Heavy Tail*) et à décroissance rapide (super-exponentielle ou en anglais *Light Tail*), avec la dernière générant des extrêmes légèrement plus faibles et moins fréquents que la première. Ces lois statistiques permettent d'assigner une période de retour à chaque quantité de pluie précipitée, et si elle est par exemple de 1000 ans, cet événement est en effet un extrême pluviométrique (Papalexiou et al., 2013). Il est à noter que le développement récent d'un test statistique basé sur une fonction de dépassement de la moyenne (Nerantzaki & Papalexiou, 2019) permet d'identifier les extrêmes hydrologiques dans les séries de mesures.

## 2.3 Procédures d'échantillonnage des extrêmes pluviométriques

En analyse fréquentielle d'événements extrêmes, une procédure d'échantillonnage sur des observations de pluies consiste à extraire à partir de données brutes, des séries (ou échantillons) des valeurs extrêmes. Ces observations brutes sont appelées des séries de durées complètes (SDC) où sont renseignés les cumuls de pluies (période humide) ainsi que les périodes sèches. Deux techniques distinctes sont communément utilisées pour extraire les valeurs extrêmes à partir d'une SDC. La première approche est celle des maxima annuels, elle consiste à sélectionner la valeur de la pluie maximale, sur une durée bien définie, pour chaque année hydrologique afin de former une série des maxima annuels (SMA) des pluies. La seconde est connue sous le nom de série POT (*Peak Over Threshold*, du fait qu'elle ait été appliquée pour la première fois sur des séries de débits) ou méthode du renouvellement, elle consiste à constituer une série de durée partielle (SDP) en sélectionnant, toutes années hydrologiques confondues, les valeurs de pluies supérieures à un certain seuil préalablement déterminé. Il existe aussi une troisième technique d'échantillonnage, dite série « gonflée », qui consiste à sélectionner les  $k$  plus grandes valeurs sur chaque année hydrologique avec  $k$  généralement variant de 1 à 3. Cependant, la légitimité d'utilisation des séries gonflées est très discutable sur le plan de la théorie statistique (Meylan & Musy, 1999), d'ailleurs cette technique a été abandonnée depuis assez longtemps. Avant de débattre sur le choix de la procédure d'échantillonnage à retenir dans ce travail, une brève description sur la nature et la provenance des séries pluviométriques brutes est donnée dans la section qui suit.

### 2.3.1 La mesure des précipitations

Disposer de données précises sur les précipitations, pour différentes échelles spatio-temporelles, est très important pas uniquement pour les prévisions météorologiques et les sciences du climat mais aussi dans la prise de décision dans divers aspects des sciences hydrologiques (inondations, crues, sécheresses, etc.). Des précipitations excessives peuvent causer d'important dommages humains et matériels. La prédiction de tels phénomènes dépend largement du degré de fiabilité des estimations de ces pluies ainsi que leurs résolutions sur différentes échelles spatio-temporelles (Teegavarapu et al., 2019). L'analyse fréquentielle des extrêmes pluviométriques peut être mise en œuvre sur des séries d'observations historiques ou simulées (générées par des modèles stochastiques de pluie) (Willems et al., 2012). Les observations historiques des précipitations peuvent être obtenues à travers des mesures au sol ou via de la télémessure (radar, satellite, etc.).

#### 2.3.1.1 Les mesures au sol : les séries de durées complètes (SDC)

##### Aperçu historique

Les mesures locales des précipitations sont très anciennes, les premières mesures attestées datent de 500 ans avant J-C en Grèce antique et de 400 ans avant J-C en Inde ancienne. Toutefois, ces observations n'ont pas survécu aux épreuves du temps. Le premier réseau de mesure pluviométrique sur la base d'un instrument de mesure standardisé en bronze (appelé

« Cheugugi »), a été développé en 1441 par Jang Yeong-Sil en Corée (Roche, Miquel, & Gaume, 2012). Quant aux plus anciennes mesures pluviométriques, dont les enregistrements sont encore disponibles, elles datent de 1770 en Corée toujours (Koutsoyiannis & Langousis, 2011). Pour l'Europe, les plus anciennes étant en Italie (1822), aux Royaume-Unis (1853), Grèce (1860) et en France (1864). En Amérique du Nord, on les retrouve aux Etats-Unis (1871) (Koutsoyiannis, 2004b). En Algérie les plus anciennes mesures pluviométriques encore opérationnelles, à notre connaissance, datent de 1859 à Yakourene (PV020902) (w. Tizi Ouzou), de 1899 pour le barrage de Meurad (PV020306) (w. Tipasa), de 1900 à Sour el Ghazlene (PV150101) (w. Bouira), de 1901 à Theniet el Had (PV011605) (w. Tissemsilt), de 1905 pour le barrage de Hamiz (PV020602) (w. Boumerdes), et de 1906 pour phare de ChercHELL (PV020312) également dans la wilaya de Tipasa.

### **Instruments de mesures**

Les observations de pluies au sol peuvent être effectuées avec deux instruments de mesures différents. Le premier, et le plus couramment utilisé, étant le pluviomètre totalisateur qui permet d'effectuer la mesure par lecture directe des cumuls précipités sur une période de temps fixe. Il existe différents modèles de pluviomètre mais respectant les mêmes standards définis par l'organisation mondiale de la météorologie (OMM, 1983). Le second, est le pluviographe, qui enregistre automatiquement l'évolution du cumul de pluie en fonction du temps. Les mesures sont fournies sous formes de bande en papier millimétré, appelée pluviogramme, ou sous format numérique pour les modèles les plus récents. Les modèles de pluviographe les plus répandus à travers le monde sont de type augets basculants, siphon (à Flotteur) et balance (Chow et al., 1988).

### **Constitution des séries d'observations**

Les mesures pluviométriques et pluviographiques fournissent toutes deux des cumuls de pluies. Cependant, la constitution de ces cumuls de pluie est substantiellement différente entre les deux instruments de mesure. En effet, le pluviomètre permet, une lecture directe sur l'appareil, des totaux cumulés de pluies pendant une journée. Généralement, le relevé se fait 1 à 2 fois par jour et à heure fixe. Ainsi, un pluviomètre fournit une série de durée complète (sur une année hydrologique) des cumuls de pluie journaliers. Par contre, un pluviographe, mesurant l'évolution du cumul de pluie en fonction du temps, fournit des cumuls sur diverses durées d'agrégation, allant de quelques minutes à 24 heures. Alors, les observations sont fournies sous forme de pluviogrammes, chacun correspondant à une averse bien définie, à dépouiller par fenêtre de temps glissante. C'est-à-dire que pour une pluie de 24 heures, le dépouillement consiste à déduire le plus fort cumul de pluie en 24 heures sur toute la durée de l'averse (qui peut durer une centaine d'heures), et de même pour les autres durées d'agrégation. Ce qui implique que le cumul pluviographique sur 24h peut s'étaler sur une période de temps différente de celle du cumul pluviométrique mesuré sur fenêtre fixe. En effet, les mesures des SDC pluviométriques sont inférieures ou égales aux mesures des SDC pluviographiques de 24 heures. Ce qui mène à une sous-estimation systématique des quantiles de pluie estimés sur la base de mesures pluviométriques (Van de Vyver, 2015b). D'où l'importance de baser les pluies de projet sur des mesures pluviographiques.



Weiss (1964) définit un coefficient correcteur pour les cumuls de pluies enregistrés sur un pas de temps et une origine fixe, ce qui est le cas des pluviomètres. Il trouve son utilité surtout pour les pluies les plus intenses (maxima annuels, POT, etc.). Généralement, ce coefficient correcteur est pris égal à 1.14 pour les pluies journalières (c'est-à-dire mesurée par pluviomètre : mesure, sur 24 heures, relevée à heure fixe). Cependant dans la pratique il varie entre 1.1 et 1.16.

$$K_w = \frac{D}{\left(D - \frac{N}{8}\right)} \quad 2.1)$$

Avec (D) la durée ou le pas d'étude d'une pluie donnée et (N) le pas de temps de la mesure. Toutefois, cette définition a tendance à systématiquement atténuer l'intensité des pluies sur la durée de référence à étudier.

### Prétraitement des données

La constitution des séries d'observations pluviométriques est un processus multi-étapes (mesures, transmissions de l'information, puis son stockage et son pré-traitement) assez long, dont il résulte des erreurs de natures différentes (Meylan & Musy, 1999; Roche, Miquel, & Gaume, 2012). Les premières, étant les erreurs aléatoires, d'importance variable et dues à de nombreux facteurs bien souvent inconnus, qu'il est impossible d'éviter. Ceci est une des raisons expliquant l'importance de la prise en considération des incertitudes dans l'inférence statistique. Les secondes, étant les erreurs systématiques et d'inconsistance. Elles peuvent être dues à des erreurs humaines dans la transcription des mesures, un défaut dans l'appareil de mesure à cause d'un manque de maintenance ou autre, d'un déplacement de la station de mesure ou encore d'un changement dans le proche environnement de la station (construction de bâtiments, croissance de la végétation aux alentours, etc.). La vérification de ce dernier type d'erreurs est une pratique très courante en hydrologie et facile à mettre en œuvre à travers la vérification des cohérences : interne (comparaison avec la radiation solaire), temporelle (analyse des mesures consécutives), spatiale (comparaison des mesures avec celles des stations voisines) pour chaque échelle temporelle séparément, puis si nécessaire leurs corrections. Cette pratique, demeure relativement sans conséquence sur la fiabilité des estimations. Toutefois, il existe en analyse statistique des pluies, une pratique qui consiste à retirer des enregistrements, les valeurs dites horsaires (cumul de pluie extrêmement fort, *Outliers* en anglais). La légitimité du rejet d'une valeur, supposée hors limites, est très discutable, particulièrement dans l'étude des événements extrêmes. En effet, il est très difficile de distinguer entre valeurs extrêmes et aberrantes, ainsi le retrait d'un horsaire, aboutit très souvent à une sous-estimation des pluies de projet (Koutsoyiannis & Langousis, 2011).

Aussi, il est à signaler que quelques fois, ces horsaires ne sont pas retirés de l'échantillon mais simplement corrigés. Une des méthodes les plus utilisées dans ce genre d'exercice est celle de la double masse. Seulement, selon Koutsoyiannis & Langousis (2011) bien souvent l'application de la méthode de la double masse et d'autres méthodes analogues, aboutit à un net effet de distorsion de données, qui à la base étaient correctes. Ce qui mène à un raisonnement assez incohérent. En effet, dans ces approches, le processus de précipitation est assumé

indépendant du temps, des manifestations de la dépendance (comme la persistance à long de Hurst-Kolmogorov, voir §2.4.8 pour plus de détails sur ce phénomène) sont interprétées comme des erreurs sur les données, et subséquemment les données sont modifiées afin de se débarrasser de l'influence de la dépendance pour aboutir à des séries encore plus proches de la fausse hypothèse initiale d'indépendance. Pour les raisons susmentionnées et afin de préserver les données historiques réelles, utilisées dans ce travail, d'une quelconque distorsion, seule une vérification qualitative sur les observations est entreprise au chapitre 3.

Parfois les observations pluviométriques comportent des lacunes, dues à une perte ou au non relevé de la mesure. Pour pallier ce désagrément, diverses techniques de comblement des pluies journalières, mensuelles et annuelles ont été développées, telles que l'interpolation du plus proche voisin, régression linéaire, transfert à base de règle flou, Krigeage ordinaire, régression linéaire multiple avec ou sans algorithme de maximisation, distribution fréquentielle bivariée, etc. (Bárdossy & Pegram, 2014; Beaulieu et al., 2007), et il a même été proposé des approches de distribution fréquentielle multivariée pour le comblement des pluies horaires sur la base de pluviométrie journalière, de la position géographique et des conditions climatiques (Western et al., 2011). Bien que, nombre de ces méthodes sont satisfaisantes et quantitativement cohérentes, il n'est procédé à aucun comblement des lacunes sur les données utilisées dans ce travail. Car, la formulation de notre modèle IDF repose sur la validité de l'invariance d'échelle simple et une certaine représentation des propriétés fractales des pluies. Or, il n'a pas été établi, à notre connaissance, que les techniques de comblement préservent les structures probabilistes des pluies. En effet, la préservation des structures de corrélation dans la modélisation stochastique des processus hydro-climatiques n'est qu'à ses débuts (Papalexiou, 2018; Papalexiou & Serinaldi, 2020). Ainsi, l'adéquation du calage d'une représentation fractale des pluies, sur des données réelles comblées (i.e. où la corrélation a été induite), peut être altérée et favoriser un biais d'estimation. De plus, Svensson et al. (2007) dans l'estimation des courbes IDF sur des séries artificiellement fragmentaires (jusqu'à 20% de lacunes) ne soulignent pas un manque de performances comparativement à des séries complètes.

### **Observations pluviométriques en Algérie**

En Algérie, il est bien connu que les postes de mesures pluviométriques sont beaucoup plus répandus que les postes pluviographiques. Ces derniers trouvent leur plus grande utilité dans les zones fortement urbanisées pour servir de base d'analyse dans les études de conception des ouvrages de gestion des eaux de pluie. A titre d'exemple, bien que le nombre moyen de poste pluviographique en France soit de 1 poste pour 500 km<sup>2</sup>, la ville de Paris compte 1 poste PG pour 10km<sup>2</sup> (Roche, Miquel, & Gaume, 2012). En comparaison, la ville d'Alger est assez faiblement instrumentée en ce qui concerne les mesures de pluies horaires. En effet, cette dernière est équipée d'un seul poste PG (Birmandreis PG020509) dont la qualité des observations est assez discutable, car localisé sur le toit d'un immeuble lui-même se trouvant dans un vallon assez étroit, en contradiction avec les recommandations de l'organisation mondiale de la météorologie (OMM, 1983). Dans ce domaine, l'Algérie n'est pas un cas isolé, bien au contraire le pays se trouve dans la moyenne en termes de densité du réseau pluviométrique/ pluviographique dans la région Méditerranéenne (voir §1.1). En effet,

concernant l'analyse fréquentielle des pluies infra-journalières, de vastes régions dans le monde peuvent être considérées comme non jaugées (Bárdossy & Pegram, 2017).

Cependant, en ingénierie hydrologique urbaine, la disponibilité de mesures pluviométriques seules est très insuffisante. En illustrant encore par la ville d'Alger, cette dernière a connu dans la nuit du 09/11/2001 au 10/11/2001 un des événements pluvieux les plus violent qui ait jamais été enregistrés en Algérie. En effet, la station pluviométrique de Bouzaréah (sur les hauteurs d'Alger) donne un cumul de 260 mm en un 1 jour, soit une intensité de 11mm/h (ISL-BRGM, 2005). Or le hyétogramme reconstitué (figure 2.1) sur la même station donne une intensité horaire maximale 62mm/h et même de 55mm/h en 30 minutes. Soit une intensité maximale de pluie jusqu'à 10 fois plus grande que l'intensité moyenne sur 24 heures. Cet exemple est sans appel et ne laisse aucun doute sur l'importance des enregistrements pluviographiques. A noter que les records mondiaux sont bien plus importants, à savoir une pluie de 2000 mm/24h sur l'île de la Réunion et de 1000 mm/24 h en rive Nord de la Méditerranée (Aiguat en 1940 et en Corse en 1993, France) (Météo France, 2020). De même le record mondial sur une durée courte est de 208 mm en 20 minutes à Curtea en Roumanie en 1947, ce qui est presque le double du record Algérien de 110 mm en 15 minutes le 10/11/2001 à la station des deux Bassins (020601) dans l'Atlas Blidéen.

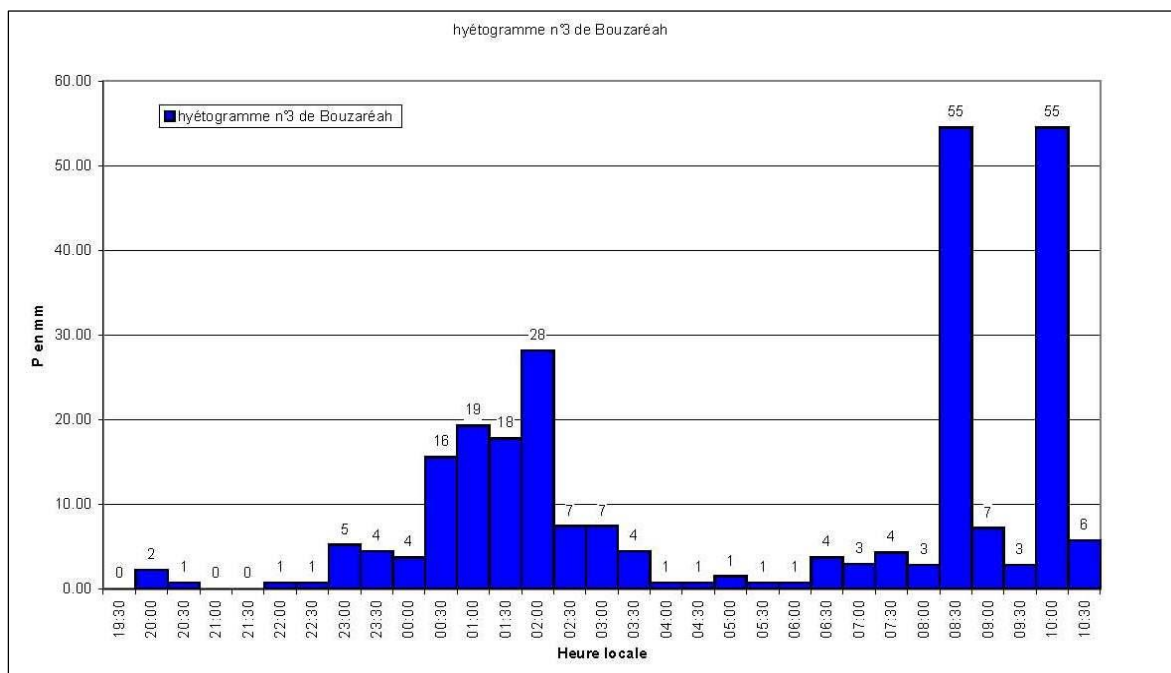


Figure 2.1 Reconstitution du hyétogramme du 09/11/2001 au 10/11/2001 à Bouzaréah (Alger) (ISL-BRGM, 2005)

### 2.3.1.2 Les mesures radar et satellite

Bien que les pluviomètres au sol soient les instruments conventionnels pour des mesures des pluies, un réseau de mesures pluviométriques de faible densité et/ou des enregistrements sur de courtes périodes d'observation ont une utilité limitée. D'où la nécessité de techniques d'observation alternatives telles que les mesures à distance (radar et satellite) (Teegavarapu et al., 2019). En effet, les radars météorologiques fournissent des mesures distribuées des

précipitations à haute résolution spatio-temporelle, c'est leurs premiers avantages par rapport aux réseaux d'observation au sol (Overeem et al., 2009). Ainsi, la télédétection permet de dépasser la faible représentativité des mesures locales des pluies dans les régions caractérisées par un fort gradient pluviométrique (Marra & Morin, 2015). Néanmoins, bien que la télémessure des précipitations offre de multiples possibilités d'applications à des échelles régionales et globales, les incertitudes y afférant et leurs implications nécessitent d'être rigoureusement analysées (Marra et al., 2017).

Dans certains pays la disponibilité des mesures radar de qualité et sur des périodes suffisamment longues rend possible de calculer les statistiques sur les extrêmes (Overeem et al., 2009). Toutefois, les résultats sont assez contrastés et ne permettent pas d'affirmer la prévalence, de la télédétection radar, sur les mesures au sol pour des estimations quantitatives des extrêmes pluviométriques. En effet, dans le cadre d'analyses comparatives sur les estimations des courbes IDF basées sur les mesures radar des pluies et enregistrements pluviographiques, un certain nombre de travaux (Marra & Morin, 2015; McGraw et al., 2019) montrent clairement une surestimation des courbes IDF radar par rapport aux courbes IDF pluviographiques, avec une accentuation de cette tendance respectivement avec l'augmentation de la période de retour et la réduction de la durée d'agrégation. A noter que cette surestimation sur les quantiles inférés sur la base de mesures radar est souvent positionnée à l'intérieur de l'intervalle de confiance à 95% de l'inférence statistique sur la base de mesures au sol. Tandis que d'autres études (A Overeem et al., 2009; Paixao et al., 2015) montrent une bonne concordance entre courbes IDF radar et PG, même si les incertitudes sur les premières sont beaucoup plus importantes que sur les deuxièmes. En effet, à moins d'avoir des mesures radar quasi parfaites, une concordance absolue est impossible. Les erreurs d'estimations radar et leurs distributions viennent interférer nécessairement sur les études statistiques. A contrario une étude plus élaborée (Peleg et al., 2018), intégrant une méthode de correction des mesures radar (en tenant compte de la variabilité spatiale à une échelle très fine) et à priori plus robuste, montre une sous-estimation des courbes IDF radar par rapport aux courbes IDF pluviographiques. A signaler, qu'il a été aussi constaté qu'une estimation des courbes IDF radar pouvait fournir des estimations irréalistes, ne convenant pas du tout l'analyse des valeurs extrêmes (Goudenhoofd et al., 2017). De plus, dans certains cas la présence de grêle rend les mesures radar inexploitable (relativement à l'identification des extrêmes pluviométriques), et l'exclusion de ces derniers peut amener à manquer de fortes intensités de pluies (Marra & Morin, 2015). Cette variabilité dans les résultats est principalement due aux incertitudes liées aux mesures radar, même si la résolution temporelle des pluviographes ne peut être négligée. En effet, les données radars corrigées sur la base d'une relation de réflectivité mènent à des estimations biaisées et bruitées des extrêmes pluviométriques, particulièrement s'agissant des courtes durées d'agrégation (Bárdossy & Pegram, 2017; Gado et al., 2017; Kim et al., 2018). Finalement, les radars pluviométriques peuvent représenter une solution au manque de mesure au sol dans certaines régions du monde. Toutefois, en raison de leurs courtes périodes de fonctionnement et des incertitudes inhérentes à leurs mesures, le calage des estimations radar en l'analyse fréquentielle des extrêmes pluviométriques requiert une évaluation approfondie de leur utilisation, qui tient compte des aspects climatiques, géographiques ainsi que des échelles spatio-temporelles des mesures.

S'agissant des mesures des précipitations à l'aide des satellites, il existe deux types de télémessures satellitaires, active et passive. La première consiste en des satellites équipés de radars semblables à ceux installés sur le sol terrestre et présentant plus ou moins les mêmes avantages et inconvénients dans l'exploitation des données fournies. La deuxième, dite passive, consiste en la télémessure dans le domaine du visible, de l'infrarouge et des portions micro-ondes du spectre électromagnétique (Koutsoyiannis & Langousis, 2011). De plus, la durée d'observation minimale pour les mesures satellitaires est de 3 heures (Schiavo Bernardi et al., 2015), ce qui requiert l'utilisation d'un modèle statistique de désagrégation pour les durées inférieures à 3 heures, et implique forcément un degré d'incertitude supplémentaire difficilement quantifiable. A noter qu'une alternative à la télédétection radar et satellite a été proposée par Rios Gaona et al. (2015). Il s'agit de l'étude de la corrélation entre l'intensité des pluies et les signaux électromagnétiques émis par les réseaux de communication de téléphonie mobile. La puissance de ce signal décroît avec l'intensité de la pluie, ainsi son enregistrement peut être converti en intensité de pluie à haute résolution spatio-temporelle pour pallier aux manques de données précipitations infra-journalières dans les études statistiques sur les pluies extrêmes. Toutefois, de même que pour les radars et satellites, les incertitudes liées à ces données sont mal connues et difficilement quantifiables. De plus, même si le principe de la mesure a été démontré, il n'existe pas à notre connaissance d'application opérationnelle. Il faudrait pour cela convaincre des opérateurs de téléphonie mobile à exploiter en temps réel, leurs données opérationnelles en ce sens ou déployer des réseaux d'antennes dédiés à cette mesure, ce qui n'est pas économiquement viable comparativement à une instrumentation conventionnelle.

Comme souligné plus haut, le principal inconvénient de l'utilisation de la télédétection radar sont les fortes incertitudes sur les estimations quantitatives des extrêmes pluviométriques, particulièrement quand ces observations à distance sont de longueurs assez courtes (moins de 10 ans). Toutefois, ces incertitudes sont difficilement quantifiables, car les mesures au sol, qui peuvent servir à une évaluation quantitative des produits de télédétection à grande échelle, ont une représentativité limitée à une petite zone d'influence. Par contre, les longues séries de mesures au sol offrent la possibilité de quantifier empiriquement les incertitudes reliées aux analyses fréquentielles basées sur les télémessures (radar et satellite) (Marra et al., 2017).

En dépit de tout ce qui a été énoncé dans cette section, l'amélioration constante des techniques de mesure par imagerie radar ouvre de nouvelles perspectives dans la caractérisation des extrêmes pluviométriques. Spécifiquement dans la caractérisation spatio-temporelle des pluies et de la diffusion immédiate des données radar pour des applications en temps réels (Ricciardelli et al., 2014; Thorndahl et al., 2017). De plus, la concordance entre IDF radar et satellite (corrélation de 70% pour  $T < 5$ ans et une tendance à la diminution avec l'augmentation de la période de retour) sous climats Méditerranéen, et dans une certaine mesure arides et semi arides, laissent présager un potentiel remarquable pour des utilisations dans des zones non-jaugées (Marra et al., 2017).

### 2.3.1.3 Les générateurs stochastiques des séries chronologiques

Pour pallier le manque de mesures pluviographiques, il est possible d'utiliser des générateurs qui vont simuler de séries chronologiques de pluies horaires. Ces modèles peuvent être locaux ou spatialisés. Il existe deux types de générateur de pluies : les agrégateurs, où chaque averse est constituée d'une suite de cellules de durée et d'intensité aléatoires et leur addition donne la pluie voulue, ainsi que les compositeurs, où une juxtaposition d'entités représente un hyétogramme (Willems, 2001).

Les processus hydro-climatiques, tels que les précipitations, sont caractérisés par une variabilité spatio-temporelle. De nos jours, il est bien établi, que vu la complexité de la dynamique de ces processus se manifestant dans l'espace et le temps, assez souvent l'utilisation des modèles déterministes n'est pas envisageable. Ainsi, la seule alternative viable, réside dans une modélisation stochastique imitant certains aspects de la variabilité spatio-temporelle sans nécessairement s'astreindre à suivre les équations régissant le processus physique (Papalexiou & Serinaldi, 2020). Toutefois, pendant assez longtemps, la question de savoir si les séries chronologiques des variables hydro-climatiques, résultent de processus stochastiques ou de processus chaotiques déterministes à faible dimension, a prêté controverse (Gaume et al., 2006).

Au cours de ces dernières années, de nombreux modèles stochastiques pour la génération de séries chronologiques des pluies sur diverses échelles temporelles ont été utilisés. Ces modèles sont basés sur différents schémas tels que: la pulsation rectangulaire de Neuman-Scott (Olsson & Burlando, 2002), les processus autorégressifs générant des champs multidimensionnels (Ferraris et al., 2003), les cascades aléatoires simples (Molnar & Burlando, 2005), les cascades aléatoires multiplicatives reliées aux propriétés fractales (Gaume et al., 2007), le modèle Bartlett-Lewis modifié (Park et al., 2019). Il existe même un modèle, tel que le modèle original Bartlett-Lewis (OBL) (Ritschel et al., 2017), capable de préserver les caractéristiques statistiques des courbes IDF historiques, avec toutefois une limitation dans la reproduction d'évènement très rare. Ainsi, chacun de ces modèles sus cités est capable de préserver ou de reproduire une, deux ou plusieurs des propriétés des séries chronologiques des pluies, avec plus ou moins d'efficacité, sur une échelle de temps bien définie. Certaines études récemment publiées (Lombardo et al., 2017; Papalexiou et al., 2018; Tsoukalas et al., 2019), montrent des modèles capables de préserver un très grand nombre des propriétés des séries chronologiques des pluies, sur différentes échelles temporelles (journalières, mensuelles et annuelles), à savoir : les distributions marginales, les structures de corrélations (i.e. variabilité spatio-temporelle), l'intermittence. Néanmoins, ce genre de procédure nécessite un très grand nombre de données avec une réparation spatiale homogène.

Au regard de ce qui a été montré plus haut, l'utilisation des générateurs stochastiques des pluies horaires comme substitut aux données historiques dans l'analyse statistique des événements extrêmes, n'est pas encore acquise. En effet, pour être performant les modèles stochastiques doivent être calés sur un très grand nombre de données, dont c'est justement le manque qui pousse vers l'utilisation de données simulées. De plus les pluies sont beaucoup plus complexes que les modèles préconisaient par les approches stochastiques ou déterministes, et aucune des deux ne peut reproduire fidèlement toutes les caractéristiques des séries

chronologiques des pluies. Le choix du modèle dépendra donc des applications mises en œuvre (Gaume et al., 2006).

#### 2.3.1.4 Conclusion

Les mesures des pluies à distance, satellites ou radars, sont de plus en plus disponibles dans le monde. Toutefois, la faible qualité de leurs mesures au temps présents, font que les mesures au sol des pluies restent dans un futur proche le premier choix pour la plupart des applications hydrologiques. En effet, les mesures pluviométriques au sol fournissent les observations de pluies les plus précises possibles comparativement aux télémessures et aux simulations (Grimaldi et al., 2011; Willems, 2012). De plus, l'utilisation de radar à haute résolution et de données satellitaires ont jusqu'ici mis en évidence d'importantes inadéquations quantitatives d'estimation comparativement à des données de référence conventionnelles (Marra et al., 2018). Cependant les mesures satellites et radars fournissent des mesures des caractéristiques locales des averses (structure spatiale et temporelle par exemple) qui peuvent aider à une conception efficace des réseaux de mesures pluviométriques (Dai et al., 2017). Ainsi, ce présent travail sera exclusivement basé sur des mesures historiques au sol, même si, dans le contexte Algérien, leur disponibilité n'est pas quantitativement optimale, elles restent de très bonne qualité (cf. Chapitre 3 et 4). Ce qui pousse à faire usage des procédures et techniques les plus efficaces possibles dans le traitement quantitatif des données disponibles et surtout de tirer avantages de leurs propriétés (notamment l'invariance d'échelle et les structures multifractales).

#### 2.3.2 Les séries des maxima annuels (SMA)

La procédure d'échantillonnage la plus répandue, en analyse statistique des événements extrêmes, est la méthode des maxima annuels (aussi appelée méthode du maximum par blocs) (Meylan & Musy, 1999; Willems et al., 2012). La composition d'une série des maximums annuels consiste : à extraire d'une SDC de pluies toutes les valeurs maximales par année hydrologique, elle considère donc un événement par année. De plus, l'échantillonnage par maxima annuels est très fréquemment utilisé pour l'établissement de courbes IDF (Blanchet et al., 2016 ; Endreny & Imbeah, 2009 ; Ghanmi et al., 2016 ; Huang et al., 2016 ; Huard et al., 2010 ; Mohymont & Demarée, 2006 ; Muller et al., 2008 ; Overeem et al., 2009 ; Panthou et al., 2014 ; Van de Vyver, 2015a).

S'agissant des SMA pluviographiques, elles sont disponibles pour diverses durées d'agrégation allant de 1 minute à 24 heures. En Algérie, les mesures pluviographiques sont disponibles pour 8 durées d'agrégation : 15, 30, 60, 120, 180, 360, 720, 1440 minutes. Avec un dépouillement des pluviogrammes considérant des fenêtres temporelles glissantes afin de capturer la plus forte intensité de pluie mesurée sur une des durées de référence. Les événements pluvieux n'ayant pas tous la même durée, la taille des séries de durées complètes est inversement proportionnelle à la durée.

En pratique, l'échantillonnage basé sur les maxima annuels des pluies jouit d'une grande popularité, en raison entre autres, de la simplicité de sa structure et de l'exploitation directe du modèle de distribution fréquentielle (la période de retour correspond à l'inverse de la probabilité

au dépassement) (El Adlouni & Ouarda, 2010; Meylan & Musy, 1999) Aussi, du fait qu'une SMA est constituée d'une valeur par bloc annuel de mesure, ces valeurs maximales de pluies sont souvent espacées dans le temps de plusieurs semaines (voire plusieurs mois), ce qui garantit l'indépendance des données de la série (pas d'autocorrélation sérielle).

### 2.3.3 Les séries de durées partielles (SDP)

La procédure d'échantillonnage par série de durée partielle, aussi appelée méthode du renouvellement, a été introduite dès le milieu des années 30 à travers un certain nombre d'auteurs : Jarvis (1936), Langebin (1949), Borgman (1963), Shane et al. (1964), Bernier (1967), Todorovic et al. (1970), Cunnane (1973) et Rosbjerg (1977) (Rosbjerg & Madsen, 2004). Avec une utilisation initiale dédiée exclusivement à l'analyse fréquentielle des pics de crues, d'où l'appellation, aujourd'hui encore très répandue, de POT (*Peak Over Threshold* : pic dépassant un seuil). Par la suite, son utilisation a été généralisée aux analyses des extrêmes pluviométriques (Stedinger et al., 1993).

La procédure d'échantillonnage par SDP consiste à extraire toutes les valeurs de pluies, à partir d'un SDC, supérieures à un certain seuil préalablement établi. Ce n'est pas tant l'augmentation de la taille de la série qui rend une SDP plus cohérente avec la théorie des valeurs extrêmes mais plutôt l'usage exhaustif, qu'elle permet, de la donnée en termes d'extrême pluviométriques. D'ailleurs, certains auteurs (Chow, 1964; Koutsoyiannis et al., 1998; Rosbjerg & Madsen, 2004) recommandent de choisir un seuil de telle sorte que le nombre moyen de dépassement annuel soit égal à 1, ce qui donne un échantillon SDP de même taille qu'un échantillon SMA mais comptant plus de valeurs fortes.

La mise en œuvre d'une procédure d'échantillonnage par SDP nécessite le respect de ses hypothèses fondamentales à savoir : les nombres de dépassement par année doivent suivre une distribution de Poisson et la série ainsi formée doit être constituée de valeurs indépendantes et identiquement distribuées suivant une loi exponentielle (même s'il a été démontré empiriquement que le calage d'autre lois telles que la GPD est tout à fait acceptable). Ceci entraîne forcément des complications méthodologiques supplémentaires comparativement à une approche SMA et explique peut-être son manque d'attrait (Meylan & Musy, 1999). Toutefois, l'échantillonnage par SDP, bien qu'étant moins répandu que celui par SMA, est aussi très utilisé dans l'établissement des courbes IDF (Ben-Zvi, 2009; Boukhelifa et al., 2011; Kingumbi & Mailhot, 2010; Madsen et al., 2009; Svensson et al., 2007; Van de Vyver & Demarée, 2010; Willems, 2000)

### 2.3.4 SMA vs SDP : Quelle approche choisir ?

Dans la mise en œuvre d'une inférence statistique l'utilisation d'une SMA ne nécessite aucune préparation particulière, étant donné que son indépendance est assurée de par sa nature, avant la mise en œuvre. Tandis que, l'indépendance d'une SDP doit être vérifiée, et cette dernière est indirectement reliée au choix du seuil. En effet, le seuil doit être suffisamment faible pour garantir des performances au moins égales à une SMA, et pas trop faible afin d'éviter



une autocorrélation sérielle (dépendance temporelle). Le seuil sélectionné doit aussi permettre une occurrence Poissonnienne des dépassements de seuil par année. Les échantillons issus des SDP doivent être identiquement distribués. De nos jours, il est bien établi que le choix d'une loi exponentielle, souvent retenue par le passé, qui n'est autre que la transformation de la loi de Gumbel pour les SDP, n'est pas idéale dans l'analyse statistique des événements extrêmes (Koutsoyiannis, 2004a, 2004b). De plus, dans des analyses comparatives sur les SDP il a été démontré qu'une distribution généralisée de Pareto (GPD), qui n'est rien de plus qu'une transformation de la GEV en SDP, était un choix plus pertinent (Madsen et al., 1993). Ainsi, il paraît clair que pour assurer la supériorité de l'approche SDP, il est essentiel de valider ses hypothèses et qu'une concordance imparfaite avec ces dernières altérerait la précision des estimations. Ce qui vient d'être souligné explique la prévalence de l'utilisation de la SMA. De même que le manque d'un consensus sur les règles à suivre et les procédures à appliquer (Lang et al., 1999).

Un très grand nombre d'études (par exemple: Katz et al., 2002; Madsen, Pearson, et al., 1997; Madsen, Rasmussen, et al., 1997) ont établi que les approches SDP dans l'analyse fréquentielle des extrêmes pluviométriques pouvaient être à minima aussi performantes. A signaler aussi, qu'un petit nombre d'études (exemple: Hu et al., 2020; Vorstermans, 2019), au contraire notent un manque de performance des SDP devant les SMA. Toutefois, le plus souvent dans des analyses statistiques sur des séries de longueurs satisfaisantes, les procédures d'échantillonnage SMA et SDP présentent des performances équivalentes. Même si, la SDP montre plus de robustesse sur des séries courtes quand il y a une forte variabilité d'échantillonnage et dans l'estimation de grande période de retour (Coles & Pericchi, 2003). Nonobstant, un strict respect des hypothèses fondamentales de la procédure d'échantillonnage SDP. De plus, la sélection du seuil introduit une sorte de subjectivité dans l'analyse des valeurs extrêmes (Willems et al., 2012), de même qu'un choix de seuil souvent établi sur la base des données à ajuster (voire: Kingumbi & Mailhot, 2010; Svensson et al., 2007) pourrait introduire un certain biais d'estimation.

En dépit des complications mathématiques inhérentes à l'échantillonnage par séries de durées partielles, leur utilisation exhaustive des extrêmes pluviométriques mesurés, permet d'ajouter des informations, et apporte certainement un degré de précision supplémentaire dans l'estimation des quantiles. Néanmoins, leur application dans l'estimation intégrée (échantillon conjoint ou unique) des courbes IDF (comme définie en chapitre introductif et développée en §2.8), pose quelques problèmes. En effet, pour avoir un échantillon conjoint équilibré (c'est-à-dire le même nombre de valeurs issues de chaque série de durée d'agrégation  $d$ ) il est nécessaire de fixer un même nombre moyen de dépassement annuel pour chacune des séries sur les diverses durées d'agrégation. Or, ce ne sera certainement pas le choix optimal pour un certain nombre de séries et impliquera systématiquement une autocorrélation sérielle qui apportera une incertitude supplémentaire. Et a contrario, choisir un nombre moyen de dépassement optimal pour chaque série de durées  $d$  induira : si par exemple les cumuls des pluies de 15 minutes deviennent les plus représentés dans l'échantillon joint, ceci impliquera un déséquilibre dans l'échantillon qui peut mener à des sur et/ou sous-estimations des quantiles.

### 2.3.5 Conclusion

Au final, cette section a présenté les différentes observations pluviométriques mises à disposition des praticiens dans les analyses fréquentielles sur les précipitations extrêmes, ainsi que les diverses techniques d'échantillonnage. Une analyse pragmatique couplée à une présentation approfondie des avantages et limitations de chaque procédure ont permis d'opter pour l'utilisation d'observations de pluie historiques issues de mesures au sol, avec une extraction des extrêmes pluviométriques à travers un échantillonnage basé sur les maxima annuels (SMA). Notons, que certaines publications (Borga et al., 2005; Svensson & Jones, 2010) préconisent de prendre en considération les effets de saisonnalité (deux périodes : sèche et humide) et de traiter les pluies d'origines convectives et non-convectives séparément. Toutefois, l'existence de saisonnalité dans les précipitations extrêmes en Algérie ne sera pas prospectée dans ce travail, de plus l'effet de la saisonnalité a un impact très limité sur la queue de la distribution des valeurs extrêmes (Iliopoulou et al., 2018). Aussi, l'invariance d'échelle multifractale a été observée aussi bien dans les pluies d'origine convective que stratiforme (Willems, 2000).

## 2.4 Modélisation statistique des valeurs extrêmes des précipitations

Bien que la prédiction à court terme des pluies soit possible à un certain degré (prévision météorologique), la prédiction déterministe à long terme des pluies n'est pas faisable. Or les conceptions de diverses structures sont basées sur l'estimation des précipitations à très long terme. Ainsi, la pluie est étudiée de façon probabiliste où elle est considérée comme une variable aléatoire régie par des lois statistiques (Papalexiou et al., 2013). En effet, au-delà d'un certain horizon temporel même le système non-linéaire le plus simple tend vers un état d'équilibre statistique, ce qui permet de négliger les dynamiques déterministes. Dans cet état d'équilibre, la distribution de probabilités des propriétés du système (conditionnées selon l'état initial) est quasiment égale à la distribution de probabilités marginales (i.e. non conditionnées) des mêmes propriétés (Koutsoyiannis & Langousis, 2011).

Dans ce qui suit, il est retenu une approche par analyse fréquentielle univariée des extrêmes hydrologiques, d'ailleurs c'est la plus couramment utilisée dans la modélisation des extrêmes (Grimaldi et al., 2011). Une analyse fréquentielle sur les extrêmes pluviométriques permet d'extrapoler vers des valeurs rares, sur lesquelles seront basées les études de dimensionnement, à partir d'observations des précipitations sur différentes échelles de temps. Cette extrapolation est réalisée en ajustant une fonction de distribution de probabilités (communément appelée loi statistique), ou une combinaison d'ajustements (metastatistiques/ super statistiques), à un jeu de données. Cet ajustement, ou plus précisément l'inférence statistique, consiste à induire les caractéristiques de la population d'une variable aléatoire (la pluie dans notre cas) à partir de celles d'un échantillon (la série d'observation). Ainsi, les techniques les plus fréquemment utilisées dans la mise en œuvre de l'ensemble des procédures reliées à la modélisation statistique des pluies extrêmes, sont évaluées, dans ce qui suit, à travers une synthèse bibliographique sur des applications similaires.

Une fois que la distribution fréquentielle la plus adaptée, la technique d'inférence la plus pertinente et la méthode d'estimation des incertitudes la plus performante sont choisies, un état de l'art des formulations des courbes IDF est revu, afin d'établir leur modélisation de manière efficiente, profitant au mieux des propriétés des pluies, et adapté au contexte de l'étude. Aussi, la méthodologie déployée ici, assume une stationnarité des données sur lesquelles elle est implémentée. C'est-à-dire, en termes plus clairs, qu'il est assumé que la fonction de densité de probabilité sur des observations de pluies restera inchangée au cours du temps. La validité de cette hypothèse, dans l'analyse des extrêmes, est discutée dans la section (2.4.8).

### 2.4.1 Approche par analyse fréquentielle classique

L'analyse probabiliste des événements extrêmes a été introduite, pour la première fois, au début du 20<sup>ème</sup> siècle pour remplacer des procédures anciennes telles que les formulations empiriques ou la méthode des courbes enveloppes (Grimaldi et al., 2011). Par la suite, avec la disponibilité d'enregistrements plus longs et l'établissement de solides développements théoriques, à travers par exemple la publication de la théorie des valeurs extrêmes (Gumbel, 1958), l'approche standard en analyse fréquentielle a très vite été établie (Klimes, 1993).

Toutes les valeurs possibles que peut prendre une variable aléatoire, telle que la pluie, constituent la population de cette dernière. La distribution fréquentielle sous-jacente à cette population est dite distribution parente : en pratique elle n'est pas connue et ne peut être connue. L'analyse fréquentielle des événements extrêmes consiste à ajuster une fonction de distribution de probabilités, via un processus statistique inférentiel, à une série d'observations classées (supposées indépendantes et identiquement distribuées : i.i.d), puis à extrapoler la queue de cette distribution au-delà des valeurs observées, c'est-à-dire vers des probabilités au dépassement les plus faibles (grandes périodes de retour). En d'autres termes, elle permet la description de la vraisemblance des événements pluviométriques extrêmes sur une période donnée, quand les occurrences sont indépendantes dans le temps (i.e. que la magnitude et l'occurrence d'un événement ne montrent aucune relation avec l'événement précédent) (El Adlouni & Ouarda, 2010). Cependant, il faut toujours garder à l'esprit que l'analyse fréquentielle est un problème d'information et tant que les observations sont suffisamment riches alors il est possible de déterminer avec précision la distribution fréquentielle. En effet, dans la pratique l'ingénieur ne dispose qu'entre 10 à 100 années de mesures ce qui n'est pas assez pour définir précisément des événements très rares. Ce qui impose aux praticiens d'avoir recours à leurs expertises et à utiliser les techniques et procédures les plus robustes et les plus efficientes (Stedinger et al., 1993).

Les trois principales étapes à mettre en œuvre dans une analyse fréquentielle univariée sur la base des SMA sont : (1) la sélection d'un échantillon sous forme d'une série de données satisfaisant à un certain nombre de critères statistiques (stationnarité, indépendance, homogénéité, etc.), (2) le choix de la fonction de distribution de probabilités (Gumbel, GEV, Log-Normale, etc.) la plus adéquate (adéquation établie postérieurement à l'inférence à travers des tests statistiques d'adéquation), (3) application de la meilleure technique d'ajustement (Moments, L-moments, maximum de vraisemblance, etc.) disponible, sur la distribution

choisie, afin d'inférer statistiquement sur la population sous-jacente puis estimer l'évènement de période de retour  $T$ .

### Période de retour

Le concept probabiliste de période de retour a été en premier défini par Fuller (1914). Bien qu'étant inapproprié, l'appellation période de retour est très populaire aussi bien auprès des initiés que du grand public. Or, cette appellation peut induire une fausse compréhension, car le processus stochastique étudié n'est en aucun cas périodique (Gaume, 2018). La notion de période de retour est utilisée pour mesurer l'occurrence aléatoire d'évènements géophysiques (tels les précipitations extrêmes) pouvant produire d'importants dommages économiques, sociaux et environnementaux. Le but d'une analyse fréquentielle étant d'estimer un évènement de période de retour donnée, susceptible de déclencher la défaillance d'une structure/ ouvrage, pour s'en prémunir.

La période de retour est notée  $T$  et est exprimée en années. Elle est définie comme étant l'inverse de la probabilité annuelle de dépassement  $(1 - F(x_T))$  d'un quantile  $x_T$ . Cette dernière définition est valable en assumant que les séries d'observations sont stationnaires, indépendantes et identiquement distribuées (i.i.d). Toutefois, il a été démontré que la condition d'indépendance n'est pas strictement nécessaire (i.e. pour un certain degré de dépendance) pour appliquer cette définition (Volpi et al., 2015). La période de retour  $T$  est à tort comprise comme étant l'évènement survenant exactement chaque  $T$  année (Volpi, 2019). En effet, en considérant  $X$  comme la variable aléatoire étudiée, le nombre d'année  $\tau$  nécessaire pour observer  $X > x_T$  est défini comme l'intervalle de récurrence ou le temps écoulé entre deux dépassements (occurrences) successives, et a une espérance  $E[\tau] = \frac{1}{(1-F)} = T$ , avec  $T$  ne dépendant pas du temps étant donné que les évènements successifs sont indépendants (pour la démonstration de cette définition : cf. chapitre 12, page 382 de Chow et al. (1988)).

### La méthode des moments

La méthode des moments a été introduite en premier par Karl Pearson en 1912, où il considère que de bonnes estimations des paramètres  $\theta$  d'une fonction de distribution de probabilité sont ceux : pour lesquels les moments d'une fonction de densité de probabilité parente sont égaux aux moments correspondants d'une série de données (Chow et al., 1988). Ainsi les moments empiriques et les paramètres sont reliés pour une distribution donnée ( $\hat{\theta} = g(\hat{\mu}_x, \hat{\sigma}_x^2, \hat{\gamma}_x \dots)$ ) où  $\hat{\theta}$  est le vecteur des paramètres à estimer et  $\hat{\mu}_x, \hat{\sigma}_x^2, \hat{\gamma}_x$  sont les estimations respectivement de la moyenne, de la variance et de l'asymétrie. Ils sont utilisés pour décrire la tendance centrale (paramètre de position) d'une variable aléatoire, sa dispersion (paramètre d'échelle) et sa symétrie (paramètre de forme). Dans ce qui suit, il est donné l'expression des quatre premiers moments et de leurs estimateurs empiriques (Stedinger et al., 1993; Vogel & Fennessey, 1993):

1. Le moment ordinaire d'ordre 1 : l'espérance (moyenne) ;

$$\mu_x = E[X] ; \hat{\mu}_x = m = \bar{X} = \sum_{i=1}^n \frac{X_i}{n} \quad (2.2)$$

2. Le moment centré d'ordre 2 : la variance ;

$$\sigma_x^2 = Var[X] = E[(X - \mu_x)^2] ; \hat{\sigma}_x^2 = s^2 = \left[ \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n} \right] \quad (2.3)$$

3. Le moment centré réduit d'ordre 3 : l'asymétrie (skewness en anglais) ;

$$\gamma_x = E \left[ \frac{(X - \mu_x)^3}{\sigma_x^3} \right] ; \hat{\gamma}_x = G = \frac{\sum_{i=1}^n (X_i - \bar{X})^3}{n s^3} \quad (2.4)$$

4. Le moment centré réduite d'ordre 4 : l'aplatissement ou l'acuité (Kurtosis)

$$\kappa_x = E \left[ \frac{(X - \mu_x)^4}{\sigma_x^4} \right] ; \hat{\kappa}_x = k = \frac{\sum_{i=1}^n (X_i - \bar{X})^4}{n s^4} \quad (2.5)$$

Actuellement, en dépit de ses multiples limitations, la méthode des moments est encore très utilisée en raison de sa simplicité de calcul et de l'interprétation communément comprise de ses 4 premiers moments, de la facile disponibilité de logiciels en statistiques (HydroLab, Hyfran, etc.) l'utilisant/proposant dans leur calcul d'inférence statistique. En effet, il est bien connu que pour des échantillons de petites tailles ( $n < 100$ ), les estimateurs de la variance  $s^2$  et de l'asymétrie  $G$  montrent une variance et un biais qui peuvent être très élevés. Ainsi, quand les échantillons étudiés ne sont pas très fournis, l'égalisation entre moments empiriques et moments théoriques ne tient plus, ce qui peut mener à de très grandes incertitudes d'estimation (Wasserman, 2004). Même si de nombreuses tentatives (Vogel & McMartin, 1991) pour établir une formulation non-biaisée ont été proposées, certaines permettent tout juste une réduction limitée du biais, car les propriétés d'échantillonnage de ces estimateurs de moments, dépendent de la distribution. De plus, les efforts pour une estimation non biaisée de l'asymétrie mènent à introduire des biais d'estimation des autres moments et conduisent à augmenter la variance des estimateurs (Wallis et al., 1974). Enfin, l'utilisation de la méthode des moments n'est pas du tout recommandée pour des ajustements de la loi GEV (Lubes & Masson, 1991), qui est souvent utilisée dans l'analyse fréquentielle des extrêmes pluviométriques. Aussi, il est à signaler que les coefficients d'asymétrie et de variation sont encore plus biaisés pour des estimations sur une population fortement asymétrique quelle que soit la taille de l'échantillon (Vogel & Fennessey, 1993). Une autre limitation est que les estimateurs des moments sont bornés, et que leur valeur maximale dépend de la taille de l'échantillon (Kirby, 1974).

### La méthode des L-moments

La méthode des L-moments a été formalisée par Hosking (1990). Comme leur nom l'indique c'est une combinaison linéaire des moments de probabilité pondérée (*Probability Weighted Moments* – PWM, en anglais), offrant une manière différente de résumer les propriétés statistiques des mesures hydrologiques. Dans ce qui suit, l'expression des estimateurs des quatre premiers PWM tels que formalisés en premier par Greenwood et al. (1979) (pour un classement décroissant de l'échantillon de taille  $n$ , avec  $i = 1$  étant la plus grande valeur et  $b_i$  l'estimateur du moment de probabilité pondérée  $\beta_i$ ) :

$$b_0 = \frac{1}{n} \sum_{i=1}^n X_i \quad (2.6)$$

$$b_1 = \sum_{i=1}^{n-1} \left[ \frac{(n-i)}{n(n-1)} \right] X_i \quad (2.7)$$

$$b_2 = \sum_{i=1}^{n-2} \left[ \frac{(n-i)(n-i-1)}{n(n-1)(n-2)} \right] X_i \quad (2.8)$$

$$b_3 = \sum_{i=1}^{n-3} \left[ \frac{(n-i)(n-i-1)(n-i-2)}{n(n-1)(n-2)(n-3)} \right] X_i \quad (2.9)$$

Les moments de probabilité pondérée utilisent des pondérations de la fonction de distribution cumulative où il est difficile d'interpréter le moment en tant que paramètre d'échelle ou de forme d'une fonction de distribution de probabilité donnée (Hosking & Wallis, 1997). Ainsi, la combinaison linéaire des PWM en L-moments permet de calculer des moments et ratios de moment qui sont faciles à interpréter et trouve aussi toute leur utilité dans le calcul des paramètres des distributions fréquentielles. En effet, les L-moments sont analogues aux moments ordinaires, car produisant le même type de mesures : position (moyenne), échelle (variance), forme (asymétrie) et aplatissement (Kurtosis). Sachant que  $X_{(i:n)}$  est la  $i^{\text{ème}}$  plus grande valeur d'un échantillon de taille  $n$  ( $i=1$  étant la plus grande valeur), les expressions des quatre premiers L-moments sont données dans ce qui suit (Hosking & Wallis, 1997) :

$$\lambda_1 = E[X] ; \lambda_1 = \beta_0 \quad (2.10)$$

$$\lambda_2 = \frac{1}{2} E[X_{(1:2)} - X_{(2:2)}] ; \lambda_2 = 2\beta_1 - \beta_0 \quad (2.11)$$

$$\lambda_3 = \frac{1}{3} E[X_{(1:3)} - X_{(2:3)} + X_{(3:3)}] ; \lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0 \quad (2.12)$$

$$\lambda_4 = \frac{1}{4} E[X_{(1:4)} - 3X_{(2:4)} + 3X_{(3:4)} - X_{(4:4)}] ; \lambda_4 = 20\beta_3 - 20\beta_2 + 12\beta_1 - \beta_0 \quad (2.13)$$

Aussi, Hosking (1990) définit les ratios des L-moments comme : L-coefficient de variation (L-C<sub>v</sub>), L-coefficient d'asymétrie (L-C<sub>s</sub>) et L-coefficient d'aplatissement (L-C<sub>k</sub>) :

$$LC_v = \tau_2 = \frac{\lambda_2}{\lambda_1} \quad (2.14)$$

$$LC_s = \tau_3 = \frac{\lambda_3}{\lambda_2} \quad (2.15)$$

$$LC_k = \tau_4 = \frac{\lambda_4}{\lambda_2} \quad (2.16)$$

Les L-moments ont été popularisés à travers l'ouvrage de Hosking & Wallis (1997) où est développée une méthodologie d'analyse fréquentielle régionale essentiellement basée sur les L-moments et leurs ratios, mais aussi en raison d'un certain nombre d'avantages comparativement aux méthodes des moments ordinaires et du maximum de vraisemblance :

1. Les PMW sur lesquels sont basés les L-moments montrent une moindre sensibilité à l'autocorrélation sérielle comparativement aux méthodes des moments ordinaires et du maximum de vraisemblance (Landwehr et al., 1979).
2. Les ratios des L-moments L-C<sub>v</sub> et L-C<sub>s</sub> sont quasiment non-biaisés contrairement aux moments ordinaires qui sont fortement biaisés et hautement variables (Hosking, 1990).
3. Les L-moments permettent une estimation très robuste en présence de valeurs très fortes où l'information fournie par le reste de l'échantillon ne sera pas occultée comme c'est le cas pour la méthode des moments classiques (c'est-à-dire qu'il n'y pas de mise au carré et au cube des valeurs d'échantillons) (Stedinger et al., 1993).
4. Les ratios L-C<sub>v</sub> et L-C<sub>s</sub> sont aussi bornés mais leurs limites ne dépendent pas de la taille de la série contrairement aux moments ordinaires (Zafirakou-Koulouris et al., 1998).
5. Ils présentent de bonnes performances sur les échantillons de petites tailles. Aussi, ils peuvent servir en tant que valeur de départ pour un calcul numérique par itération du maximum de vraisemblance (Katz et al., 2002).
6. Ils présentent une grande robustesse en présence d'horsains et utilisables sur une très large gamme de lois statistiques (Kyselý et al., 2007).
7. Une moindre sensibilité à la variabilité d'échantillonnage comparativement aux moments ordinaires et mieux adaptés, du moins théoriquement, aux lois statistiques à trois paramètres (Gaume, 2018).

### Maximum de vraisemblance

La méthode du maximum de vraisemblance (MLEs : *Maximum Likelihood Estimators*, en anglais) a été développée par Ronald Aymler Fisher (1922). Selon lui la meilleure estimation des paramètres – qui sont inconnus –, d'une fonction de distribution de probabilités, sont ceux qui maximisent la vraisemblance (ou le produit des probabilités conjointes des densités de chaque élément de l'échantillon i.i.d) d'apparition de l'échantillon observé. En admettant  $\theta$  comme les paramètres d'une fonction de densité  $f$  d'une loi de probabilité, la vraisemblance s'écrit :

$$L = \ell(X|\theta) = \prod_{i=1}^n f_{\theta}(X_i) \quad (2.17)$$

Le passage sous forme logarithmique, qui atteint son maximum simultanément, est plus pratique du fait que beaucoup de fonctions de densité ont une forme exponentielle, aussi la substitution du produit par la somme permet de simplifier les opérations. La maximisation du log-vraisemblance s'écrit alors :

$$\ln \ell(X|\theta) = \sum_{i=1}^n \ln f_{\theta}(X_i) \quad (2.18)$$

La méthode du maximum de vraisemblance est théoriquement l'approche la plus correcte pour l'ajustement d'une fonction de distribution de probabilité à un jeu de données, dans le sens où elle produit des estimations de paramètres non biaisées et très efficaces (c'est-à-dire ceux

permettant d'estimer les paramètres de la population avec l'erreur moyenne la plus faible possible). Toutefois, pour quelques distributions fréquentielles, il n'existe pas de solution analytique pour tous les paramètres sur la base des caractéristiques de l'échantillon, et le log-vraisemblance doit être maximisé numériquement, ce qui peut être compliqué (Chow et al., 1988). Par exemple en considérant une loi de Gumbel, le calcul de la log-vraisemblance revient à résoudre un système à deux équations aux dérivées partielles, de la log-vraisemblance, respectivement par rapport aux paramètres de position et d'échelle. Une des premières solutions a été établie par Kimball (cf. Gumbel (1958)), et elle consiste en un calcul itératif par tabulation ou par ordinateur. Pour ce qui est de la loi GEV, une des solutions numériques les plus efficaces dans l'estimation des paramètres par maximum de vraisemblance est celle proposée par Smith (1985) et testée par Laio (2004) dans le cadre de la théorie des valeurs extrêmes. Cette dernière solution peut être implémentée sous le logiciel de programmation statistique R à l'aide du package nsRFA (Viglione et al., 2020).

Les MLEs produisent de très bonne estimation des caractéristiques statistiques sur des échantillons de grandes tailles. Par contre, quand l'échantillon est de petite taille l'estimation des paramètres peut être biaisée (Meylan & Musy, 1999). Aussi, ils sont quelques fois peu performants quand la distribution expérimentale (observations) dévie significativement de la distribution théorique à ajuster (Stedinger et al., 1993). De plus, Gumbel (1958) note que les MLEs donnent plus de poids aux faibles valeurs de l'échantillon.

#### **Autres méthodes : Graphique, Régression, K-moments.**

Parmi les plus anciennes procédures d'ajustement de lois statistiques à un jeu données, il est possible de citer la méthode graphique qui consiste à : reporter sur du papier semi-logarithmique les fréquences expérimentales (Hazen, Weibull, Gringorten, etc.) des valeurs observées, ensuite à transformer l'expression d'un quantile (selon une distribution préalablement sélectionnée) en une relation de droite passant au plus près des points expérimentaux, puis en déduire la pente et l'intercept qui correspondent alors aux paramètres de la loi statistique. Bien entendu cette approche n'est utilisable que pour les lois statistiques à deux paramètres (Gauss, Log-Normale, Gumbel, etc.). La subjectivité, les imprécisions et les larges incertitudes résultant de l'application de cette approche ne sont plus à démontrer. Même, si elle a rendu grand service par le passé, aujourd'hui elle n'est pratiquement plus utilisée.

D'autres méthodes dont l'utilisation a été très largement répandue sont celles basées sur les techniques de régression-corrélation. En effet, il est possible d'estimer des quantiles de fréquence donnée, par régression linéaire (de type moindres carrés) ou non linéaire. Ce genre d'approche, mène très souvent vers un croisement de quantile (Ouali et al., 2016). Quand la progression de la régression est différente de celle de la distribution fréquentielle, par exemple : il est possible que le quantile décennal estimé par régression prenne une valeur supérieure au quantile double décennal estimé à travers la distribution de probabilité. Certes, au prix d'une plus grande complexité mathématique certaines techniques de régression peuvent pallier ce problème, telle que la *monotone composite quantile regression neural network* (MCQRNN)

(Cannon, 2018). Cependant, l'utilisation d'une régression dans l'inférence statistique accentue les incertitudes d'estimation. En effet, la méthode des moindres carrés présente des



intervalles de confiance, sur les paramètres des distributions fréquentielles, beaucoup plus larges que quelconques autres techniques d'ajustement paramétriques.

A signaler que les méthodes des moments et des L-moments comptent une multitude de dérivés, tels que les PWM partiels ; LL-moment ; LH-moment, etc., apportant plus ou moins une plus-value en fonction du contexte de leur application. Toutefois, l'approche proposée par Koutsoyiannis (2019) mérite d'être explicitée plus en détail. En effet, l'auteur part du principe que les moments ordinaires et centrés expriment d'importantes propriétés théoriques des distributions de probabilités, et qu'il est difficile d'en faire des estimations précises, à partir d'échantillons hydrologiques, pour des ordres supérieurs à 2. Et que les L-moments, fournissent de meilleures estimations que les moments ordinaires, mais sont toutefois tous du premier ordre en terme du processus de modélisation stochastique. En piochant des concepts des moments et L-moments, l'auteur introduit les K-moments (*knowable* – moments) qui combinent les avantages (et contiennent virtuellement) les deux catégories. En quelques sortes les K-moments sont les L-moments d'ordre supérieur. Ils permettent l'estimation des statistiques d'ordres élevés qui trouvent toute leur utilité dans les processus stochastiques ainsi que dans l'étude des distributions marginales et conjointes.

#### **2.4.2 Perspective Bayésienne et procédure Markov Chain Monte Carlo (MCMC)**

Bien qu'étant basée sur un solide raisonnement statistique, la méthode du maximum de vraisemblance, résolue par calcul numérique itératif, éprouve quelques difficultés envers l'inférence sur des échantillons relativement courts. En effet, selon Martins & Stedinger (2000), plusieurs études faites dans les années 1990, montrent que l'estimation des paramètres de la GEV, sur des échantillons de petites tailles, par la méthode du maximum de vraisemblance est souvent instable et il est plutôt recommandé une estimation basée sur les L-moments (supposément moins biaisés et de plus faible variance). Par la suite d'autres études ont montré que pour un paramètre de forme allant de -0.25 à +0.3 les estimations par méthodes des moments donnaient une erreur quadratique moyenne plus faible que les L-moments et maximum de vraisemblance pour de petits échantillons

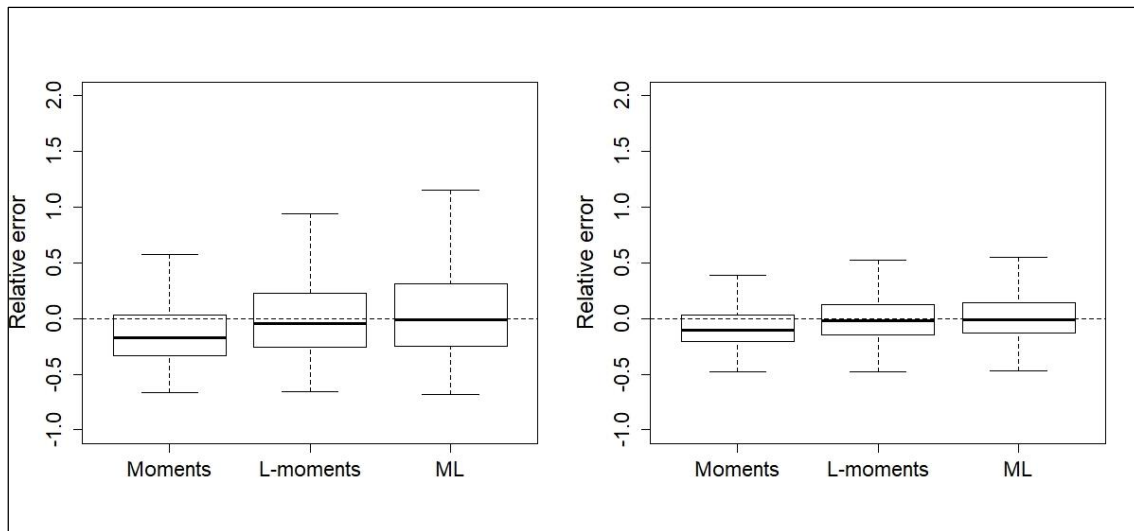


Figure 2.2 Boxplot de l'erreur relative sur l'estimation des quantiles 100 ans pour trois méthodes standard d'inférence. Un total de 1000 séries tirées dans une distribution parente GEV (10, 10, -0.2). A gauche  $N=30$  ans et à droite  $N=100$  ans.

De même, Gaume (2018) en calculant les erreurs relatives entre un quantile centennial, issu d'une distribution parente, prise comme référence, et des quantiles centennaux issus d'un jeu de séries simulées (1000 échantillons tirés dans la distribution de référence) de deux longueurs différentes (30 et 100 ans) (figure 2.2) a trouvé que : (1) les estimateurs basés sur le maximum de vraisemblance sont quasiment non-biaisés, tandis que ceux basés sur les moments ordinaires sont biaisés avec une tendance à la sous-estimation du quantile 100 ans pour cette distribution spécifique, (2) les estimateurs basés sur le maximum de vraisemblance ont la plus grande variance pour les échantillons de 30 ans, mais celle-ci décroît très vite avec l'augmentation de la taille de l'échantillon (c'est asymptotiquement l'estimateur avec la plus faible variance), (3) une contradiction avec le canon établi que les L-moments ont une variance plus faible que les moments ordinaires, toutefois ceci peut s'expliquer par le fait de la perte de la faible sensibilité à la variabilité d'échantillonnage des L-moments en raison de la fonction non linéaire qui les relie avec les paramètres et les quantiles de la distribution sous-jacente, (4) le fait le plus notable sur les MLEs est que l'erreur relative d'estimation atteint 50% et plus pour les jeux de données de 30 ans (relativement courts).

Au regard des complications, en termes de biais, de sensibilité à la variabilité d'échantillonnage et d'incertitudes que peut avoir l'utilisation des méthodes d'inférence standards (Moments, L-moments, Maximum de vraisemblance), il est primordial voire avantageux de dépasser ces limitations. En effet, il existe une autre approche incorporant une philosophie complètement différente. Il s'agit de l'inférence Bayésienne.

### Bayesian MCMC

En analyse fréquentielle sur les variables hydrologiques, la perspective Bayésienne combine le théorème de Bayes, la formulation de la vraisemblance (voir §2.4.1) et une procédure inférentielle Monte Carlo Markov Chain (MCMC) (cf. l'ouvrage de Robert & Casella (2004)) pour une résolution numérique optimale de l'ensemble. Ainsi, la procédure

Bayésienne MCMC permet d'explorer de façon complète la distribution fréquentielle des paramètres  $\theta$  d'une fonction de distribution de probabilité donnée. En effet, comme définie précédemment, la méthode du maximum de vraisemblance consiste à trouver le jeu de paramètres  $\theta$  qui maximise la valeur de la vraisemblance  $l(D|\theta)$ . En considérant le théorème de Bayes, il est possible d'écrire l'expression de la distribution des paramètres conditionnellement au jeu d'observation  $D$  (Gaume et al., 2010; Olivier Payraastre et al., 2013):

$$p(\theta|D) = \frac{l(D|\theta)p(\theta)}{p(D)} \quad (2.19)$$

Avec  $p(\theta|D)$  représentant la distribution à posteriori ou plus précisément la densité de probabilité marginale du vecteur de paramètres  $\theta$  sachant le jeu de donnée  $D$ , la probabilité  $p(\theta)$  correspond à une information alternative disponible à priori sur la distribution des paramètres  $\theta$ , la constante  $p(D)$  est la probabilité de l'échantillon  $D$  et n'influe pas sur le résultat (elle est inconnue). Dans l'inférence Bayésienne mise en œuvre dans ce travail il est considéré un Prior non informatif, c'est-à-dire qu'il n'y a pas d'information a priori disponible sur la distribution des paramètres  $\theta$ ,  $p(\theta)$  est choisie uniforme et égale à 1. Ainsi, ce cas de figure implique que  $p(\theta|D)$  est proportionnelle à  $l(D|\theta)$  (Gaume et al., 2010).

Le résultat de cette inférence n'est pas un vecteur optimal mais sa fonction de densité à posteriori. L'algorithme MCMC (sur l'équation 2.19) permet d'échantillonner un grand nombre de vecteurs de paramètres  $\theta$ . Alors, si le problème d'inférence est correctement posé, l'étendue de la distribution postérieure devrait diminuer à mesure que le contenu en information de  $D$  augmente (Tanner, 1996). Cette dernière caractéristique est justement exploitée dans la formulation des relations IDF (§2.8.4) – pour une estimation efficiente de l'incertitude –, où le modèle sera calé sur un jeu de données plus grand (unique), car constitué des mesures de cumuls de pluies sur toutes les durées d'agrégation. Toutefois, l'utilisation d'un échantillon conjoint ne peut être possible que si les observations ne rejettent pas l'hypothèse d'invariance d'échelle et montrent un comportement fractal (cf. Chapitre 4 de la présente thèse).

En inférence Bayésienne, les paramètres  $\theta$  sont considérés en tant que variables aléatoires et l'incertitude qui leur est associée peut être explicitement modélisée, et donc permettre d'assigner des intervalles de crédibilité (qui représentent l'analogie Bayésienne aux intervalles de confiance en statistique fréquentiste) aux paramètres de la distribution fréquentielle considérée et aux quantiles correspondants (Viglione et al., 2013). La constante de proportionnalité  $p(D)$  n'influence pas le résultat, dans le sens où le calcul de  $l(D|\theta)$  ou de  $l(D|\theta)p(\theta)$  (quand une distribution a priori est retenue) est suffisant pour échantillonner des vecteurs  $\theta$  à partir d'une distribution avec une fonction de densité  $p(\theta|D)$ . Le résultat final d'une procédure MCMC est l'ensemble de jeux de vecteurs  $\theta$ , typiquement de l'ordre de la dizaines de milliers, avec une densité  $p(\theta|D)$  (Gaume et al., 2010). Aussi, la constante  $p(D)$ , qui représente la probabilité d'un échantillon  $D$  peut être estimée numériquement dans des cas simple mais s'avère beaucoup plus difficile à calculer dans des cas plus compliqués (Payraastre, 2005). D'où l'intérêt d'un algorithme MCMC qui permet d'échantillonner des jeux sans avoir à calculer la constante  $p(D)$  (Reis & Stedinger, 2005).

L'approche Bayésienne MCMC décrite présentement permet de combiner l'utilisation de données observées localement (stations de mesure PV, PG, hydrométriques, etc.) avec des

informations supplémentaires indépendantes des mesures, tels que les niveaux d'eau exceptionnels sous les ponts, les laisses des crues, etc., les lecteurs intéressés peuvent se référer à plusieurs publications qui traitent de cet aspect (Gaume et al., 2010; Halbert et al., 2016; Nguyen et al., 2014; Payrastra et al., 2011; Viglione et al., 2013). Dans ce travail de thèse, cette possibilité n'est pas explorée, car l'objectif demeurant de trouver un modèle de courbes IDF qui puisse être calé sur des échantillons de données relativement courts avec un degré de fiabilité acceptable ou du moins avec une estimation de l'incertitude relativement précise.

### Algorithme de Metropolis-Hastings

Les algorithmes MCMC, combinant la marche aléatoire (*random walk*, en anglais) de la méthode Monte Carlo et les chaînes de Markov, constituent la dernière étape de l'inférence Bayésienne ; ils sont définis comme une classe d'algorithmes pour un échantillonnage efficace à partir d'une distribution aléatoire multivariée (Gaume et al., 2010). Le principe de la méthode MCMC a été proposé par Metropolis et al. (1953) et Hastings (1970). Payrastra (2005), résume le principe de simulation par algorithme de Metropolis-Hastings pour une procédure MCMC dans une inférence Bayésienne comme suit :

« Dans un espace à  $N$  dimensions, considérant une fonction de transition  $g$  symétrique telle que  $g(x|y) = g(y|x)$ , et une densité de probabilité  $\pi(x)$ , cet algorithme consiste à construire une chaîne de Markov suivante :

- A chaque itération  $n$ , on génère une valeur candidate  $y$  pour  $X_{n+1}$  avec la fonction  $g$ , conditionnellement à  $X_n = x$ .
- La probabilité d'effectuer la transition vers la valeur candidate est égale à  $\alpha(x, y) = \min\left(\frac{\pi(y)}{\pi(x)}, 1\right)$ . Pour ce faire on génère une valeur aléatoire  $t$  sur  $[0, 1]$  à partir d'une loi uniforme, et le saut est effectué si  $\alpha(x, y) > t$ . Si la transition est retenue,  $X_{n+1} = y$ ,  $X_{n+1} = x$  dans le cas contraire.

L'algorithme ainsi défini renvoie une série de valeurs de  $x$  dont on peut montrer que la distribution statistique à l'équilibre a pour densité  $\pi(x)$ . La vitesse de convergence dépend étroitement de la fonction de transition  $g$ .

Dans notre travail, en considérant que  $x$  est le vecteur des paramètres de la distribution à ajuster, et que la densité  $\pi(x)$  est la densité de probabilité a posteriori des ces paramètres  $p(\theta|D)$ , l'algorithme va permettre de générer une série de vecteurs de paramètres dont la densité est  $p(\theta|D)$ .

La fonction de transition, pour sa part, peut être définie de la façon suivante : à partir d'un vecteur  $\theta$  de paramètres, la nouvelle valeur de chaque paramètre est générée par un tirage dans une loi normale centrée sur la valeur initiale du paramètre. Un écart type différent peut être retenu pour chacun des paramètres. En adaptant la valeur de ces écarts types, on modifiera la vitesse de convergence de l'algorithme.

Une fois l'échantillonnage des vecteurs de paramètres ainsi effectué, il est aisé de calculer les valeurs des quantiles associées à chaque jeu de paramètres, puis d'en déduire l'intervalle de crédibilité, sur les valeurs des quantiles.

Dans ce travail l'algorithme Bayesian MCMC est mis en œuvre sous le logiciel de programmation statistique R en utilisant la librairie nsRFA (Viglione et al., 2020).

### Convergence de l'algorithme MCMC

L'utilisation dans une simulation numérique par algorithme MCMC nécessite la vérification de sa convergence pour chaque application. S'il est bien établi que l'algorithme MCMC ne pose que très rarement des problèmes de convergence, dans l'estimation des 3 paramètres d'une distribution GEV par exemple, ce n'est pas le cas d'une relation de vraisemblance avec 4 paramètres (voir §2.8.4). Le programme MCMC utilisé ici a connu plusieurs phases de développement (Gaume et al., 2010; Halbert et al., 2016; Nguyen et al., 2014; Payrastra et al., 2011), et sa convergence a été testée pour un nombre d'itération entre 40 000 et 300 000, pour une vraisemblance à 4 paramètres (3 pour la GEV et 1 pour le modèle régional) (Nguyen et al., 2014). Les auteurs ont conclu que l'algorithme MCMC convergeait pour un minimum de 30 000 itérations. Etant donné, que dans ce travail de thèse le modèle de courbe IDF proposé à une vraisemblance analogue, quelques essais (de 30 000 à 200 000 itérations) sur la convergence de l'algorithme MCMC, ont permis de conclure que la recommandation d'un minimum de 30 000 itérations représente un bon compromis.

### 2.4.3 Approches alternatives

#### 2.4.3.1 Analyse fréquentielle méta-statistique

L'approche des valeurs extrêmes metastatistiques (MEV : *metastatistical extreme value*) est définie comme la distribution des paramètres statistiques décrivant l'occurrence et l'intensité des pluies ordinaires (c'est à dire la distribution statistique sur toute la période d'observation des distributions statistiques des pluies ordinaires sur une année) (Marani & Ignaccolo, 2015). Un évènement ordinaire peut être défini comme étant le maximum de pluie, sur une certaine durée (de quelques minutes à un jour), sur une période de temps sans aucun 0 de cumul de pluie et suffisamment espacé avec les autres périodes pour assurer une indépendance (Marra et al., 2018).

Les étapes d'application des metastatistiques consistent à identifier les évènements de pluies ordinaires pour chaque année d'observations, ajuster les évènements de chaque année séparément à une distribution de Weibull, établir une distribution cumulative des pluies journalières maximales correspondant à chaque année en considérant trois paramètres statistiques (Forme, échelle, nombre de jour humide), déduire une expression approximative qui est la moyenne de distributions cumulées de chaque année, inverser la distribution cumulative moyenne (distributions metastatistiques) pour calculer les quantiles. Tester sur des séries de pluies journalières, la MEV semble surpasser la théorie des valeurs extrêmes pour l'estimation de quantile de période de retour supérieure à la durée d'observation des pluies

(Zorzetto et al., 2016). Toutefois, quand seulement moins de 20 années d'observation sont disponibles la MEV a tendance à sous-estimer les quantiles de périodes de retour 100 ans mais à réduire l'incertitude (jusqu'à moins 30%) comparativement aux SMA et SDP. La méthode MEV est moins sensible, comparativement à la théorie des valeurs extrêmes, aux lacunes dans les séries et aux erreurs multiplicatives/ additives (erreurs d'observation) pour des estimations sur des séries courtes (jusqu'à 5 ans) (Marra et al., 2018).

Les premières études sur les approches metastatistiques ont fait ressortir que ces dernières présentent une modélisation plus stable que l'approche classique d'analyse fréquentielle sur la base d'un SMA. Cependant, stabilité ne signifie pas forcément plus de précision ou de fiabilité, aussi l'incohérence théorique de vouloir estimer des extrêmes journaliers (ou infra journaliers) sur la base de valeurs courantes (tous les enregistrements) reste très discutable. En effet, dans une étude comparative sur les effets de la paramétrisation stochastique sur les approches statistiques appliquées aux valeurs extrêmes, montrent que les performances de la paramétrisation stochastique doivent précisément et spécifiquement être évaluées sur les événements extrêmes (Hu et al., 2019)

#### 2.4.3.2 *Analyse fréquentielle super-statistique*

La variabilité temporelle des paramètres d'une distribution fréquentielle, est depuis une vingtaine d'années reconnue dans littérature scientifique (Beck & Cohen, 2003) comme étant une distribution super-statistique (ou composée ou contagieuse). C'est-à-dire les statistiques des statistiques : ce qui signifie qu'il y a un ajout d'une distribution de probabilité qui caractérise les paramètres d'une première distribution de probabilité qui traite d'une variable aléatoire. C'est-à-dire que les paramètres, d'une loi statistique ajustée à un jeu de données, sont considérés comme étant des variables aléatoires pouvant être décrites par une distribution de fréquence.

La modélisation statistique des précipitations par approche super-statistique a été introduite par De Michele & Avanzi (2018). La procédure proposée est en deux étapes : la première consiste à identifier la distribution exacte des SMA à partir d'une chaîne de Markov dans une forme analytique fermée, sous hypothèse d'indépendance stochastique. La deuxième étape consiste à formuler la conjecture super-statistique des pluies journalières, ce qui signifie qu'on assume que les paramètres de distribution exacte varient d'une année à l'autre selon la probabilité de distribution (ce qui est soutenu par des preuves empiriques). Les premiers résultats montrent que la distribution super-statistique est plus robuste que la distribution GEV. Selon les auteurs (Carlo De Michele & Avanzi, 2018) la GEV a tendance à sous-estimer les quantiles de grande périodes de retour (qui ne sont d'aucun intérêt pratique) de 7-13%, tandis que la distribution super-statistique tend à les surestimer.

#### 2.4.3.3 *Autres méthodes : formulation non paramétrique, graphique et stochastique*

Les intensités, durées et fréquences des pluies extrêmes peuvent être obtenues par d'autres méthodes telles que : pluie maximum probable, courbes enveloppes, générateurs stochastiques d'évènement pluvieux (pulsation de Neuman-Scott, Bartlett-Lewis, réduction d'échelle

temporelle sur la base des propriétés fractales des pluies, etc.). Dans ce qui suit, il est décrit de façon brève, quelques approches alternatives au processus statistique inférentiel pour la caractérisation des extrêmes pluviométriques.

### **Formulation non paramétrique**

Occasionnellement, des méthodes non paramétriques sont utilisées dans l'estimation des distributions fréquentielles, et ne font pas la supposition que les pluies découlent d'une famille particulière de lois statistiques. Seulement, elles sont très peu utilisées en pratique (Plotting position). Dans ce qui suit une description succincte de quelques approches non paramétriques qui peuvent être utiles dans certaines circonstances particulières.

Le concept de période de retour et son estimation étant un concept pivot dans la gestion des risques pour diverses applications géophysiques, Volpi et al. (2019) proposent de baser les estimations de valeurs de pluies d'une période de retour sur la base des séries de durées complètes (*complete time series analyses CTA*) sans décimation des données. Les auteurs mènent une étude comparative entre les fonctions de distributions de probabilités empiriques entre les séries des maxima annuelles et les CTA. Cette approche repose sur l'hypothèse que le temps d'occurrence moyen (période de retour) entre deux événements potentiellement dommageables n'est pas affecté par la structure de dépendance du processus sous-jacent. La CTA est plus conservatrice que les SMA dans le sens qu'à période de retour égale, la CTA donne des valeurs plus faibles avec une plus grande incertitude. Les auteurs trouvent que la CTA donne des valeurs plus précises relativement à la distribution parente. La SMA mène systématiquement à des surestimations des périodes de retour, petites à moyennes, et à une sous-estimation de la variabilité de l'échantillonnage. La CTA et SMA sont supposées fournir les mêmes résultats pour la queue supérieure de la distribution (*upper tail*) soutenant l'utilisation de l'analyse fréquentielle sur les valeurs extrêmes dans l'estimation des grandes périodes de retour. Toutefois, l'utilisation de CTA dans l'inférence des distributions statistiques dans le cadre d'application d'ingénierie n'a pas été testée pour mettre en évidence les potentiels limitations et avantages de cette approche.

### **Méthodes graphiques**

Quelques auteurs (Castellarin et al., 2009 ; Viglione et al., 2012) ont étudié le concept de courbes enveloppes régionales intensités-durées. Ces courbes sont une représentation graphique de la borne supérieure des pluies observées dans une région. C'est en quelque sorte une redéfinition du concept de borne supérieure analogue à la pluie maximale probable (PMP) proposée par Hershfield (1961). Ces courbes lient la pluie maximale annuelle sur une certaine durée et la pluie moyenne annuelle. Cependant, l'utilisation de la notion de limite supérieure sur extrêmes hydrologiques est hautement contestée. En effet, ce postulat présente une incohérence théorique (Koutsoyiannis, 1999), du fait que ces approches font appel à des méthodes statistiques, qui supposent une limite supérieure non bornée des pluies pour déterminer la PMP, or son fondement repose sur la limitation de la borne supérieure des pluies.

Il existe aussi une représentation graphique dite courbes Huff (Dolšak et al., 2016; Pan et al., 2017). Ces dernières relient les cumuls de pluies adimensionnels (pourcentage cumulatif d'une pluie) et les durées adimensionnelles d'une averse, 110 à 140 averses sont nécessaires pour définir des courbes Huff stables. Les averses sont réparties en 4 groupes des durées. Les courbes Huff sont une représentation probabiliste des structures temporelles d'une averse, elles permettent de produire des hyétoigrammes synthétiques assez réalistes qui peuvent servir comme pluie de projet. La méthode (Pilgrim & Cordery, 1975) permet aussi d'établir la distribution temporelle des pluies.

### Approches Stochastiques

Palynchuk & Guo (2008) proposent une alternative au modèle statistique IDF, pour la conception des réseaux d'eaux pluviales, basée uniquement sur l'analyse des tempêtes. Ils constituent leurs jeux de données avec uniquement les événements dépassant un certain seuil et font une analyse fréquentielle conjointe qui considère la durée de la tempête et son cumul de pluie. Cette procédure a l'avantage de traiter avec des durées d'évènements réalistes (contrairement à l'approche conventionnelle). Cette procédure peut mener à des plus faibles estimations en terme du ruissellement et d'intensité, comparativement à l'approche conventionnelle, pour les grands bassins versants.

Il existe une approche stochastique qui trouve sa justification dans l'utilisation limitée des données en échantillonnage SDP et SMA. Il s'agit d'une modélisation stochastique qui permet de faire une utilisation complète des enregistrements de pluies continues. Cette approche permet d'établir les courbes IDF sans passer par une formulation paramétrique et elle consiste à : (a) caler un modèle stochastique aux données continues de pluies, (b) utiliser le modèle stochastique ainsi ajusté pour générer, par simulation Monte Carlo, de très longs ensembles de réalisations stochastiques, (c) extraire à partir de l'ensemble synthétique ainsi formé des séries des maxima annuelles pour chaque durée, puis les classer et appliquer une formule de fréquence expérimentale pour affecter une fréquence au non dépassement à chaque valeur de SMA puis déduire les courbes IDF (graphiquement établies) (Chow et al., 1988). Néanmoins, cette approche est plus complexe qu'une estimation directe des courbes IDF sur la base de SMA ou de SDP, le nombre de paramètres à caler est plus important et plus laborieux (Veneziano et al., 2007). Cependant, cette approche peut être simplifiée en considérant un modèle stochastique des pluies intégrant une invariance d'échelle multifractale (Schertzer & Lovejoy, 1987) ce qui aboutit à une réduction du nombre de paramètres, une simplification des ajustements et l'obtention de la distribution des intensités de pluies sur diverses durées par des méthodes semi-analytiques (Langousis & Veneziano, 2007; Veneziano et al., 2006).

Le concept de la transposition stochastique des averses (SST : Stochastic Storm Transposition, en anglais) est assez ancien (Alexander, 1963; Fontaine & Potter, 1989). La méthode consiste à aléatoirement transposer le champ de pluies en entier, pour des averses mesurées, en conservant leurs structures spatio-temporelles (intensité, surface, déplacement). La SST peut être considérée comme un Bootstrap, c'est à dire un re-échantillonnage à partir d'un catalogue d'averses. La transposition se fait dans une région pluviométriquement homogène (Wright et al., 2020).



### Approche déterministe

La méthode de la pluie maximum probable (PMP) (Hershfield, 1961) et son extension crue maximum probable (CMP) a été très largement utilisée en hydrologie. Etant donné que les pluies sont moins dépendantes que les écoulements par rapport à la topographie, et que leurs séries sont généralement plus longues, et qu'elles permettent une flexibilité dans la transposition spatiale : La transposition déterministe des pluies (PMP/CMP) est devenue une étape dans l'estimation des débits dans les bassins non-jaugés plutôt que l'approche probabiliste de Fuller (1914) (qui est en quelque sorte la première analyse fréquentielle régionale sur les écoulements), pour la transposition des écoulements. Tandis que l'analyse fréquentielle au site a continué à être utilisée pour des estimations de période de retour allant de 10 à 1000 ans. Les défenseurs de la PMP mettent en avant qu'elle offre la possibilité d'intégrer la structure des pluies. Par contre, ces limitations les plus importantes sont : (a) l'utilisation d'une valeur unique limite sans probabilité de dépassement ce qui rend la méthode inadéquate pour une application en analyse des risques hydrologiques, (b) le calcul de la plus large averse possible implique nécessairement la subjectivité d'analyse du concepteur.

### Méthodes basées sur l'intelligence artificielle : Apprentissage automatique

La prévision hydrologique sur la base d'intelligence artificielle peut être basée sur des processus déterministes tels que les réseaux de neurones (formulation conceptuelle) ou même sur des formulations mécanistes. Actuellement, les algorithmes de Machine Learning sont très diversifiés : réseaux de neurones (RN), machines à vecteurs de support (MVS), Random Forest (RF). L'intelligence artificielle sur une base déterministe sera limitée à la prévision à court termes (Koutsoyiannis & Langousis, 2011; Papalexiou & Serinaldi, 2020). Tandis que l'apprentissage automatique (*Machine Learning*, en anglais) nécessitera une très grande quantité de données de pluies pour être aussi performant qu'une modélisation stochastique (exemple : Papacharalampous, Tyralis, & Koutsoyiannis (2019)). Ou bien l'utilisation de données de tailles usuelles mais sur plusieurs bassins et en intégrant aussi des données hydrométriques, d'évapotranspiration, de température, ce qui n'est pas toujours évident et à plus forte raison en milieu urbain (Papacharalampous, Tyralis, Langousis, et al., 2019).

#### 2.4.3.4 Conclusion

Une revue des approches alternatives à une analyse fréquentielle classique pour l'estimation des pluies de projets, permet de comprendre que ces techniques peuvent effectivement être satisfaisantes dans certains cas, nécessitant une grande quantité de données dans d'autres, etc. Toutefois, l'ensemble de ces approches ne permet pas de caractériser l'incertitude liée à leurs estimations. Dans ce travail, il est retenu, comme méthode d'estimation des paramètres d'une distribution fréquentielle donnée, une approche Bayésienne MCMC. Car cette dernière, ne permet pas d'inférer seulement les paramètres d'une distribution donnée, mais tous les paramètres de la relation de courbes IDF, à travers l'écriture de la vraisemblance du modèle établi, en une seule étape et intégrant les incertitudes d'estimation dans la procédure de

calcul. En effet, les procédures d'inférence fréquentielle Bayésienne sont théoriquement cohérentes, numériquement précise, et maintenant la mise en œuvre de leurs calculs est abordable (Gaume, 2018).

#### **2.4.4 Vers l'hégémonie de l'inférence fréquentielle Bayésienne**

Tel qu'expliqué par Efron (2005) : « D'une manière générale, le 19<sup>ème</sup> siècle a été Bayésien et le 20<sup>ème</sup> siècle fréquentiste, du moins du point de vu de la plupart de praticiens. Cependant au 21<sup>ème</sup> siècle, les scientifiques vont soumettre aux statisticiens de plus grands problèmes à résoudre, souvent comprenant des millions de données et des milliers de paramètres. Quelle philosophie va dominer la pratique courante ? Mon avis, en m'appuyant sur des exemples récents, est qu'une combinaison d'idée Bayésienne et fréquentiste sera nécessaire pour traiter avec un environnement scientifique de plus en plus intense. »

L'expérience montre, à travers des analyses fréquentielles classiques, que les intervalles de confiances sur des distributions Normale, Log-Normale et Pearson III, etc., échouent à contenir l'évènement hydrologique centennal réel, contrairement à une approche basée sur les procédures de simulation Monte Carlo (Stedinger, 1983a). Aussi, l'inférence Bayesian MCMC présente une flexibilité grâce à la vraisemblance et une robustesse, que d'autres méthodes ne possèdent pas nécessairement (Payrastré, 2005). En effet, elle permet, en plus des séries de mesures, de prendre en considération toutes autres informations quantitatives sur le phénomène étudié, mais aussi d'augmenter la taille des échantillons analysés à travers la notion d'invariance d'échelle spatio-temporelle.

Pour ce qui est de la fiabilité des estimations en analyse fréquentielle, l'inférence Bayésienne intègre les incertitudes dans le processus de calcul sur les paramètres des distributions étudiées, ce qui permet de fournir l'incertitude, non seulement, sur les paramètres, mais aussi sur les quantiles leurs correspondant. Ainsi, l'approche Bayésienne est considérée comme la plus adaptée pour ce genre d'exercice (Coles & Pericchi, 2003; Reis & Stedinger, 2005). En effet, l'inférence fréquentielle Bayésienne s'est progressivement développée dans tous les domaines de l'hydrologie statistique et aussi dans l'estimation des courbes IDF (Chandra et al., 2015; Cheng & AghaKouchak, 2015; Huard et al., 2010; C. H. R. Lima et al., 2016; Mèlèse et al., 2018; Muller et al., 2008; Rosbjerg & Madsen, 1996; Van de Vyver, 2015a). De plus, il est clairement établi que les marges d'erreur sur les estimations doivent être considérées quand le résultat de l'inférence statistique a pour finalité la conception et l'analyse du risque (Gaume, 2018).

#### **2.4.5 La distribution GEV : Le choix adéquat**

Le choix d'une loi statistique repose sur l'équilibre entre biais et variance. Sachant que le premier diminue et le second augmente respectivement avec l'augmentation du nombre de paramètres de la loi (Grimaldi et al., 2011). La procédure typique pour le choix d'une loi statistique à ajuster à une série de pluie est de (1) sélectionner à priori un certain nombre de lois statistiques, (2) d'estimer les paramètres de ces lois en utilisant une des nombreuses techniques

d'ajustement disponibles, (3) choisir celles qui sont compatibles avec l'échantillon (ce qui peut être vérifié par des tests statistiques d'adéquation) (Papalexiou et al., 2013).

L'analyse fréquentielle des précipitations extrêmes sur la base des SMA inclue l'utilisation de diverses lois statistiques, les plus couramment utilisées étant : GEV, Gumbel, Log-Normale, Pearson III. Bien évidemment l'utilisation de la GEV et sa variante Gumbel étant dominantes (Willems et al., 2012), même si en Amérique du Nord et au Royaume-Uni la Pearson III est très populaire. Aussi, la théorie des valeurs extrêmes requiert de larges échantillons de données indépendantes et identiquement distribuées et une hypothèse de convergence à une distribution limite dans le but d'avoir une convergence asymptotique des SMA et SDP respectivement à la GEV et GPD. Dans la pratique la distribution identique est implicite et l'indépendance est imposée par la méthode d'échantillonnage (SMA ou SDP). Dans ce qui suit il est débattu des limites de la distribution GEV et de ses avantages.

### Définitions

La distribution des extrêmes pluviométriques peut être obtenue à partir des observations les plus larges possibles sur une année si : le nombre de jours de pluie par année (N) est donné et si la distribution parente des pluies journalières est aussi connue. Or N est une variable aléatoire et la distribution parente est inconnue et ne peut être connue. Pour résoudre cette problématique la théorie asymptotique a été développée. Selon la théorie des valeurs extrêmes de Fisher & Tippett (1928), la fonction de distribution de probabilité d'une variable aléatoire ajustée à une série d'observation, converge vers l'un des trois types de distribution des valeurs extrêmes : type I, type II et type III, respectivement quand le nombre de valeurs extrêmes est grand. Les propriétés de ces trois formes limites, ont été par la suite développées par : Gumbel en 1941 la loi des valeurs extrêmes de type I (EVI : *extreme value type I*), Fréchet en 1927 pour la loi des valeurs extrêmes de type II (EVII : *extreme value type II*), et Weibull en 1939 pour la loi des valeurs extrêmes de type III (EVIII : *extreme value type III*) (Chow et al., 1988).

Dans la plupart des manuels hydrologiques de référence (Chow et al., 1988), il est indiqué que Jenkinson (1955) a montré que les trois formes limites étaient des cas spéciaux d'une distribution unique appelée distribution générale des valeurs extrêmes (GEV : *General Extreme Value*). Toutefois, on retrouve la même dans Gnedenko (1943). La fonction de distribution de probabilité de la GEV à la forme suivante :

$$F_{\theta}(x) = \exp \left[ - \left( 1 - \frac{k(x - \xi)}{\alpha} \right)^{1/k} \right]_{\alpha > 0} \quad (2.20)$$

Et la fonction de densité correspondante a l'expression suivante :

$$f_{\theta}(x) = \frac{\partial F_{\theta}(x)}{\partial x} = \frac{1}{\alpha} \left( 1 - \frac{k(x - \xi)}{\alpha} \right)^{1/k-1} \exp \left[ - \left( 1 - \frac{k(x - \xi)}{\alpha} \right)^{1/k} \right]_{\alpha > 0} \quad (2.21)$$

$\xi, \alpha, k$ , étant respectivement les paramètres de position, d'échelle et de forme. Les trois cas limites de la GEV sont : (1) la EVI pour  $k = 0$ , connue aussi comme loi de Gumbel, (2) la EVII pour  $k < 0$  où  $(\xi + \alpha/k \leq x \leq +\infty)$ , et (3) la EVIII pour  $k > 0$  où  $(-\infty \leq x \leq \xi + \alpha/k)$ , connue aussi comme loi de Weibull. A noter que la différence entre une distribution EVII et Fréchet est que la seconde a une borne inférieure nulle (Meylan & Musy, 1999).

### Avantages, inconvénient et limitations

Il est bien connu que la pluie maximale annuelle (SMA) est une variable hydrologique fondamentale qui n'atteint pas les conditions asymptotiques : car le nombre de valeurs de jour de pluie par années doit être très large. Or, la théorie classique des valeurs extrêmes (théorème de Fisher-Tippett) suppose que les données à ajuster doivent satisfaire au régime asymptotique. Dans ces conditions, il n'y a tout simplement pas de garantie que la distribution statistique étudiée est approximée par l'une des lois GEV pour la gamme de périodes de retour considérée, car la théorie est asymptotique. Ainsi, l'utilisation de la distribution GEV dans l'estimation des quantiles à des fins de dimensionnement peut être inappropriée (Carlo De Michele & Avanzi, 2018). En effet, l'applicabilité de la théorie des valeurs extrêmes asymptotique pour l'ajustement des pluies maximales journalières a été questionnée (Marani & Ignaccolo, 2015). A titre d'exemple, même si l'ajustement de la distribution GEV est adéquate sur une SMA, du fait de la faible taille des échantillons étudiés ou une forte variabilité d'échantillonnage, le paramètre de forme de la distribution ajustée peut ne pas représenter le paramètre de forme de la distribution parente. Pour cette raison il existe des recommandations spécifiques sur le calcul du paramètre de forme, et des précautions telles que la comparaison avec des postes voisins sous même régimes pluviométrique. Par exemple, Koutsoyiannis (2004a, 2004b) recommande de prendre un paramètre de forme fixe dans certaines régions du monde dont les distributions fréquentielles sur les extrêmes ont été profondément analysées ( $k=-0.15$  dans la région Méditerranéenne). Mais aussi, il est possible de faire une estimation robuste, en augmentant la taille de l'échantillon ce qui concorde mieux avec l'hypothèse des extrêmes asymptotiques, à travers une analyse fréquentielle régionale. Car, quand l'effet de la longueur d'observation est corrigé le paramètre de forme varie dans un petit intervalle.

Cependant, en analysant les ajustements de la distribution GEV à des jeux de données, très longs, provenant des différents endroits à travers le monde (Norvège, USA et autres), Ragulina & Reitan (2017) montrent que le paramètre de forme ne peut être considéré comme constant. Ils constatent que le paramètre de forme varie probablement en fonction de l'altitude tandis que l'effet de la latitude est incertain. Ils confirment aussi un lien avec la classification de Köppen-Geiger. En effet, dans une analyse avec un très large échantillons hydrométriques, Tyralis et al. (2019) concluent que le paramètre de forme de la distribution GEV dépend principalement des indices climatiques. Cette limitation, trouve une solution dans l'approche Bayesian MCMC qui permet de considérer un Prior qui va limiter les valeurs possibles du paramètre de forme et ainsi garantir une estimation réaliste.

Cette inconsistance théorique, sur l'hypothèse des extrêmes asymptotiques, est aussi valable pour l'échantillonnage SDP, car dans le cas le plus quantitativement favorable, une SDP

sera tout au plus 3 fois plus large qu'une SMA, car le nombre de dépassement moyen pour un seuil optimal dépasse rarement 3 (Todorovic, 1978). Ainsi, pour atteindre le régime asymptotique il faut un très grand nombre de valeur maximale de pluie par année, or en une année le nombre des jours de pluies est inférieur à 365 (intermittence) et est loin de tendre vers l'infini. Cette exigence théorique ne peut être satisfaite et les analyses des valeurs extrêmes sur la base SMA/GEV ou SDP/GPD ne sont qu'une approximation satisfaisante (Koutsoyiannis, 2004a). En effet, pour dépasser cette limitation, la théorie des super statistiques (De Michele & Avanzi, 2018) a été proposée or elle ne montre de différences significatives avec SMA/GEV que pour les très grandes périodes de retour qui ne sont d'aucune utilité pratique et dont la pertinence de leurs grandeurs ne peut être vérifiée.

### **Justification de la GEV à travers quelques exemples**

Une explication potentielle à la fréquente utilisation de la GEV en analyse fréquentielle des extrêmes, est qu'elle assez souvent adéquate avec les séries régionales homogènes. En effet, la définition et les structures des L-moments conduisent artificiellement à choisir la distribution GEV (Klemes, 2000). Toutefois, il y a un très grand nombre de confirmations empiriques sur l'adéquation et la nécessité d'utiliser la GEV en analyse fréquentielle sur les extrêmes hydrologiques. Elle est même considérée comme le meilleur choix de distribution par Papalexiou & Koutsoyiannis (2013) à la suite de l'étude 15317 séries pluviométriques réparties à travers le monde. A dire vrai les distributions basées sur la théorie des valeurs extrêmes (GEV) des SMA ne peut pas satisfaire au régime asymptotique, car la convergence vers la limite est très lente (nombre de jours de pluies par année inférieur à 365, là où les échantillons des pluies annuelles devraient être très larges pour être plus concordant avec la théorie des valeurs extrêmes).

Dans le bassin méditerranéen beaucoup d'études statistiques (Aronica & Freni, 2005; Ben-Zvi, 2009; Blanchet et al., 2016; García-Marín et al., 2019; Ghanmi et al., 2016; Haktanir et al., 2010), sur les extrêmes pluviométriques infra journaliers et horaires confirment que le choix de la distribution GEV semble le plus approprié pour les SMA sur diverses durées (de 5 min à 1440 min). Le même constat est fait en Algérie dans l'étude des distributions statistiques des précipitations extrêmes (Boucefiane & Meddi, 2019; Habibi et al., 2013; Meddi et al., 2017; Meddi & Toumi, 2015). A noter que, quelques fois dans les ajustements de distributions fréquentielles à des extrêmes pluviométriques, le comportement avec queue à décroissance lente est caché du fait de la faible longueur des échantillons étudiés (Papalexiou et al., 2013). Ceci explique pourquoi dans certaines régions du monde, dont l'Algérie, où les échantillons ne sont pas très longs, la distribution GEV (queue à décroissance lente) est discriminée au profit de la loi de Gumbel et de la loi Log-Normale (qui a une queue à décroissance lente moins marquée que la GEV), particulièrement dans le cadre d'application d'ingénierie pratique. Un des objectifs connexes de ce travail de thèse est d'œuvrer à l'acceptation de l'utilisation systématique de la distribution GEV dans les estimations des pluies de projets.

### 2.4.6 Contrôle des ajustements : tests d'adéquation

Le contrôle des ajustements, ou statistiquement parlant les tests d'adéquation, consistent à vérifier si l'ajustement d'une fonction de distribution de probabilité considérée à un jeu de données, est adéquate ou non. C'est-à-dire que ces tests permettent d'évaluer si oui ou non une série d'observation aurait pu être tirée à partir d'une distribution fréquentielle spécifique. Il existe plusieurs familles de tests statistiques, tels que :

1. Tests par analyse graphique : Inspection visuelle des ajustements (Meylan & Musy, 1999), diagrammes des moments et des L-moments (Vogel & Fennessey, 1993), plotting position (Stedinger et al., 1993), etc. Toutefois, ces tests peuvent être assez trompeurs.
2. Tests basés sur la régression – corrélation : mesure de la concordance entre période de retour empirique et observations à travers le coefficient de corrélation ou la régression simple, etc., (D'Agostino & Stephens, 1986).
3. Tests basés sur le principe d'entropie maximale afin de mesurer la discordance entre modèle réel et approximé: Critère d'information d'Akaike en 1973 (AIC) et critère d'information Bayésienne de Schwarz en 1978 (BIC) (Laio et al., 2009).
4. Tests basés sur la fonction de distribution expérimentale (FDE): Anderson-Darling et Kolmogorov-Smirnov, etc., (D'Agostino & Stephens, 1986).

Les tests d'adéquation font appel au principe d'hypothèse nulle  $H_0$ , qui suppose que  $F(x) = F_0(x)$  où  $F(x)$  est la fonction de répartition de la distribution fréquentielle à ajuster à la série d'observation étudiée et  $F_0(x)$  est la distribution parente de la population dont est issue cette série. L'hypothèse alternative est dite  $H_1$ , elle est vraie quand il y a inégalité. Ces tests ne peuvent qu'accepter ou rejeter  $H_0$ . En effet, les tests d'adéquation des distributions fréquentielles ne sont valables que quand ils peuvent démontrer la conformité (ou plus à proprement parler quand ils ne rejettent pas l'hypothèse d'égalité) d'un jeu de données à une distribution hypothétique ou d'une manière équivalent sa discordance par rapport à la distribution. C'est-à-dire qu'ils sont peu puissants, dans le sens où ils ne permettent pas de comparer, entre eux, deux modèles fréquentiels différents (Meylan & Musy, 1999; Stedinger et al., 1993).

Un test souvent utilisé est celui du  $\chi^2$  de Pearson, il consiste à vérifier si la fréquence cumulée empirique est identique à la fréquence cumulée théorique. C'est un test notoirement peu puissant, c'est-à-dire qu'il ne peut réfuter avec suffisamment de certitude l'hypothèse nulle  $H_0$  quand c'est l'hypothèse alternative  $H_1$  qui est vraie. De plus le groupement en classe, des valeurs observées, en fait un test non exhaustif qui entraîne une perte d'information, particulièrement pour les classes extrêmes (Meylan & Musy, 1999). Toutefois, ce test reste d'une grande utilité pratique surtout sur la famille de distributions à deux paramètres (position et échelle) ou pour des applications particulières telles que sur les données discrètes, multivariées ou même tronquées (D'Agostino & Stephens, 1986).

Il y a aussi le test de Kolmogorov-Smirnov (K-S) un des plus anciens (1933) basé sur la fonction de distribution empirique, où il est question de mesurer le plus grand écart entre fréquence expérimentale et théorique. Comparativement au test du  $\chi^2$ , la série observée est

testée dans son intégralité et le problème de petitesse d'un échantillon pour la définition de la classe ne se pose pas (Ang & Tang, 2006).

Un autre test d'adéquation basé sur la fonction de distribution empirique, est celui d'Anderson & Darling (1954) (AD). Il est considéré comme l'un des tests statistiques d'adéquation le plus puissants (Stephens, 1976). En effet, dans le test K-S, les probabilités sont définies sur une échelle arithmétique, ce qui fait que les fonctions de distributions cumulées théorique et expérimentale sont relativement plates aux queues des distributions fréquentielles, ainsi la déviation maximale dans le test K-S surviendra rarement dans la queue de la distribution. Tandis, que dans un test de  $\chi^2$  les fréquences expérimentales à la queue doivent être groupées ensemble, en conséquent les deux tests K-S et  $\chi^2$ , ne révéleront aucune discordance entre les fréquences théorique et empirique à la queue de la distribution fréquentielle sous-jacente (Ang & Tang, 2006). Par contre le test AD place une plus grande pondération, comparativement aux autres tests, sur la queue de la distribution fréquentielle (Laio, 2004). De plus il est spécifiquement conçu pour la prise en charge des échantillons de petites tailles et des distributions fréquentielles fortement asymétriques (Laio et al., 2009). Ce qui est, en fait, un test particulièrement bien adapté pour vérifier l'adéquation de fonction de distribution de probabilités sur des mesures de valeurs extrêmes (D'Agostino & Stephens, 1986).

A signaler que les trois tests précédemment discutés ne fournissent pas une information absolue sur l'adéquation d'une distribution spécifique. Ainsi, l'adéquation d'un ajustement est en fonction du seuil de signification  $\alpha$  choisi. Toutefois, ces tests demeurent parmi les plus couramment utilisés dans le choix d'une loi statistique. En effet, ils comptent de nombreuses applications en analyse fréquentielle hydrologique et dans l'estimation des courbes IDF (Ben-Zvi, 2009; Chandra et al., 2015; Gargouri-Ellouze & Chebchoub, 2008; Panthou et al., 2014; Van de Vyver & Demarée, 2010).

En plus des tests susmentionnés il existe des tests non paramétriques tels que les critères AIC et BIC, qui sont très souvent utilisés en modélisation hydrologique mais très peu dans la vérification de l'adéquation des lois statistiques. Les critères AIC, BIC et le test d'adéquation d'Anderson-Darling fournissent des résultats de qualité comparable dans l'identification des distributions de probabilité des extrêmes hydrologiques (distribution parente fortement asymétrique et échantillons hydrologiques courts) (Laio et al., 2009). Ainsi, dans ce travail de thèse, le test AD sera utilisé pour la vérification de l'adéquation de la distribution GEV, non seulement sur des échantillons de pluie pour une durée donnée, mais aussi sur un échantillon conjoint (intégrant tous les enregistrements sur diverses durées). Pour ce dernier cas de figure une précaution particulière est prise sur le critère de rejet, ce point est détaillé plus tard dans le chapitre 5.

#### 2.4.7 Evaluation des incertitudes d'estimation

Les champs d'observation et les séries chronologiques de pluies sont caractérisés par une grande variabilité et une irrégularité, plus que n'importe quelles autres variables météorologiques. Ainsi, une grande variabilité implique de grandes incertitudes (Koutsoyiannis

& Langousis, 2011). Aussi, l'analyse de 4000 stations pluviographiques (Papalexiou, AghaKouchak, et al., 2018) montre que les queues des distributions fréquentielles des pluies horaires sont plus importantes que celles usuellement utilisées (GEV, GPD, etc.) avec d'importantes implications et principalement une sous-estimation des extrêmes. En gardant ces deux dernières remarques en tête, la nécessité d'une bonne définition des incertitudes d'estimation, est plus qu'évidente.

Les méthodes statistiques appliquées actuellement en analyse fréquentielle hydrologique nécessitent de longues périodes d'observations particulièrement pour les grandes périodes de retour. De plus, le choix du modèle statistique et de la technique de calcul des paramètres peut introduire de l'incertitude dans les estimations. A titre d'exemple, Hu et al. (2020), ont étudié la sensibilité de l'analyse fréquentielle des variables hydrologiques, par rapport à la taille des échantillons, au choix du modèle statistique, et à la méthode d'estimation des paramètres. Cette analyse a été menée sur 6 régions hydrologiques à travers les Etats-Unis. Les résultats montrent qu'il y a convergence entre une station de référence, de 70 années de mesures (SMA), et des estimations faites sur des séries de 35 années d'observations. Par contre, l'incertitude sur les quantiles est 50% plus importante pour les périodes de retour de 100 ans et plus.

En plus des incertitudes issues de la méthodologie d'estimation appliquée, les instruments de mesures et d'observations des variables hydrologiques sont aussi sources d'incertitudes. En effet, le procédé de mesures pluviographiques par auget basculant (*Tipping Bucket, en anglais*) qui fournit les observations des intensités ou cumuls de pluie sur un intervalle de temps peut engendrer des incertitudes (sous-estimation) des SMA sur une durée d quelconque. Morbidelli et al. (2017) montrent qu'en moyenne l'erreur d'estimation des SMA (pour de grandes valeurs de  $d \gg 1$  minute) est de 17% et qu'elle peut atteindre 50% dans la région de Lazio au centre de l'Italie.

Ainsi, les incertitudes peuvent être définies comme une indication quantitative de la fiabilité d'une quelconques grandeur hydrologique (mesurée ou inférée). Elles peuvent être aléatoires ou épistémiques et sont inférées d'une manière probabiliste et non-probabiliste (Montanari, 2011). En effet, il est communément admis que les incertitudes peuvent être regroupées en deux catégories : variabilité naturelle (incertitudes structurales, aléatoire, externe, objective, inhérente, irréductible, stochastique), incertitudes connues (épistémique, interne, subjective, fonctionnelle, réductible). Dans les faits, l'incertitude d'un modèle peut être caractérisée de quatre façons : totalement structurale, partiellement structurale et partiellement épistémique, totalement épistémique. Dans l'évaluation des performances d'un modèle, il est toujours préférable de séparer les différents types d'incertitudes. Toutefois, très souvent il n'est pas possible de séparer les deux types d'incertitudes et elles sont évaluées de façon intégrée. Une manière alternative de considérer les incertitudes est de les subdiviser selon les catégories suivantes : hasard inhérent, incertitude structurale reflétant l'incapacité d'un modèle à reproduire précisément le comportement d'un système, incertitude sur les paramètres du modèle, incertitudes sur les données, incertitudes opérationnelles. Le modèle proposé dans ce travail étant un modèle purement statistique, les incertitudes qui font objet de l'étude sont des incertitudes de diverses origines (partiellement structurale et partiellement épistémique) mais sont prises en charge à travers une estimation efficiente des intervalles de crédibilité sur les paramètres de la modélisation statistique des courbes IDF et les quantiles correspondants. Dans



ce qui suit il est brièvement présenté les trois principales méthodes d'estimation des incertitudes en inférence statistique.

Quand l'analyse fréquentielle est mise en œuvre sur des enregistrements de pluies relativement courts et/ou à forte variabilité d'échantillonnage, il est plus prudent de considérer la borne supérieure de l'intervalle de confiance sur les quantiles de pluies extrêmes (ou autres variables hydrologiques), dans les projets de conception et d'analyse du risque. Dans la pratique c'est la borne supérieure de l'intervalle de confiance à 70% qui est souvent retenue, car même si l'intervalle à 95% correspond à une probabilité plus forte de recouvrir la valeur de quantile, il est en général d'amplitude trop grande pour avoir une signification pratique (Bernier & Veron, 1964). Généralement l'aléa pluviométrique et l'incertitude d'échantillonnage peuvent être combinés en un seul aléa. Cependant, l'incertitude d'échantillonnage n'est pas, et de loin, la seule incertitude à craindre. Ce dernier constat, est une des raisons expliquant les performances de l'inférence fréquentielle Bayésienne MCMC dans l'estimation des incertitudes sur les quantiles. En effet, l'algorithme MCMC simule un très grand nombre d'échantillons dans son processus d'implémentation, ce qui permet quelque peu de réduire l'effet de la variabilité d'échantillonnage, de plus la vraisemblance favorise un calcul en une seule étape, ce qui réduit les incertitudes opérationnelles. En plus de la procédure Bayesian MCMC, pour le calcul des intervalles de crédibilité, les deux autres procédures (Erreur standard et Bootstrap) les plus utilisées pour cerner les incertitudes d'estimation sur les quantiles, sont résumées dans ce qui suit.

#### 2.4.7.1 Erreur standard

Une mesure simple de la précision d'une estimation de quantile est sa variance, qui est égale au carré de l'erreur standard :  $SE^2 = Var(x_p)$ . L'intervalle de confiance en est une description alternative. Les intervalles de confiance sont calculés en supposant une distribution d'échantillonnage normale. C'est-à-dire que si l'on a à disposition, par exemple, quelques milliers d'échantillons, la loi normale est supposée être la fonction de distribution de probabilités des caractéristiques statistiques (moyenne, variance, quantiles, etc.) de chacun de ces échantillons. Ainsi, une estimation de la variance de la distribution d'échantillonnage, sur un quantile de période de retour  $T$ , permet de construire l'intervalle de confiance pour un seuil de signification  $\alpha$  donné.

L'acceptation de la normalité dans la distribution d'échantillonnage trouve sa justification dans le fait que la loi normale est la forme asymptotique d'un très grand nombre de distributions. Aussi, si l'on ne connaît que l'espérance et la variance, la loi normale reste le meilleur choix possible. Kite (1975) montre que pour les distributions d'applications courantes en hydrologie, la distribution d'échantillonnage d'un quantile n'est pas significativement différente d'une loi normale. Cependant, il arrive à ses conclusions sur la base de tests d'adéquation de K-S et  $\chi^2$  de Pearson. Toutefois, il a été souligné précédemment que ces tests étaient peu révélateurs sur les queues des distributions, or c'est justement l'incertitudes sur les quantiles de la queue de la distribution qui sont d'intérêt pratique. Ainsi, l'utilisation injustifiée d'une loi normale dans le calcul des intervalles de confiance, au lieu d'une autre, mènerait à une erreur sur le calcul de l'incertitude dont la portée est inconnue. C'est surtout, entre autres, la commodité d'utilisation

de la loi normale dans la construction des intervalles de confiance, qui explique son usage fréquent (Meylan & Musy, 1999).

#### 2.4.7.2 *Bootstrap*

La technique du Bootstrap pour le calcul des incertitudes sur les paramètres des distributions fréquentielles et les quantiles correspondants a été introduite par Efron & Tibshirani (1986). Il existe deux variantes de Bootstrap pour la construction des intervalles de confiance sur les quantiles des variables hydrologiques. La première est dite rééchantillonnage non-paramétrique. Elle consiste à faire un tirage avec remise, à partir de la série de mesures, à l'aide d'une procédure d'échantillonnage de Monte Carlo afin de construire un certain nombre (typiquement autour de 10 000) d'échantillon de même taille que la série originelle. Il s'agit ensuite d'ajuster la loi statistique voulue à chacun des échantillons simulés, puis en déduire les quantiles de période de retour  $T$  donnée. Finalement, un classement croissant de ses quantiles simulés permet de connaître l'intervalle de confiance pour un seuil de signification donné (s'il est de 5%, alors les quantiles de rang 250 et 9750 représentent l'intervalle de confiance à 95%). La deuxième variante est un rééchantillonnage paramétrique, où la loi statistique considérée est ajustée à une série de mesure, afin de tirer aléatoirement dans cette distribution un certain nombre d'échantillons. Une fois les échantillons simulés constitués, la même procédure, que pour le Bootstrap non-paramétrique, permet de construire les intervalles de confiances.

La technique du Bootstrap est d'utilisation assez courante dans la définition des intervalles de confiance sur les paramètres des distributions et les quantiles d'intensité de pluies (Ng et al., 2019; Uboldi et al., 2014; Van de Vyver & Demarée, 2010).

#### 2.4.7.3 *Maximum de vraisemblance et Bayésien MCMC*

Les méthodes numériques combinant les simulations Monte Carlo et la marche aléatoire des chaînes de Markov (MCMC), ont été développées dans le but d'explorer l'hyperespace des variables (les paramètres  $\theta$  dans le cas de la perspective Bayésienne) suivant leurs probabilités conjointes (Gaume, 2018).

Il s'agit donc, de récupérer la distribution a posteriori  $p(\theta|D)$  et de calculer pour chaque jeu de paramètres les quantiles de période de retour  $T$  voulue. Pareillement que pour le Bootstrap, un classement croissant permet de construire les intervalles de crédibilité pour le seuil de signification désiré. Dans une étude comparative sur des données simulées (non réelles), Gaume (2018) montre que les deux procédures MCMC et Bootstrap non-paramétrique de calcul des intervalles de crédibilité sont précises, au sens que l'échantillon analysé est distribué conformément à la loi statistique étudiée. Cette comparaison montre aussi des intervalles plus larges pour le MCMC, ce qui signifie que le Bootstrap a plus de difficulté à atteindre les valeurs d'incertitudes assez fortes.

Par contre, Mélése et al. (2018), aboutissent à un résultat opposé en comparant les incertitudes sur les courbes IDF obtenues par MCMC et Bootstrap. En effet, les intervalles de confiance du Bootstrap sont beaucoup plus larges, particulièrement la borne supérieure, que les

intervalles de crédibilité MCMC. Toutefois, les auteurs concluent que l'estimation des intervalles de crédibilité par MCMC est plus performante du fait que le Bootstrap donne des intervalles de confiance aberrants. Néanmoins, les techniques d'inférence étant foncièrement différentes (maximisation numérique de la vraisemblance par algorithme d'optimisation), il est difficile de tirer des conclusions fermes. Etant donnée la prédominance établie de l'analyse fréquentielle Bayésienne, c'est bien évidemment une construction des intervalles de crédibilité à partir de la distribution à postériori qui est retenue dans ce travail.

#### 2.4.8 Analyse fréquentielle régionale des précipitations extrêmes

Les statistiques des extrêmes jouent un rôle très important dans les pratiques d'ingénierie de conception et de gestion des ressources en eau. Sans surprise, une des plus importantes contributions à la théorie des valeurs extrêmes depuis son émergence est la technique d'analyse régionale (Katz et al., 2002). En effet, l'utilisation de données issues de réseaux de mesure de faibles densités et de répartition irrégulière mène à des modélisations hydrologiques globales avec de fortes incertitudes sur les paramètres du modèle et sur les estimations (Zeng et al., 2018).

Une des références majeures dans la régionalisation en hydrologie reste très certainement l'ouvrage de Hosking & Wallis (1997) où ils décrivent et passent en revue différentes méthodes et procédures d'analyses en régionalisation, avec comme socle de départ, la méthode de l'indice de crue de (Dalrymple, 1960), qui suppose que la distribution fréquentielle de tous les sites dans une même région est identique à un facteur d'échelle près. En effet, une analyse régionale suppose que le paramètre de position est variable dans une région et que les paramètres de dispersion (échelle) et de forme sont constants dans la région (Katz et al., 2002). Cette approche par analyse fréquentielle régionale inclut l'utilisation des L-moments et la dispersion des ratios des L-moments, dans quasiment toutes les procédures de calcul (Willems et al., 2012). L'hypothèse de base de la méthode de l'indice de crue s'écrit comme suit :

$$Q_i^{(j)} = \mu_i q^{(j)}, i = 1, \dots, M \quad (2.22)$$

Où  $i$  est l'indice de la station,  $M$  le nombre total de stations, dans une région homogène,  $Q_i^{(j)}$  le  $j^{\text{ième}}$  quantile au  $i^{\text{ième}}$  site,  $q^{(j)}$  le  $j^{\text{ième}}$  quantile adimensionnel (courbe de croissance régionale) et  $\mu_i$  est le facteur d'échelle au site (l'indice de crue pour les débits ou indice d'averse pour les pluies) qui ne dépend pas de  $j$  (le principe d'invariance d'échelle simple). La condition d'homogénéité régionale susmentionnée est implicite dans la définition des régions homogènes dans la régionalisation par L-moments. Cette condition signifie que tous les sites peuvent être décrits par une seule distribution de probabilité (équation 2.22) ayant des paramètres communs après que les données de chaque site aient été renormées par le facteur d'échelle au site à savoir la moyenne ou la médiane (qui est plus robuste à la variabilité d'échantillonnage) (Wallis et al., 2007). Bien que cette méthode soit plus populaire dans l'analyse fréquentielle des débits, elle est maintenant très utilisée en analyse fréquentielle régionale des pluies (Kysely et al., 2007).

Deux hypothèses sous-jacentes inhérentes à l'analyse fréquentielle sont la stationnarité des données sur la période d'observation et d'utilisation, et l'indépendance sérielle des données au

niveau de la station. La pierre angulaire d'une analyse fréquentielle régionale est que les données des sites provenant d'une même région homogène peuvent être mélangées (échantillon conjoint ou *pooled* en anglais) pour améliorer la fiabilité des estimations des fréquences élevées pour tous les sites (Hosking & Wallis, 1997). A noter que généralement la régionalisation des pluies est plus simple que pour les débits et les débits d'étiage, particulièrement dans les zones géographiques sans relief topographique remarquable (Stedinger et al., 1993).

L'idée d'utiliser une information régionale pour améliorer la robustesse des estimations de quantiles est assez ancienne. Un des premiers exemples connus est la formule de calcul des débits de période de retour  $T$  donnée par Fuller (1914), dans le cadre d'une étude de dimensionnement d'un évacuateur de crue pour  $T=1000$ ans :

$$Q_T = \bar{Q} (1 + 0.8 \log_{10} T) \quad (2.23)$$

Avec  $Q_T$  de pointe de période de retour  $T$ , et  $\bar{Q}$  le débit moyen au site de jaugeage et 0.8 représente un coefficient calculé sur la base d'observation régionale des écoulements. Au final, un des premiers modèles hydrologiques probabilistes est une analyse fréquentielle régionale (Wright et al., 2020).

#### 2.4.8.1 Apport de l'analyse fréquentielle régionale dans l'étude des extrêmes

L'approche régionale part de l'idée que dans une région relativement homogène, l'estimation des quantiles en un site peut être améliorée en utilisant les informations sur les extrêmes disponibles au niveau des autres sites (Buishand, 1991). C'est-à-dire qu'elle permet de réduire les incertitudes d'inférence. Il s'agit d'échanger du temps contre de l'espace. Toutefois, cet avantage peut être limité par la nécessité d'homogénéité sur de vastes régions (Paixao et al., 2015). De même, cette approche permet de réduire la sensibilité des estimations à la variabilité d'échantillonnage, particulièrement quand les échantillons sont de petites tailles (Muller et al., 2009). Néanmoins, l'enrichissement d'un jeu de données dans la modélisation hydrologique globale n'est plus efficient au-delà d'un certain seuil, mais il permet toujours de réduire l'incertitude sur les paramètres des modèles (Zeng et al., 2018).

Effectivement, quand les estimations fréquentielles locales sont mises en œuvre sur des séries relativement courtes, le résultat devient peu fiable. L'analyse régionale est une solution acceptable pour réduire les incertitudes de ces estimations. Ribatet et al. (2007) fait une analyse de précision sur les quantiles régionaux estimés avec un degré d'homogénéité variable de l'échantillon conjoint, ce qui montre que plus la région est homogène et l'échantillon large, plus la précision sur l'estimation des quantiles augmente. L'analyse fréquentielle régionale des débits (ou même des pluies) souffre d'un certain nombre de critiques du fait qu'elle repose sur des hypothèses solides. Une de ces hypothèses implicites est que le coefficient de variation des extrêmes doit être constant dans toute la région homogène, ce qui ne se vérifie pas en pratique et qui ne peut être justifié physiquement. Un assouplissement, dans une certaine mesure, de ces critères d'homogénéité offre un équilibre positif en permettant d'intégrer un plus grand nombre de données aux régions ainsi formées.

L'analyse fréquentielle régionale des précipitations extrêmes, dans l'estimation des courbes IDF, a déjà fait ses preuves, qu'elle soit appliquée sur des échantillonnages SMA ou SDP

(Rosbjerg & Madsen, 1996). De même, en Algérie, elle a connu plusieurs applications (Benhattab et al., 2014; Kondratieva & Amarchi, 2015; Meddi & Toumi, 2015), sur les pluies maximales journalières, qui ont grandement amélioré les estimations comparativement à des analyses locales.

L'analyse fréquentielle régionale sur les précipitations extrêmes a connu de nombreuses approches dérivées, que ce soit l'utilisation de techniques Bayésiennes dans la délimitation des régions homogènes (Renard, 2011), ou une analyse bivariée intégrant des caractéristiques statistiques de pluies sur différentes échelles (Caporali et al., 2008, 2018), ou encore une combinaison de techniques de régression et d'analyse régionale incorporant de l'information physico-météorologique (Ouali et al., 2016). Toutefois, dans ce travail ces techniques dérivées ne sont pas explorées, car elles considèrent de multiples interpolations, des analyses multivariées où les incertitudes ne sont pas considérées, plusieurs étapes de calcul, ce qui fait qu'elles incorporent trop de bruit. Ainsi, dans le chapitre 6, c'est plutôt l'applicabilité d'une analyse fréquentielle régionale intégrant des observations de pluies extrêmes sur diverses échelles temporelles et spatiales qui est étudiée.

#### 2.4.8.2 Délimitation des régions pluviométriques homogènes

En analyse fréquentielle régionale des précipitations plusieurs techniques et procédures ont été utilisées dans la délimitation des régions homogènes, telles que : l'analyse par grappe (Clustering), l'analyse en composante principale (ACP), l'analyse par corrélation canonique, l'analyse par régression multivariée, rapport des lames, Gradex, etc. (St-Hilaire et al., 2003). Toutefois, l'application de ces techniques requiert l'identification de mesures de similarité hydrologiques. Identification qui reste assez subjective, bien qu'il existe des méthodes permettant d'identifier les attributs physiques les plus indicatifs sur les extrêmes hydrologiques, qui serviront d'entrants dans les procédures de délimitation (c'est-à-dire jusqu'à quelle étendue l'homogénéité hydrologique peut être expliquée par l'homogénéité physique) (Ilorme & Griffis, 2013).

Une fois les régions supposées homogènes délimitées, il s'agit d'en valider le découpage. Un certain nombre de tests d'homogénéité régionale, spécifiques à l'analyse fréquentielle régionale, ont été développés (Lu & Stedinger, 1992). Cependant, l'utilisation des techniques de validation des régions homogènes via les tests d'homogénéité et de discordance de Hosking & Wallis (1997), est prédominante. Ces derniers sont basés sur la comparaison des L-moments, ratios des L-moments locaux et régionaux. En d'autres termes, ces tests consistent à estimer le degré d'hétérogénéité d'un groupe de sites (ou station) et d'évaluer s'ils peuvent raisonnablement être considérés comme une région homogène.

#### Tests d'homogénéité (H) et de discordance (D) de Hosking et Wallis.

La procédure de validation, d'une région supposée homogène, selon la procédure de Hosking & Wallis (1997), consiste à :

1. Délimiter les régions homogènes d'une manière subjective ou en appliquant une technique de classification hiérarchique de type grappe (Clustering de Ward) ;
2. Tester l'homogénéité via les tests de discordance (D) puis d'homogénéité (H). A l'issue de cette étape les régions finales sont formées ;
3. Ajuster une distribution régionale sur un échantillon unique où chaque série a été renormée par sa moyenne ou médiane ou toutes autres caractéristiques du bassin versant (facteur d'échelle) ;
4. Tester l'adéquation par la méthode Z statistique proposée par Hosking & Wallis (1997) ;
5. Enfin estimer les quantiles locaux en multipliant le quantile régional de période de retour T par le facteur d'échelle.

Les deux avantages majeurs de l'approche régionale est qu'elle cerne mieux la queue de la distribution (chaque échantillon régional contient plus de valeurs extrêmes) comparativement à l'approche locale et qu'elle permet d'estimer des distributions en des sites non-jaugés (Kysely et al., 2007). En complément de ces tests, Requena et al. (2017), proposent l'utilisation de l'indice de Gini qui est non paramétrique (et qui en économie mesure le degré d'inégalité des revenus), afin de mesurer le degré d'hétérogénéité. En effet, il permet de faire un classement de performance sur différentes variantes de délimitations de régions homogènes dans une même zone, ce qui ne peut être fait sur la base de la statistique d'homogénéité H1, car cette dernière assume une distribution bien définie (Kappa).

Quelques fois, certaines variables reliées aux écoulements, tel que le pic de crue ou le volume de crue, peuvent être déclarées hétérogène au sens de la statistique H1 de Hosking et Wallis. Ce qui a mené au développement de tests d'homogénéité régionale, basés sur les L-moments, multivariés (Chebana & Ouarda, 2007; Chebana et al., 2009).

Il est généralement admis que l'homogénéité statistique survient en réponse à l'homogénéité physique (en hydrométrie). Ainsi, le découpage en régions homogènes peut se faire sur la base d'attributs physiques (et même météorologiques) plutôt que sur les statistiques des écoulements. Ce qui a comme avantage de classer les bassins non-jaugés, sur la base des caractéristiques physiographiques, à l'intérieur d'une région homogène. Toutefois, la similarité de la réponse hydrologique sur la base de ces caractéristiques n'est pas garantie étant donnée la complexité des interactions en jeu. De plus, il faut toujours garder à l'esprit que le découpage d'une zone est hautement dépendant des mesures de similarités retenues et que le choix de retirer ou non un site discordant (au profit d'une autre région) est totalement subjectif (Burn, 1990b; Hosking & Wallis, 1997) et représente une des limitations de la méthode des L-moments régionaux et nuit significativement à la précision des estimations en sites non-jaugés (Ilorme & Griffis, 2013). En effet, Burn (1990b) argue que l'utilisation des statistiques des séries d'observations, comme mesures de similarité, dans la délimitation des zones homogènes peut compromettre la validité des tests d'homogénéité régionale étant donné qu'ils sont basés sur les mêmes statistiques d'observations. Ainsi, cette façon de procéder peut entraîner des estimations des quantiles régionaux peu précises bien que les régions soient statistiquement homogènes. Cependant, Hosking & Wallis (1997) rejettent cet argumentaire, et clament que c'est plutôt la méthode de la région d'influence (ROI : Region Of Influence en anglais) (Burn, 1990b, 1990a)

qui est sujette à des critiques du fait qu'elle fait appel à un grand nombre de paramètres en fonction de considérations subjectives.

### **Délimitation des régions homogènes par grappe (clustering)**

La technique de délimitation des régions homogènes par analyse en grappe ou classification (*Cluster Analysis*, en anglais) repose sur le principe de mesure de distance statistique reflétant la similarité (ou pas) parmi le jeu d'attributs choisis pour chacune des stations d'observation. L'algorithme hiérarchique agglomératif de Ward est la technique la plus couramment utilisée dans les approches de clustering. Cette technique tend à délimiter des régions homogènes d'approximativement la même taille ce qui est plus approprié pour la régionalisation hydrologique (Ilorme & Griffis, 2013). Cette procédure a de nombreux dérivés, particulièrement dans les zones où la densité du réseau de mesure est assez élevée (Debbarma et al., 2019; Olsson et al., 2019; A. Smith et al., 2015).

### **Délimitation par région d'influence**

La délimitation des régions homogènes sur la base d'une région d'influence (ROI), introduit un concept fondamentalement différent de l'homogénéité régionale de Hosking & Wallis (1997). En effet, elle consiste à faire des estimations locales en se basant sur les mesures d'autres sites avoisinants et présentant une similarité (géographique, topographique, statistique, etc.) avec le site sous-jacent. Ce qui implique que pour chaque site une région homogène différente est considérée (i.e. chaque site a sa propre région homogène). Selon Burn (1990a, 1990b), cette méthode montre une plus forte homogénéité dans les régions formées comparativement à la méthode d'analyse régionale basée sur l'indice de crue (ou d'averse) avec régions fixes, de même qu'une estimation des quantiles plus précise. Dans un point non-jaugé cette méthode fait des estimations en se basant sur la similarité à travers des caractéristiques géographiques/topographiques d'un choix totalement subjectif. Pour des estimations dans des zones non-jaugées l'AFR de Hosking et Wallis reste une meilleure alternative car basée sur des critères statistiques.

### **Délimitation par analyse en composante principale (ACP)**

L'analyse en composante principale (ACP) est une technique qui permet de réduire la taille d'un ensemble d'informations (à plusieurs variables) en se contentant d'un nombre réduit de variables (de dimensions) tout en maintenant la majeure partie de la variance de l'ensemble à travers une transformation en fonctions linéaires non-corrélées des variables originales. Les variables avec la plus grande variance peuvent être retenues pour former la région homogène. Il existe une technique analogue à l'ACP, à savoir : la décomposition orthogonale aux valeurs propres. Elle consiste à décomposer des données avec des fonctions orthogonales déterminées à partir des données (en anglais : empirical orthogonal functions, abrégé en EOF). L'application de cette dernière revient à faire une ACP, à l'exception près que les EOF permettent d'obtenir à la fois des répartitions temporelles et spatiales (Bernard et al., 2013).

### **Délimitation des régions homogènes par régression – corrélation**

Une multitude de techniques et de procédures de régression – corrélation ont été appliquées pour identifier les mesures de similarité les plus adéquates dans la délimitation des régions homogènes dans l'analyse fréquentielle régionale (Johnson & Wichern, 2007; Marquínez et al., 2003; Pagliero et al., 2019). L'analyse par corrélation canonique (ACC) étant la fréquemment utilisée. Dans une approche basée sur cette dernière Ouali & Cannon (2018) clament que l'utilisation de la régression non linéaire sur des quantiles (QR) pour la délimitation des régions homogènes et l'estimation des courbes IDF en des sites non jaugés mène à de meilleurs résultats qu'une analyse régionale classique. Toutefois, les variables physiographiques retenues (latitude, longitude, élévation, pente, rugosité de la pente, aspect), et utilisées comme prédicateur pour caler ce modèle IDF (ACC/QR), n'ont pas la même influence sur les pluies en fonction de leurs durées, ce qui renvoie à une incohérence théorique globale du modèle.

### **Conclusion sur les techniques de délimitations des régions homogènes**

Bien que les tests d'homogénéité et de discordance de Hosking & Wallis (1997) aient été préconisés pour valider la région homogène, bien souvent ils servent comme technique de délimitation à travers de multiples essais. Dans une études comparative entre la ROI, les tests de L-moments et une régression Bayésienne, dans la délimitation de régions homogènes à fortes variabilités topographique et climatique, avec une densité de réseau de mesures pluviométriques assez élevée, Haddad et al. (2015) trouve que les trois méthodes sont d'une précision similaire. Tandis que, Viglione et al. (2007) dans une analyse de performance de différents tests d'homogénéité statistique, trouve que les tests d'homogénéité (H) et de discordance (D) de Hosking & Wallis (1997) sont les plus performants quand les échantillons sont légèrement asymétriques et inversement les tests de rang (K-échantillons d'Anderson-Darling et de Durbin-Knott) sont plus performants dans le cas de forte asymétrie.

#### **2.4.9 Stationnarité vs non-stationnarité**

Un processus stochastique est strictement stationnaire quand la distribution de probabilité conjointe, d'un nombre arbitraire de ses variables aléatoires, ne change pas quand elle est déplacée dans le temps ou l'espace, et par extension les propriétés statistiques ne changent pas aussi au cours du temps (Montanari, 2011). Les mesures chronologiques en hydrologie peuvent être considérées comme la partie observable du processus stochastique avec une distribution de probabilité définie. Deux aspects peuvent être étudiés : le premier est de définir des critères objectifs permettant de savoir si l'hypothèse de stationnarité peut être adoptée. Tandis que le deuxième concerne les effets de la non stationnarité dans l'estimation des paramètres de la distribution de probabilité et de ses quantiles (Totaro et al., 2020).

Au moment présent il n'y a aucun consensus méthodologique sur l'estimation des variables de dimensionnement sous changement climatique (François et al., 2019). Ainsi, beaucoup de tests de tendances sont mal utilisés et mal compris particulièrement quand des inférences non stationnaires sont basées sur leurs sorties (Serinaldi et al., 2018). En effet, Iliopoulou & Koutsoyiannis (2020) montrent que les approches basées sur les tendances locales sont moins



performantes que les approches basées sur les moyennes locales (segment de 30 années) et globales (sur toute la période de mesures : jusqu'à 150 années). Or la mise en œuvre des modèles non stationnaires, dans l'analyse des phénomènes de réchauffement climatique, est dominée dans la pratique par l'étude des tendances linéaires locales. De plus, la mauvaise interprétation des différents segments d'un processus de persistance à long terme (dynamique de Hurst-Kolmogorov) peut éventuellement expliquer pourquoi il était si commun de clamer l'existence de tendances, sur les précipitations extrêmes, ces dernières années (Iliopoulou & Koutsoyiannis, 2020).

La conséquence de ce qui vient d'être souligné plus haut, est que les études d'estimations de quantiles sur les extrêmes pluviométriques, sur la base de données historiques et de projections futures sont assez contrastées. Ainsi, certaines publications (Yilmaz et al., 2014) ne détectent aucune non-stationnarité, même s'ils relèvent une légère augmentation des intensités pour la période projetée. D'autres (Cannon & Innocenti, 2019; Ragno et al., 2018; Westra et al., 2014), au contraire prévoient une augmentation de l'intensité et de la fréquence des événements extrêmes sous hypothèse de non-stationnarité. Une troisième catégorie de publications (De Luca & Galasso, 2018; Ganguli & Coulibaly, 2017; Mondal & Daniel, 2019), ne trouvent pas de différences notables entre modèle stationnaire et non stationnaire pour l'estimation des pluies/débits de projets. A souligner que Ouarda et al. (2019) trouvent que les modèles non-stationnaires s'ajustent mieux au données (historiques et projections futures) au sens des tests AIC et BIC dans la validation d'un modèle de courbes IDF.

L'utilisation d'un modèle non-stationnaire, dans les études de conception, implique que la variation temporelle de la structure du modèle s'étale sur toute la durée de vie de l'ouvrage. Un modèle probabiliste non-stationnaire ne peut être basé, uniquement, sur des mesures au site. Quand le modèle non-stationnaire est inféré de manière inductive cela engendre une source supplémentaire d'incertitude et n'apporte aucune amélioration sur la fiabilité d'estimation (Serinaldi & Kilsby, 2015). En effet, La non-stationnarité devrait être invoquée uniquement quand le futur peut être prédit de façon déterministe, et associée à la réduction des incertitudes, dans le cas contraire il en résulte une sous-estimation de la variabilité des incertitudes et donc du risque (Koutsoyiannis & Montanari, 2015). Ainsi, quand la structure de modèle non-stationnaire est incertaine le modèle stationnaire avec une évaluation convenable des incertitudes est une solution plus cohérente et fiable pour des applications pratiques dans le cadre de la conception et la gestion des ouvrages.

## **2.5 Invariances d'échelle simple et propriétés multifractales des extrêmes pluviométriques**

La théorie mathématique des processus stochastiques a été fondée par Kolmogorov (1931, 1933 et 1938) de même que les processus d'invariance d'échelle (Koutsoyiannis et al., 2018). Ainsi, l'utilisation des propriétés statistiques des observations dans un processus inférentiel est d'un intérêt primordial dans le développement d'analyses hydrologiques avec peu ou sans mesures directes. Durrans (2013) définit les notions d'invariance d'échelle sur les pluies maximales journalières et horaires comme étant spécifiquement développées pour pallier

l'incohérence théorique des modèles IDF : incohérence due au fait d'une modélisation séparée des observations de pluies sur les diverses durées d'agrégation.

L'origine de l'invariance d'échelle des pluies n'est pas bien comprise, toutefois il est souvent fait le lien entre multifractalité et turbulence atmosphérique. Néanmoins, l'invariance d'échelle multifractale a été observée aussi bien dans les pluies d'origine convective que stratiforme (Willems, 2000). Quand c'est la convection diurne qui domine, la gamme de durée sur laquelle s'étend l'invariance d'échelle est réduite à quelques heures. En dehors de cette période, les fluctuations des intensités de pluies sont plus faibles que sous la multifractalité (Veneziano et al., 2007). À noter que, l'intermittence des pluies peut rendre difficile l'application du principe d'invariance d'échelle. Toutefois, le problème ne se pose pas dans notre cas, car seules les pluies maximales annuelles sont l'objet de cette étude (Klemes, 2000).

### 2.5.1 Qu'est-ce que l'invariance d'échelle simple (*Simple Scaling*)

Gupta & Waymire (1990) présentent l'invariance d'échelle simple (ou multiple) comme ayant deux niveaux de définition, en relation avec les caractéristiques statistiques considérées, dont l'un induit l'autre. La première définition est l'invariance d'échelle simple (ou multiple) au sens strict qui fait référence aux propriétés d'échelle de la distribution fréquentielle. La deuxième est l'invariance d'échelle simple (ou multiple) au sens large qui fait référence aux propriétés d'échelles des moments. Le sens strict de l'invariance d'échelle multiple (ou simple) a été utilisé en premier pour fournir une représentation générale des cascades multiplicatives dans les processus stochastiques sur les pluies. Le sens large de l'invariance d'échelle multiple (ou simple) : connaît des applications nombreuses dans la théorie statistique des pluies. C'est-à-dire dans l'étude spatio-temporelle des structures probabilistes des pluies (le comportement multifractal). En effet, c'est Burlando & Rosso (1996) qui ont étendu le principe d'invariance d'échelle sur les structures probabilistes spatiales des pluies aux structures probabilistes temporelles des pluies.

Soit  $Y$  une variable aléatoire. La fonction  $F(y) = P(Y \leq y)$  est appelée fonction de densité de probabilité cumulative de la variable  $Y$ . Si  $Y$  correspond aux cumuls des pluies maximales annuelles, la période de retour s'écrit :  $T(y) = 1/[1 - F(y)]$ . Le quantile  $y_T$  est la valeur de  $y$  qui est défini à travers la fonction  $T(y)$ , qui est égale à  $T$  :  $y_T = F^{-1}(1 - 1/T)$ .

La notion d'invariance d'échelle au sens strict qui permet de considérer que les distributions fréquentielles, des cumuls de pluies maximales annuelles sur diverses durées  $d$ , sont toutes identiques à un facteur d'échelle près dépendant de la durée  $d$ , s'exprime à travers la relation suivante (Menabde et al., 1999):

$$Y_{d,T} = d^{k(q)} y_T \quad (2.24)$$

Où,  $d^{k(q)}$  est le facteur d'échelle et  $k(q)$  représente la fonction d'exposant d'échelle qui sera définie à la section suivante. Le sens large de l'invariance d'échelle, qui est confondu avec les propriétés d'échelle multifractales sur les séries des maxima annuels de pluies, implique la relation suivante (Burlando & Rosso, 1996; Schertzer & Lovejoy, 1987; Veneziano & Furcolo, 2002) :

$$E[Y_d^q] = \alpha_q d^{k(q)} \quad (2.25)$$

Avec  $E[Y_d^q]$  représentant les moments d'ordre  $q$  sur des durées  $d$ . Sachant que, l'équation (2.24) est la définition même du principe d'invariance d'échelle au sens strict, et que la relation (équation 2.25) sur les échelles des moments est confondu avec l'invariance d'échelle au sens large, il en résulte qu'elle découle de la relation (2.24) mais que l'inverse n'est pas vrai.

Ainsi, il est clair que la définition des propriétés d'échelle multifractale correspond à la même définition que l'invariance d'échelle multiple au sens large, en faisant référence aux propriétés d'échelle des moments statistiques. L'étude des propriétés d'échelle des moments statistiques est l'une des méthodes les plus couramment utilisée pour identifier les propriétés multifractales des observations de pluies.

### 2.5.2 Caractère multifractal des séries des maxima annuels sur diverses durées

Veneziano et al. (2006) est le premier à avoir plaidé pour une définition unique de la multifractalité et soutient que la définition la plus intéressante et la plus générale est que la multifractalité soit considérée comme étant une propriété d'invariance d'échelle (tel qu'explicité à la section précédente).

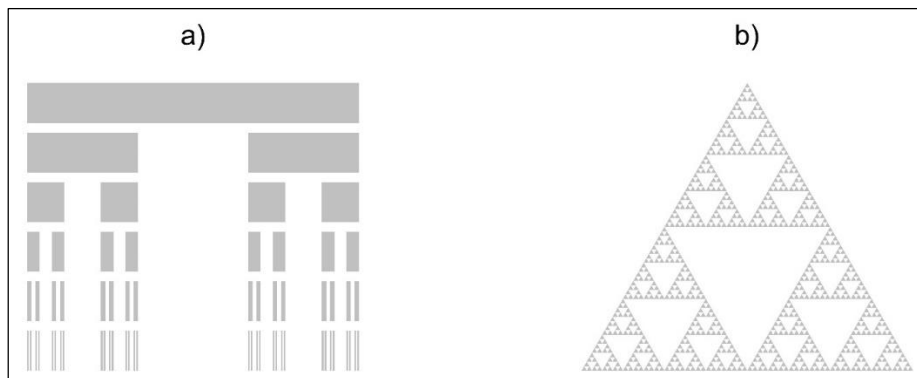


Figure 2.3 Les 6 premiers niveaux de : (a) l'ensemble de Cantor (1984) et du triangle de Sierpinski (1915) (Panthou, 2013).

Bien que les premiers objets dits « fractals » soient apparus au 19<sup>ème</sup> siècle (figure 2.3), c'est Mandelbrot (1974) qui donna son nom à cette notion, au départ géométrique, sur les objets dont la structure se retrouve à plusieurs échelles. Par la suite, l'utilisation de cette notion a été étendue à divers domaines scientifiques, et elle est aujourd'hui fréquemment utilisée en hydro-climatologie. En effet, dès qu'un objet d'étude possède les mêmes propriétés pour différentes échelles (dans une gamme d'échelle définie), la théorie des fractales peut être utilisée pour le décrire.

Sivakumar (2001) note l'existence de quatre méthodes pour prospecter la nature multifractale des pluies : (1) la fonction d'autocorrélation ; (2) le spectre de puissance ; (3) une fonction de distribution de probabilités empirique ; (4) une analyse d'échelle des moments statistiques (MSA de l'anglais *Moments Scaling Analysis*).

Toutefois, la méthode d'identification des propriétés multifractales la plus couramment utilisée est la méthode d'échelle des moments statistiques (MSA) introduite par Gupta &

Waymire (1990). Elle consiste à vérifier deux conditions que sont la log-log linéarité entre les moments d'ordre  $q$  des pluies avec leurs durées correspondantes et la linéarité (ou non) de la fonction d'exposant d'échelle  $k(q)$  avec l'ordre  $q$  des moments, pour justifier l'utilisation du principe l'invariance d'échelle. La figure 2.4 montre que si la fonction d'exposant d'échelle  $k(q)$  est définie : (1) suivant une relation graphique non-linéaire, elle est dite multifractale ; (2) suivant une relation graphique de droite ne passant pas par l'origine, elle est dite monofractale ; (3) suivant une relation graphique linéaire (ne passant donc pas par l'origine), elle est dite autosimilaire (García-Marín et al., 2013).

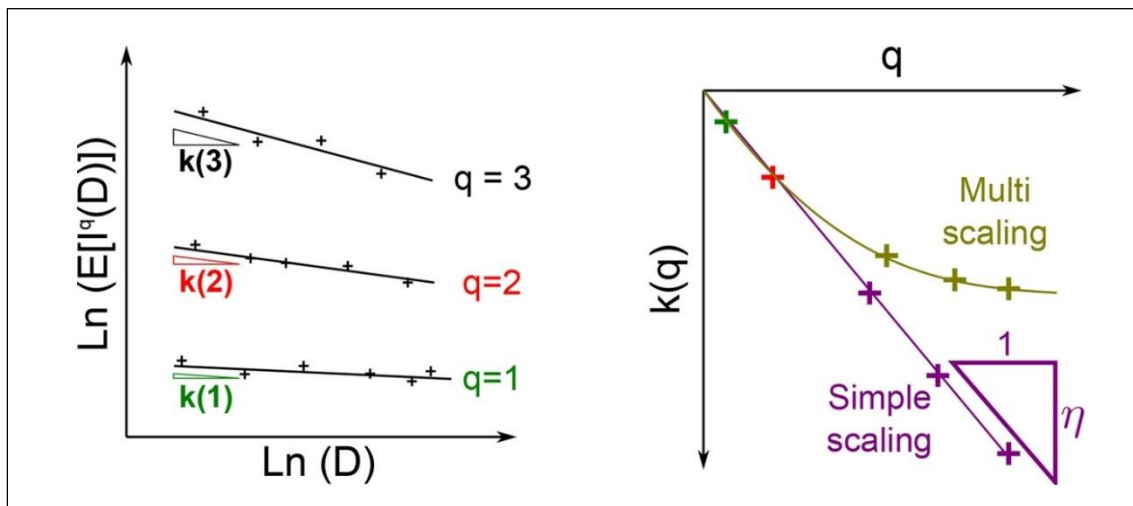


Figure 2.4 Schématisation de l'analyse d'échelle des moments : (à gauche) Log-log linéarité entre les moments et les durées, (à droite) Relation d'exposant d'échelle (Panthou et al., 2014)

### 2.5.3 Invariance d'échelle simple et multifractalité dans la modélisation des précipitations extrêmes

Les notions d'invariance d'échelle, de multifractalité et de cascade multiplicative ont surtout été utilisées pour générer des séries simulées à des pas de temps inférieures à 24 heures (Olsson & Berndtsson, 1998; Serinaldi, 2010). Cependant, au cours des dernières années, elles ont de plus en plus été utilisées en analyse des structures probabilistes des pluies à des fins de modélisation statistique des extrêmes pluviométriques sur la base de données historiques (Casas-Castillo et al., 2018.; Innocenti et al., 2017; M.I.P. de Lima & de Lima, 2009). Aussi, un comportement fractal peut être décelé, aussi bien sur des données réelles, que sur des données simulées, toutefois les séries simulées ont plus de mal à rendre compte de la non-linéarité de la fonction d'exposant d'échelle (García-Marín et al., 2008).

Le fait que la pluie ne soit pas un processus multifractal simple augmente la complexité des modèles, toutefois il a été montré que même des modèles grossiers suffisent pour évaluer les extrêmes pluviométriques (Veneziano et al., 2006). Les concepts utilisés en analyse fractale sont stochastiques, ainsi les techniques fractales peuvent directement être intégrées dans la théorie des processus stochastiques (Koutsoyiannis et al., 2018).

De nos jours, il est bien établi que les séries des maxima annuels des pluies satisfont à la notion d'invariance d'échelle, sur une certaine gamme de durée, et que même les modèles

statistiques empiriques les plus simples, tels que les courbes IDF classiques présentent une structure multifractale (Veneziano & Furcolo, 2002). En effet, le modèle de Montana incorpore un facteur d'échelle et donc une fonction d'exposant d'échelle calée empiriquement sur des estimations de quantiles de diverses durées.

Ainsi, le principe d'invariance d'échelle et le comportement fractal des observations de pluies, ont connu diverses applications en modélisation statistique sur les extrêmes pluviométriques. Les plus fréquentes étant un enrichissement du jeu de données de manières spatiales (régionalisation) ou temporelle (Blanchet et al., 2016; Burlando & Rosso, 1996; Menabde et al., 1999; Muller et al., 2008) dans l'estimation des courbes IDF. Mais aussi en tant que test de validation de modèles statistiques, où leurs performances est basée sur leurs capacités à reproduire des structures fractales (García-Marín et al., 2013), et même en tant que critère de similarité statistique dans le regroupement en régions homogènes pour des analyses fréquentielles régionales (García-Marín et al., 2015). A signaler que certains travaux, considérant l'invariance d'échelle et les propriétés fractales des pluies, n'en ont pas fait une utilisation optimale en maintenant une incohérence théorique dans la formulation (Pizarro et al., 2015) ou bien en ne profitant pas de toutes les potentialités offertes et même en complexifiant inutilement le modèle (Langousis et al., 2009 ; Langousis & Veneziano, 2007).

#### **2.5.4 Invariance d'échelle simple vs invariance d'échelle multiple**

Burlando & Rosso (1996) montrent que quand l'invariance d'échelle multiple (l'exposant d'échelle a une relation multifractale) existe, mais que l'invariance d'échelle simple reste plausible, l'utilisation de l'invariance d'échelle multiple dans la modélisation des intensités de pluies (plutôt que l'invariance simple) améliore très peu les performances du modèle.

Van de Vyver (2018) suggère la formulation d'une relation IDF sur la base des propriétés d'invariance d'échelle multiple dans le cadre d'une estimation Bayésienne intégrée, ce qui conduit à un modèle ayant plus de paramètres que celui proposé ici. Il montre que le modèle multi-échelles est plus performant que le modèle simple échelle. Ceci s'explique très certainement par le fait que les deux modèles ont été calés sur une très large gamme de durées allant de 1 heure à 48 heures. Pour des applications nécessitant la connaissance des pluies sur plus de 24 heures il est certainement préférable d'utiliser un modèle basé sur l'invariance d'échelle multiple. Par contre pour des usages d'intensité de pluies de durées inférieures à 24 heures, le modèle basé sur l'invariance d'échelle simple reste un compromis assez performant.

## **2.6 Cartographie des précipitations extrêmes**

La cartographie des extrêmes pluviométriques peut concerner différentes caractéristiques des pluies telles que : moyenne, médiane, les paramètres des lois statistiques ajustées, quantile estimé au site, coefficient de Montana, etc.

Globalement, la cartographie consiste à connaître en tous points d'une carte, ou plus spécifiquement d'une grille (dont le maillage est variable), une valeur caractéristique de la

variable étudiée. Ainsi, il revient à transférer une information en des points de mesures (par exemple : station pluviométriques) connus vers l'ensemble de l'espace étudié. Cette spatialisation est réalisable à travers des techniques d'interpolation spatiale dites déterministes telles que la méthode du polygone de Thiessen ou la pondération par l'inverse de la distance (IDW : *inverse distance weighting*), ou bien via des méthodes statistiques dites géostatistiques comme le Krigeage. Les analyses comparatives de performance entre Krigeage et IDW sont assez contrastées (Ly et al., 2011). Toutefois, pour une spatialisation sur la base de faible densité de réseau de mesure, le Krigeage est nettement plus performant (Borga & Vizzaccaro, 1997). Aussi, ce dernier est conçu pour donner la meilleure estimation linéaire non-biaisée (Szolgay et al., 2009). De plus, le Krigeage a la capacité de fournir une incertitude sur les prévisions (C'est-à-dire une incertitude d'interpolation en tous points de la grille) et peut permettre d'éviter les valeurs négatives de prévisions (Diggle & Ribeiro, 2007; Kitanidis, 1997). A signaler que l'interpolation spatiale par Krigeage permet l'incorporation de données secondaires qui peuvent substantiellement améliorer les prévisions. Etant donné la faible densité du réseau de mesure la zone d'étude c'est le Krigeage qui est retenu pour illustrer les apports du modèle IDF Bayesian intégrée proposé ici. Dans ce qui suit il est donné une définition succincte du Krigeage et de ses outils, de même qu'un aperçu sur l'utilisation des analyses multivariées intégrant des paramètres morphométriques et sur les apports des covariables dans le krigeage.

### 2.6.1 Interpolation spatiale par Krigeage

La géostatistique a été initiée par l'ingénieur des mines Sud-Africain Danie G. Krige, néanmoins la méthode et le terme Krigeage ont été formalisé par (Matheron, 1971). La technique géostatistique la plus fréquemment utilisée en spatialisation des extrêmes pluviométriques est très certainement le Krigeage (Song et al., 2015). Il se base sur l'analyse structurelle des points de mesure, aussi appelée variographie, c'est-à-dire que le variogramme quantifie la variabilité spatiale d'un phénomène spatial. Par définition, pour une fonction aléatoire stationnaire  $Z$  avec une moyenne  $m$  et une variance  $\sigma^2$  connues, le semivariogramme est définie comme étant la moitié de la variance des différences d'observations entre deux positions séparées par  $h$ , où  $h$  est la distance et  $x$  un vecteur de position (Kebaili Bargaoui & Chebbi, 2009) :

$$\gamma(h) = \frac{1}{2} \text{Var}[Z(x) - Z(x - h)] \quad (2.26)$$

En assumant que la moyenne et la variance sont indépendantes de la position, le semivariogramme est la fonction de  $h$  (équation 2.27) correspondant à la différence espérée aux carrés entre deux positions séparées par  $h$ .

$$\gamma(h) = \frac{1}{2} E\{[Z(x) - Z(x - h)]^2\} \quad (2.27)$$

Où  $\gamma(h)$  est une fonction croissante positive. Le semivariogramme peut être calculé s'il y a anisotropie dans le champ aléatoire. La détermination du semivariogramme expérimental implique un calcul entre les valeurs des emplacements couplés. C'est-à-dire pour chaque paire d'observations (mesure à la station) sur une distance et une direction spécifique. Le plus souvent

chaque paire d'emplacement a une distance unique, ce qui rend difficile de tracer rapidement toutes les paires. Alors, il est plus commode de classer les différentes paires dans un groupe de distances (par exemple : toutes les paires de points se trouvant à plus de 2km et à moins de 3km). Le semivariogramme expérimental ainsi formé affiche les valeurs de semi-variance sur l'axe des ordonnées et la distance (ou le décalage) sur l'axe des abscisses (figure 2.5).

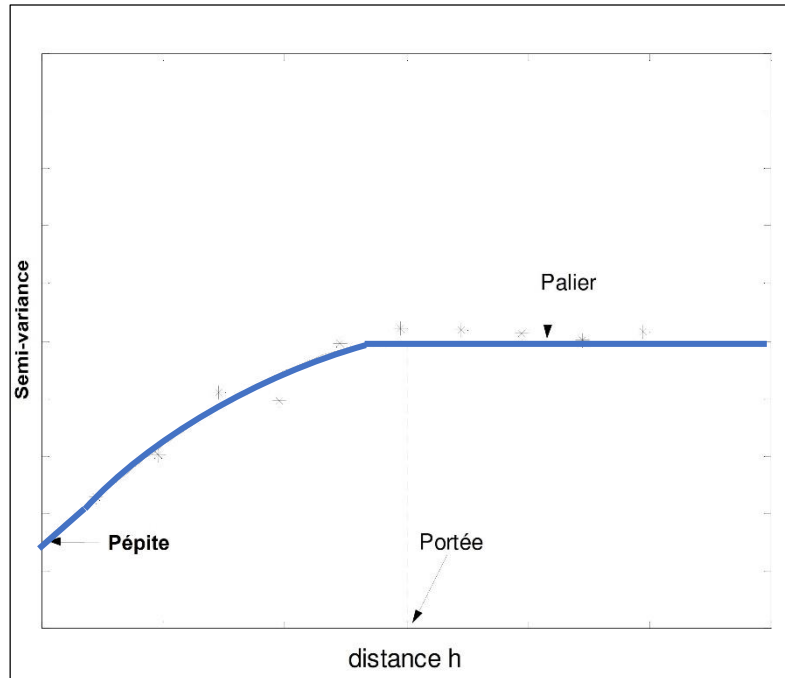


Figure 2.5 Exemple d'un modèle sphérique ajusté à semivariogramme expérimental

Le semivariogramme expérimental donne des informations sur l'autocorrélation spatiale des observations, sans toutefois indiquer les directions et les distances possibles. Ainsi, pour s'assurer, entre autres, que les prévisions présentent des écarts de Krigeage positifs, il est nécessaire d'ajuster un modèle au semivariogramme. Il existe un choix assez large de modèles à ajuster (Circulaire, sphérique, Gaussien, linéaire, exponentiel, etc.). Si par exemple il est considéré une fonction sphérique, qui est très couramment utilisée, cette dernière va s'élever dans un premier temps, puis se stabiliser (palier) pour les distances les plus grandes, au-delà de la portée (Figure 2.5). C'est-à-dire que le modèle sphérique indique une réduction progressive de l'autocorrélation (augmentation de la semi-variance) jusqu'à une certaine distance, au-delà de laquelle l'autocorrélation est théoriquement nulle.

L'interprétation du semivariogramme expérimental est très importante afin de saisir son comportement. Autrement dit, un variogramme est caractérisé par un seuil (ou palier) et une portée qu'il faut connaître. Le seuil (ou le palier) du variogramme est défini comme valeur de semi-variance qui correspond à zéro corrélation et elle est égale à la variance  $\sigma^2$  du champ. La portée peut être définie comme la distance sur laquelle les données sont corrélées. La première approximation de la portée est déduite graphiquement. Ainsi, la valeur à laquelle le semivariogramme atteint la portée est appelée seuil. Et un seuil partiel correspond au seuil moins une pépite (figure 2.5). Un semivariogramme présente un effet pépite quand deux positions séparées par une distance infiniment petite ont une valeur de semi variance non nulle (théoriquement pour une distance de séparation nulle, la semi-variance est nulle). Cet effet

pépité peut être attribué à des erreurs de mesures et/ou à des sources de variations spatiales à une distance inférieure à l'intervalle d'échantillonnage.

Le Krigeage génère des pondérations à partir des valeurs mesurées avoisinantes pour prédire les emplacements non jaugés. Les observations les plus proches des emplacements non mesurés auront bien-sûr la plus forte influence. La valeur estimée par krigeage est alors une combinaison linéaire des observations à  $N_v$  stations voisines (Kebaili Bargaoui & Chebbi, 2009):

$$\hat{Z}(x_0) = \sum_{i=1}^{N_v} \lambda_i Z(x_i) \quad (2.28)$$

Avec :

- $Z(x_i)$  : valeur relevée à l'emplacement  $i$  ;
- $\lambda_i$  : pondération inconnue de la valeur relevée à l'emplacement  $i$  ;
- $x_0$  : emplacement de krigeage (prévision).

Les pondérations  $\lambda_i$  s'appuient sur les distances entre les points de relevé et l'emplacement de la prévision, ainsi que sur l'organisation spatiale générale des points de relevé (de mesure), ces dernières informations sont déduites du variogramme théorique ajusté.

En conclusion, la méthode du Krigeage donne des estimations non biaisées et minimise l'écart-type de l'erreur d'estimation. Le Krigeage prend en considération, via le variogramme, la structure spatiale du processus (*data driven*), c'est-à-dire qu'il utilise les relations statistiques (autocorrélations) entre les points de mesures ce qui permet de produire une surface de prévision. Aussi, il offre la possibilité de prendre en considération une tendance dans la moyenne et/ou l'incertitude sur les données interpolées. Le Krigeage ordinaire est la version la plus utilisée (Song et al., 2015; Szolgay et al., 2009).

### 2.6.2 Analyse fréquentielle locale vs régionale

Les différentes approches d'analyse fréquentielle se subdivisent en quatre catégories : la première est l'analyse fréquentielle locale univariée, la seconde est l'analyse fréquentielle régionale univariée, la troisième est l'analyse fréquentielle locale multivariée (ou bivariée), et la quatrième (et la moins étudiée) est l'analyse fréquentielle régionale multivariée (Chebana & Ouarda, 2007; Chebana et al., 2009). Dans ce travail de thèse, la modélisation Bayésienne intégrée des courbes IDF est basée sur une analyse locale univariée puis il est prospecté la possibilité d'étendre son application en analyse fréquentielle régionale univariée. Ce sont les deux seules approches d'analyse fréquentielle retenues. Un des atouts majeurs de l'analyse fréquentielle régionale est qu'elle permet de substituer le temps par de l'espace. En effet, Hosking & Wallis (1997) ont montré que des observations d'une longueur de 600 années équivalent stations (20 stations avec en moyenne 30 années d'enregistrements) permettaient de bonnes estimations pour une gamme de période de retour allant jusqu'à 1000 ans. Ce qui est plus concordant avec les règles de 5 fois T et 3 fois T respectivement de Institute of Hydrology (1999) et du CERTU (2003), pour des estimations de quantiles supérieures à 100 ans. En



définitif, l'analyse fréquentielle locale reste applicable pour des estimations jusqu'à des période de retour de 100 ans, tandis que l'analyse fréquentielle régionale est préférable pour des estimations de période de retour allant de 100 à 1000 ans.

Dans des études comparatives entre approches locales (calage séparé et intégré) combinées à des interpolations spatiales et approches régionales pour l'estimation des courbes IDF en des sites non jaugés, l'analyse fréquentielle régionale paraît être plus appropriée (García-Marín et al., 2019; Kyselý et al., 2011; Panthou et al., 2012). Particulièrement sur des séries d'observation de petites tailles : variabilité d'échantillonnage et gestion difficile des horsains en analyse locale.

### 2.6.3 Méthode par estimation régionale des quantiles des pluies

Au cours de ces 25 dernières années la pratique la plus courante d'analyse régionale sur les variables hydrologiques est celle définie par Hosking & Wallis (1997), et qui consiste à normaliser (ou réduire) les observations, à l'intérieur de régions supposées homogènes, par un facteur d'échelle (moyenne, médiane ou toutes autres propriétés statistiques ou caractéristiques morphométriques) afin de les intégrer à un échantillon conjoint unique (*pooled sample*) qui sert de base à une inférence statistique régionale dont les issues finales sont les quantiles régionaux adimensionnels. Ainsi, il est possible de déduire, les quantiles sur une période de retour  $T$  donnée, en chaque point où le facteur d'échelle est connu. Cette pratique peut être étendue sur une très haute résolution spatiale, particulièrement pour les précipitations, en faisant appel aux techniques géostatistiques qui permettent de spatialiser le facteur d'échelle considéré en tous point d'une grille d'interpolation, et pas seulement aux sites où le facteur d'échelle est connu, afin de constituer une cartographie d'une variable hydrologique quelconque observée sur une échelle  $d$  et pour une période de retour  $T$  spécifique. C'est la cartographie régionale des variables hydrologiques.

Néanmoins, l'analyse régionale des variables hydrologiques couvre un large spectre de techniques et de procédures, et pas uniquement l'analyse fréquentielle régionale de Hosking & Wallis (1997). En effet, Durrans & Kirby (2004) définissent la régionalisation comme étant toute méthode cherchant à substituer le temps par l'espace. C'est-à-dire quand les observations hydrologiques sont limitées car elles s'étalent sur de courtes périodes de temps, elles peuvent être augmentées et améliorées en utilisant les observations disponibles au niveau d'autres sites hydrologiquement similaires.

Même si l'utilisation de l'analyse fréquentielle régionale, dans l'élaboration des courbes IDF reste la plus fréquente pour extrapoler les résultats d'estimation de quantiles (Brath et al., 2003; Burn, 2014; Das, 2019; Endreny & Pashiardis, 2007), ou même les paramètres du modèle IDF (Blanchet et al., 2016; Ombadi et al., 2018; Sane et al., 2018), vers des sites non-jaugés, d'autres pratiques sont aussi très répandues. Telle que la régionalisation des précipitations extrêmes à travers l'établissement de relation de régression (linéaire simple et/ou multiple et non-linéaire) sur les ratios des pluies permet de faire des estimations en des sites ne disposant que de mesures pluviométriques (absence de mesures horaires ou infra journalières) (Alila, 2000; Yu & Chen, 1997). Mais aussi les méthodes basées sur un facteur de réduction surfacique

(ARF : Areal reduction factor) pour des averses centrées (mesures radar) ou surface fixe (mesures au sol), puis de trouver des relations analytiques entre les ARF afin d'extrapoler vers les sites non-jaugés (De Michele et al., 2011; Kang et al., 2019; Mèlèse et al., 2019; Panthou et al., 2014). Toutefois, même si l'utilité des ARF est établie depuis assez longtemps (Sivapalan & Blöschl, 1998), certaines études récentes (Pavlovic et al., 2016) comparant différentes procédures d'utilisation des ARF, dans des zones fortement fournies en mesures pluviographiques et radar, montrent que toutes les approches ARF étudiées aboutissaient à des incertitudes significatives dans leurs estimations.

#### **2.6.4 Méthode par interpolation des quantiles de pluie au site**

La manière la plus simple d'étendre les estimations des quantiles de pluies en des sites non-jaugés est basée sur les techniques d'interpolation spatiale des résultats obtenus en des points mesurés. Cette approche peut être mise en œuvre sur les quantiles mêmes (Di Baldassarre, Castellarin, et al., 2006; Meddi & Toumi, 2015) ou sur les paramètres des distributions fréquentielles et/ou du modèle IDF (Borga et al., 2005; Soltani et al., 2017; Uboldi et al., 2014). Toutefois, ce genre d'approche échoue bien souvent à correctement reproduire la répartition spatio-temporelle des précipitations particulièrement en cas d'observations sur de courtes périodes et de faible densité des réseaux de mesure (Grundmann et al., 2019). Ce qui peut avoir comme conséquence : une augmentation directe des incertitudes d'estimation (cf. chapitre 6 de la présente thèse).

#### **2.6.5 Analyses multivariées dans la spatialisation des précipitations**

##### *2.6.5.1 Approche basée sur la régression multiple : Précipitations et morphométrie*

L'utilisation de relation de régression (linéaire simple ou multiple, ou non-linéaire), dans la spatialisation (ou la régionalisation) des extrêmes pluviométriques, a été appliquée de façons assez variées. Ceci, en combinant les propriétés statistiques des pluies elles-mêmes avec les caractéristiques morphométriques des bassins ou des zones étudiés, particulièrement dans les zones à relief contrasté et nécessitant de fortes densités de réseaux de mesures (Meersmans et al., 2016). Par exemple, il s'agit d'appliquer des relations de régression linéaire multivariée pour l'indentification de la relation entre les pluies de courtes durées et le relief (altitude, distance à la mer, etc.), puis d'implémenter une cartographie prenant en compte la structure spatiale des résidus de la régression multivariée (Kieffer-Weisse & Bois, 2001; Mebarki, 2003; Zahar & Laborde, 2007). Ces dernières techniques peuvent être assez utiles quand les séries d'observations sont courtes et que le réseau de mesures est peu dense, particulièrement dans les zones à relief homogène. Ces méthodes sont aussi utilisées pour établir des relations entre les paramètres des courbes IDF, tel que l'exposant d'échelle, et les pluies maximales journalières dans le but d'estimer des courbes IDF en des sites ne mesurant pas les pluies horaires (Yu et al., 2004). Aussi, certains auteurs proposent de combler les séries de mesures et/ou de les étendre afin d'avoir des échantillons de même longueur sur lesquels est basée une analyse fréquentielle régionale pour l'établissement des courbes IDF. Le comblement et l'extension des échantillons peuvent être mis en œuvre à travers des régressions non-linéaires Haddad et al.,

2011; Haddad & Rahman, 2014) ou via une interpolation spatiale par Krigeage (Libertino et al., 2018; Pappas et al., 2014). Toutefois, les incertitudes apportées par les techniques de comblement/ extension n'ont pas été prises en compte, de même il est assez difficile de les quantifier.

### 2.6.5.2 Approche basée sur le processus max-stable

La géostatistique classique est basée sur les processus et distributions Gaussiennes, bien que ces derniers ne soient pas appropriés aux extrêmes hydrologiques, des techniques comme le krigeage sont quasi systématiquement utilisées dans l'interpolation spatiale sur, par exemple, les précipitations extrêmes (Davison et al., 2013). Les processus max-stables sont des modèles plus adaptés lorsque l'intérêt porte sur la modélisation des valeurs extrêmes, c'est-à-dire sur les queues de la distribution, la théorie des valeurs extrêmes nous dicte quelles distributions considérer. Ces dernières doivent être max-stables et imposent donc des contraintes sur les copules adéquates. Bien qu'il paraisse étrange au premier abord de parler de copules pour les processus stochastiques, leur utilisation peut être adéquate puisque les processus sont souvent observés en un nombre fini de positions et la procédure d'estimation est alors intrinsèquement multivariée (Ribatet & Sedki, 2013). Ainsi, l'utilisation des processus max-stable dans la cartographie des extrêmes pluviométriques est de plus en plus utilisée (Sebille et al., 2017; Tawn et al., 2018; Tye & Cooley, 2015; Yoon et al., 2015), offrant un meilleur cadre de spatialisation des quantiles des valeurs extrêmes, car elle permet de faire le lien avec la théorie des valeurs extrêmes, dans un domaine de dimensions infinies où il est possible d'interpoler et d'extrapoler de façon directe vers des points non-jaugés. Toutefois, l'application des processus max-stable dans la modélisation spatiale des extrêmes est assez compliquée du fait de la non disponibilité des fonctions de densité multivariées. Actuellement les méthodes basées sur la vraisemblance sont loin de fournir un cadre flexible pour l'inférence. Certains auteurs (Padoan et al., 2010) proposent une vraisemblance par paire, dérivée de la vraisemblance composite afin de réduire les incertitudes d'estimation. Cependant, des publications récentes (Huser et al., 2019) montrent des avancées prometteuses pour la mise en œuvre d'inférences sur une vraisemblance complète dans les distributions multivariées max-stables.

## 2.6.6 Quelle approche retenir dans la cartographie des pluies extrêmes ?

Les approches de régionalisation basées sur les régressions entre propriétés statistiques et caractéristiques morphométriques ne sont pas retenues dans la suite de ce travail, car ces dernières ne sont pas utiles en cas de faible densité des réseaux de mesures sur des reliefs assez homogènes, ce qui n'est pas le cas ici. En effet, pour les régions avec topographie contrastée, la relation entre pluie extrême et topographie est plus complexe (Johnson et al., 2016). De plus, définir la pondération de l'asymétrie d'une distribution des valeurs extrêmes en un point donné, à partir de l'asymétrie d'une carte est fortement contreproductif (McCuen, 2001). A signaler aussi, que des publications très récentes (Grundmann et al., 2019) proposent de reproduire la répartition spatio-temporelle des précipitations dans un bassin versant à travers une

modélisation pluie-débit stochastique inverse. Cette approche nécessite toutefois des mesures hydrométriques de grandes qualités.

Finalement, c'est une combinaison d'analyses fréquentielles univariées locale ou régionale avec une procédure géostatistique d'interpolation spatiale qui est retenue pour la cartographie des quantiles de pluies estimés sur diverses durées  $d$  et pour plusieurs périodes de retour  $T$  d'usage courant. Bien que les techniques de Co-Krigeage permettant d'inclure des données secondaires (par exemple sur le relief) aient montré leurs supériorités par rapport au Krigeage ordinaire, c'est ce dernier qui est, dans un premier temps, choisi comme méthode d'interpolation par souci de simplicité du modèle. Car le but ce travail, n'est pas de prospecter la meilleure technique de spatialisation, mais d'illustrer les avantages d'une estimation Bayésienne intégrée des quantiles de pluies - qui apportent précision, robustesse et prise en compte des incertitudes associées - dans la cartographie des précipitations extrêmes. L'accent sera donc mis dans cette thèse sur l'évaluation des incertitudes associées aux estimations de quantiles de pluies.

## **2.7 Caractérisation du risque : période de retour, probabilité de défaillance et précision d'estimation.**

### **2.7.1 Période de retour vs. Probabilité de défaillance**

Dans le cadre des projets de conceptions et d'évaluation du risque, la mise en œuvre d'analyse fréquentielle sur les extrêmes pluviométriques nécessite une définition claire des notions de stationnarité et de non-stationnarité, de période de retour et de risque de défaillance, pour une connaissance précise du risque (Serinaldi & Kilsby, 2015). En effet, des publications (Serinaldi, 2015) montrent que la notion de période de retour mène à des défauts de conception et induit de mauvaise utilisation. Ces travaux plaident pour l'utilisation des notions de risque (ou de probabilité) de défaillance, qui intègre la durée de vie de l'ouvrage, et de probabilité de dépassement. Car ces outils permettent une évaluation du risque plus cohérente et une meilleure communication

Pour rappel, et tel qu'expliqué en section 2.4.1, en hydrologie, il est communément assumé que la fréquence au dépassement  $p$  est constante d'une année à une autre, ce qui mène à une période de retour moyenne qui est égale à  $1/p$ . Cette expression est beaucoup plus complexe sous hypothèse de non-stationnarité, et même en assumant un processus stationnaire, cette définition n'est pas satisfaisante car, elle ne prend pas en considération les horizons de calcul et donc ne peut être représentative du temps entre évènements successifs de même magnitude. Certaines publications (Read & Vogel, 2015) recommandent de considérer plutôt la notion de fiabilité (qui tient compte de l'horizon de calcul) au lieu de la période de retour pour les analyses stationnaire et non-stationnaire. A noter que la notion de probabilité de défaillance d'un ouvrage est équivalente à la fiabilité et elle est égale à 1, moins la fiabilité. Alors, la probabilité de défaillance est la probabilité qu'une structure subisse un dépassement de son seuil critique de dimensionnement, c'est à dire sa période de retour, durant sa durée de vie (Meylan & Musy, 1999). Ainsi, la fiabilité considère une probabilité conjointe sur la fréquence de dépassement  $p$

et l'horizon de calcul  $N_c$  (durée de vie de l'ouvrage), contrairement à la période de retour qui s'appuie sur une probabilité moyenne. La fiabilité découle donc d'une loi binomiale, et elle est donnée par (Chow et al., 1988; Meylan & Musy, 1999; Read & Vogel, 2015; Stedinger et al., 1993) :

$$Fiabilité = (1 - p)^{N_c} \quad (2.29)$$

Ou de manière équivalente la probabilité (ou risque) de défaillance est donnée par l'expression suivante :

$$Risque\ de\ défaillance = 1 - (1 - p)^{N_c} \quad (2.30)$$

Par exemple, si un réseau d'assainissement pluvial, avec une durée de vie de 30 années, est dimensionné pour une période de retour de 10 ans, son risque de défaillance sera de  $1 - (1 - 0.1)^{30} = 0.96$ . Pareillement pour un évacuateur de crue d'un barrage de durée de vie 100 ans et dimensionné pour une crue centennale, son risque de défaillance sera de  $1 - (1 - 0.01)^{100} = 0.64$ . Ces chiffres de risque de défaillances permettent de mettre en perspective la fiabilité des dimensionnements afin de mieux considérer les périodes de retour à adopter.

### 2.7.2 Précision et incertitude sur les estimations de quantiles

Les méthodes d'analyse fréquentielle classiques sous-estiment systématiquement les risques. Leurs mises en œuvre sur des échantillons relativement court sont généralement affectées par de grandes incertitudes (Gaume, 2018). En effet, les propriétés des petits échantillons sont ignorées par les estimateurs de quantiles, l'évènement ainsi retenu pour la conception avec une fréquence de dépassement donnée correspondra réellement à une fréquence de dépassement plus grande (et donc une plus faible période de retour que celle retenue pour la conception) que celle attendue (Stedinger, 1983b).

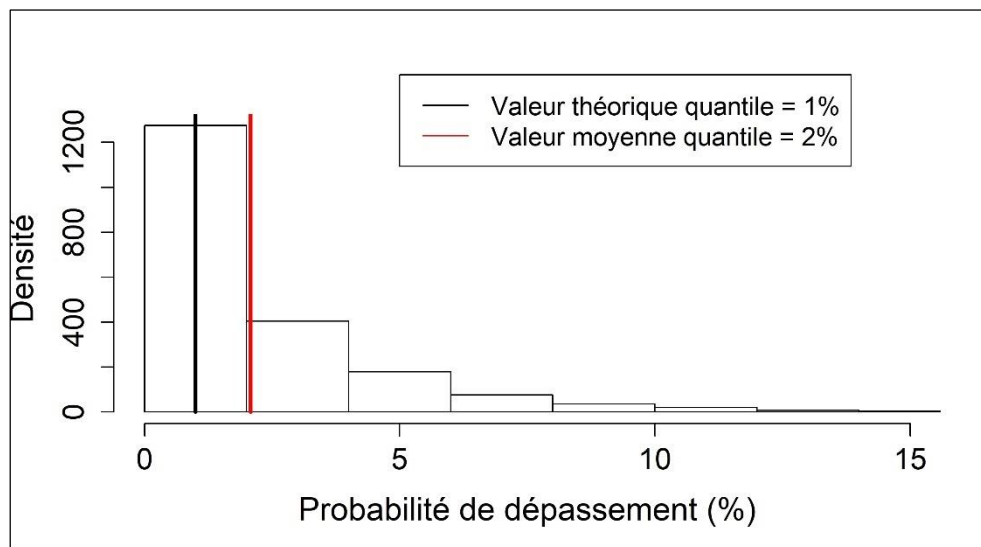


Figure 2.6 Distribution des fréquences de dépassement de quantiles centennaux de 2000 échantillons de 30 valeurs tirés dans une distribution GEV de référence (station 020506 de Bouzeréah)

A titre d'exemple, un quantile de pluie journalière centennal estimé à partir d'une distribution GEV de référence (station 020506 de Bouzeréah) ajustée à une série de 30 années, prend une valeur de 153mm ( $p=1\%$ ), tandis que le quantile centennal moyen sur 2000 échantillons de 30 valeurs tirés à partir de cette même distribution de référence prend une valeur 136mm, ce qui correspond à une période de retour  $T=50$  ( $p=2\%$ ) ans pour la distribution de référence (voir figure 2.6). A signaler que ce genre d'exemple a déjà été montré dans des publications de références (Gaume, 2018; Stedinger, 1983b). Ceci, devrait systématiquement inciter à l'utilisation de techniques permettant l'augmentation des tailles des séries à travers l'analyse régionale (Dalrymple, 1960; Hosking & Wallis, 1993, 1997) ou la considération d'informations historiques dans le processus inférentiel (Gaume et al., 2010; Payrastre et al., 2011; Viglione et al., 2013). En tout état de cause, cet état de fait devrait inciter à prendre en compte systématiquement les incertitudes d'estimation dans le dimensionnement des ouvrages.

Au regard de ce qui vient d'être montré, en analyses fréquentielles classiques les quantiles correspondants à une fréquence de dépassement donnée sont systématiquement sous-estimés et donc inférieurs aux quantiles de probabilité espérée (qui est la fréquence de défaillance de la structure hydraulique). Ainsi, la probabilité de défaillance est plus importante que celle correspondant à  $1/T$ . La probabilité annuelle de défaillance est quelque fois appelée distribution prédictive. Kuczera (1999) souligne l'importance de la probabilité de défaillance (espérée) et est le premier à recommander un cadre Bayésien à travers des simulations Monte Carlo pour l'estimation numérique de la distribution prédictive, car la formulation de sa densité de probabilité est inconnue et diffère probablement de la distribution GEV (Gaume, 2018). Un des principaux avantages de l'estimation des quantiles hydrologiques sur la base d'une distribution prédictive, est qu'elle permet de prendre en considération les incertitudes sur les paramètres des modèles. De plus, elle donne une estimation équivalente aux estimations d'incertitudes dans les méthodes d'analyse fréquentielle classique (Stedinger, 1997).

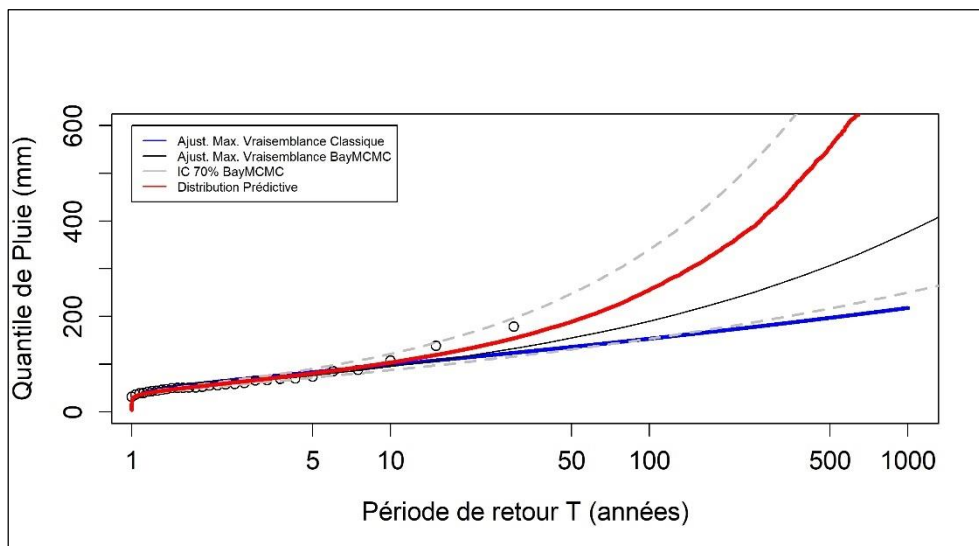


Figure 2.7 Illustration de la distribution prédictive (en rouge) avec la distribution GEV de référence ajustée par Bayesian MCMC (en noir) et par maximum de vraisemblance classique (bleu). Les intervalles de crédibilité à 70% (en gris) sont déduits de la distribution à posteriori de Bayesian MCMC.

Ainsi, il est clairement établi qu'il est préférable de baser les études de conception et d'analyse du risque sur les limites des erreurs d'estimation, vu les complications d'ordre

numérique qu'implique le calcul de la distribution prédictive. La borne supérieure de l'intervalle de crédibilité à 70% paraît être un bon compromis. En effet, en reprenant l'exemple précédant d'ajustement d'une distribution GEV à la station de Bouzeréah (020506), la distribution prédictive correspondante se positionne juste en dessous de la borne supérieure de l'intervalle de crédibilité à 70% (voir figure 2.7).

## 2.8 Courbes IDF : Outil de synthèse des extrêmes pluviométriques

Les courbes Intensité-durée-fréquence (IDF) sont utilisées depuis presque un siècle (Bernard, 1932), et les différentes méthodologies proposées pour leur établissement ainsi que les différentes formulations étudiées à travers les années sont des preuves indiscutables de leur importance. Elles représentent un outil d'intérêt primordial en hydrologie ayant un rôle crucial dans la conception efficace et sécuritaire d'infrastructures hydrauliques majeures et mineures (barrages, réseaux urbains, aménagements fluviaux, etc.) (El Adlouni & Ouarda, 2010). A noter que dans ces applications d'ingénierie les précipitations sont modélisées comme un processus autonome sans aucune référence à la dynamique atmosphérique (Koutsoyiannis & Langousis, 2011).

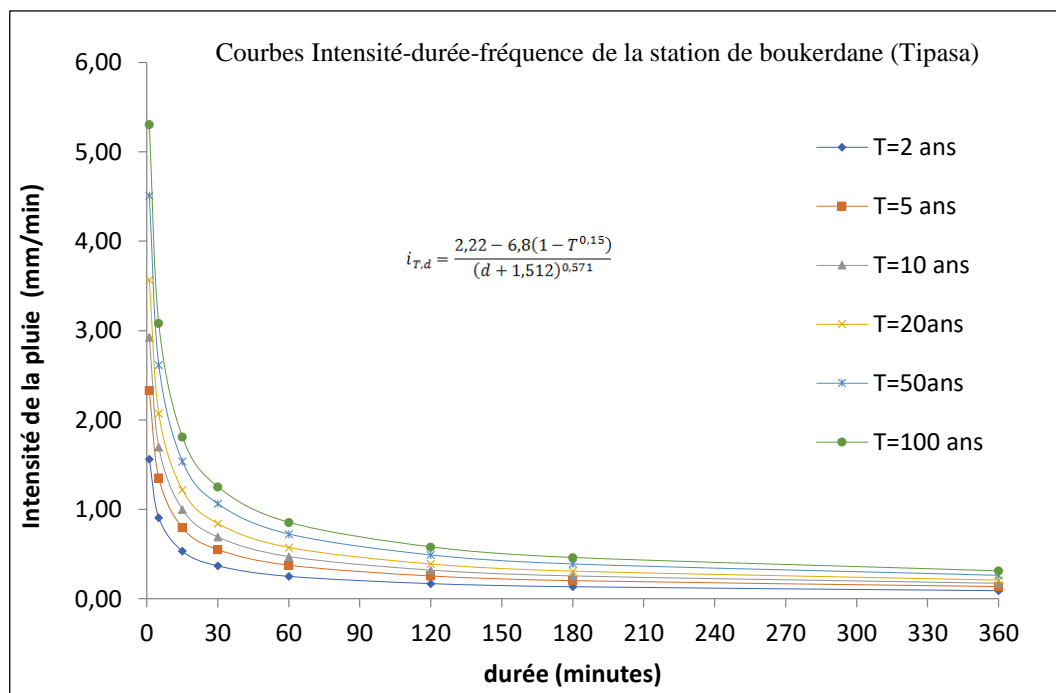


Figure 2.8 Courbes IDF modélisées par formulation semi-empirique ( Boukhelifa et al., 2011)

L'appellation courbes Intensité-Durée-Fréquence, peut être considérée comme générique car résumant les résultats d'analyses statistiques des valeurs extrêmes de pluies (cumuls ou intensités) pour différentes fréquences au dépassement (ou période de retour) sur diverses durées d'agrégation. L'intensité de la pluie maximale annuelle étant une quantité non-négative qui augmente de façon monotone avec la réduction de la fréquence et de la durée (figure 2.8). Leurs formulations ont évolué d'une forme purement empirique à une forme théoriquement plus cohérente, ainsi aujourd'hui encore leur étude reste un champ de recherche très actif

(Grimaldi et al., 2011). La non-négativité et la monotonie sont des contraintes majeures dans la formulation des courbes IDF (Cannon, 2018). Ceci explique, entre autres, que pendant très longtemps, les expressions des courbes IDF ont été établies de façon empirique ou semi empirique. Toutefois, cela ne signifie nullement que ces formulations empiriques ne fournissent pas une bonne représentation (relativement réaliste) des intensités de pluies. En effet, la plupart des modèles IDF sont par nature autosimilaires et donc satisfont à la relation d'invariance d'échelle simple (au sens strict), et que souvent les pluies temporelles montrent un comportement multifractal (Veneziano & Furcolo, 2002). Par contre, l'utilisation généralisée de ces modèles empiriques qui n'est pas possible. Bien souvent les modèles IDF empiriques sont calés sur des séries assez longues, et leurs formulations rendent compte des conditions climatiques pour lesquelles elles ont été développées. Dans ce qui suit, il est discuté des fondements théoriques des différentes formulations et de leurs limites d'application, aussi il est développé une formulation des courbe IDF mathématiquement cohérente et hydrologiquement fiable.

### **2.8.1 Utilisation en analyse hydrologique : Conception des systèmes de gestions des eaux de pluie.**

Traditionnellement, la pluie de projet considérée dans le dimensionnement des ouvrages de gestion des eaux de pluies (réseaux d'assainissement, bassins de rétention, déversoirs d'orage etc.) est une averse uniforme ayant une intensité constante sur toute sa durée. Cette dernière est fréquemment prise égale au temps de concentration du bassin versant étudié, car à égalité d'intensité, les pluies occasionnant le débit maximum en un point du réseau d'écoulement sont celles dont la durée de précipitation est au moins égale au temps de concentration de ce dernier (Remenieras, 1986). Cependant, très souvent il est nécessaire de connaître l'évolution des intensités de pluie en fonction du temps. Cette distribution temporelle de la pluie est obtenue via un hyétogramme (McCuen, 2004). Ces hyétogrammes synthétiques sont construits à travers plusieurs techniques telles que l'averse composite (Chow et al., 1988), le double triangle (CERTU, 2003), type Chicago (Keifer & Chu, 1957). L'ensemble de ces procédures, permettant de proposer une structure temporelle des pluies, découle des courbes IDF, c'est d'ailleurs une utilisation typique de ces dernières. Ce qui dénote de l'importance des courbes IDF en ingénierie hydraulique, particulièrement dans les sous-bassins ou les petits bassins urbains où la connaissance des structures temporelles des pluies est primordiale, comparativement à la structure spatiale, dans l'étude des écoulements (Zhu et al., 2018).

En plus de la connaissance des quantiles d'intensités de pluie sur des durées bien définies, à travers les courbes IDF, dans les études de conception et d'analyse du risque, certains auteurs (Carlo De Michele et al., 2001; Laborde, 1986) préconisent l'utilisation de facteur de réduction (ou d'abattement) spatial des intensités de pluies. Cela est très discutable, car la grande variabilité spatio-temporelle des pluies fait qu'il est relativement peu probable que le réseau de mesure pluviographique capture l'extrême pluviométrique durant l'averse. Ainsi, la méthode du facteur de réduction spatiale des pluies ne peut être considérée comme la réduction du maximum. L'utilisation d'une telle définition du facteur de réduction spatiale aboutit à une sous-estimation des extrêmes pluviométriques de l'ordre de 30 à 50% ,ce qui a pour



conséquence inévitable la sous-estimation du risque d'inondation urbaine (Bárdossy & Pegram, 2018).

### 2.8.2 Formulation empirique des courbes IDF

Traditionnellement la formulation des courbes IDF, en considérant un échantillonnage SMA, était basée sur des constatations empiriques (Chow et al., 1988). La procédure typique d'estimation des paramètres de courbes IDF consiste en trois étapes de calcul :

1. Ajustement d'une fonction de distribution de probabilité, préalablement choisie, à chacune des séries des maxima annuels de pluies sur durée  $d$  ;
2. Calcul des intensités (ou les cumulés) de pluies pour chaque durée  $d$  et pour une gamme de période de retour  $T$  (généralement : 5, 10, 20, 50 et 100 ans) en utilisant les distributions ajustées en étapes 1 ;
3. Etablissement d'une formulation des courbes IDF de deux manières différentes :
  - a. Pour chaque période de retour  $T$ , les intensités calculées dans l'étape 2 pour chaque durée  $d$ , sont formalisées, par moindres carrés, dans une relation d'intensité en fonction de la durée pour chacune des périodes de retour  $T$ .
  - b. Les intensités calculées dans l'étape 2 pour chaque durée  $d$  et pour toutes les périodes de retour  $T$ , sont formalisées, par moindres carrés à trois variables, dans une relation unique d'intensité en fonction de la durée.

Les formulations empiriques des courbes IDF, à travers la bibliographie, peuvent être résumées sous les trois expressions suivantes :

$$i(d, T) = \frac{a}{(d^\zeta + \theta)^\eta} \quad (2.31)$$

$$i(d, T) = \frac{aT}{(d^\zeta + \theta)^\eta} \quad (2.32)$$

$$i(d, T) = \frac{a + B \log T}{(d^\zeta + \theta)^\eta} \quad (2.33)$$

Avec  $i(d)$  et  $i(d, T)$  les intensités de pluies sur une durée donnée  $d$  et pour une période de retour  $T$ . Où  $a > 0$  et  $B > 0$  et représentent les paramètres d'échelle, et  $\theta \geq 0$ ,  $0 < \eta < 1$  et  $\zeta$  représentent les paramètres de formes ( $\zeta$  est souvent pris égal à 1, autrement il en résulte une sur-paramétrisation). Dans l'équation (2.31) en posant  $\zeta = 1$  et  $\theta = 0$ , il est obtenu le modèle empirique de Montana qui est très largement utilisé. De même, en posant  $\zeta = 1$  et  $\eta = 1$ , il est obtenu le modèle de Talbot recommandé pour des durées de moins de 1 heure.

La méthodologie classique d'établissement des courbes IDF n'est pas précise et fiable, car elle nécessite des ajustements séparés, ce qui aboutit à un grand nombre de paramètres à estimer et leur formulation n'est pas indépendante de la durée (Bougadis & Adamowski, 2006). Les paramètres de modèles IDF empiriques sont généralement calculés par moindre carrés ou régression multiple, où les incertitudes d'estimation ne sont pas prises en compte dans la procédure de calcul en trois étapes.

### 2.8.3 Formulation semi-empirique des courbes IDF

Koutsoyiannis et al. (1998) proposent une formulation semi-empirique où il y a une séparation de la durée et de la période de retour en deux fonctions distinctes (équation 2.34).

$$y_{d,T} = \frac{a(T)}{(d + \theta)^\eta} = \frac{a(T)}{b(d)} \quad (2.34)$$

Avec  $a(T)$  et  $b(d)$  respectivement les dépendances fonctionnelles sur la période de retour et sur la durée, et les paramètres numériques  $\eta$ ,  $\theta$  dépendent de la position géographique et des conditions climatiques. Cette séparation permet que la dépendance fonctionnelle  $a(T)$  sur la période de retour dérive explicitement de la fonction de distribution de probabilité sous-jacente des cumuls ou intensités de pluies extrêmes, ce qui apporte de la cohérence théorique comparativement aux formulations purement empiriques (équations de 2.31, 2.32 et 2.33). Même si elle apparaît convenable, cette séparation n'est pas la plus adaptée aux données d'observation (Veneziano et al., 2007) (voir aussi 2.8.4). Cette formulation semi-empirique compte comme la relation de courbe IDF la plus largement utilisée aux cours de ces vingt dernières années (Ben-Zvi, 2009; Boukhelifa et al., 2011; Endreny & Imbeah, 2009; Mohymont et al., 2004; Mohymont & Demarée, 2006; Van de Vyver & Demarée, 2010).

Koutsoyiannis et al. (1998) proposent trois approches pour le calcul des paramètres du modèle fréquentiel global sur les IDF :

1. En 3 étapes, tel que dans la procédure classique.
2. En 2 étapes, méthode dite robuste. Qui consiste en premier, à déterminer empiriquement les paramètres de la dépendance fonctionnelle sur la durée en minimisant la statistique de Kruskal-Wallis. Dans une deuxième étape, les séries de pluies sur diverses durées sont renormées par la dépendance fonctionnelle sur la durée afin de former un échantillon unique (invariance d'échelle simple implicite) sur lequel sera ajustée une fonction de distribution de probabilité par L-moments.
3. En une seule étape en combinant moindres carrés multiples et algorithme d'optimisation numérique.

Bien qu'elle présente plus de cohérence, la formulation semi-empirique reste entachée de la même limitation où les incertitudes ne sont pas prises en compte et la précision du résultat de calcul dépendent de l'ampleur de la variabilité d'échantillonnage (étant donné l'estimation empirique partielle). Même si la procédure d'estimation des paramètres en une étape réduit le nombre de calculs, la méthode des moindres carrés est moins robuste que toutes autres techniques d'inférence.

Dans une étude comparative entre différentes formulations empiriques et semi-empiriques avec mise en œuvre classique en trois étapes, Di Baldassarre et al. (2006) montrent que le modèle IDF type Montana n'est valable que si les données satisfont au principe d'invariance d'échelle simple et dans le cas contraire, le modèle a généralement du mal à reproduire avec fiabilité les pluies de projet de courtes durées (<1h). Et que le modèle fréquentiel global de Koutsoyiannis et al. (1998) demeure la meilleure alternative.

### 2.8.4 Formulation mathématique des courbes IDF

Dans ce qui suit, il sera montré l'existence d'une relation entre une possible forme de courbes IDF et quelques propriétés statistiques sur les séries de pluies observées, à savoir l'invariance d'échelle simple et la structure multifractale. Il existe un très grand nombre de formules reliant entre les quantiles d'intensité ou de cumul de pluie  $y_{d,T}$  sur une durée  $d$  et pour une période de retour  $T$ . La formule (équation 2.34) présentée par Koutsoyiannis et al. (1998) est d'une utilisation courante. Toutefois, il est préférable de considérer, comme point de départ, une formulation plus générale (équation 2.35) des relations de courbes IDF :

$$y_{d,T} = a(T)g(T, d) \quad (2.35)$$

En cas d'invariance d'échelle simple (au sens strict), les distributions fréquentielles des intensités ou cumuls de pluies sur diverses durées, en un site spécifique, sont identiques à un facteur d'échelle près :

$$y_{d,T} = a(T)g(d) \text{ ou } \frac{y_{d,T}}{g(d)} = a(T) \quad (2.36)$$

En d'autres termes, en cas d'invariance d'échelle simple, la fonction  $g$  ne dépend pas de la période de retour. A noter que la formulation IDF de Koutsoyiannis et al. (1998) (équation 2.34) est une forme particulière de l'équation (2.36) avec  $g(d) = (d + \theta)^{-\eta}$ . Son utilisation entraîne une hypothèse d'invariance d'échelle simple implicite. Dans l'équation (2.36),  $g(d)$  est le facteur d'échelle dépendant de la durée  $d$ , et  $a(T)$  est la distribution fréquentielle réduite (adimensionnelle).

Sachant que les séries des maxima annuels montrent des propriétés d'échelles multifractales, à savoir que les moments d'ordre  $q$   $E[Y_d^q]$  sur les intensités ou cumuls de pluies pour diverses durées ont une relation de proportionnalité avec le facteur d'échelle et subséquemment la fonction d'exposant d'échelle (équation 2.25). Alors, si l'équation (2.36) est valide, la distribution de  $\frac{y_d}{g(d)}$ , et ses moments, ne dépend pas de la durée  $d$ .

$$E\left[\left(\frac{Y_d}{g(d)}\right)^q\right] = \alpha_q d^{k(q)} g(d)^{-q} \quad (2.37)$$

Ceci implique que le terme  $d^{k(q)} g(d)^{-q}$  ne dépend pas de  $d$  :

$$d^{-k(q)} g(d)^q = h(q) \Rightarrow g(d) = h(q)^{1/q} d^{\frac{k(q)}{q}} \quad (2.38)$$

Mais par définition  $g(d)$  ne dépend pas de  $q$ , ainsi  $h(q)^{1/q} = \text{cst}$  et plus important encore  $k(q)/q$  est constant et égal à  $k(1)$ . Deux conclusions peuvent être tirées de la combinaison de ces deux hypothèses concernant les propriétés statistiques des intensités ou cumuls de pluies (invariance d'échelle simple et multifractalité) :

1. Premièrement la fonction  $g(d)$  est nécessairement de forme  $g(d) = \beta d^b$  avec  $b = k(1)$  et  $\beta = h(q)^{1/q}$ . Ainsi, la relation IDF prend la forme suivante :

$$y_{d,T} = a'(T)d^b, \text{ avec } a'(T) = \beta a(T) \quad (2.39)$$

2. Deuxièmement la fonction d'exposant d'échelle multifractale doit être linéaire :  $k(q) = q k(1)$ .

Finalement, l'équation (2.39) est la conséquence naturelle de la combinaison de deux hypothèses concernant les propriétés statistiques des séries d'intensités ou des cumuls des pluies : invariance d'échelle simple et multifractalité. Menabde et al. (1999) est arrivé à la même conclusion sur la base d'une démonstration légèrement différente. L'équation (2.39) correspond exactement à l'équation (2.34) quand  $\theta = 0$ . Plusieurs applications sur le modèle en équation (2.34) sont déjà arrivées à la même conclusion que le paramètre  $\theta$  est souvent proche de 0 (Blanchet et al., 2016; Koutsoyiannis et al., 1998; Menabde et al., 1999; Van de Vyver, 2015a).

Sur la base de ce résultat théorique, la validation locale du modèle IDF (équation 2.39) est fait en trois phases :

1. Evaluer si les deux hypothèses d'invariance d'échelle simple et de multifractalité sont raisonnablement cohérentes avec les données disponibles, afin de justifier l'utilisation de l'équation (2.39) en tant que modèle de courbe IDF. Dans la foulée, il peut être vérifié que c'est bien une fonction linéaire d'exposant d'échelle qui ajuste les données, ceci représente une indication supplémentaire de l'adéquation de l'hypothèse d'invariance d'échelle (cf. chapitre 4).
2. Le calage du modèle de courbes IDF par approche Bayésienne intégrée (cf. chapitre 5).
3. Evaluation de la concordance et de la valeur ajoutée de l'estimation Bayésienne intégrée comparativement à une approche Bayésienne séparée (c'est-à-dire : application du processus inférentiel sur chaque durée séparément) (cf. chapitre 5).

#### 2.8.4.1 *Quelques remarques sur la formulation mathématique proposée*

Dans la forme, le modèle proposé dans ce travail est identique à celui de Montana. Ainsi, seule l'autosimilarité est parfaitement cohérente avec l'invariance d'échelle simple. Dans le sens où l'hypothèse d'invariance d'échelle simple, implique que la fonction d'exposant d'échelle ne peut qu'être linéaire (c'est-à-dire que  $k(q)$  et  $q$  décrivent une relation linéaire passant par l'origine).

Le modèle proposé fait les hypothèses suivantes : stationnarité, indépendance des SMA, indépendance des pluies sur diverses durées d'agrégation (ce qui n'est pas tout à fait valide). En effet, l'incorporation des dépendances entre pluie sur diverses durées complique fortement la modélisation (Davison et al., 2012; Ribatet & Sedki, 2013). Toutefois, il y a nécessité d'incorporer les dépendances dans la modélisation des pluies sur diverses durées, seulement quand les distributions marginales sont sujet d'intérêt (Sebille et al., 2017 ; Veneziano et al., 2007).

#### 2.8.4.2 *Quelques exemples de modélisation mathématique des courbes IDF*

Plusieurs auteurs (Blanchet et al., 2016; Innocenti et al., 2017; Menabde et al., 1999; Panthou et al., 2014) utilisent le modèle proposé (équation 2.39) dans l'estimation des courbes

IDF. Cependant, la fonction d'exposant d'échelle est déduite d'une analyse d'échelle des moments (MSA) ce qui mène à une estimation erronée quand il y a forte variabilité d'échantillonnage (cf. chapitre 4). Et les paramètres de la distribution de probabilité sous-jacente sont inférés dans une deuxième étape. D'autres auteurs (C. H. R. Lima et al., 2018), retiennent le principe d'invariance d'échelle simple pour la modélisation des courbes mais préconisent une relation non linéaire pour la fonction d'exposant d'échelle, ce qui est incohérent avec le principe d'invariance d'échelle simple lui-même.

#### 2.8.4.3 Apport du modèle IDF retenu

D'un côté un nombre important de paramètres permet une plus grande flexibilité dans les ajustements. Mais de l'autre, dans un modèle à plusieurs paramètres l'effet des incertitudes sur ce dernier est amplifié (Di Baldassarre et al., 2006). Ainsi, l'utilisation des propriétés multifractales et du concept d'invariance d'échelle simple pour l'estimation des courbes IDF est très attractive du fait qu'elle présente un nombre significativement réduit de paramètres (i.e. que les statistiques des pluies sur diverses durées sont toutes liées par un facteur d'échelle). D'un point de vue mathématique, l'invariance d'échelle simple et la multifractalité permettent de faire une caractérisation des courbes IDF sous des conditions asymptotiques et pré-asymptotiques ce qui permet d'éviter des procédures de simulations Monte Carlo intensives (Tyrallis & Langousis, 2019; Veneziano et al., 2006, 2007; Veneziano & Furcolo, 2002; Veneziano & Langousis, 2010).

De plus, en analyse fractale, il y a souvent un manque d'appréciation des incertitudes sur le résultat. Il est bénéfique d'étudier les biais et les incertitudes issus des analyses fractales intégrées dans les processus stochastiques : dans un cadre théorique consistant. Un processus stochastique offre une base solide pour la modélisation et l'interprétation d'un phénomène naturel et peut directement incorporer, de manière rigoureuse, les concepts de fractalité, en dépassant l'ambiguïté qui caractérise plusieurs études fractales. Mettre en œuvre des calculs statistiques sans considérer les incertitudes et biais qui y sont liés peut conduire à l'identification de phénomène qui n'existe pas (Koutsoyiannis et al., 2018). Pour ces raisons, la validité et les hypothèses sur lesquelles repose le modèle IDF Bayésien intégré sont intensivement testées en chapitre 4 et 5.

Dans ce travail, la formulation d'une relation de vraisemblance intégrant la relation IDF (équation 2.39) dans sa globalité (ce qui n'est faisable que sous hypothèse d'invariance d'échelle simple) permet de calculer l'ensemble des paramètres (paramètres de la distribution de probabilité et de la relation d'échelle) en une seule étape, ce qui réduit les incertitudes opérationnelles. De plus une implémentation à travers un processus inférentiel Bayésien fournit une estimation des paramètres qui tient compte des incertitudes et de même permet de connaître les intervalles de crédibilité sur les quantiles estimés de la manière la plus efficiente possible (Reis & Stedinger, 2005).

## 2.9 Conclusion du chapitre 2

La formulation des courbes IDF retenue ici est le produit d'une fonction de distribution de probabilité adimensionnelle (réduite) et d'une relation d'échelle temporelle (ou facteur d'échelle), mise en œuvre sur des séries des maxima annuels de pluies sur diverses durées. A travers l'écriture de la vraisemblance du modèle et l'utilisation d'un algorithme Bayesian MCMC, et en considérant une distribution GEV, l'ensemble des paramètres (3 pour la distribution et 1 pour la relation d'échelle) du modèle IDF proposé (équation 2.39) sont inférés en une seule étape ce qui, en plus de l'incorporation des incertitudes dans le processus de calcul, permet de grandement réduire les incertitudes d'estimation sur les quantiles. Cette modélisation de relations de courbes IDF est mise en œuvre à travers le chapitre 4 et 5 où les hypothèses sont vérifiées, de même que la validité du modèle est intensivement testée afin de mettre à jour toutes les possibilités offertes par la méthodologie développée. Aussi, il est prospecté au chapitre 6 la possibilité d'étendre l'utilisation du modèle IDF proposé à des applications régionales à travers des procédures de cartographie permettant l'incorporation des incertitudes d'estimation et d'interpolation.

Cependant, dans la modélisation des extrêmes pluviométriques, le problème réel est comment l'intensité de pluie augmente quand la probabilité au dépassement diminue. C'est-à-dire que quand la probabilité tend vers zéro l'intensité de la pluie tend vers l'infini. Il existe une limite inférieure, déterminée mathématiquement, à ce type relation et elle est représentée par une décroissance exponentielle (Gumbel) ou alternativement de type puissance (GEV). Et la seule question qui reste en suspens c'est la notion de l'infini (Koutsoyiannis & Langousis, 2011). Ou plutôt, quelle est la limite physique pour les intensités de pluie ?

Aussi, bien que le concept et les applications des périodes de retour soient bien établis dans la bibliographie : les estimations exploratoires (i.e. au-delà des périodes d'observations) des périodes de retour restent une tâche difficile. En effet, un évènement de période de retour  $T$  est par définition une estimation dynamique, basée sur des observations historiques, sujet à redéfinir dès qu'il y a de nouvelles observations. Le nombre limité d'observations, a des conséquences sur la fiabilité des estimations en termes de précision et d'incertitude particulièrement pour les extrapolations au-delà des périodes d'observations (Volpi, 2019).



## **Chapitre 3**

### **Description de la zone d'étude et données utilisées**



### 3.1 Introduction

Le présent chapitre a pour objet de présenter la zone d'étude retenue dans ce travail de thèse. Il y sera donné un aperçu général de la géographie et de l'hydrologie de la région étudiée, ainsi que son importance socio-économique. Mais plus particulièrement il sera présenté l'ensemble des données pluviométriques servant de base à la méthodologie développée dans le cadre de ce travail. Aussi, il est mis en avant les relations entre les caractéristiques statistiques des pluies maximales annuelles et quelques caractéristiques topographiques de la zone d'étude.

### 3.2 Situation géographique de la zone d'étude

La zone d'étude couvre les régions centrales du Nord de l'Algérie (Figure 3.1). Elle s'étend sur une superficie de 25000 km<sup>2</sup>. Elle est comprise entre la mer Méditerranée au Nord et les limites Sud de la grande chaîne montagneuse de l'Atlas Tellien. En effet, la zone d'étude est une sorte de succession de vallées comprises entre deux chaînes de montagnes quasi parallèles suivant une direction Ouest-Est. Au Nord, s'érigent plusieurs massifs montagneux importants qui sont d'Ouest en Est : la Dahra dont le plus haut sommet est représenté par le Mont Zeccar à 1550 m, l'Atlas Blidéen et la Djurdjura comptant tous deux de nombreux pics, les plus élevés étant respectivement le Mont Sidi Abdelkader à 1629 m et le Mont Lalla Khadidja à 2308 m. Dans le Sud, les massifs montagneux les plus importants sont l'Ouarsenis, qui culminent à 1985 m au Mont Sidi Amar, et le massif du Titteri à l'Est de ce dernier et dont le plus haut sommet se trouve à Dirah à 1810 m.

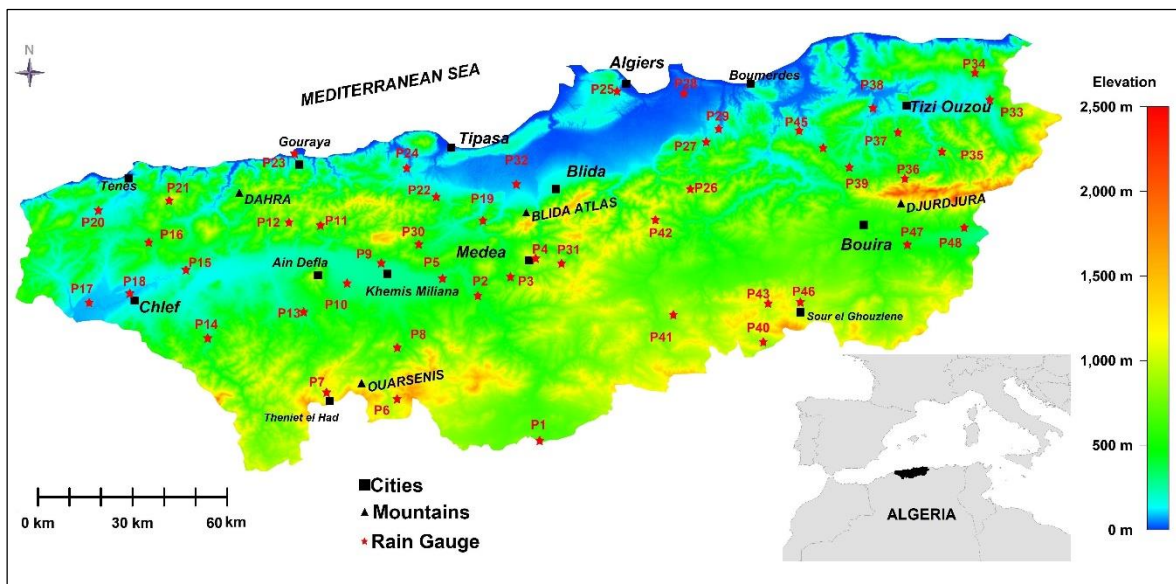


Figure 3.1 Zone d'étude, topographie et postes pluviographiques (PG)

Au centre, la chaîne de montagne septentrionale s'écarte des côtes pour laisser place à la Mitidja, une plaine littorale séparée de mer par la chaîne collinaire du Sahel (Mont Chenoua à 905m d'altitude) à l'Ouest et ouverte sur la mer à l'Est, avec une altitude moyenne

de 50 m. Cette dernière, avec la plaine du Cheliff à l'Ouest (entre Dahra et Ouarsenis), représentent deux des plus importantes plaines agricoles de l'Algérie. Aussi, à l'Est de la zone d'étude, se trouve les plaines de Beni Slimane et du Sebaou.

### 3.3 Importance démographique et socio-économique

La zone d'étude compte plusieurs agglomérations densément urbanisées. Administrativement parlant elle couvre dix Wilayas qui sont de l'Ouest vers l'Est : Chlef, Tipasa, Ain Defla, Nord de Tissemsilt, Alger, Blida, Médéa, Boumerdes, Tizi Ouzou et l'Ouest de Bouira. Cette zone concentre actuellement (projection démographique de 2018 sur la base du RGPH 2008) plus de 10 millions d'habitants, soit le quart de la population Algérienne. Aussi, elle compte une activité économique importante comparativement à d'autres régions du pays. Cette activité est relativement diversifiée : industrie légère, agriculture, agro-alimentaire, pharmaceutique, tourisme, etc.

### 3.4 Synthèse hydrologique

Du point de vue hydrologique, la zone d'étude englobe, au Nord, le grand bassin hydrologique des côtiers Algérois (code ANRH 02), entrecoupé à son extrémité orientale par le grand bassin d'Isser (code ANRH 09). Au Sud une grande partie du grand bassin du Cheliff (code ANRH 01 : les sous-régions haut et du moyen Cheliff), et au Sud-Est elle comprend toute la partie occidentale du grand bassin de la Soummam (code ANRH 15) (figure 3.2).



Figure 3.2 Extrait de la carte du réseau d'observations hydrologiques de l'ANRH

Les bassins versants (Mazafran, El Harrach, Sebaou, etc.) que comptent les côtiers Algérois et le Isser sont orientés selon plusieurs directions, toutefois, les écoulements se font globalement du sud vers le Nord, et se jettent en Mer Méditerranée. De même que l'Oued Cheliff principal cours du bassin du même nom et plus long cours d'eau d'Algérie (745 km).

Cependant, ce dernier prend naissance dans l'Atlas Saharien (au Mont Ammour) et s'écoule vers le Nord-Est, et devient parallèle à la Mer Méditerranée, après avoir contourné la chaîne Montagneuse du Ouarsenis (dans l'Atlas Télien), selon un axe d'écoulement d'Est vers l'Ouest, puis se jette en Mer au-delà des limites de la zone d'étude à la frontière des deux grands bassins hydrologiques des côtières Algérois et du Cheliff. Les apports sont relativement importants, et de nombreux cours d'eau ont des écoulements pérennes, à titre d'exemple les bassins hydrologiques des côtières Algérois et de l'Isser ont un apport moyen annuel de 3500 Hm<sup>3</sup> (ABH-AHS, 2002). Ainsi, de nombreux ouvrages hydrauliques de captages des eaux ont été implantés dans la zone d'étude, à savoir : 21 grands barrages (soit le quart des barrages Algériens), plus de 200 retenues collinaires et plus de 1500 forages d'eau.

### 3.5 Données utilisées : enregistrements pluviographiques

Les données pluviométriques utilisées ici ont été fournies par l'ANRH (Agence Nationale des Ressources Hydrauliques). Elles consistent dans les enregistrements de 50 postes pluviographiques (voir tableau 3.1 et figure 3.1) répartis sur toute la zone d'étude à partir desquels sont tirées les séries des maximums annuels (SMA) des pluies sur huit durées d'agrégation (15, 30, 60, 120, 180, 360, 720 et 1440 minutes). Le nombre moyen d'années de fonctionnement est de 22 ans avec un maximum de 38 années et un minimum de 8 années. La liste des codes ANRH des 50 stations pluviographiques est donnée en annexe 2.1

Tableau 3.1 Coordonnées des stations pluviographiques du Centre Nord de l'Algérie

NB : Coordonnées géographique selon la projection UTM WGS84 (zone S31)

Id.	Station	X(km)	Y(km)	Altitude (m)	Longueur (années)	Id.	Station	X(km)	Y(km)	Altitude (m)	Longueur (années)
1	Boughzoul	478	3964	643	24	26	Deux Bassins	522	4042	890	24
2	Ghrib	453	4008	460	15	27	Hamiz	526	4057	590	12
3	Barrage Tamezguida	466	4016	490	26	28	Rouiba	519	4070	20	25
4	Medea	473	4021	935	24	29	Keddara	532	4063	150	13
5	Djnane Ben Oudhah	446	4014	320	8	30	Oued Zeboudj	439	4020	400	38
6	Bordj Amir Abdelkader	428	3972	1080	29	31	Ouzra	487	4019	930	32
7	Theniet EL Had	410	3977	1160	28	32	Mouzaia	472	4038	140	14
8	Deurdeur	434	3987	575	14	33	Azazga	617	4071	500	29
9	Khemis Miliana	421	4018	300	25	34	Aghrib	612	4079	680	11
10	Harraza	411	4011	312	16	35	Beni Yenni	602	4054	850	29
11	Arib Ebda	410	4026	280	28	36	Ait Djema	594	4044	940	10
12	El Anneb	393	4030	1075	29	37	Bouassem	588	4061	600	26
13	El Touaibia	400	4002	335	30	38	Argoub	580	4067	100	16
14	Oued Fodda	373	3992	500	25	39	Ain Zaouia	578	4049	260	12
15	Oum Drou	364	4010	153	32	40	Souk el Djemaa	544	3992	829	24
16	Benaria	346	4023	390	30	41	Souagui	518	3998	850	24
17	Domaine Si Tayeb	330	4005	107	29	42	Mahterre	514	4029	875	22
18	Chlef	348	4006	130	28	43	Dechmia	547	4005	1144	20
19	Berkhissa	453	4031	400	15	44	TiziGhenif	564	4058	415	28
20	Abou el Hassene	332	4027	165	28	45	ChabetELAmeur	559	4057	260	24
21	Bissa	354	4037	1100	8	46	Sour El Ghozlene	558	4004	882	23
22	Meurad	440	4037	300	17	47	Bouira	579	4028	510	14
23	Gouraya	398	4048	34	17	48	M'Chedalla	609	4031	450	29
24	Boukerdane	431	4044	110	20	49	Soumaa	490	4041	177	12
25	Birmandreis	497	4072	140	16	50	Ouled Fayet	495	4066	162	19

En Algérie, depuis 2007 les bandes pluviographiques ne sont plus numérisées à cause du manque de la ressources humaines et de l'obsolescence du matériel utilisé lors des premières campagnes de numérisation. Toutefois, cette situation critique peut être rapidement dépassées, surtout avec la disponibilité de logiciels open-source ne nécessitant qu'un ordinateur de bureau et un scanner standard. En effet, certains auteurs (Burboa et al., 2020) ont développés des logiciels libres qui permettent de numériser puis dépouiller les enregistrements pluviographiques afin de fournir les lames de pluies sur divers pas de temps.

### 3.5.1 Sélection des séries d'observations

Les observations pluviographiques retenues dans ce travail sont issues de 50 postes pluviographiques (PG) répartis dans la région centre Nord de l'Algérie, ce qui correspond à l'ensemble des pluviographes encore fonctionnels ou ayant fonctionné dans toutes la zone d'étude. Les plus anciens postes ont commencé à opérer à la fin des années soixante, à savoir Ouzra (021112), Oued Zeboudj (021014), Oum Drou (012203), El Touaibia (011901). Les observations numérisées ne sont disponibles que de 1968 à 2007, néanmoins un certain nombre de stations comportent des lacunes (voir figure 3.3). La station ayant la plus grande durée de fonctionnement sans aucune interruption est celle d'Oued Zeboudj (021014) avec 38 années de mesures, et inversement celles comptant les plus petites périodes de fonctionnement sont Bissa (020224) et Djenane Ben Oudhah (011514) avec 8 années. Plus des deux tiers des postes pluviographiques se trouvent en milieu rural : 52% (26 PG) en zone d'habitat rural et 20% (10 PG) au niveau des barrages gérés par l'ANBT, et seulement 28% (14 PG) en zone d'habitat urbain dense.

Les critères de sélection des stations PG à utiliser dans ce travail, sont similaires à ceux proposés par Papalexiou & Koutsoyiannis (2013). Ils consistent à retenir les séries d'observations : (a) de plus de 50 années de longueurs, (b) avec moins de 20% de lacunes, (c) ayant un pourcentage de mesure douteuse inférieure à 0.1%. Toutefois, le critère (a) n'est pas considéré étant donné qu'aucune des stations analysées ne dépassent les 50 années de fonctionnement, de plus la méthodologie qui sera mise en œuvre ici tend à la meilleure utilisation possible de la moindre données disponibles (cf. chapitre 5). De même, le seuil minimal de lacunes est relevé à 30% étant donné que l'absence de mesures, pour les stations avec plus de 20% de lacunes (voir figure 3.3), correspond bien souvent à une période sèche. Pour ce qui est du critère (c) chaque mesure fournie est assignée d'une bonne qualité selon l'ANRH. Cependant, afin de s'assurer de ce dernier aspect, une vérification supplémentaire de la concordance des observations des séries maximales annuelles d'une station donnée avec les stations voisines, montre que la station de Ouled Fayet (020502) donne des mesures absurdes dans le sens où plus de 50% des cumuls maximaux journaliers dépassent les 300 mm (et atteignant même 800 mm à cinq reprises) sur 19 années d'enregistrement, ce qui est totalement irréaliste pour cette région du monde. Ainsi, une seule station (Ouled Fayet 020502) est rejetée. Il est retenu alors 49 stations répondant toutes aux critères susmentionnés.

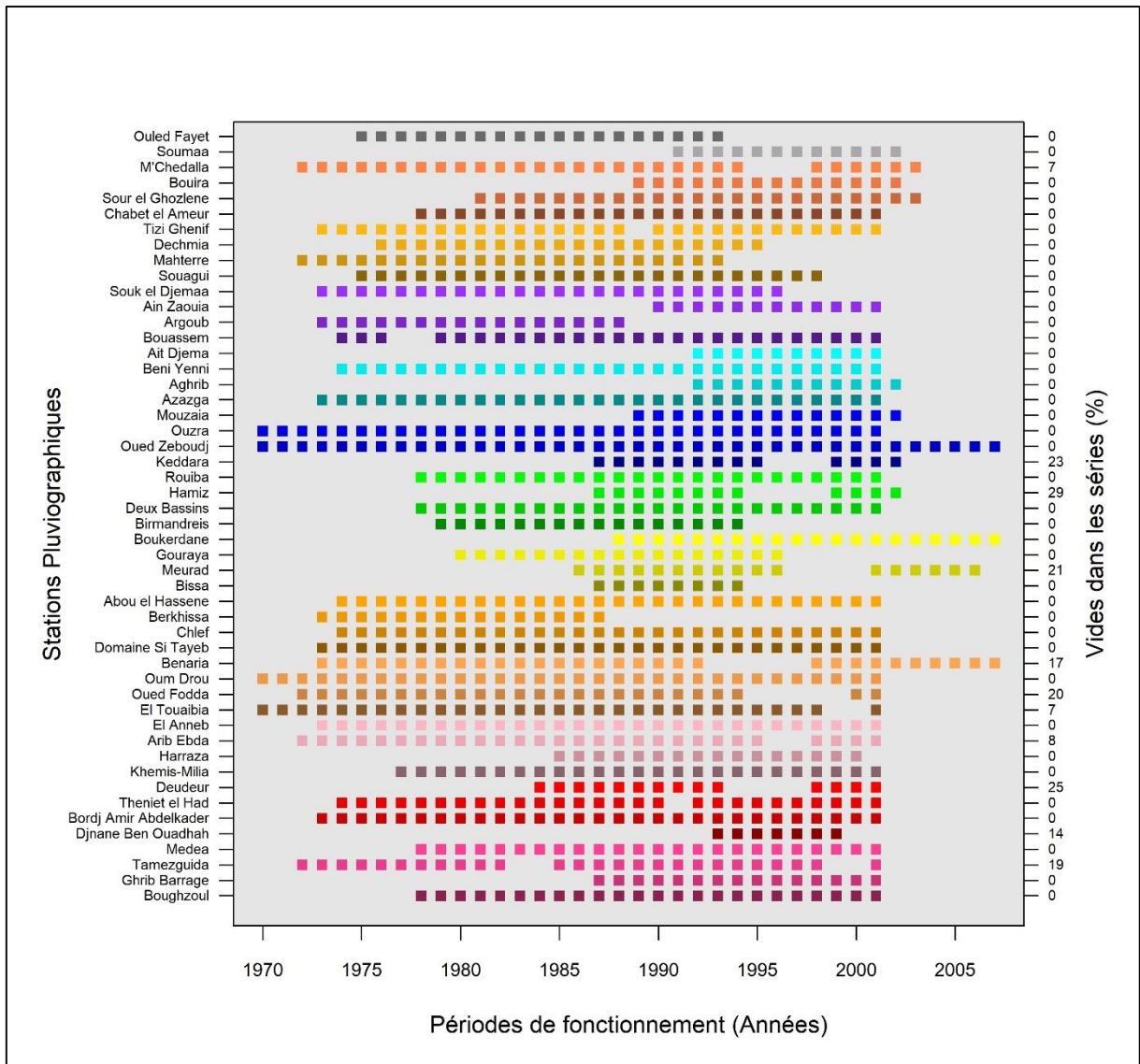


Figure 3.3 Période de fonctionnement des stations pluviographiques de la région centre Nord de l'Algérie.

### 3.5.2 Extraction et validation des séries des maxima annuels de pluies

La méthode d'échantillonnage retenue dans ce travail étant celle des maxima annuels, l'extraction des données pour constituer les séries des maxima annuels (SMA), puis leur validation, à partir d'observation des séries de durées complètes sans lacunes reste triviale. Toutefois, pour des SDC avec des années d'observations comprenant des vides, se pose la problématique de savoir si le cumul maximal annuel extrait est le véritable maximum. Dans notre cas d'étude, même si les lacunes coïncident le plus souvent avec la période dite sèche, il reste utile, voire nécessaire de s'assurer de la précision de la démarche qui a été entreprise. Ainsi, pareillement que pour la sélection des stations PG, c'est la procédure proposée par Papalexiou & Koutsoyiannis (2013) qui est appliquée pour valider l'extraction des cumuls maximaux annuels. Elle repose sur le principe que plus le pourcentage de lacunes, sur une année, est important plus la probabilité que le maximum extrait ne soit pas le véritable maximum, et que si ce dernier fait partie des 40% des valeurs les plus faibles dans la SMA constituée, la probabilité du faux maximum n'en est que plus grande. Donc, cette procédure

consiste à rejeter une année incomplète avec plus de 4 mois de lacunes (tiers de l'année) ou si son cumul maximal fait partie des 40% des valeurs les plus faibles dans la SMA ainsi constituée. Cette procédure a été jugée satisfaisante et a permis de valider les 49 stations sélectionnées.

### **3.6 Climatologie des régions centrales du nord de l'Algérie**

Le climat du Nord de l'Algérie a souvent, et pendant très longtemps, été présenté de type Méditerranéen. Cependant, la récente étude de Zeroual et al. (2019) sur le climat passé, présent et futur de l'Algérie montre plutôt que le climat du centre Nord de l'Algérie est de type cfa selon la classification de Köppen-Geiger. La classe cfa correspond à un climat tempéré chaud, un régime pluviométrique plutôt humide et un été chaud. En effet, la pluie moyenne interannuelle varie entre 300 et 1100 mm, même si les précipitations sont moins fréquentes en saison estivale, elles ne sont pas non plus rares.

#### **3.6.1 Précipitations moyennes et maximales annuelles**

La pluie moyenne interannuelle de la région d'étude est de 600 mm. Elle est calculée sur la base des totaux journaliers de pluie mesurés par pluviomètre au niveau des stations qui en disposent. L'évolution de la pluie annuelle de chaque station est représentée par un diagramme en boîtes à moustaches sur la figure 3.4. Cependant, cette valeur de 600 mm ne reflète pas la réalité du régime pluviométrique de la région. En effet, 19 stations parmi celles analysées ne présentent quasiment aucune mesure de pluie annuelle supérieure à 600 mm, avec quelques enregistrements de pluies inférieurs à 200 mm par an. Ces stations sont plutôt situées au Sud de la zone étudiée. A contrario, un petit groupe de 6 stations ont des moyennes interannuelles des pluies supérieures à 800 mm, où il est noté un dépassement des 1400 mm de pluies annuelles assez fréquemment avec comme valeur record une pluie annuelle de 1833 mm au cours de l'année hydrologique 1938/1939 à la station de Aghrib à 680 d'altitude. Ces six stations ont comme points communs une proximité à la Mer Méditerranée (<35 km) et une altitude de plus de 500 m. Le reste des stations ont une pluviométrie moyenne interannuelle variant entre 600 et 800 mm, avec toutefois quelques valeurs records sur les totaux annuels entre 1200 et 1400 mm aux stations de Hamiz (020602), deux bassins (020601) et Tizi Ghenif (050906). Ce dernier groupe de stations présente aussi une proximité avec la mer Méditerranée à l'exception de station de Theniet el Had (011605), dans le massif montagneux de l'Ouarsenis à 1160 m d'altitude, qui est localisée à la bordure Sud de la zone d'étude.

Cette première étape d'analyse laisse à penser que les stations au Nord de la zone d'étude sont fortement influencées par la Méditerranée, et présentent un régime pluviométrique plus important, tandis que celles situées au Sud subissent plus fortement l'influence du climat désertique du Sahara Algérien. A signaler aussi la probable influence de l'altitude. Ainsi, ce contraste dans le régime pluviométrique est très vraisemblablement relié à un certain nombre de caractéristiques géographiques (relief, localisation, exposition, etc.), les effets de ces

différents aspects sont plus amplement développés dans la section suivante. Néanmoins, le constat le plus important, statistiquement parlant, issu des remarques précédentes est qu'une variabilité dans les mesures de pluies aux niveaux local et régional induit forcément une grande variabilité d'échantillonnage dans les séries d'observations. Ceci, nécessitera une attention particulière dans l'interprétation des résultats de la mise en œuvre du développement méthodologique établi dans ce travail de thèse.

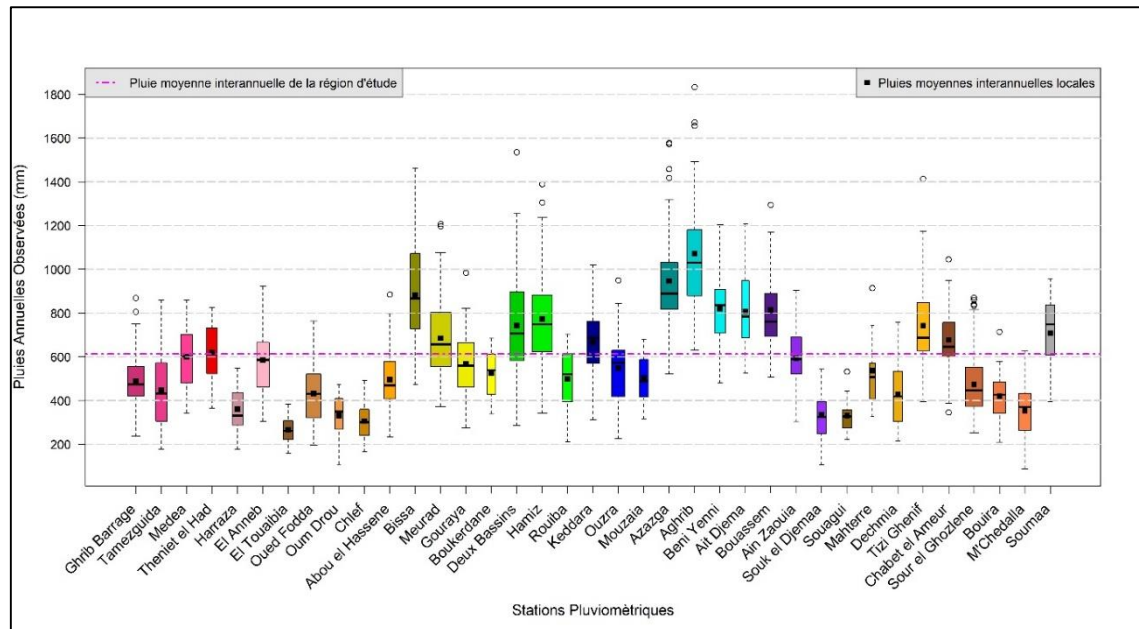


Figure 3.4 Totaux pluviométriques annuels du centre Nord de l'Algérie

Concernant les séries des maxima annuels, le cumul de pluie maximal moyen pour une durée de 1440 minutes (1 jour) est de 50 mm, toutefois des pluies dépassant les 100 mm par jour sont fréquentes et peuvent survenir en tout lieu de la zone d'étude, et des cumuls journaliers de plus de 150 mm sont aussi à signaler dans une dizaine de stations particulièrement dans la partie Nord-Est (figure 3.5c). Les observations records sont de l'ordre de 250 mm par jour dans trois stations différentes : Benaria (012205) dans la partie Nord-Ouest et Ain Zaouia (021924) et Tizi Ghenif (090506) dans la partie Nord-Est. Toutefois, certains pluviographes enregistrent des cumuls records pour les pluies de moins de 1 heure (voir figure 3.5a), à savoir plus de 50 mm en 15 minutes pour les stations Tamezguida, Bordj el Amir Abdelkader, El Anneb, Oum Drou, Benaria, Berkhissa, Oued Zeboudj, Ouzra, Deux Bassins, Hamiz, Rouiba, Azazga, Aghrib, Beni Yenni, Bouassem, Argoub, Ain Zaouia et Tizi Ghenif, et c'est très probablement des évènements pluvieux convectifs qui en sont à l'origine. En effet, la quasi-totalité de ces stations susmentionnées sont situées en plaine ou en montagne avec une exposition directe vers la plaine. A noter la singularité de la station des Deux Bassins qui a enregistré un cumul de 110 mm en 15 minutes le 10-11-2001 (soit le jour de l'inondation de Bab el Oued à Alger), et qui est clairement dominée par des pluies convectives, étant donné que quasiment tous les maxima annuels qui y ont été observés en 2 heures de temps correspondent au minimum à 50% des cumuls journaliers maximum. Aussi, toutes les stations du centre de la zone d'étude ont enregistré des cumuls horaires autour de 90 mm le 10-11-2001 (figure 3.5b). L'évolution des SMA pour les durées non présentées en figure 3.5 sont renseignées en annexe 2.2.

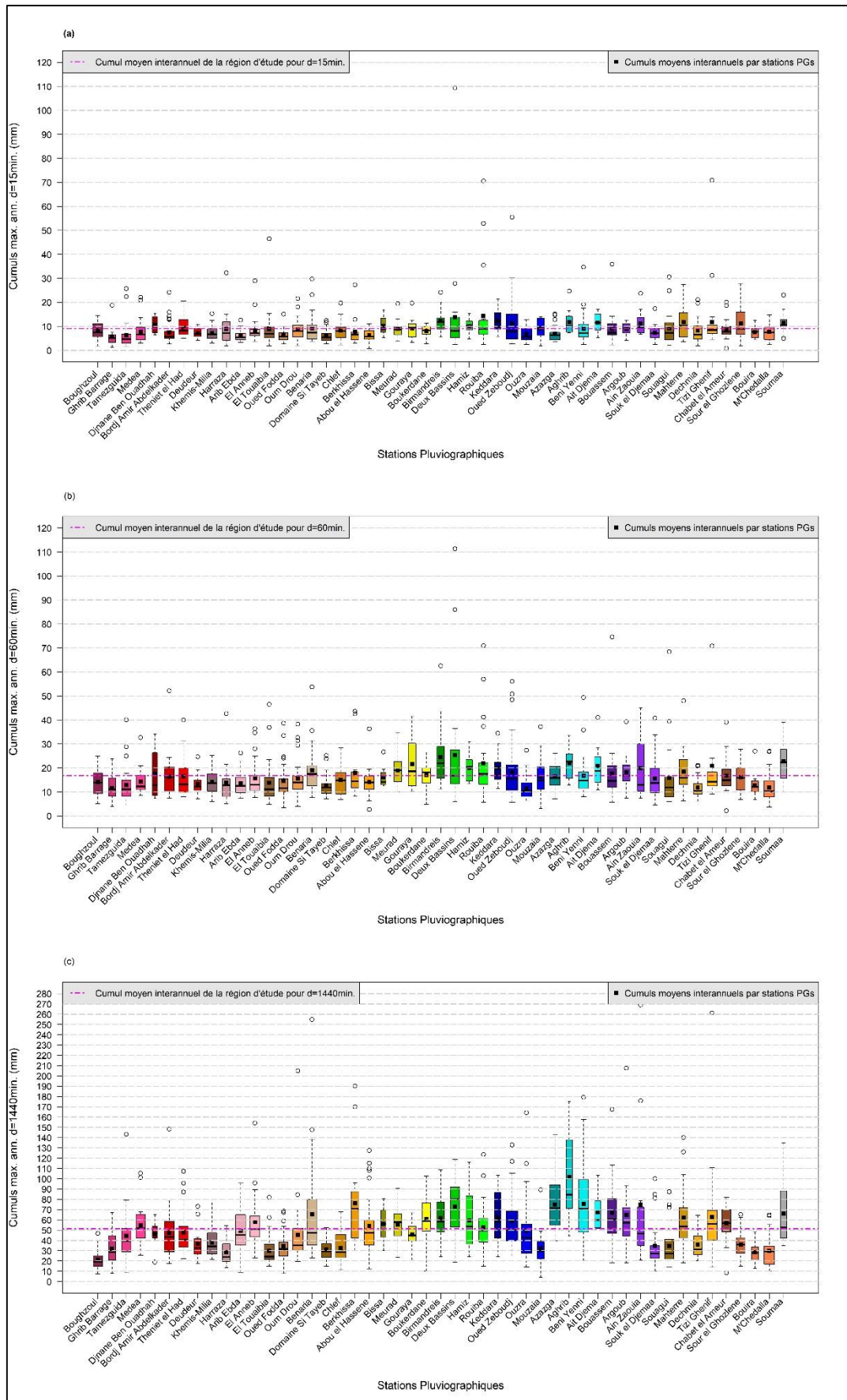


Figure 3.5 Totaux pluviométriques pour des durée de : (a) 15 minutes, (b) 60 minutes et 1440 minutes pour 49 stations PG dans le centre nord de l'Algérie



A titre de comparaison, les valeurs de cumuls de pluie précédemment citées sont proches des valeurs des pluies décennales (Arnaud et al., 2008), observées dans le sud de la France, qui est une des régions Européennes les plus exposées aux crues éclair (Gaume et al., 2009). Ce qui laisse à penser que notre zone d'étude encourt vraiment des événements pluvieux extrêmes. D'ailleurs, plusieurs lieux, et particulièrement les plus urbanisés, subissent régulièrement des inondations d'ampleurs variées d'où la nécessité d'une meilleure caractérisation de l'aléa pluviométrique à travers cette étude.

En conclusion, il apparaît que ce sont des événements convectifs qui donnent ces horsains sur les durées assez courtes, et de même il est très probable que les pluies extrêmes de durées courtes soient plutôt dominées par la convection. En effet, les cumuls de pluies maximales, sur quasiment toutes les stations, pour les durées de moins de 2 heures proviennent d'événements pluvieux différents comparativement aux cumuls de pluies sur les durées de 2 à 24 heures. D'ailleurs Di Baldassarre et al. (2006) font le même constat dans l'analyse des pluies maximales annuelles dans le Centre Est de l'Italie. Ainsi, pour les durées de 2 à 24 heures le régime pluviométrique est plutôt gouverné par des pluies orographiques à l'exception de la station des Deux Bassins (cf. chapitre 4).

### **3.6.2 Effet de la topographie sur les précipitations maximales annuelles**

Du fait d'un relief assez divers, l'incidence des précipitations est assez contrastée dans tout le Nord de l'Algérie (Meddi & Toumi, 2015). Toutefois, il est possible d'identifier une certaine tendance avec des pluies augmentant d'Ouest vers Est et diminuant du Nord au Sud (Taibi et al., 2017; Zeroual et al., 2017). En effet, la figure 3.6 semble confirmer ce dernier constat. De plus, les pluies moyennes interannuelles sont nettement plus importantes, comparativement aux autres stations, dans la partie Nord-Est de la zone d'étude, à l'exception de la station de Bissa (0202024) à 1100 m d'altitude dans le Nord-Ouest de la zone d'étude. Ceci laisse paraître qu'il y aurait une corrélation entre l'altitude et le régime pluviométrique, tel que considéré par Khelfi et al. (2017) dans la cartographie du coefficient de Montana dans la même région d'étude. Toutefois, d'autres stations (Souk el Djemaa, Souagui, Mahterre, Dechmia et Sour el Ghozlene) dans la partie Sud-Est de la zone d'étude ont des altitudes allant de 800 à 1150 mètres mais présentent des pluies moyennes interannuelles et de cumuls de pluies maximales annuelles parmi les plus faibles de la région d'étude (figures 3.6 et 3.7). Aussi, les stations ayant les latitudes les plus au Nord (c'est-à-dire les plus proches de la Mer) semblent avoir les pluies annuelles les plus importantes ainsi qu'un plus grand nombre d'averses moyennes par années (figure 3.6). Cette dernière constatation est encore plus affirmée s'agissant des cumuls annuels maximum (figure 3.7 : présentation sur les durées de 15, 60, 120 et 1440 minutes, le même graphique pour les 4 autres durées est présenté en annexe 2.3). Néanmoins, il est juste de signaler que les stations de Theniet el Had (011605) et Bordj el Emir Abdelkader (011603), bien qu'étant à la frontière Sud de la zone d'étude, présentent une pluviométrie plus importante que les autres stations situées à la même latitude. Ceci, est certainement dû à l'étendue de la chaîne montagneuse de l'Ouarsenis et aux altitudes importantes des deux stations (1160 et 1080 m) ce qui favorisent la formation d'un micro climat local.

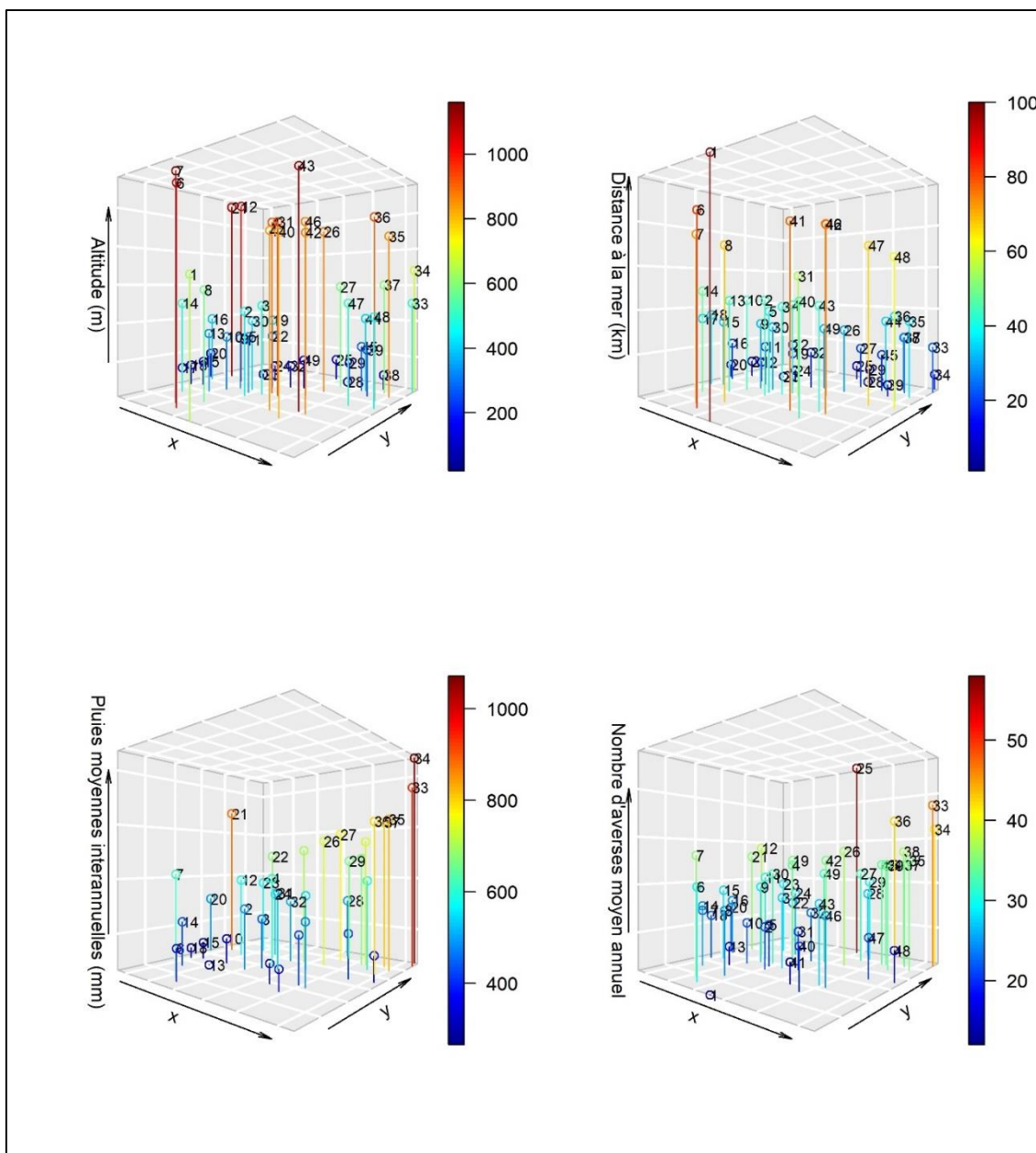


Figure 3.6 Diagramme de dispersion de quelques caractéristiques géographiques et statistiques des stations pluviographiques du Centre Nord de l'Algérie.

Les constatations précédentes ne représentent guère plus qu'une interprétation assez sommaire des effets de la topographie sur les précipitations extrêmes. En effet, les relations entre caractéristiques géographiques et statistiques des pluies maximales annuelles sont très complexes et un grand nombre de propriétés géographiques peuvent avoir une influence sur le régime pluviométrique. A titre d'exemple, Kieffer-Weisse & Bois (2001) ont considéré 12 caractéristiques topographiques dans leur analyses des effets du relief sur les caractéristiques statistiques des précipitations extrêmes dans les Alpes Françaises. Ainsi, ils ont abouti à des relations linéaires multivariées intégrant différentes caractéristiques topographiques en fonction de la durée d'agrégation des précipitations. Les paramètres les plus pertinents étant : la direction des pentes des montagnes, la distance à la mer, le relief

entourant la station de mesure et l'effet barrière des montagnes. L'altitude a été jugée comme variable non pertinente.

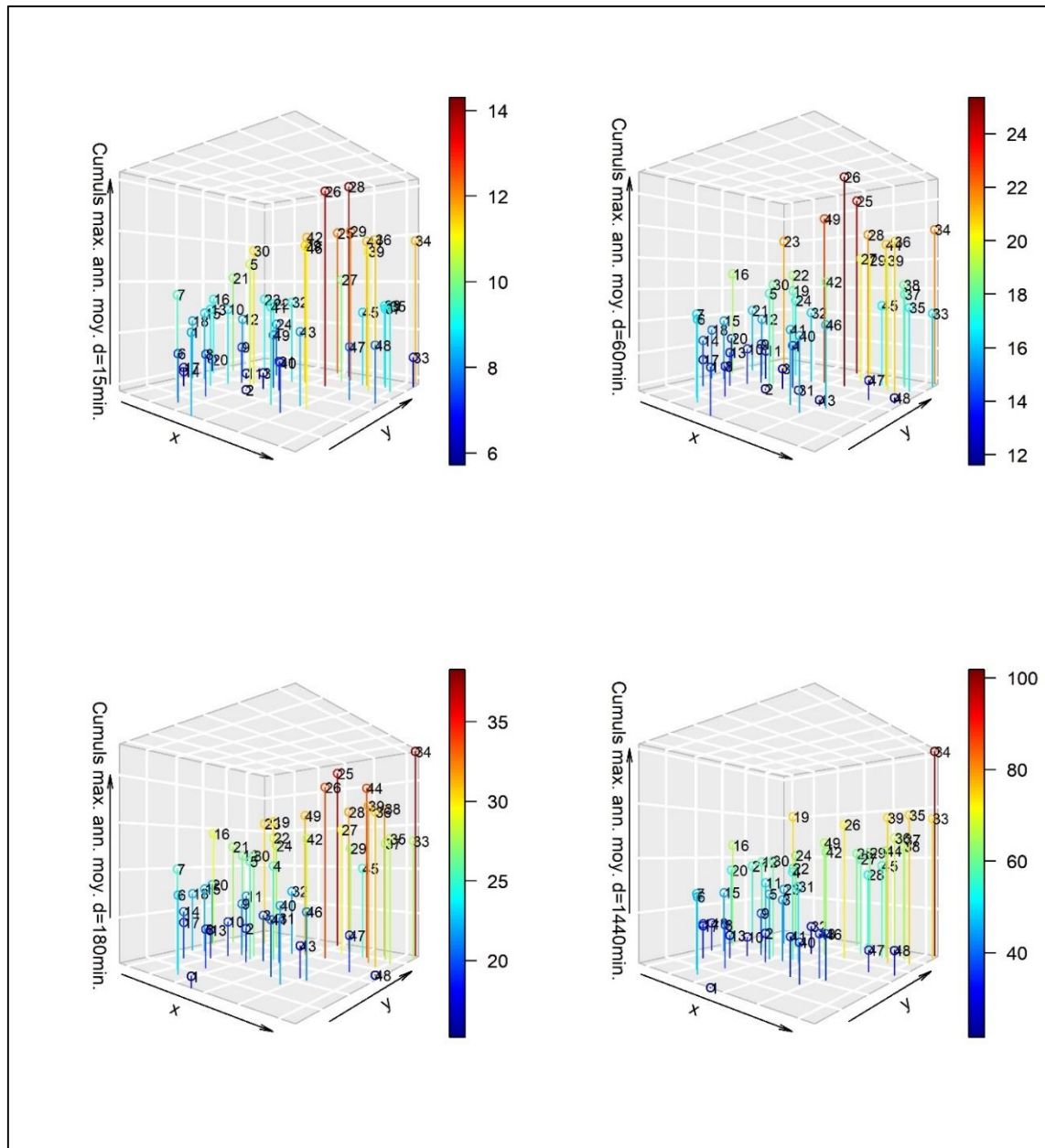


Figure 3.7 Diagramme de dispersion des cumuls maximum annuels moyen sur 4 durées d'agrégation (15, 60, 180 et 1440 minutes) pour 49 stations PG du centre Nord de l'Algérie.

En considérant une approche similaire à celle de Kieffer-Weisse & Bois (2001) et de Zahar & Laborde (2007), il a été prospecté un modèle linéaire multivariable permettant d'identifier une corrélation entre les cumuls de pluies maximales médians, les coordonnées géographiques, l'altitude et la distance à la Mer. Après avoir testé plusieurs combinaisons, des modèles parcimonieux et statistiquement significatifs ont été identifiés uniquement pour les durées de 180, 360, 720 et 1440 minutes. Ceci, dénote de la complexité des processus régissant les précipitations extrêmes. En effet, aucune des caractéristiques topographiques retenues ne montrent de corrélation avec les pluies de courtes durées (15, 30, 60 et 120 minutes), ce qui confirme que ces dernières sont issues d'un phénomène météorologique

différent comparativement aux pluies de longues durées (>3 heures). A signaler que la latitude est plus pertinente pour décrire l'influence de la Mer Méditerranée que la distance aux traits de côtes. Les relations obtenues sont renseignées dans le tableau 3.2.

Tableau 3.2 Résultats de la régression linéaire multivariée entre caractéristiques géographiques et statistiques des précipitations extrêmes du centre Nord de l'Algérie.

Durées (minutes)	Modèle linéaire multivariable	Coefficient de corrélation R
180	$\text{Cumul}_{\text{Median}_{\text{moy}}} = 0.189 \text{ coord}_Y - 0.0241 \text{ coord}_X + 0.00549 \text{ Altitude} - 729.8$	0.81
360	$\text{Cumul}_{\text{Median}_{\text{moy}}} = 0.293 \text{ coord}_Y - 0.0326 \text{ coord}_X + 0.0105 \text{ Altitude} - 1139$	0.82
720	$\text{Cumul}_{\text{Median}_{\text{moy}}} = 0.421 \text{ coord}_Y - 0.0446 \text{ coord}_X + 0.0177 \text{ Altitude} - 1644$	0.84
1440	$\text{Cumul}_{\text{Median}_{\text{moy}}} = 0.478 \text{ coord}_Y + 0.0165 \text{ Altitude} - 1888$	0.81

### 3.7 Quelques caractéristiques statistiques

Les moyennes et les valeurs maximales des séries des maxima annuels (SMA) de pluies sur les 8 durées d'agrégation pour les 49 stations pluviographiques de la zone d'étude sont reportées sur un graphique log-log en figure 3.8. Ce dernier graphique montre clairement

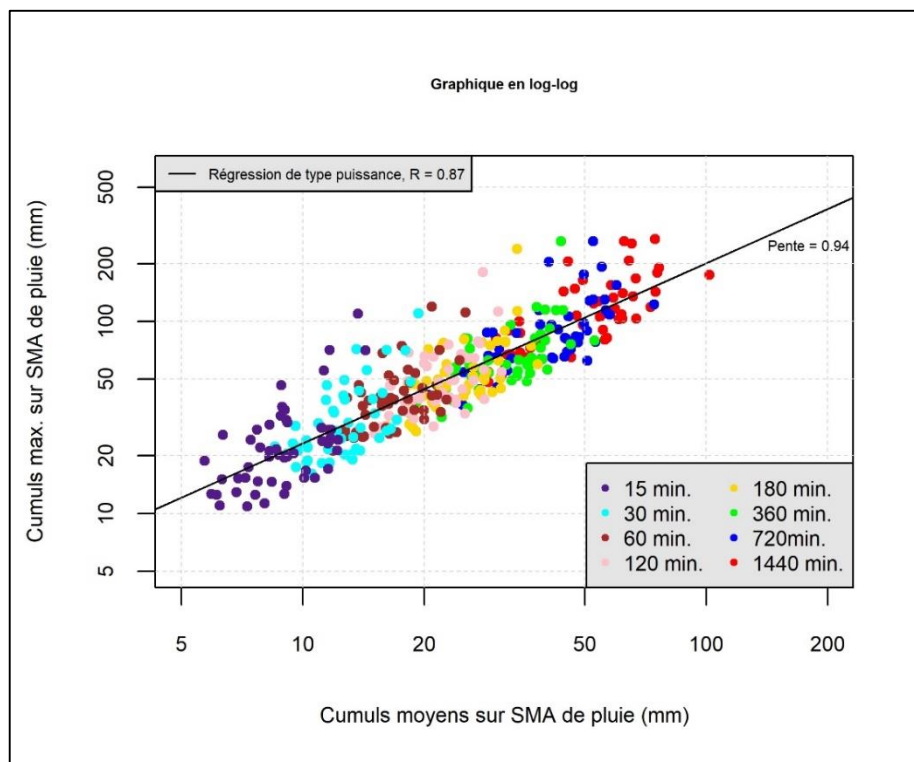


Figure 3.8 Graphique des cumuls de pluies maximum vs. moyens pour les 49 SMA sur 8 durées d'agrégation

l'existence d'une relation entre moyenne et maximum des cumuls de pluies des SMA indépendamment de la durée d'agrégation. En raison de la variabilité d'échantillonnage des SMA, les points projetés forment un nuage relativement épais autour de la relation ajustée. Néanmoins, il est assez visible que le groupe de points correspondant aux durées de 15, 30, 60 et 120 minutes sont légèrement plus dispersés (particulièrement pour 15 et 30 minutes). Encore une fois, ceci peut constituer une indication supplémentaire sur la nature différente des processus pluviométriques analysés sur différentes durées. Toutefois, le coefficient de corrélation  $R$  est de 0.87, ce qui est statistiquement assez significatif. Cette relation est approximée par un modèle de type puissance avec un exposant légèrement inférieur à 1 (0.94), et elle est donnée par l'expression suivante :

$$Y_{max} = 0.43 Y_{med}^{0.94} \quad (3.1)$$

En approximant l'exposant 0.94 à 1, l'équation (3.1) est de la même forme que l'équation (2.22) de l'indice de crue (ou d'averse) à la base de l'analyse fréquentielle régionale. Ou plus généralement elle correspond à l'expression de l'invariance d'échelle simple qui permet de mener des inférences statistiques sur la base d'un échantillon conjoint (*pooled simple*). Comme il sera montré en chapitre 4 et 5, l'invariance d'échelle simple est plausible et peut être beaucoup plus qu'un artifice mathématique, offrant la possibilité d'une intégration des observations dans un échantillon unique, mais une réalité en termes de structure probabilistes spatio-temporelle des précipitations extrêmes.

Dans un autre registre, la figure 3.9 renseigne sur la variance et l'asymétrie des SMA. Les valeurs de L-CV varient globalement de 0.15 à 0.35 avec une plus faible variation pour les stations les plus au sud de la région d'étude, ce qui indique une relative homogénéité des SMA, mais plus encore la tendance à avoir moins de valeurs exceptionnelles comparativement aux zones septentrionales. Autre fait marquant, la grandeur de la variance des échantillons SMA paraît être indépendante de la durée. Ceci, est possiblement l'indication que de la relative indépendance des SMA sur diverses durées, même si l'autocorrélation sérielle reste significative entre les paires de durées les plus proches dans le temps (cf. chapitre 5). Concernant l'asymétrie des SMA, ces derniers montrent une grande variabilité avec un L-CS variant de -0.1 à +0.5. Les stations localisées au Nord de la zone d'étude, et donc ayant un régime pluviométrique plus important, semblent être plus fortement asymétriques et indépendamment de la durée.

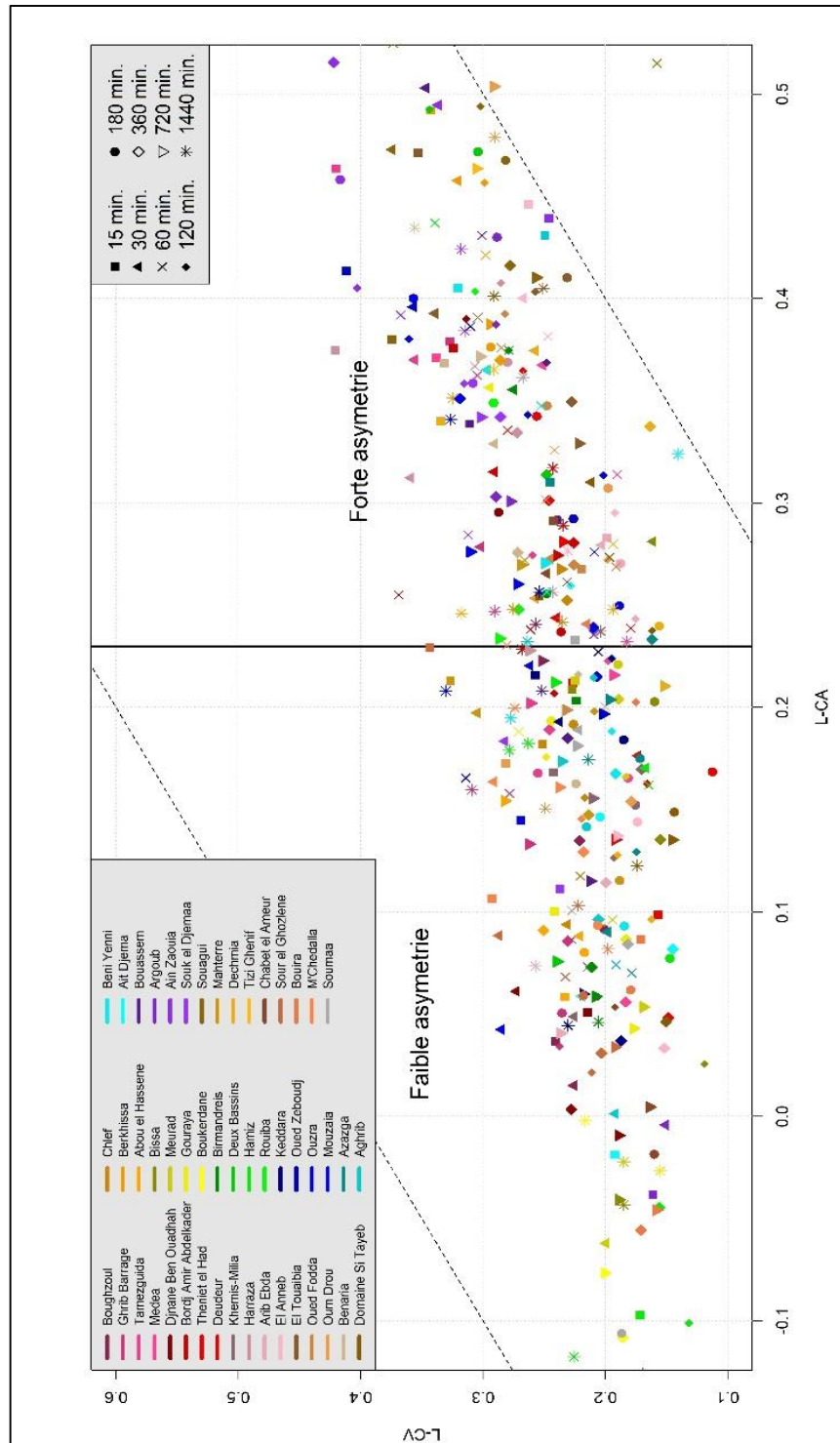


Figure 3.9 Graphique du L-coefficient d'asymétrie vs. L-coefficient de variation sur 8 durées d'agrégation pour 49 stations PG dans le centre Nord de l'Algérie.

La prédominance de l'asymétrie dans les échantillons SMA peut constituer un bon indicateur sur les distributions fréquentielles candidates à la modélisation statistique des précipitations extrêmes. Dans cette optique, la figure 3.10 représente le diagramme des L-moments pour les SMA de la zone d'étude. Un certain nombre d'échantillons se positionnent directement sur la GEV et LN3, la Loi de Gumbel paraît potentiellement aussi comme étant un bon choix, tandis que la loi de Pearson III est plutôt discriminée. Par contre la grande

variabilité d'échantillonnage, toutes durées confondues, ne permet pas d'identifier la distribution fréquentielle la plus appropriée pour l'ensemble des SMA. Il a été indiqué au chapitre précédent que c'est la loi GEV qui est retenue, toutefois sa pertinence sera systématiquement vérifiée à l'aide du test d'adéquation d'Anderson-Darling (cf. chapitre 5).

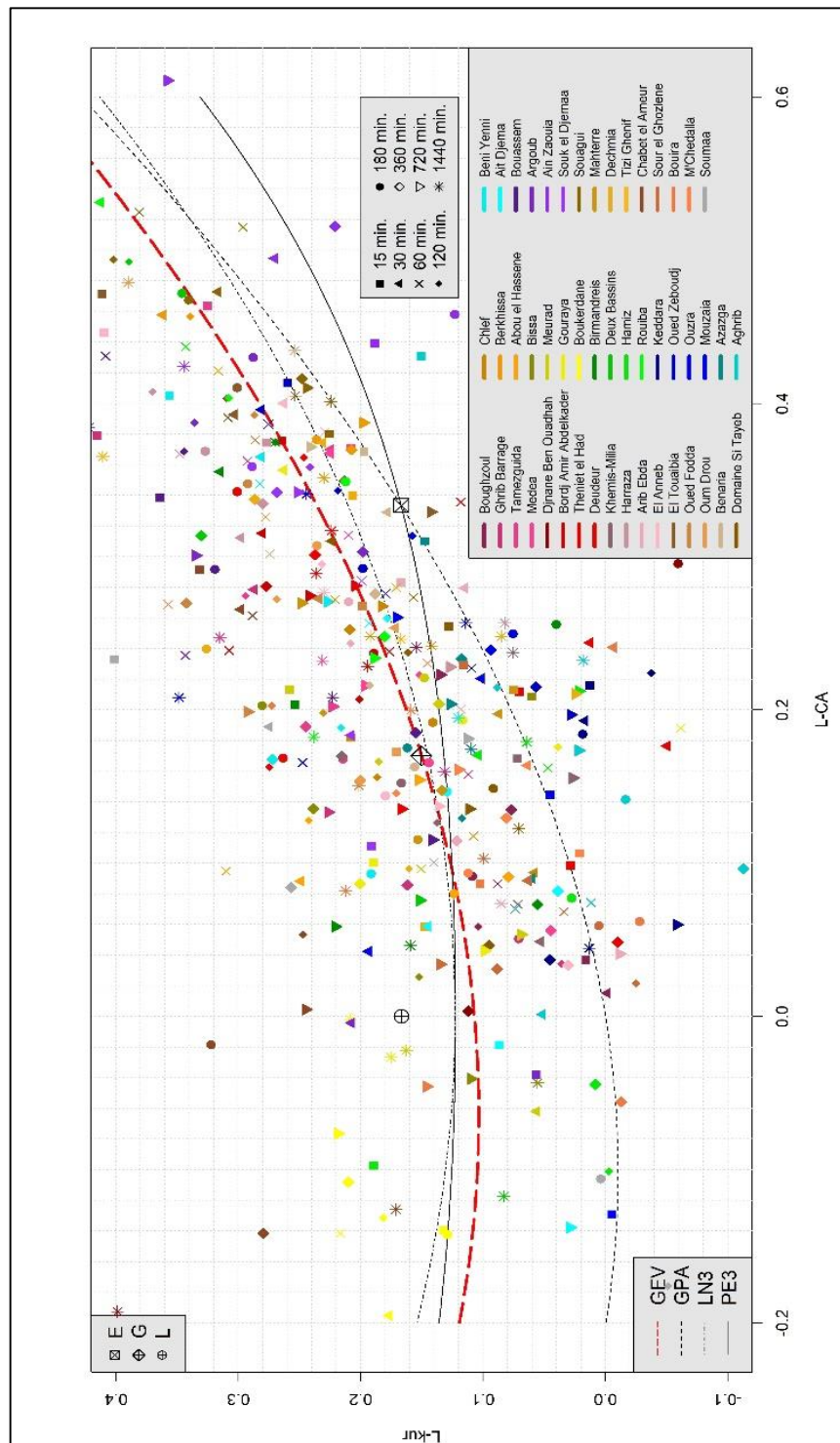


Figure 3.10 Diagramme des L-moments pour les SMA de 49 PG du centre Nord de l'Algérie.

### **3.8 Conclusion chapitre 3**

Certaines des séries des maxima annuels de pluies issues des mesures pluviographiques comportent des lacunes sur quelques années (entre 20 et 30%). Toutefois, elles ont quand même été incluses dans ce travail en considérant les critères de validation sur SMA des pluies proposés par Papalexiou & Koutsoyiannis (2013). Ainsi, il est mis à disposition dans cette étude 49 postes pluviographiques répartis inégalement sur toute la zone d'étude. Aussi, des analyses exploratoires, sur les SMA et les caractéristiques géographiques, ont permis de confirmer l'existence d'un gradient pluviométrique Nord-Sud et la vraisemblable prédominance de la latitude sur le régime pluviométrique de la région centre Nord de l'Algérie, ce qui sera d'une aide certaine dans la formation de régions homogènes. Enfin, il paraît évident que les données analysées restent, plus que toutes autres pratiques usuelles, le meilleur juge sur le choix de la fonction de distribution de probabilités à considérer.





## **Chapitre 4**

### **Etude des structures probabilistes des séries des maxima annuels**

## 4.1 Introduction

Dans ce qui suit, les hypothèses d'invariance d'échelle simple et de comportement fractal ; des observations des pluies maximales annuelles sur diverses durées, sont vérifiés localement (pour chaque station seule) et régionalement (deux variantes : 1. intégration spatiale de plusieurs sites pour chaque durée séparément, 2. Intégration spatio-temporelle : toutes durées confondues sur plusieurs sites). Cette vérification est menée à travers deux approches complémentaires que sont les tests d'homogénéité et de discordance de Hosking & Wallis (1997) et l'analyse d'échelle des moments MSA (Gupta & Waymire, 1990).

## 4.2 Vérification locale de l'invariance d'échelle simple et des structures fractales

Le modèle de courbe IDF proposé dans ce travail (chapitre 2, section 2.8.4, p.74) est relié à deux propriétés des séries de pluie, à savoir l'invariance d'échelle simple et les structures multifractales. Ainsi, avant le calage du modèle IDF intégré (équation 2.39), il est absolument nécessaire de vérifier si le jeu de données concorde, raisonnablement bien, avec les deux propriétés. Deux approches complémentaires sont utilisées :

1. Vérification de la log-log linéarité de la relation entre les moments d'ordre  $q$  et les durées d'agrégation  $d$  (Burlando & Rosso, 1996; Gupta & Waymire, 1990), de même que la linéarité de la fonction d'exposant d'échelle  $k(q)$  avec l'ordre  $q$  des moments, signe d'une éventuelle invariance d'échelle simple. Suivant la proposition de Innocenti et al. (2017). Cette analyse sera appelée dans ce qui suit analyse d'échelle des moments (MSA : *Moment Scaling Analysis*). S'il y a linéarité, la pente de la fonction d'exposant d'échelle peut être considérée comme une première estimation du paramètre  $b$  dans l'équation (2.39).
2. La mise en œuvre du test d'homogénéité proposé par Hosking & Wallis (1997) afin de vérifier si les distributions fréquentielles des cumuls de pluies sur diverses durées d'agrégation  $d$ , peuvent être considérées comme similaires, si elles sont correctement réduites. A noter que ce test est habituellement utilisé en analyse fréquentielle régionale pour vérifier l'homogénéité statistique entre différents sites (ou stations). Mais fondamentalement, c'est un test d'homogénéité, qui est utilisé dans ce qui suit pour comparer les distributions sur diverses durées d'agrégation pour chaque site.

### 4.2.1 Analyse d'échelles des moments

La MSA compte de nombreuses applications (Blanchet et al., 2016; Casas-Castillo et al., 2018; Innocenti et al., 2017; Panthou et al., 2014) pour vérifier si les données observées ne rejettent pas l'invariance d'échelle simple et si effectivement elles montrent un comportement fractal. Cependant, certains auteurs (García-Marín et al., 2013; García-Marín et al., 2019; Soltani et al., 2017; Yu et al., 2004) considèrent dans leur analyse d'échelle des moments, des

valeurs d'ordre  $q$  allant jusqu'à 6. Sachant que les moments ordinaires sont extrêmement sensibles à la variabilité d'échantillonnage, et donc peu fiables, à partir de l'ordre 3, ceci peut poser problème et amener à décrire des propriétés statistiques qui n'existent pas réellement.

Lombardo et al, (2014) montrent que l'utilisation de moments statistiques ordinaires d'ordre supérieur à 3, dans l'étude des structures multifractales en hydrologie, est contestable. En effet, les biais et les incertitudes d'estimation, sur les moments statistiques d'ordre élevé, sont conséquents et ne peuvent être ignorés dans l'analyse d'un quelconque processus hydrologique. Leur recommandation est de se limiter aux deux premiers moments, car seuls ces derniers peuvent être estimés d'une manière sûre (i.e. avec suffisamment de fiabilité). Toutefois, restreindre l'analyse des échelles des moments à l'ordre 2 ne permet pas de rendre compte d'une éventuelle non linéarité de la fonction d'exposant d'échelle  $k(q)$  (Lovejoy & Schertzer, 2013).

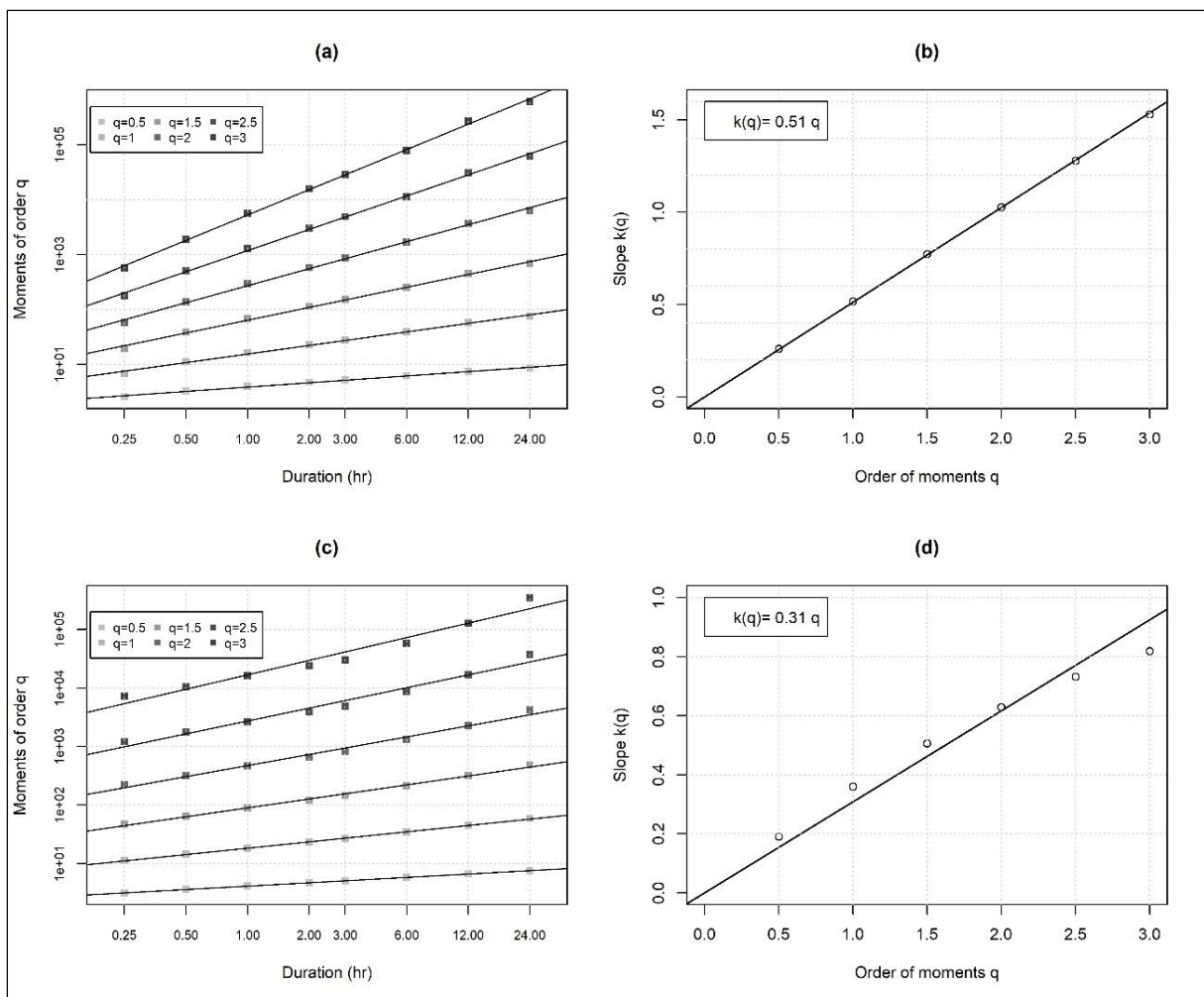


Figure 4.1 (à droite) : Régression log-log linéaire entre les moments d'ordre  $q$  et les durées d'agrégation  $d$ . (à gauche) Ajustement d'une régression linéaire à la fonction d'exposant d'échelle  $k(q)$  et les ordres  $q$  des moments. (a) et (b) pour la station 33. (c) et (d) pour la station 30.

Ainsi, dans ce travail, la MSA est implémentée jusqu'aux moments d'ordre 3. Toutefois, l'utilisation de seulement trois valeurs, reste insuffisante pour bien décrire des relations graphiques. La solution est d'utiliser des moments statistiques d'ordre non-entier (fractionnel et négatif). Cette pratique est très courante (Bairwa et al., 2016; Casas-Castillo et al., 2018;

Gupta & Waymire, 1990; M. I. P. de Lima & de Lima, 2009; Lombardo et al., 2014; Lovejoy & Schertzer, 2013; Sivakumar, 2001), et permet de bien rendre compte de la courbure éventuelle (i.e. la non-linéarité) de la fonction d'exposant d'échelle  $k(q)$ .

Signalons quand même, qu'il est possible de déterminer les valeurs critiques, minimale et maximale, des ordres  $q$  des moments au-delà desquelles l'estimation des moments statistiques est incertaine (Waymire & Ossiander, 2000).

La figure 4.1 illustre les résultats d'une MSA pour deux stations représentatives (Station 30 d'Oued Zeboudj et Station 33 de Azazga) (les résultats pour le restant des stations sont donnés en annexe 3). Les coefficients de détermination  $R^2$  entre  $\log E[Y_d^q]$  et  $\log d$ , prennent des valeurs supérieures à 0.9 pour quasiment toutes les stations étudiées et pour tous les moments d'ordre  $q$  considérés (figure 4.2).

Tableau 4.1 Valeurs du coefficient de détermination  $R^2$  entre  $k(q)$  et  $q$  pour les cas A (15 min- 24 h) et le cas B (1h – 24h)

id	Stations	$R^2$ (A)	$R^2$ (B)	id	station	$R^2$ (A)	$R^2$ (B)
1	Boughzoul	1.00	1.00	25	Birmandreis	1.00	0.99
2	Ghrib Barrage	1.00	1.00	26	Deux Bassins	<b>0.70</b>	<b>0.87</b>
3	Tamezguida	1.00	1.00	27	Hamiz	1.00	1.00
4	Medea	1.00	1.00	28	Rouiba	0.96	0.99
5	Djnane Ben Oudhah	1.00	0.99	29	Keddara	1.00	1.00
6	Bordj Amir Abdelkader	1.00	1.00	30	Oued Zeboudj	0.99	1.00
7	Theniet EL Had	1.00	1.00	31	Ouzra	1.00	1.00
8	Deurdeur	1.00	1.00	32	Mouzaia	1.00	1.00
9	Khemis Miliana	1.00	1.00	33	Azazga	1.00	1.00
10	Harraza	0.98	0.98	34	Aghrib	1.00	1.00
11	Arib Ebda	1.00	1.00	35	Beni Yenni	1.00	1.00
12	El Anneb	1.00	1.00	36	Ait Djema	1.00	1.00
13	El Touaibia	0.96	0.99	37	Bouassem	0.99	0.99
14	Oued Fodda	1.00	1.00	38	Argoub	1.00	1.00
15	Oum Drou	1.00	0.99	39	Ain Zaouia	1.00	1.00
16	Benaria	1.00	1.00	40	Souk el Djemaa	1.00	1.00
17	Domaine Si Tayeb	1.00	1.00	41	Souagui	0.99	0.95
18	Chlef	1.00	1.00	42	Mahterre	1.00	1.00
19	Berkhissa	1.00	1.00	43	Dechmia	1.00	1.00
20	Abou el Hassene	1.00	1.00	44	TiziGhenif	1.00	0.97
21	Bissa	1.00	1.00	45	ChabetElAmeur	1.00	1.00
22	Meurad	1.00	1.00	46	Sour El Ghozlène	1.00	1.00
23	Gouraya	1.00	1.00	47	Bouira	1.00	1.00
24	Boukerdane	1.00	1.00	48	M'Chedalla	1.00	1.00

De même, les  $R^2$  entre  $k(q)$  et  $q$  prennent tous des valeurs supérieures à 0.94 (voir tableau 4.1), à l'exception de la station 26 des deux Bassins. Néanmoins, pour un très petit nombre de cas, telle la station 30 d'Oued Zeboudj (voir figure 4.1d), il apparaît une légère déviation du modèle linéaire (courbure) de la fonction d'exposant d'échelle  $k(q)$ . Il est assez difficile, en raison de la taille limitée des échantillons, d'affirmer si ceci révèle clairement une déviation du modèle théorique d'invariance d'échelle, ou bien si cela peut être attribué à la sensibilité des moments d'ordre supérieurs, et par conséquent la valeur associée de  $k(q)$ , à la variabilité d'échantillonnage. Toutefois, il apparaît que les événements pluvieux, donnant les plus fortes

quantités de précipitation, sont de type convectif pour quelques stations. Alors, la quantité totale de pluie précipitée est concentrée en une ou deux heures (maximum journalier quasiment égal au maximum de pluie en 2 heures). Une telle situation, réduit la pente de  $\log E[Y_d^q]$  avec  $\log d$  (c'est-à-dire la valeur de  $k(q)$ ), particulièrement pour les valeurs supérieures d'ordre des moments. La prédominance de tels évènements dans les séries des maxima annuels est particulièrement marquée à la station 26.

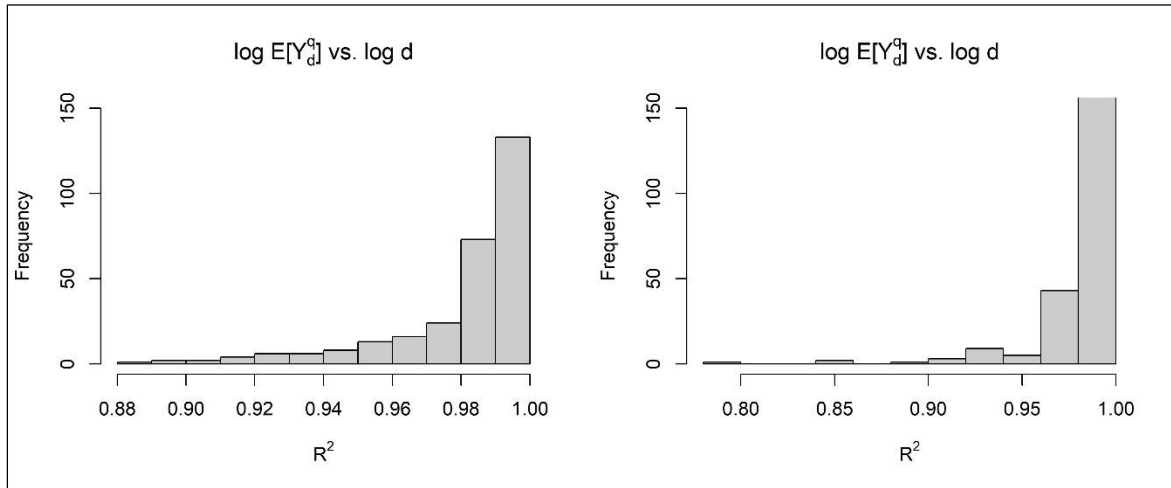


Figure 4.2 Valeurs des coefficients de détermination  $R^2$  entre  $\log E[Y_d^q]$  et  $\log d$ , pour 48 stations pluviographiques. Avec  $q=(0.5, 1, 1.5, 2, 2.5, 3)$ . (à gauche) : Cas A (15 min – 24h). (à droite) : Cas B (1h – 24h).

Comme première conclusion partielle, à travers la MSA, la multifractalité et l'invariance d'échelle simple semblent clairement cohérentes avec le jeu de données. Innocenti et al. (2017) arrivent à la même conclusion basée sur l'analyse d'un très grand jeu de données de 2700 stations pluviographiques en Amériques du Nord : l'hypothèse d'invariance d'échelle y semble cohérente avec les observations, particulièrement pour les durées supérieures à 1 heure.

Cependant, il est important de garder à l'esprit la possible limitation susmentionnée. Si, les queues des distributions fréquentielles sont contrôlées par des évènements pluvieux convectifs avec des propriétés d'échelle quelque peu spécifiques, les observations peuvent significativement s'éloigner de l'hypothèse d'invariance d'échelle simple. Ainsi, l'ajustement d'une distribution fréquentielle réduite à un jeu de données conjoint (pooled) de cumuls de pluies réduits, introduirait des biais. Ceci est vérifié dans l'analyse d'implémentation du modèle IDF proposé (intégré).

#### 4.2.2 Validation de la relation d'exposant d'échelle

Afin de confirmer que le calage de la relation de facteur d'échelle  $d^b$  est raisonnablement adapté aux données, les médianes empiriques de séries d'observations sur diverses durées sont comparées aux médianes calculées via le modèle IDF intégré. Les médianes ont été retenues, dans cette analyse, en lieu et place des moyennes ou quelconques autres valeurs de quantiles estimés, en raison de leur sensibilité limitée à la variabilité d'échantillonnage.

Les intervalles de crédibilités à 80%, sont obtenus par simulation Monte Carlo, sur les médianes déduites des distributions calées (Figure 4.3). Pour chaque station pluviographique et pour chaque durée, une série de cumuls de pluie de même longueur que l'échantillon observé, est tirée aléatoirement pour chaque jeu de paramètres échantillonnés à travers une procédure de calage Bayesian MCMC. Ainsi, les intervalles de crédibilités comptent à la fois pour : les incertitudes sur les valeurs calées des paramètres du modèle IDF et la variabilité d'échantillonnage. Notons que cette dernière analyse est très similaire à celle présentée en Figure 4.1 pour le moment d'ordre 1 :  $b=k(1)$ . Néanmoins, ici, les intervalles de crédibilité aident à évaluer l'écart entre valeurs observées et modélisées.

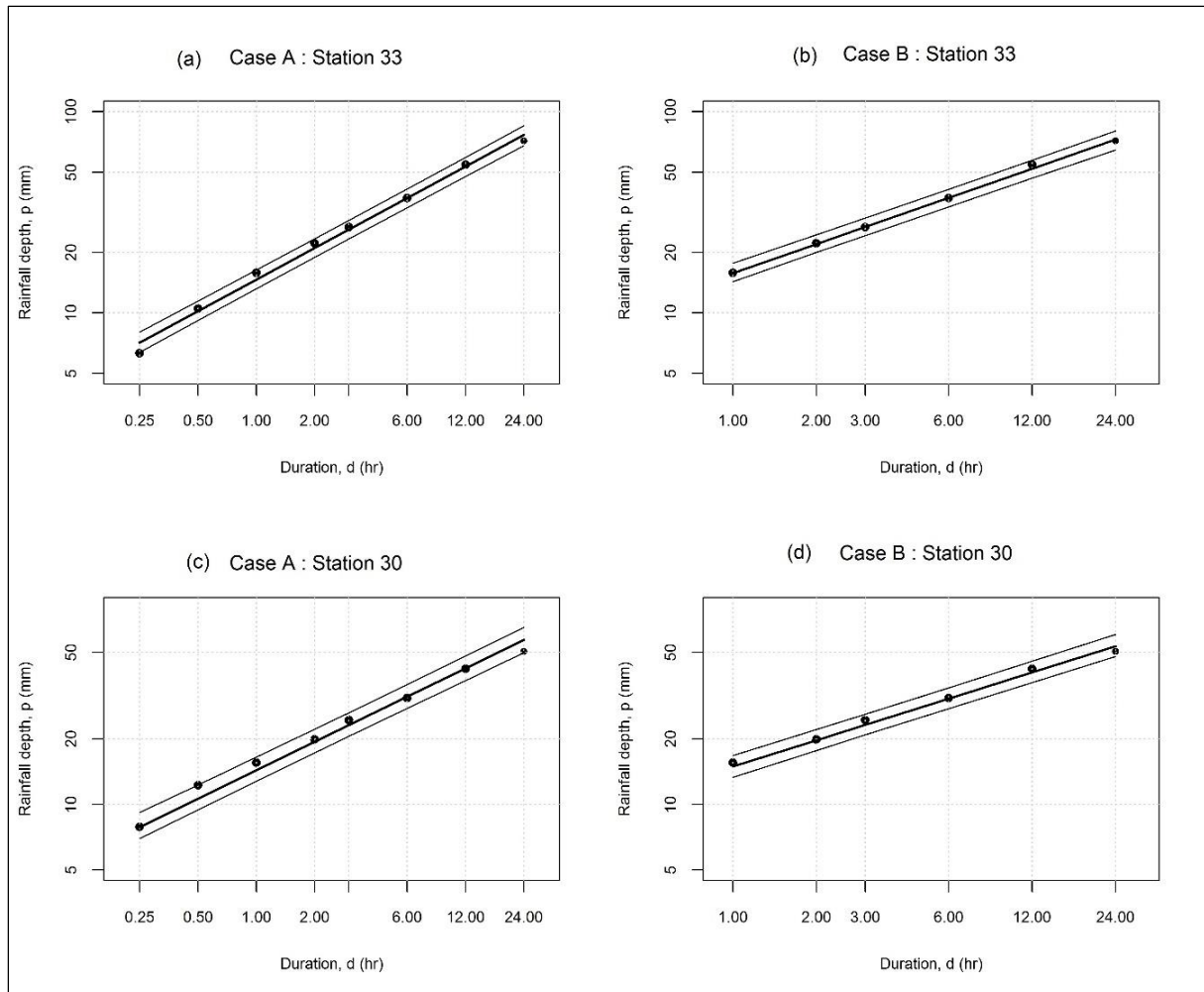


Figure 4.3 Evolution des médianes de cumuls de pluies avec les durées (sur une échelle log-log). Points noirs : Valeurs empiriques des médianes pour chacune des séries de durées  $d$ . Trait en gras : Médianes théoriques calculées par le modèle IDF Bayésien Intégré. Traits fins : Intervalles de crédibilités à 80% sur les médianes

Comme le montre bien la Figure 4.3, une forte proportion des valeurs de médianes sur les séries observées est à l'intérieur des intervalles de crédibilités à 80% pour les deux cas A (durées considérées de 15 minutes à 24 heures) et B (durées considérées de 1 heure à 24 heures). De plus les résultats de calage du facteur d'échelle pour deux cas A et B sont très similaires. Ce résultat est confirmé pour toutes les stations pluviographiques considérées (voir annexe 4). Ceci, en même temps que la valeur limitée de la largeur de l'intervalle de crédibilité à 80%

confirme l'adéquation de la relation d'échelle proposée, et écarte le besoin d'introduire une relation d'échelle plus complexe, comprenant un plus grand nombre de paramètres, tel que cela a été fait dans d'autres travaux (Koutsoyiannis et al., 1998; Muller et al., 2008).

#### 4.2.3 Vérification supplémentaire sur l'applicabilité de l'invariance d'échelle simple

Le test d'homogénéité proposé par Hosking & Wallis (1997) donne un autre point de vue sur la plausibilité de l'hypothèse d'invariance d'échelle (Tableau 4.2). Théoriquement, la valeur de la statistique  $H1$  devrait être une variable normale  $N(0,1)$ . Si la notion d'invariance d'échelle est supportée par données de pluie maximales annuelles, la statistique  $H1$  devrait prendre une valeur moyenne autour de 0 pour les 48 stations pluviographiques et dépasser dans 10% des cas, le quantile à 90% de la distribution  $N(0,1)$  : 1.28. Les valeurs de  $H$  dépassant ce seuil de 1.28 sont en gras dans le tableau 4.2. Dans le cas B où les durées d'agrégation considérées se situent entre 1 et 24 heures, ce seuil critique est dépassé pour 5 stations sur 48 (c'est-à-dire autour de 10% des échantillons). La gamme de durée de 1-24 heures est souvent la période sur laquelle l'invariance d'échelle simple est considérée comme valide (Blanchet et al., 2016; Burlando & Rosso, 1996; Panthou et al., 2014; Yu et al., 2004).

Tableau 4.2 Résultats des tests d'homogénéité de Hosking et Wallis, Statistique  $H1$  pour les deux cas A (15 min-24 h) et le cas B (1h – 24h)

id	Stations	H1 (A)	H1 (B)	id	station	H1 (A)	H1 (B)
1	Boughzoul	-1.06	-0.59	25	Birmandreis	-1.09	-0.48
2	Ghrib Barrage	-1.17	-0.96	26	Deux Bassins	<b>1.30</b>	0.81
3	Tamezguida	0.49	-1.29	27	Hamiz	<b>1.76</b>	<b>2.57</b>
4	Medea	<b>2.52</b>	-2.28	28	Rouiba	0.18	-1.21
5	Djnane Ben Oudhah	1.14	0.58	29	Keddara	-0.43	-1.18
6	Bordj Amir Abdelkader	-0.71	-1.60	30	Oued Zeboudj	<b>5.10</b>	<b>1.58</b>
7	Theniet EL Had	-1.97	-1.70	31	Ouzra	<b>3.93</b>	<b>1.71</b>
8	Deurdeur	-0.05	0.06	32	Mouzaia	-1.66	-1.78
9	Khemis Miliana	0.02	-0.76	33	Azazga	0.80	-0.43
10	Harraza	-0.31	-1.54	34	Aghrib	-0.94	-0.67
11	Arib Ebda	-0.12	0.48	35	Beni Yenni	0.60	0.14
12	El Anneb	0.63	0.75	36	Ait Djema	<b>2.03</b>	<b>2.44</b>
13	El Touaibia	-0.03	-0.93	37	Bouassem	-0.16	-0.80
14	Oued Fodda	-1.31	-1.21	38	Argoub	0.27	-1.19
15	Oum Drou	-0.08	0.02	39	Ain Zaouia	-0.51	-1.83
16	Benaria	0.51	0.42	40	Souk el Djemaa	-1.48	-2.02
17	Domaine Si Tayeb	<b>3.10</b>	0.14	41	Souagui	-0.75	-0.95
18	Chlef	-2.29	-2.33	42	Mahterre	0.82	-0.40
19	Berkhissa	-2.19	-2.13	43	Dechmia	<b>2.75</b>	-1.14
20	Abou el Hassene	0.27	0.75	44	TiziGhenif	NAN	NAN
21	Bissa	-1.09	-1.38	45	ChabetElAmeur	<b>1.43</b>	0.21
22	Meurad	-1.90	-2.30	46	Sour El Ghazlene	<b>2.95</b>	-1.29
23	Gouraya	<b>1.45</b>	<b>1.69</b>	47	Bouira	-1.80	-1.92
24	Boukerdane	-2.13	-1.87	48	M'Chedalla	0.88	0.16



Concernant le cas A (où les durées considérées vont de 15 minutes à 24 heures), la valeur moyenne de  $H1$  (0.2) reste petite mais 11 de ses valeurs calculées (presque le  $\frac{1}{4}$  de l'échantillon) dépassent 1.28. La validité de l'invariance d'échelle simple sur cette gamme de durées est assez douteuse ( $<1h$ ). Ce qui est aussi en accord avec plusieurs travaux précédents (Marani, 2003; Olsson & Burlando, 2002). Aussi, il est important de mentionner que le test  $H1$  ne tient pas compte des dépendances existantes entre les séries de cumuls de pluies annuels maximum sur différents pas de temps. Ainsi, les résultats de ce test devraient être considérés comme quelques peu optimistes. Même si le cas A est moins cohérent avec l'hypothèse d'invariance d'échelle du modèle IDF intégré, il sera pris en considération, dans un but comparatif, dans le reste de ce travail.

De plus, en analysant la structure probabiliste des séries de durées complètes des pluies, (Marani, 2003) identifie trois régimes : régime interne, régime de transition, régime d'invariance d'échelle. Il montre qu'une fonction d'exposant d'échelle de type puissance (invariance d'échelle simple) n'est pas valable sur toute la gamme de durées (de quelques minutes à 2 jours), et que près de l'origine le régime interne a une variance qui est une fonction quadratique de la durée. Il y a aussi un régime transitoire qui explique que les cumuls sur de petits intervalles de temps (entre quelques minutes et 1 heure) suit quelquefois une fonction d'échelle de type puissance. Ce qui dénote que l'utilisation du principe d'invariance d'échelle simple dans la modélisation statistique des extrêmes pluviométriques sur une gamme de durées 15 minutes à 24 heures n'est pas si irréaliste, bien au contraire.

### 4.3 Vérification de l'homogénéité régionale des séries des maxima annuels de pluies

Comme indiqué au chapitre 2 (section 2.4.8, p.50) la procédure d'analyse fréquentielle régionale de Hosking et Wallis (1997) (HW) compte 4 étapes : la délimitation subjective ou par classification hiérarchique de type grappe (Clustering de Ward), la validation des régions homogènes à travers un test de discordance  $D_i$  puis d'homogénéité  $H1$ , l'ajustement d'une loi statistique régionale sur l'échantillon agrégé, puis le test d'adéquation de la distribution ajustée avec l'échantillon. Dans ce chapitre, les deux premières étapes sont mises en œuvre. Les deux étapes restantes sont étudiées en profondeur dans le chapitre 5.

La discordance  $D_i$  est mesurée à chaque site. Elle estime la dispersion des ratios de L-moments d'un site  $i$  considéré par rapport aux ratios moyens des autres sites formant la région. Hosking & Wallis (1997) donnent une région comme discordante si  $D_i \geq 3$  pour un nombre de sites  $N \geq 15$ , pour des régions formées par moins de 15 sites, le critère d'exclusion d'une station varie en fonction du nombre de sites étudiés (voir annexe 1.3).

La mesure d'hétérogénéité  $H$  a été développée par Hosking & Wallis (1993, 1997) en tant qu'indicateur de la quantité d'hétérogénéité dans les ratios L-moments pour un groupe de site/stations. La statistique  $H1$  mesure la variabilité relative des L-Cv (et  $H2$  celle des L-Cs,  $H3$  celle L-Ck, mais  $H2$  est moins discriminante que  $H1$  car elle prend rarement des valeurs supérieures à 2 même pour des régions grossièrement hétérogènes) observés. Plus précisément,

la mesure de  $H_1$  compare la variabilité observée avec celle espérée (attendue) d'un large échantillon tiré, dans une région homogène, à partir d'une distribution Kappa ayant des ratios L-moments moyens pondérés observés dans la même région. Les recommandations initiales de cette méthode étant que les régions où  $H_1 < 1$  sont acceptablement homogènes, pour  $1 < H_1 < 2$  sont possiblement hétérogènes,  $H_1 > 2$  sont certainement hétérogènes et qu'une redéfinition de la région doit être envisagée. Ce critère  $H_1$  mesure l'hétérogénéité statistique pour une distribution connue et ne tient pas compte de la variabilité d'une autre origine. Ces variabilités dans la mesure des précipitations peuvent être dues au déplacement de l'instrument de mesure PG ou PV, au changement d'opérateur, à différent niveau d'assiduité dans le relevé de la mesure, au taux de valeurs manquantes dues à la transcription, au manque d'attention dans la précision de la mesure, au changement de l'environnement de la station de mesure (bâtiment, vent, arbres, etc.). Reconnaisant cette variabilité additionnelle Wallis et al. (2007) suggèrent que les pluies annuelles maximales avec  $H_1 < 2$  peuvent être considérées comme acceptablement homogènes et  $H_1 > 3.00$  indiquerait une hétérogénéité.

Aussi, une valeur négative de la statistique  $H$  du test d'homogénéité est probablement due à des corrélations positives parmi les sites de la région. Et de grandes valeurs négatives  $H < -2$  sont certainement à associer à d'importantes corrélations croisées (Hosking & Wallis, 1997).

Généralement, en analyse fréquentielle régionale, les variables hydrologiques mesurées sur une même échelle temporelle sont regroupées pour former des régions spatialement homogènes. En plus de cette définition, il est considéré dans ce travail une homogénéité régionale sur l'espace et le temps. C'est-à-dire que les observations de pluies maximales annuelles sur huit durées d'agrégation (15, 30, 60, 120, 180, 360, 720 et 1440 minutes) et pour différents sites sont regroupées pour former un échantillon unique, c'est l'intégration spatio-temporelle, sous réserve de la validité de l'homogénéité spatio-temporelle. En d'autres termes c'est la plausibilité d'une invariance d'échelle simple spatio-temporelle qui est analysée. Ainsi, il est défini deux niveaux de régionalisation :

1. Une régionalisation spatiale pour diverses durées d'agrégation, valable sous réserve d'homogénéité spatiale. Les zones homogènes formées sont dites régions
2. Une régionalisation spatio-temporelle (intégration des observations dans l'espace et le temps), valable sous réserve d'homogénéité spatio-temporelle. Les zones homogènes formées sont dites sous-régions

Dans ce qui suit, après avoir formé les régions supposément homogènes, il sera mis en œuvre les tests de discordance  $D_i$  et d'homogénéité  $H_1$  de Hosking & Wallis (1997) pour les deux niveaux de régionalisation, et dans un but comparatif une analyse d'échelles des moments spatio-temporelle est aussi exécutée.

#### 4.3.1 Délimitation initiale des régions homogènes

Dans ce qui suit, la délimitation des régions homogènes spatiales (1<sup>er</sup> niveau de régionalisation) est mise en œuvre sur la base d'interprétation intuitive de la topographie et du positionnement géographique, ainsi que de certaines caractéristiques statistiques de chaque station, telles que : les cumuls de pluies moyennes/ médianes annuelles sur 24 heures et 1 heure,

les  $L-C_v$  des séries de 1440 minutes et  $L-C_v$  des séries de 60 minutes, le ratio de cumul moyen entre les durées 60 et 1440 minutes. Les régions, ainsi formées sont indiquées en figure 4.4.

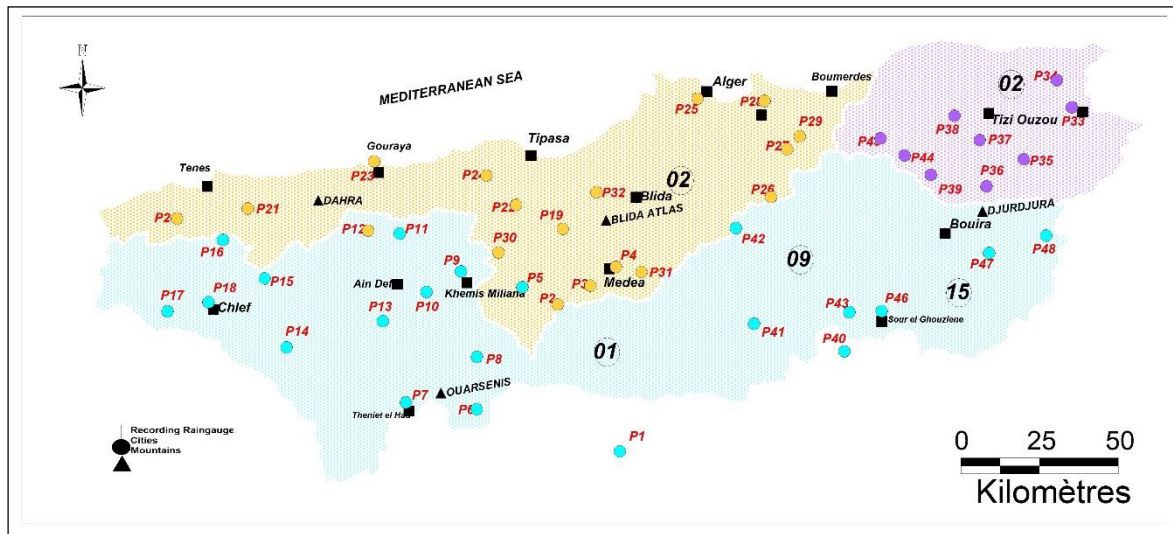


Figure 4.4 Délimitation initiale des Régions pluviométriques homogènes

#### 4.3.2 Tests de discordance et d'homogénéité HW (intégration spatiale)

La plupart des régions formées présentent une homogénéité spatiale quelle que soit la durée d'agrégation considérée.

Tableau 4.3 Résultats des tests de Discordance  $D_i$  et d'homogénéité  $H1$  de HW pour une régionalisation spatialisée sur différentes durées d'agrégation.

Régions		15	30	60	120	180	360	720	1440
1	H1	<b>2.48</b>	1.66	1.41	1.08	1.20	0.58	0.89	<b>-0.14</b>
	$D_i$	St5	St5	St5	St5	St5		St5	St5
2	H1	2.80	1.25	0.74	1.34	2.14	1.50	0.76	1.25
	Stations rejetées / $D_i$	-	-	-	-	<b>St21</b>	<b>St32</b>	-	-
3	H1	0.40	1.40	0.86	1.29	<b>2.34</b>	4.4	<b>3.26</b>	2.19
	Stations rejetées / $D_i$	-	-	-	-	-	-	-	-

Toutefois, des remarques s'imposent sur quelques valeurs particulières :

- **Région 1** : Le test de discordance discrimine la station 5 Djnane Ben Oudhah sur les durées 3 et 6 heures, cependant au regard de sa longueur d'enregistrement qui n'est que de 8 années, cette discordance est très vraisemblablement due à la variabilité d'échantillonnage. Au sens du test d'homogénéité toutes régions sur les diverses durées sont déclarées homogènes. Néanmoins, pour la durée 15 minutes il a été considéré un critère H1 assoupli ( $H1 > 3$  et non  $H1 > 2$ ) pour se soustraire de l'hétérogénéité. Aussi, pour la durée de 1440 minutes (1hour), H1 prend une valeur négative, ce qui renseigne sur une possible corrélation spatiale inter-site sur cette durée. Il est vrai que cette corrélation peut réduire la précision de l'estimation, mais toutefois, n'accentue pas un

éventuel biais d'estimation contrairement à une possible hétérogénéité (Hosking & Wallis, 1988).

- **Région 2** : Pareillement que pour la région 1, la région sur les différentes durées sont déclarées homogènes au prix d'un assouplissement du critère de rejet H1 (séries de 15 et 180 minutes). Par contre, les stations 21 de Bissa et 32 de Mouzaia sont rejetées par la mesure de discordance, mais leur conservation dans les régions considérées maintient l'homogénéité au sens de la statistique H1
- **Région 3** : En retirant les stations 44 de Tizi Ghenif et 45 de Chabet el Aneur, la région 3 présente une homogénéité au sens des deux tests de HW sur toutes les durées à l'exception de la durée de 720 minutes.

Toutes les régions homogènes pour les diverses durées d'agrégation sont indiquées sur le tableau 4.5. Toutefois, les critères d'homogénéité ont été assouplis au-delà des recommandations de HW pour une ou deux séries d'observations, dans le but que l'enrichissement du jeu de données à travers leur intégration contrebalancera positivement l'homogénéité imparfaite. Ceci est vérifié au chapitre 5.

### 4.3.3 Tests de discordance et d'homogénéité HW (intégration spatio-temporelle)

La constitution de régions homogènes avec intégration spatio-temporelle des données, appelées sous-régions dans ce qui suit, s'avère beaucoup plus compliquée que la régionalisation d'une variable hydrologique sur une même échelle temporelle. En effet, ni la mise en œuvre d'une ACP ou d'un algorithme de classification de Ward, ou encore moins une délimitation subjective n'ont été concluants pour former des régions homogènes avec intégration spatio-temporelle des observations.

Tableau 4.4 Résultats des tests de Discordance  $D_i$  et d'homogénéité H1 de HW pour une régionalisation avec intégration spatio-temporelle.

Sous-régions	A	B	C	D	E	F	G
Nombre de sites	7	4	7	7	5	7	6
H1 (cas A)	3.62	2.09	2.92	1.98	1.85	2.01	3.03
H1 (cas B)	4.31	1.74	1.95	3.41	2.5	1.61	3.17
Stations rejetées / $D_i$ (cas A)	-	-	-	-	-	-	-
Stations rejetées / $D_i$ (cas B)	-	-	-	-	-	-	-

Afin de constituer ces sous-régions (intégration spatio-temporelle), il est mis en place une procédure de *trial-and-error* basée sur la convergence de l'algorithme Bayesian MCMC dans le calage du modèle IDF intégré régional (calage sur un échantillon unique composé d'observations variant dans le temps et l'espace. En termes plus claires, quand le calcul Bayésien MCMC ne converge pas, l'échantillon conjoint est déclaré hétérogène. Même si les

mécanismes régissant cette façon de faire n'ont pas été explorés, les sous-régions avec intégration spatio-temporelle ainsi formées (voir tableau 4.5 et figure 4.5), sont quasiment toutes déclarées homogènes au sens des tests  $D_i$  et  $H_1$  de Hosking et Wallis (Tableau 4.4). La seule exception notable étant la sous-région A. Notons que les cas A intégrant les durées sur une gamme de 15 minutes à 24 heures passent plus facilement les tests de HW comparativement au cas B (1h – 24h), cela ne peut s'expliquer que par un nombre de données intégrées plus important dans le cas A par rapport au cas B.

Tableau 4.5 Délimitation et description des régions et sous-régions homogènes du centre Nord de l'Algérie.

Nota : les stations en gras n'ont pas été considérées pour former les régions homogènes.

Identifiant régions	Description régions	Stations
1	Hinterland centre Nord de l'Algérie (19 stations) : - Plaine du Chélif - Versant Sud de l'Atlas Tellien (Djurdjura, Atlas Blidéen, Dahra). - Versant Nord Ouarsenis - Amont Bassin versant Chélif (Haut et Moyen) (code ANRH 01) - Amont Bassin versant Isser (code ANRH 09)	<u>Sous-région E (3809 km<sup>2</sup>) : 5 stations</u> 13(El Touaibia); 14(Oued Fodda); 15(Oum Drou); 17(Domaine si Tayeb); 18 (Chlef);
		<u>Sous-région F (3889 km<sup>2</sup>) : 8 stations</u> 2(Barrage Ghrib) ; 5(Djnane Ben Oudhah) ; 6 (Bordj Amir Aek) ; 7(Theniet el Had) ; 8(Deurdeur) ; 9(Khemis-Miliana) ; <b>10(Harraza)</b> ; 11(Arib Ebda).
		<u>Sous-région G (4246 km<sup>2</sup>) : 6 stations</u> 40(Souk el Djemaa) ; 41(Souagui) ; 43(Dechmia) ; 46(Sour el Ghozlene) ; 47(Bouira) ; 48(M'Chedalla).
2	Côte Méditerranéenne du centre Nord de l'Algérie (20 stations) : - Région Littorale de Boumerdès jusqu'aux environs de Ténès (Limite w. Chlef) - Plaine de la Mitidja - Versant nord de l'Atlas Tellien (Atlas Blidéen; et Dahra).	<u>Sous-région A (3264 km<sup>2</sup>) : 7 stations</u> 12(El Anneb) ; 16(Benaria) ; 20(Abou el Hassene) ; 21(Bissa) ; 22(Meurad) ; 23(Gouraya) ; 24(Boukerdane).
		<u>Sous-région B (2072 km<sup>2</sup>) : 6 stations</u> 3(Tamezguida) ; 4(Medea) ; 26(Deux Bassins) ; 27(Hamiz) ; 31(Ouzra) ; 42(Mahterre).
		<u>Sous-région C (3205 km<sup>2</sup>) : 7 stations</u> 19(Berkhissa) ; 25(Birmandreis) ; 28(Rouiba) ; 29(Keddara) ; 30(Oued Zeboudj) ; 32(Mouzaia) ; 49(Soumaa).
3	Nord Grande Kabylie (9 stations) : - Massif côtier de Taguemount - Versant nord Djurdjura - Bassin versant du Sebaou - Aval bassin versant Isser (rive droite)	<u>Sous-Région D (4619 m<sup>2</sup>) :</u> 33 (Azazga) ; 34(Aghrib); 35(Beni Yenni); 36(Ait Djema); 37(Bouassem); 38(Argoub); 39(Ain Zaouia); <b>44(Tizi Ghenif)</b> ; <b>45(Chabet el Ameur)</b>

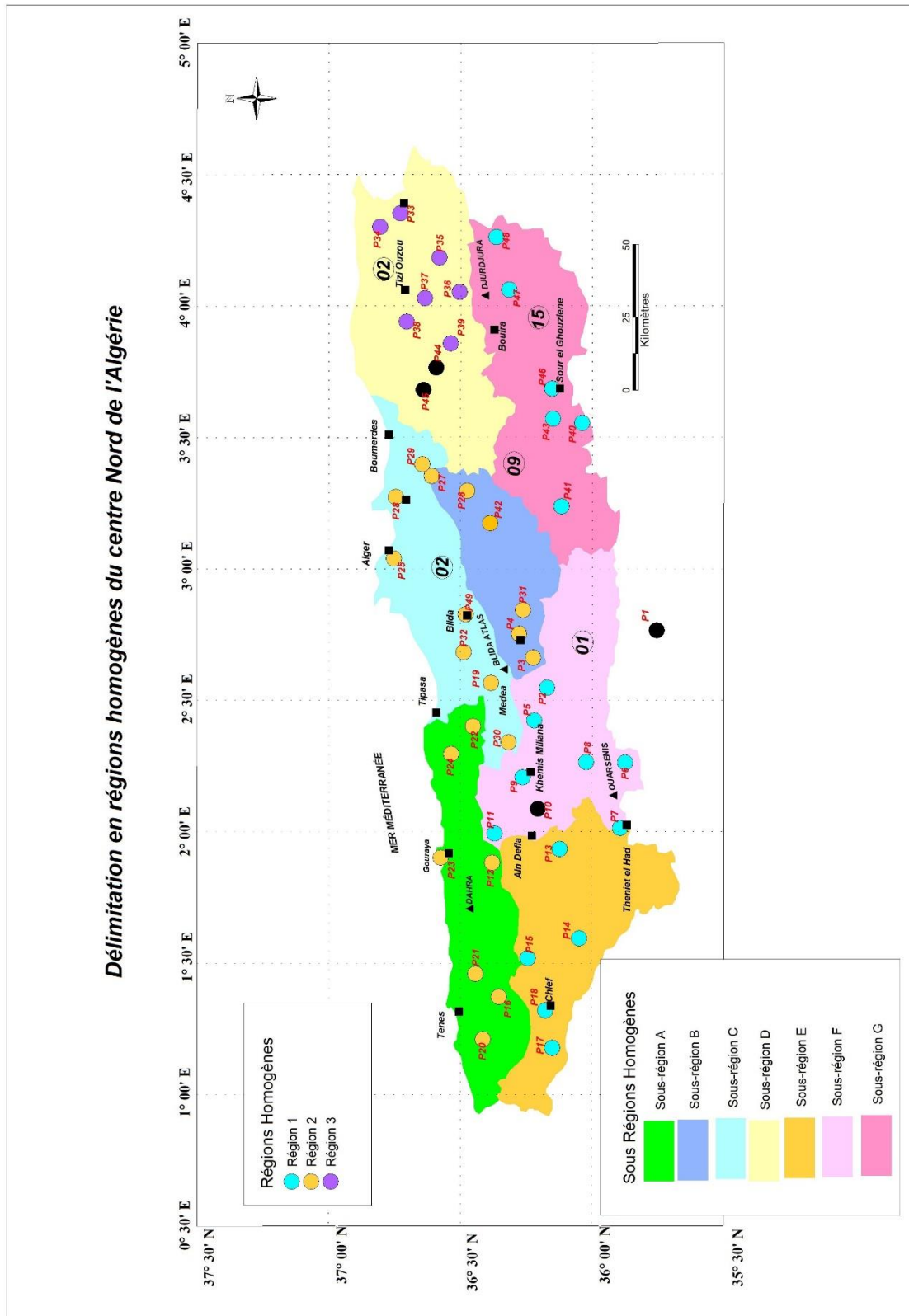


Figure 4.5 Carte des régions et sous-région homogènes du centre Nord de l'Algérie

#### 4.3.4 Analyse supplémentaire sur l'invariance d'échelle spatio-temporelle

Afin de compléter les résultats obtenus sur les sous-régions formées, une analyse d'échelle des moments est implémentée sur les échantillons conjoints, intégrant des observations sur les huit durées d'agrégation de plusieurs sites présentant une supposée homogénéité spatiale des 7 sous-régions (tableau 4.5). Dans la figure 4.6 est présenté le résultat des relations graphiques entre le logarithme des moments de cumulés de pluie  $\log E[Y_d^q]$  et le logarithme des durées  $\log d$  et ainsi que la relation entre la fonction d'exposant d'échelle  $k(q)$  et l'ordre  $q$  des moments pour deux sous-régions répondant différemment aux tests H1 de Hosking et Wallis (1997). En effet la sous-région A est hétérogène  $H1=3.62$  (figure 4.6a) et la sous-région D est homogène  $H1=1.98$  (figure 4.6b). Toutefois, la MSA laisse paraître que les deux sous régions (A et D) ne rejettent pas l'invariance d'échelle simple et montrent un comportement fractal avec une relation d'exposant d'échelle linéaire (autosimilarité : monofractale passant par l'origine). Ceci confirme les résultats obtenus par quelques auteurs (Ribatet et al., 2007), où il est constaté qu'un assouplissement mesuré des critères d'homogénéité régionale permet l'intégration d'un plus grand nombre de données, induisant une amélioration des précisions d'estimation.

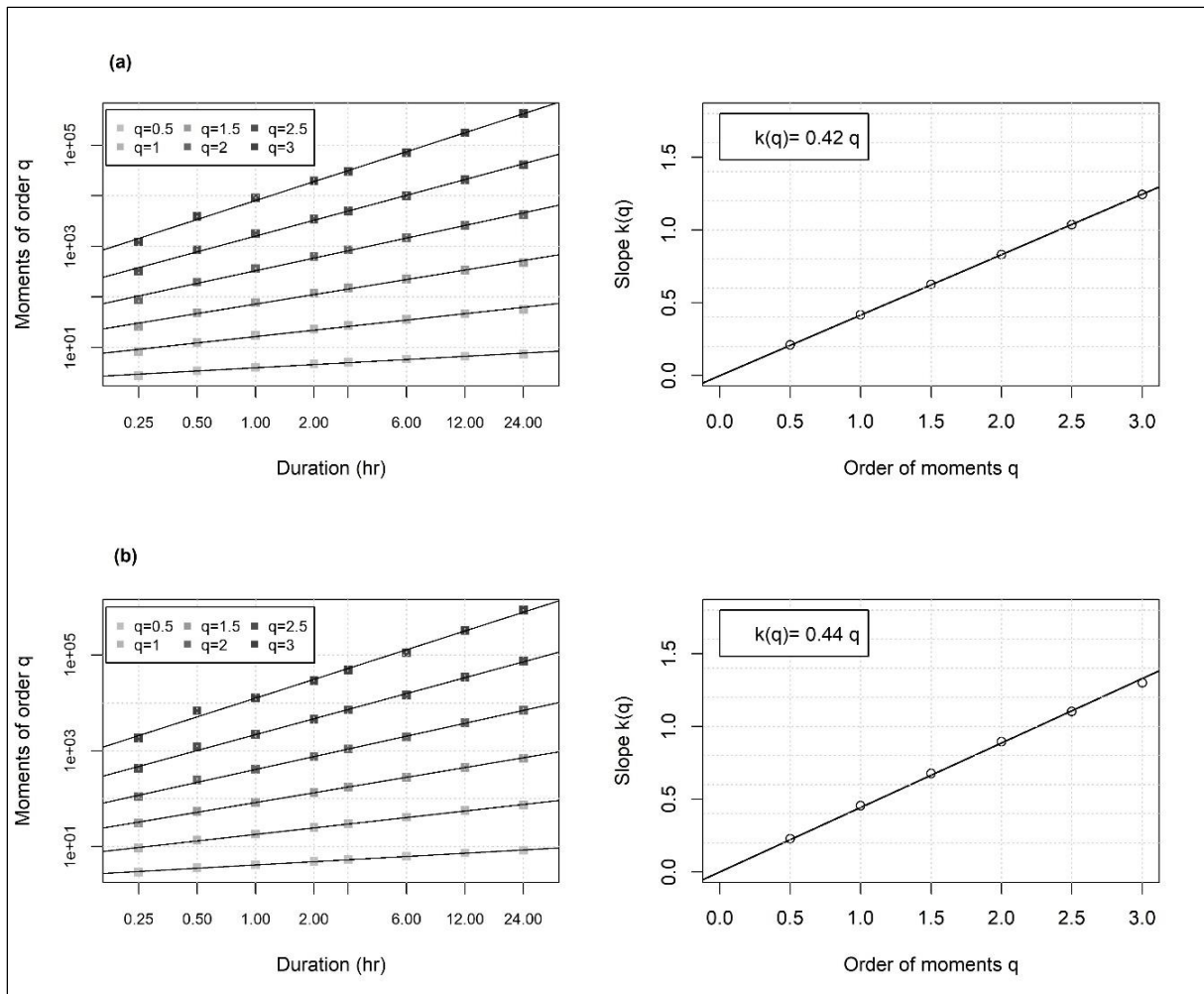


Figure 4.6 MSA en considérant le modèle M3A (intégration spatio-temporelle de 15minutes à 24heures).  
(a) Sous-région A et (b) sous-région D

## 4.4 Conclusion du chapitre 4

Les tests d'homogénéité (H) et de discordance (D) de Hosking & Wallis (1997) ont été utilisés pour la validation de région homogène mais aussi en tant que simple test d'homogénéité pour la vérification de la validité d'un échantillonnage intégré (conjoint). De plus, une ACP et un algorithme de classification (cluster) de Ward ont été mis en œuvre sur les données disponibles dans notre région d'étude. Les tests sur les échantillons régionaux intégrés n'ont pas été concluants. Par contre pour une application sur les durées de pluie séparément, les résultats de groupement étaient différents pour chaque durée d'agrégation, ce qui ne permet pas d'appliquer le modèle IDF Bayésien intégré proposé. Ce qui a amené à ne pas présenter le résultat de ces tests (ACP et Ward clustering) et à se focaliser uniquement sur le test d'homogénéité et de discordance.

Après avoir clairement établi que les données observées localement ne rejettent pas l'invariance d'échelle simple et la linéarité de la fonction d'exposant d'échelle (voir chapitre 5 pour plus de détails sur la mise en œuvre), il apparaît que l'extension de ce modèle IDF Bayésien intégré pour des estimations régionales est raisonnablement possible. En effet, une délimitation très fine (7 stations au maximum) des sous-régions (intégration spatio-temporelle) assurant une homogénéité, même imparfaite, permet une extrapolation des estimations, via le modèle Bayésien intégré dans l'espace et le temps, vers des sites non-jaugés. Ainsi, la robustesse du modèle à travers l'incorporation de plus de données peut contrebalancer les biais d'estimation dus à de possibles hétérogénéités (issues de l'assouplissement des critères d'homogénéité). Ceci doit être impérativement vérifié pour chaque cas d'application (voir chapitre 6).





## **Chapitre 5 :**

### **Etude des Estimations Bayésiennes intégrées des courbes Intensités-Durées-Fréquences (IDF)**

## 5.1 Introduction

La méthode rationnelle développée à partir du milieu du 19<sup>ème</sup> siècle (Kuichling, 1889; Lloyd-Davies, 1906; Mulvaney, 1850) repose sur le fait que le débit de pointe à l'exutoire, en considérant un coefficient de ruissellement constant, est proportionnel à l'intensité moyenne maximale des pluies sur une durée égale au temps de concentration du bassin versant. Dès lors, la méthode rationnelle s'est imposée comme une méthode de référence dans la conception des structures de gestion des eaux de pluies, particulièrement en milieu urbain. Son utilisation requiert la connaissance de la fréquence d'occurrence des intensités de pluie sur diverses durées. Cette information est synthétisée sous forme de courbe Intensité-Durée-Fréquence (IDF) (Chow et al., 1988). Au-delà des aspects de dimensionnement, ces dernières années, les courbes IDF ont aussi été le centre d'intérêt de plusieurs travaux de recherches, particulièrement pour l'analyse de l'impact des changements climatiques sur les précipitations extrêmes (Agilan & Umamahesh, 2016; Chandra et al., 2015; Cheng & AghaKouchak, 2015; Herath et al., 2016; C. H. R. Lima et al., 2016; Mailhot et al., 2007; Mondal & Mujumdar, 2015; Peck et al., 2012; Shahabul Alam & Elshorbagy, 2015).

Classiquement, les courbes IDF sont établies en ajustant une distribution statistique, séparément, sur les séries des maxima annuels des pluies sur diverses durées. Les paramètres des relations IDF sont calculés empiriquement dans une deuxième étape. Toutefois, l'estimation des courbes IDF est exécutée en deux ou trois étapes de calcul. Des incertitudes sont associées à chacune de ces étapes d'estimation (Bernard Mohymont & Demarée, 2006; Aart Overeem et al., 2008; Hans Van de Vyver & Demarée, 2010). Elles sont particulièrement reliées à la taille limitée des jeux de données observées disponibles pour le calage des courbes IDF (Koutsoyiannis et al., 1998). Ces incertitudes d'estimation sont difficiles à évaluer dans la procédure classique d'estimation des courbes IDF, étape par étape et sont souvent ignorées. Cependant, ces incertitudes d'estimation font partie intégrante de l'information statistique issue de l'analyse des observations et doivent être prises en considération dans l'évaluation du risque et la conception des ouvrages (Coles et al., 2003; Gaume, 2018).

La méthodologie proposée et testée vise deux objectifs : (i) faire la meilleure utilisation possible des mesures et de leurs propriétés dans le calage des courbes IDF, (ii) évaluer les incertitudes reliées aux paramètres estimés du modèle IDF et donc aux quantiles pluies leurs correspondant.

Au regard du premier objectif, plusieurs travaux de recherche ont souligné le lien entre la formulation de la relation de courbes IDF et les propriétés d'échelle des séries pluviométriques (Bendjoudi et al., 1997; Burlando & Rosso, 1996; Menabde et al., 1999). En particulier, en cas d'invariance d'échelle simple, les fonctions de distribution des probabilités des intensités/ cumuls de pluie sont identiques sur une certaine gamme d'échelle de temps (durée) à un facteur d'échelle près, facteur dépendant des durées considérées. Les distributions sont alors le produit d'une distribution adimensionnelle réduite et d'un facteur d'échelle dépendant de la durée. Comme dans le cas d'une analyse statistique régionale

(Hosking & Wallis, 1997), cette propriété rend possible de caler la distribution fréquentielle réduite à partir d'un échantillon conjoint (pooled) des séries des maxima annuels d'intensités/ cumuls de pluie observés sur différentes durées. L'échantillon à ajuster étant enrichi, on espère à travers cette approche d'obtenir des estimations plus précises. Depuis les travaux de Koutsoyiannis et al. (1998) et Menabde et al. (1999), plusieurs travaux de recherches ont utilisé ces propriétés d'invariance d'échelle simple dans le calage de courbes IDF à partir d'échantillons conjoints (Aronica & Freni, 2005; Bara et al., 2009; Blanchet et al., 2016; Ghanmi et al., 2016; Innocenti et al., 2017; Menabde et al., 1999; Muller et al., 2008; Nhat et al., 2007; Panthou et al., 2014; Xu & Tung, 2009; Yu et al., 2004). La plupart de ces travaux procèdent en deux étapes, calant en premier la relation facteur d'échelle et durée, puis dans une deuxième étape en ajustant la distribution fréquentielle réduite avec l'échantillon des intensités/ cumuls de pluie normalisés (c'est-à-dire divisés par le facteur d'échelle). Cette approche ne permet pas de facilement évaluer les incertitudes associées à l'ensemble de la procédure de calage des courbes IDF. Quelques études seulement ont proposé des approches intégrées, permettant un calage combiné des paramètres de la fonction d'échelle et de la distribution fréquentielle réduite, ouvrant la possibilité à une évaluation des incertitudes associées (Blanchet et al., 2016; Muller et al., 2008).

Au regard du deuxième objectif d'évaluation des incertitudes, les approches Bayésiennes, considérées comme les plus adaptées (Coles & Pericchi, 2003; Reis & Stedinger, 2005), se sont progressivement développées dans tous les domaines de l'hydrologie statistique et même dans l'estimation des courbes IDF (Chandra et al., 2015; Cheng & AghaKouchak, 2015; Huard et al., 2010; Rosbjerg & Madsen, 1996; Van de Vyver, 2015a). La possibilité de combiner une inférence statistique Bayésienne et une approche intégrée dans l'estimation des courbes a été montrée dans quelques travaux (Muller et al., 2008; Van de Vyver, 2015a). Toutefois, ces publications se sont limitées à quelques sites d'application, choisis dans un but illustratif. Mais l'approche intégrée repose sur une hypothèse. L'hypothèse que les éventuels biais introduits via un jeu de données contraint à travers un calage intégré et une expression mathématique nécessairement simplifiée des courbes IDF, vont largement être compensés par l'augmentation de la taille des séries sur lesquelles l'inférence statistique est menée. La pertinence de cette hypothèse peut être vérifiée par simulation numérique : cf. par exemple, chapitre 7 de Hosking & Wallis (1997) sur l'analyse statistique régionale ou Halbert et al. (2016). Mais il est aussi primordial de vérifier empiriquement de telles approches intégrées et leurs valeurs ajoutées, sur de nombreuses séries de données, ce qui est proposé ici.

## 5.2 Formulation Bayésienne intégrée des courbes IDF

La procédure d'inférence Bayésienne utilisée dans ce travail est très similaire à celle proposée par Nguyen et al. (2014) et Halbert et al. (2016) dans l'analyse fréquentielle régionale des débits, où la surface des bassins versants étant substituée par la durée d'agrégation. Cette procédure est incluse dans le package nsRFA du logiciel de modélisation

statistique R (Viglione et al., 2014) et correspond aux fonctions BayesianMCMC et BayesianMCMCreg.

Soit  $Y_d$  un échantillon de  $n$  observations d'intensités/ cumuls de pluies maximales annuelles sur des durées  $d$  ( $Y_d = \{y_{d_1}^1, y_{d_1}^2, \dots, y_{d_1}^n\}$ ) et  $Y$  l'échantillon correspondant de  $n$  valeurs annuelles maximales pour  $l$  différentes durées considérées  $\{d_1, d_2, \dots, d_l\}$ .

$$\mathbf{Y} = \begin{Bmatrix} y_{d_1}^1 & \cdots & y_{d_1}^n \\ \vdots & \ddots & \vdots \\ y_{d_l}^1 & \cdots & y_{d_l}^n \end{Bmatrix}$$

La forme des relations de courbes IDF ayant été définie au chapitre 2 (équation 2.39), les paramètres de modèle statistique correspondant restent à estimer : c'est-à-dire le paramètre  $b$  aussi bien que les paramètres d'une distribution fréquentielle pour les intensités/ cumuls de pluies réduits  $F_\theta(y_d/d^b)$ . Telle que justifiée au chapitre 2 (§ 2.4.5) c'est la distribution GEV à 3 paramètres qui est utilisée dans ce qui suit. Cette distribution est souvent considérée comme bien adaptée dans la description des occurrences de variables hydrologiques extrêmes (Koutsoyiannis, 2004a, 2004b), et a déjà été testée avec succès dans notre région d'étude (Meddi & Toumi, 2015). Pour rappel, sa fonction de répartition prend la forme suivante, avec le vecteur  $\theta$  des paramètres de position, d'échelle et de forme :  $\theta = \{\xi, \alpha, k\}$ .

$$F_\theta(x) = \exp \left[ - \left( 1 - \frac{k(x - \xi)}{\alpha} \right)^{1/k} \right]_{\alpha > 0} \quad (5.1)$$

Et la fonction de densité correspondant à l'expression suivante :

$$f_\theta(x) = \frac{\partial F_\theta(x)}{\partial x} = \frac{1}{\alpha} \left( 1 - \frac{k(x - \xi)}{\alpha} \right)^{1/k-1} \exp \left[ - \left( 1 - \frac{k(x - \xi)}{\alpha} \right)^{1/k} \right]_{\alpha > 0} \quad (5.2)$$

Une fois le type de distribution choisi, la vraisemblance des échantillons  $Y_d$  ou  $Y$  sachant les valeurs des paramètres  $\xi, \alpha, k$  et  $b$  peut être calculée si les divers intensités/ cumuls de pluies observés sont considérés comme indépendants (équation 5.3) :

$$\ell(Y_d | \theta) = \prod_{i=1}^n f_\theta(y_d^i) \quad \text{et} \quad \ell(Y | \theta, b) = \prod_{i=1}^n \prod_{j=1}^l f_\theta \left( \frac{y_{d_j}^i}{d_j^b} \right) \quad (5.3)$$

Les conséquences de l'hypothèse simplificatrice d'indépendance entre les maxima annuels d'intensités/ cumuls calculés sur diverses durées  $d_j$  sont discutées plus tard en section (5.6).

Une fois l'expression de vraisemblance définie, l'algorithme Bayesian Monte Carlo Markov Chaîne (MCMC) peut être implémenté afin de déduire la fonction de densité des probabilités postérieure des paramètres  $\xi, \alpha, k$  et  $b$  et des quantiles d'intensités/ cumuls correspondants  $(y_{d,T} = d^b F_\theta^{-1}(1 - 1/T))$ , suivant l'échantillon observé  $Y$ . Ce genre d'algorithme est devenu très commun en hydrologie, les lecteurs intéressés peuvent se

reporter au chapitre 2 (section 2.4.2) ou pour de plus amples détails vers des publications précédentes (Gaume et al., 2010; Reis & Stedinger, 2005; Viglione et al., 2013).

La procédure proposée a deux avantages majeurs comparativement à de nombreux travaux précédents sur les courbes IDF. Premièrement, les paramètres des courbes IDF ne sont pas estimés en plusieurs étapes successives mais en une unique étape intégrée. Deuxièmement, les incertitudes reliées aux estimations, sous forme de distributions postérieures, sont évaluées.

### 5.3 Méthodologie d'évaluation des performances

Deux aspects principaux sont considérés dans l'évaluation de la procédure de calage du modèle IDF Bayésien Intégré, et la comparaison des résultats obtenus avec l'approche intégrée (méthode M1) et l'approche classique consistant à caler les distributions statistiques pour chaque durée séparément (méthode M0). Premièrement, il est vérifié que les distributions calées avec la méthode M1, c'est-à-dire contraintes par les valeurs observées sur toutes les durées, sont adéquates avec les échantillons observés sur diverses durées d'agrégation. C'est le test d'Anderson-Darling (AD) qui est utilisé pour comparer les échantillons observés pour chaque durée et les distributions calées intégrées correspondant au maximum de vraisemblance. AD est connu comme étant un test puissant (Stephens, 1976), plaçant plus de poids, que les autres tests, sur les queues des distributions (Laio, 2004), qui sont d'un intérêt primordial dans les analyses sur les IDF. Comme indiqué au chapitre 2, le test AD a été souvent utilisé dans des précédentes études sur les courbes IDF (Ben-Zvi, 2009; Chang et al., 2016; Gargouri-Ellouze & Chebchoub, 2008; Panthou et al., 2014; Van de Vyver & Demarée, 2010). Plus la valeur de la statistique  $A^2$  du test AD est faible, plus l'ajustement est adéquat (Ben-Zvi, 2009).

Selon Stephens (1977), la valeur critique de  $A^2$  pour un seuil de signification de 10% est de 1.933. Néanmoins, cette valeur critique devrait être considérée comme un guide plutôt que comme une référence précise de comparaison et cela pour deux raisons. Premièrement, elle a été établie pour une distribution de Gumbel et non pas pour la distribution GEV, précisément une distribution EVII (Fréchet) avec une variabilité et une asymétrie plus importante, qui est considérée ici. Ainsi, la valeur critique qui devrait être considérée est probablement plus grande que cette valeur de référence 1.933. Deuxièmement, étant donné que la distribution du maximum de vraisemblance est calée sur les données observées, mais pas seulement, vu l'utilisation de données observées sur d'autres durées d'agrégation, la valeur critique  $A^2$  devrait être ajustée, c'est-à-dire réduite, pour une comparaison juste (Stephens, 1977). En l'absence d'un quelconque travail sur les valeurs critiques de  $A^2$  en des situations comparables à celle présentée dans ce travail et en considérant la complexité de la problématique, il a été décidé de garder la valeur critique de référence standard  $A^2$ .

Deux autres critères sont considérés dans l'évaluation de la cohérence et de la performance de la méthode M1. Ils sont basés tous deux sur les intervalles de crédibilité calculés par Bayesian MCMC. Le premier critère rend compte de la proportion des intervalles de crédibilité à 90% du modèle M1 qui est incluse dans les intervalles de

crédibilité à 90% du modèle M0 (OP : pourcentage de chevauchement). Ce premier critère est calculé pour un test supplémentaire sur la cohérence de l'approche intégrée. Le deuxième critère représente le ratio entre les intervalles de crédibilités à 90% des modèles M1 et M0 (PR : pourcentage de réduction). Ce second critère est calculé pour évaluer la performance (valeur ajoutée sur la précision d'estimation) de l'approche intégrée proposée. L'influence de possibles dépendances, entre les séries observées pour différentes durées, sur les intervalles de crédibilités de M1 est discutée en section 5.6.

## 5.4 Calage local du modèle IDF Bayésien Intégré

Selon le développement en section 5.2, quatre paramètres doivent être estimés à travers une procédure d'inférence statistique Bayésienne : les trois paramètres de la distribution GEV réduite ( $\xi$ ,  $\alpha$ ,  $\kappa$ ) et le paramètre d'échelle  $b$ . La procédure d'inférence consiste fondamentalement en un échantillonnage aléatoire des séries de jeux de paramètres probables, typiquement quelques centaines de milliers, sur la base des données observées. Les quantiles de cumuls de pluies correspondants à ces jeux de paramètres peuvent être par la suite calculés de même que les intervalles de crédibilité sur la base de ces centaines de milliers de valeurs de quantiles échantillonnés. Notons que le processus d'inférence Bayésienne MCMC est exécuté avec un Prior « non-informatif », bien que l'utilisation d'information a priori, en bornant par exemple les valeurs possibles du paramètre de forme  $\kappa$ , soit possible et permet éventuellement d'éviter des estimations irréalistes. Toutefois, il n'a pas été considéré comme nécessaire d'introduire un niveau de complexité supplémentaire. En effet, les estimations (Prior non-informatif) des paramètres de forme sont apparues cohérentes avec des estimations par L-moments dans la même région d'étude (Meddi & Toumi, 2015; Papalexiou & Koutsoyiannis, 2013).

Le tableau 5.1 donne les résultats obtenus sur la cohérence et les critères de performances du modèle IDF Bayésien intégré, tels que définis dans la section précédente, pour toutes les stations pluviographiques et pour les deux cas de calcul A et B. Les deux critères OP et PR, présentés en tableau 5.1, sont calculés sur la base du quantile 100 ans, mais les résultats sont consistants pour tous les autres quantiles. Dans le même esprit, dans un but de clarté, seuls les résultats pour des durées 1h et 24 heures sont présentés. La figure 5.3 illustre les résultats pour la station 33 de Azazga, qui passe parfaitement les tests d'homogénéité (HW et MSA), et la station 30 d'Oued Zeboudj qui montre une homogénéité douteuse et une invariance d'échelle simple imparfaite (les résultats des stations restantes sont données en annexe 5).

### Notice explicative du Tableau 5.1

- **N** : Longueurs des séries d'observation.
- **Valeurs entre parenthèses correspondent au cas B.**
- **Les valeurs  $A^2$  en gras rendent compte du dépassement de la valeur critique 1.933 pour un seuil de signification de 10%.**
- **PR** : Pourcentage de réduction de l'amplitude de l'intervalle de crédibilité T=100ans.
- **OP** : Pourcentage de chevauchement entre les intervalles de crédibilité des modèles M0 et M1.

Tableau 5.1 Statistique AD, cohérence et critère d'évaluation des performances

ID.	N	PR/d=1h	PR/d=24h	OP/d=1h	OP/d=24h	A <sup>2</sup> /d=1h	A <sup>2</sup> /d=24	Id.	N	PR/d=1h.	PR/d=24h.	OP/d=1h.	OP/d=24h.	A <sup>2</sup> /d=1h	A <sup>2</sup> /d=24h
1	24	60 (62)	73 (79)	86 (65)	92 (39)	1.02 (0.3)	1.21 (0.47)	25	16	96 (96)	11 (6)	82 (73)	75 (87)	0.38 (0.29)	0.78 (0.61)
2	15	70 (65)	47 (48)	100 (85)	100 (86)	0.28 (0.35)	0.94 (0.73)	26	24	86 (81)	-438 (-540)	85 (73)	0 (0)	0.5 (0.51)	1.3 (1.26)
3	26	82 (71)	42 (23)	99 (85)	100 (92)	0.47 (0.39)	0.71 (0.38)	27	12	83 (51)	97 (93)	100 (93)	57 (64)	0.74 (0.29)	1.7 (1.33)
4	24	85 (78)	48 (33)	100 (100)	100 (100)	0.42 (1.1)	<b>2.03</b> (0.7)	28	25	79 (58)	35 (-15)	100 (93)	100 (100)	0.31 (0.6)	0.83 (0.44)
5	8	100 (100)	- 13 (-9)	78 (69)	66 (99)	1.41 (1.33)	0.51 (0.45)	29	13	96 (96)	66 (61)	100 (77)	100 (81)	0.23 (0.31)	0.6 (0.65)
6	29	93 (89)	75 (68)	87 (90)	100 (94)	1.66 (0.96)	1.25 (0.79)	30	38	78 (74)	70 (69)	100 (74)	100 (100)	0.33 (0.75)	1.23 (0.66)
7	28	96 (95)	69 (67)	100 (98)	100 (100)	0.77 (0.67)	0.56 (0.6)	31	32	- (-)	86 (73)	0 (22)	81 (76)	0.9 (0.63)	0.62 (1.5)
8	14	90 (76)	96 (90)	76 (82)	69 (71)	0.28 (0.3)	1.69 (1.17)	32	14	71 (62)	70 (66)	100 (98)	100 (93)	0.36 (0.22)	0.47 (0.38)
9	25	64 (42)	97 (96)	93 (100)	100 (82)	0.75 (0.59)	1.53 (0.7)	33	29	27 (12)	74 (72)	100 (100)	100 (88)	0.93 (0.28)	1.48 (1.01)
10	16	91 (89)	74 (71)	100 (87)	100 (100)	0.31 (0.33)	0.94 (0.47)	34	11	91 (90)	83 (82)	100 (99)	100 (89)	0.3 (0.29)	0.71 (0.72)
11	28	84 (77)	46 (31)	78 (95)	100 (96)	0.64 (0.24)	1.52 (0.93)	35	29	85 (79)	69 (58)	100 (88)	100 (91)	0.29 (0.5)	0.98 (1.19)
12	29	89 (86)	64 (58)	90 (74)	100 (93)	0.63 (1.93)	<b>2.09</b> (1.31)	36	10	100 (99)	100 (100)	72 (85)	0 (59)	0.5 (0.63)	0.56 (0.59)
13	30	86 (78)	84 (77)	100 (85)	100 (100)	0.37 (0.91)	1.86 (1.07)	37	26	80 (76)	42 (32)	91 (81)	100 (100)	0.4 (0.48)	0.36 (0.28)
14	25	73 (64)	38 (34)	74 (86)	100 (100)	1.13 (0.75)	0.94 (0.36)	38	16	71 (-)	92 (75)	100 (82)	65 (98)	0.55 (0.72)	0.66 (0.42)
15	32	52 (32)	88 (85)	100 (100)	97 (83)	0.82 (0.77)	1.85 (1.03)	39	12	99 (99)	97 (95)	100 (100)	100 (100)	0.49 (0.29)	0.59 (0.33)
16	30	56 (21)	92 (89)	100 (100)	85 (75)	0.92 (0.64)	1.8 (0.73)	40	24	80 (67)	77 (69)	100 (100)	100 (100)	0.57 (0.36)	1.45 (0.65)
17	29	90 (86)	49 (44)	63 (51)	100 (97)	1.13 (1.33)	<b>3.41</b> (1.04)	41	28	96 (93)	83 (74)	82 (95)	100 (100)	0.28 (0.76)	1.93 (0.81)
18	28	70 (33)	59 (26)	100 (100)	100 (98)	0.79 (0.94)	<b>2.14</b> (0.9)	42	22	77 (70)	70 (60)	100 (89)	100 (95)	0.32 (0.39)	0.47 (0.71)
19	15	96 (91)	97 (95)	100 (100)	100 (100)	0.43 (0.44)	1.1 (0.66)	43	20	77 (68)	84 (80)	100 (100)	100 (100)	0.32 (0.54)	0.85 (0.98)
20	28	37 (-52)	83 (63)	83 (66)	81 (82)	1.34 (1.04)	<b>2.5</b> (1.76)	44	28	93 (90)	48 (34)	100 (98)	100 (100)	1.14 (1.54)	0.66 (0.58)
21	8	100 (100)	74 (60)	0 (20)	100 (100)	0.76 (0.66)	0.65 (0.81)	45	24	62 (33)	-139 (-305)	65 (69)	0 (14)	0.85 (1.14)	0.81 (0.74)
22	17	90 (81)	40 (9)	82 (86)	100 (94)	0.71 (0.33)	0.55 (0.28)	46	23	68 (54)	68 (50)	100 (76)	100 (80)	0.25 (0.47)	0.31 (0.62)
23	17	98 (98)	- 91(-47)	67 (50)	0 (58)	1.59 ( <b>1.99</b> )	<b>2.31</b> (0.69)	47	14	93 (90)	75 (65)	76 (77)	100 (90)	0.35 (0.74)	0.7 (0.37)
24	20	19 (-19)	46 (26)	84 (84)	99 (85)	0.46 (0.34)	1.02 (0.76)	48	29	77 (73)	80 (76)	99 (68)	100 (77)	0.26 (0.46)	0.62 (0.87)



A l'exception de quelques très rares cas, l'approche intégrée testée (M1) paraît fournir de plus petits intervalles de crédibilité que l'inférence basée sur des durées séparées (M0) : un pourcentage de réduction PR presque toujours positif (tableau 5.1). Le pourcentage de réduction des intervalles de crédibilités sur les quantiles 100 ans, entre les méthodes M0 et M1 est en moyenne égal à 70% pour différentes durées considérées. Ceci n'est évidemment pas étonnant. Les intervalles de crédibilités calculés à travers une procédure d'inférence Bayésienne reflètent le contenu en information des jeux de données analysés et ainsi ils sont reliés à leurs tailles. Ceci explique aussi pourquoi le pourcentage de réduction est généralement plus grand sur le Cas A que le cas B. Mais cette tendance générale, de réduction des intervalles de crédibilité avec la quantité de données, pouvait néanmoins être contrebalancée par une inadéquation entre le jeu de données et le modèle statistique calé, par exemple si le jeu de données conjoint était statistiquement hétérogène. Les résultats montrent que même si le modèle statistique d'invariance d'échelle simple proposé n'est pas parfaitement concordant avec les jeux de données, l'équilibre entre quantité de données et adéquation du modèle paraît positif au sens de la réduction des intervalles de crédibilité calculés. Cet état de fait est particulièrement bien révélé par le cas A, qui montre en moyenne de meilleures performances que le cas B, en dépit de sa concordance imparfaite avec l'hypothèse d'invariance simple. Néanmoins, il est impératif de garder à l'esprit que de possibles dépendances, aussi bien que les hétérogénéités sur les distributions, entre les cumuls de pluies maximales annuelles observés sur diverses durées, n'ont pas été prises en considération. La formulation de la vraisemblance utilisée dans la méthode intégrée assume indépendance et homogénéité. Ainsi, les intervalles de crédibilités calculées avec la méthode M1 sont certainement sous-estimés : voir discussion en section 5.6 ou les résultats basés sur des échantillons simulés hétérogènes fournis par Halbert et al. (2016).

Plus important encore, les distributions calées à travers la méthode intégrée M1 paraissent cohérentes avec les distributions calées sur des durées séparées (le pourcentage de chevauchement OP est généralement bon), et avec les mesures de pluies telle que montrées par les résultats du test AD (Tableau 5.1). Le chevauchement atteint souvent 100% : c'est à dire que l'approche intégrée améliore la précision des estimations de quantiles comparativement à la méthode de référence M0. Pour une faible proportion de stations (<10%) le OP est significativement inférieur à 50% et quelques fois 0%. Notons que les comparaisons sont basées sur les intervalles de crédibilité à 90% calculés. Les quantiles effectifs inconnus, que la procédure d'inférence statistique cherche à évaluer, ont théoriquement 10% de chance de se trouver en dehors des intervalles calculés. Il n'est ainsi pas étonnant que l'approche intégrée, en fournissant des informations supplémentaires, peut quelques fois significativement réorienter la distribution ajustée. Ce dernier constat, combiné avec la possible sous-estimation des intervalles de crédibilité dans la méthode M1, peut mener à ce qu'il n'y ait pas de chevauchement pour quelques cas. De plus, il a été vérifié que le quantile 100 ans estimé avec la méthode intégrée est concordant avec des valeurs de quantiles estimés au voisinage de la station.

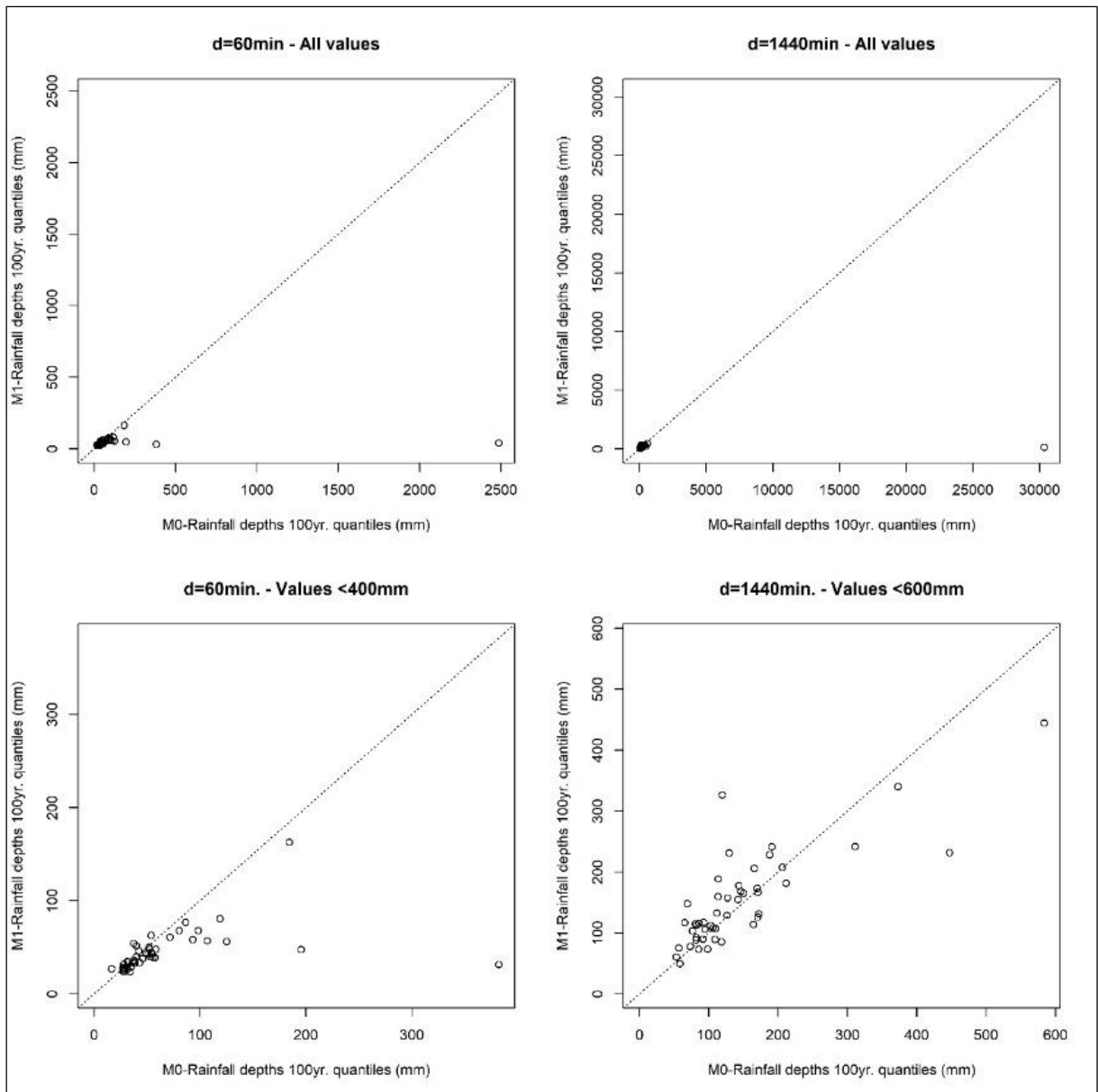


Figure 5.1 Quantiles 100 ans des cumuls de pluie : Model M0 vs. Model M1

Pour quelques stations avec des séries d'enregistrements courts, la méthode d'inférence statistique traditionnelle M0 peut mener à des estimations de quantiles clairement absurdes (voir par exemple la station 5 de Djane Ben Ouadhah -8ans- et 36 de Ait Djema -10an- pour des durées respectives de 1h et 24h dans la figure 5.1). La méthode intégrée M1 ne fournit pas de telles estimations absurdes ou surréalistes. Des tests supplémentaires en section 5.5 vont confirmer la relative robustesse de la méthode intégrée. Le test d'Anderson-Darling comparant le maximum de vraisemblance des distributions obtenues via la méthode intégrée M1 avec les données mesurées, donne aussi des résultats extrêmement satisfaisants (Tableau 5.1). En effet, en dépit des jeux de données contraints (pooled) du modèle IDF intégré, les distributions ajustées restent cohérentes avec les jeux de données observées sur toutes les durées. Le nombre de valeurs de  $A^2$  dépassant le seuil critique de rejet à 10% est un peu élevé pour les cumuls de pluie de 24 heures pour le cas A (prise en considération dans le calcul des pluies de moins de 1heure) : 6 valeurs en gras dans le tableau 5.1. Ceci, peut être interprété comme une indication que l'hypothèse d'invariance d'échelle simple sur

toute la gamme de durées (cas A de 15 minutes à 24 heures) n'est pas parfaitement adaptée aux données observées.

D'un autre côté, la méthode intégrée ne montre pas une tendance claire de réduction des quantiles estimés de 1 heure et une augmentation des quantiles de 24 heures telle qu'illustrée en figure 5.3. Ce dernier constat est aussi confirmé pour les stations où le seuil de rejet de  $A^2$  est dépassé (statistique du test AD). La possibilité d'un biais dû à la déviation du modèle d'invariance d'échelle simple, reliée à la grande proportion d'évènements convectifs pour les grandes périodes de retour, peut exister mais n'est pas détectable dans les résultats et a probablement une influence limitée dans les estimations de quantiles de période de retour de 10 à 100ans.

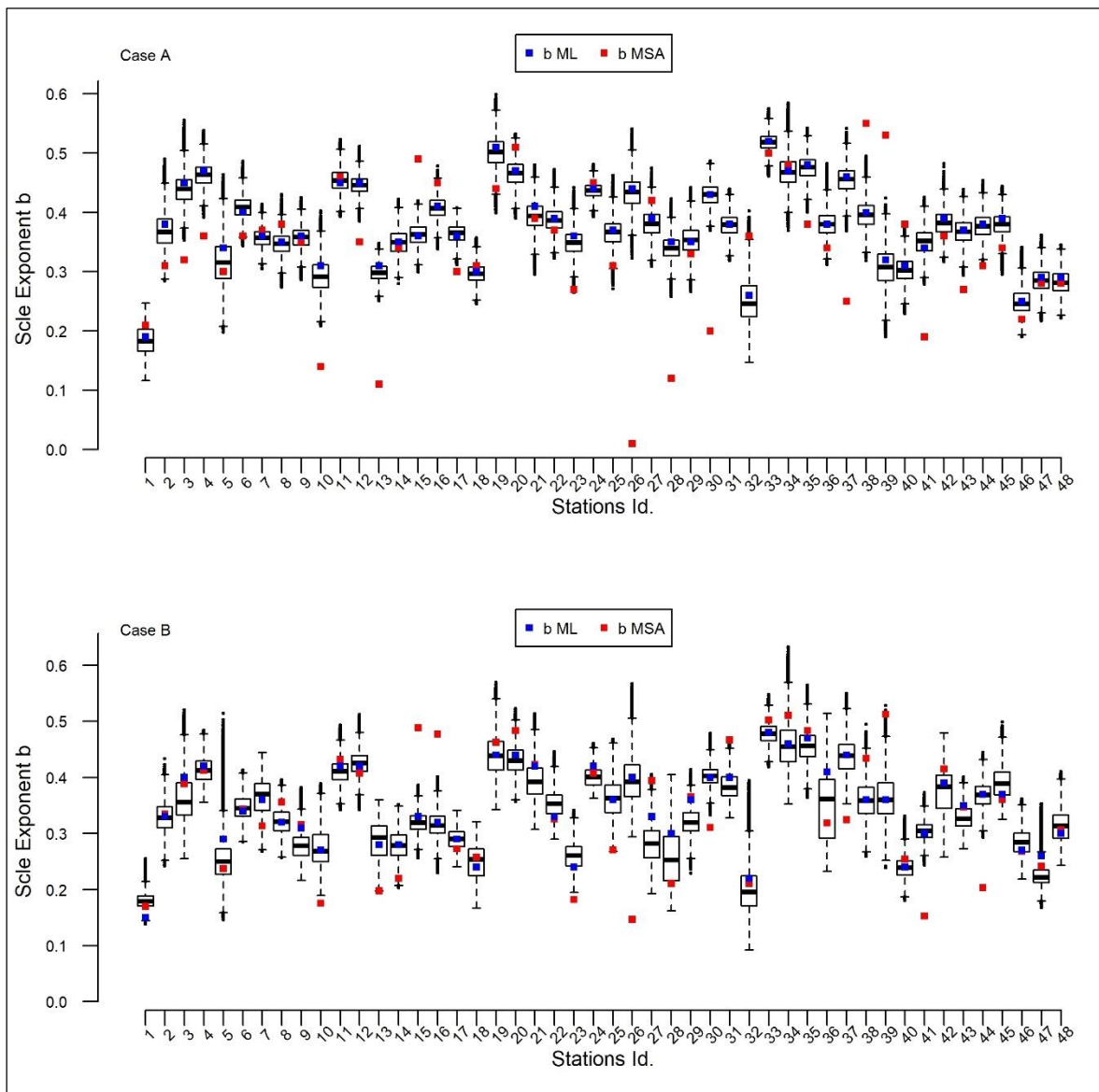


Figure 5.2 Les Boxplots représentent la distribution Postérieure (Bayesian MCMC) sur l'exposant d'échelle  $b$  estimé pour les cas A et B. Carrés Bleu (Estimation par Maximum de vraisemblance Bayesian MCMC). Carrés rouges (estimation par MSA)

Finalement, la figure 5.2 montre les distributions a posteriori pour le facteur d'échelle  $b$  calé, comparées avec les pentes  $k(q)$  (relation d'exposant d'échelle) issues de l'analyse des

échelles des moments. Les deux résultats semblent globalement cohérents mais les résultats de la MSA sont beaucoup plus variables : dans le sens où ils sont très sensibles à la variabilité d'échantillonnage et à la méthode d'estimation. Quelques fois, les estimations données par la MSA se positionnent en dehors de la gamme de valeurs possibles selon la méthode intégrée. C'est-à-dire qu'ils sont inconsistants avec la distribution GEV et le jeu de données. Ceci montre, que la méthode intégrée en une étape devrait certainement être préférée à la méthode en deux étapes, comme celle récemment proposée par Innocenti et al. (2017), et consistant à caler en premier le facteur d'échelle  $b$  sur la base d'une MSA puis la distribution fréquentielle réduite. La méthode intégrée fournit des résultats plus consistants et robustes. Les facteurs d'échelle calés sont difficiles à comparer du fait des incertitudes d'estimations. La plupart de leurs valeurs sont comprises entre 0.3 et 0.4. La valeur de la station 1 de Boughzoul s'écarte complètement des autres valeurs de  $b$ , avec une valeur significativement plus faible, signe que les pluies maximales annuelles sont dues à des événements convectifs courts. Ce qui est cohérent avec sa localisation dans la zone sud de la région d'étude. Pareillement, les stations de mesures avec apparemment de fortes estimations des valeurs de  $b$  (événements pluvieux prolongés) sont plus proches de la mer Méditerranée. Cette cohérence spatiale dans les résultats suggère que le modèle intégré proposé peut être régionalisé.

Les figures précédemment présentées (Figures 5.1, 5.2 et 5.3), basées sur un jeu de données relativement large mais néanmoins fini, laisse quelques questions ouvertes. Premièrement, l'adéquation de la relation du facteur d'échelle  $d^b$  en équation (2.39), a été indirectement évaluée à travers la MSA et quelques démonstrations théoriques. Une déviation de cette relation log-linéaire peut être une source majeure d'hétérogénéité et peut limiter la valeur ajoutée de l'approche intégrée. Toutefois, la vérification à travers des simulations sur les médianes empiriques, théorique (modèle M1 intégré) et leurs intervalles de crédibilité à 80% (figure 4.3 au chapitre 4), montre clairement l'adéquation de la relation d'échelle proposée. Pareillement, la formulation de la vraisemblance est basée sur une hypothèse d'indépendance qui n'est probablement pas garantie pour des échantillons conjoints de cumuls réduits de pluies maximales observés sur diverses durées. Quelle est alors l'ampleur de la sous-estimation des intervalles de crédibilité théoriques calculés par l'approche intégrée. Finalement, l'approche intégrée semble fournir des estimations robustes de quantiles (c'est-à-dire moins sensible à la variabilité d'échantillonnage). C'est clairement montré pour des stations avec des données de tailles limitées, telle la station 36 de Ait Djema (figure 5.4a-b) et par la faible proportion d'estimations aberrantes des quantiles (Figure 5.1). Est-il possible de confirmer ces propriétés de façon plus systématique ? Ces questions sont amplement discutées dans les deux sections suivantes.

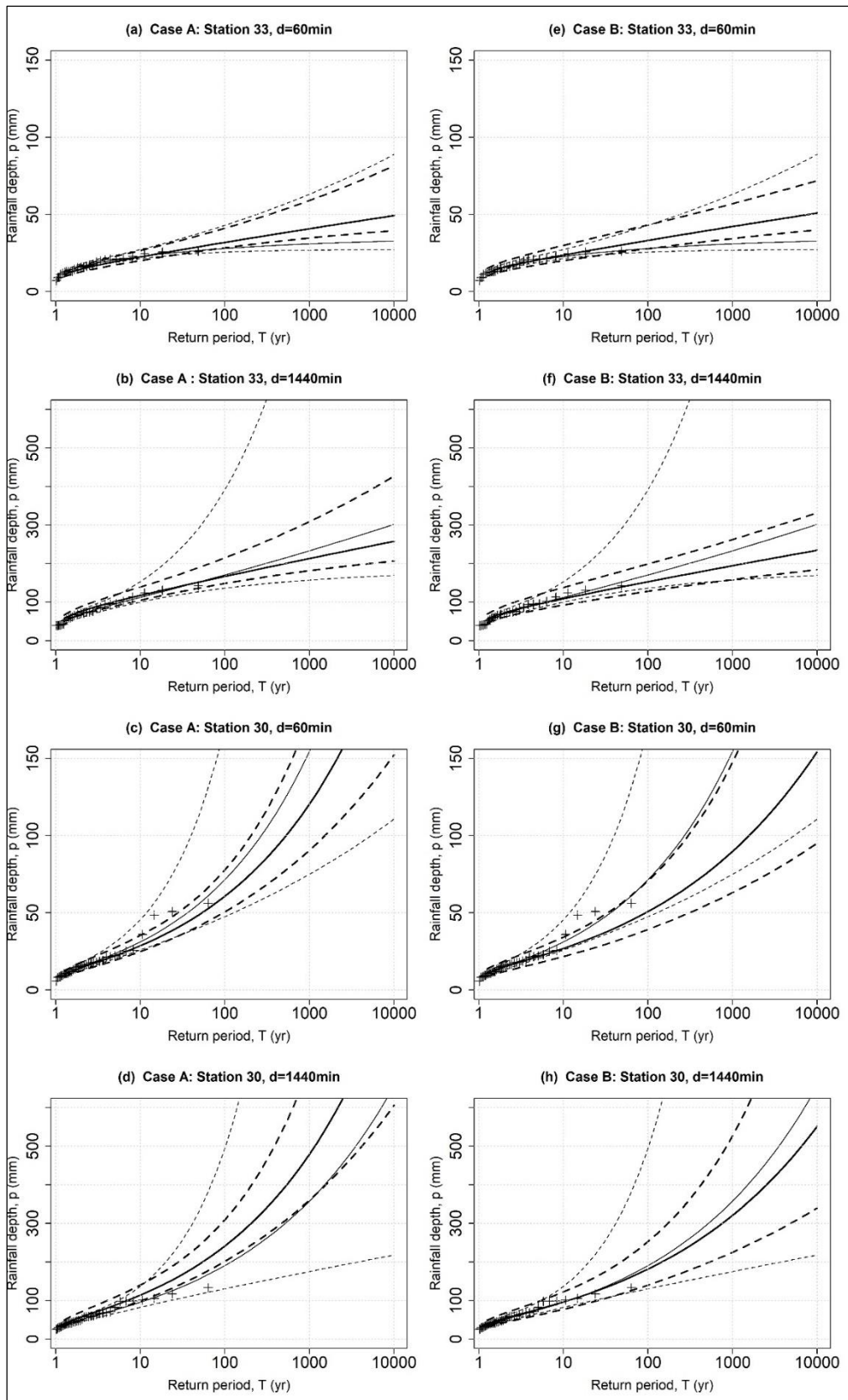


Figure 5.3 Séries des maxima annuels de cumuls de pluie ajustées à une distribution GEV. (à gauche : a, b, c, d) : Pour le cas A (15 min – 24h), (à droite : e, f, g, h) : Pour le cas B (60 min – 24 h). Les croix pour les observations. Traits fins : modèle M0. Trait gras : Modèle M1. Trait en tiret : intervalles de crédibilité.

## 5.5 Robustesse envers les courtes séries d'observations

Les quantiles estimés, aussi bien que les intervalles de crédibilité, sont dépendants des échantillons. L'approche intégrée de calage des courbes IDF mène quelques fois à des réorientations significatives des distributions calées particulièrement pour les séries de courtes durées avec un chevauchement limité des intervalles de crédibilité, résultant des modèles M0 et M1 (par exemple ; les stations 5, 21, ou 36 en tableau 5.1 et la figure 5.4).

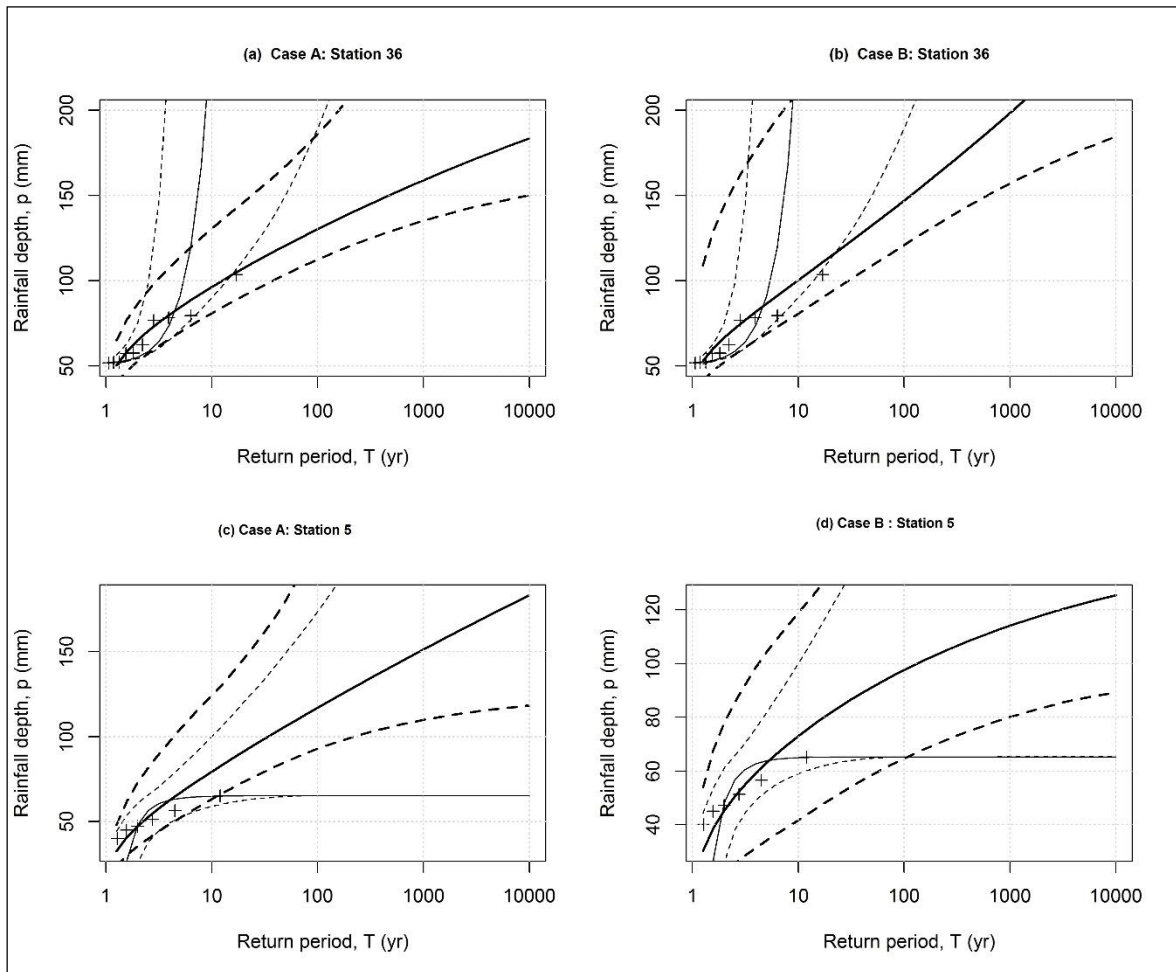


Figure 5.4 Séries des cumuls de pluies maximales annuelles ajustées avec une distribution GEV. Les croix pour les observations. Traits fins : modèle M0. Trait gras : Modèle M1. Trait en tiret : intervalles de crédibilité.

Le calage sur la base des durées séparées (M0) mène quelques fois à des estimations de quantiles qui sont irréalistes (beaucoup trop fortes) ou incohérentes avec les valeurs estimées au voisinage de la station. La seule exception étant la pluie sur 24heures à la station 26 ; il y a toujours une exception. L'approche intégrée M1 paraît fournir des estimations raisonnables pour tous les sites, même quand les séries d'observations sont de taille limitée (voir figure 5.4 : station 5 ou 36). Cette apparente robustesse, est une propriété qui semble confirmer ce qui a été montré à travers le calcul des intervalles de crédibilité : les cumuls mesurés sur une durée donnée fournissent une information utile pour conditionner la distribution fréquentielle des cumuls maximums sur d'autres durées. Des simulations Monte Carlo ont

été menées afin de confirmer cette apparente robustesse. Trois stations avec des séries d'enregistrement assez longues, ont été sélectionnées dans différentes sous-régions : station 15 de Oum Drou, 30 d'Oued Zeboudj et 33 de Azazga (voir figure 3.1 pour la localisation), avec respectivement des périodes d'enregistrement de 32, 38 et 29 années. Les quantiles 100 ans de maximum de vraisemblance de référence ont été estimés avec les méthodes M0 et M1 sur la base des SMA de pluies complètes. Pour chaque station, 100 séries de 10 années de longueur ont été aléatoirement tirées avec remise dans les séries d'observations. Deux cas sont considérés pour le modèle M1. Le cas 1, les séries sur les diverses durées ont été tirées indépendamment l'une de l'autre. Le cas 2, les séries sur diverses durées ont été tirées en même temps pour une année donnée afin de maintenir une éventuelle dépendance entre les cumuls de pluies de différentes durées. Les quantiles 100 ans, correspondant au maximum de vraisemblance, ont été estimés encore une fois mais sur la base des séries de 10 années tirées aléatoirement ainsi que les erreurs relatives avec les estimations de références.

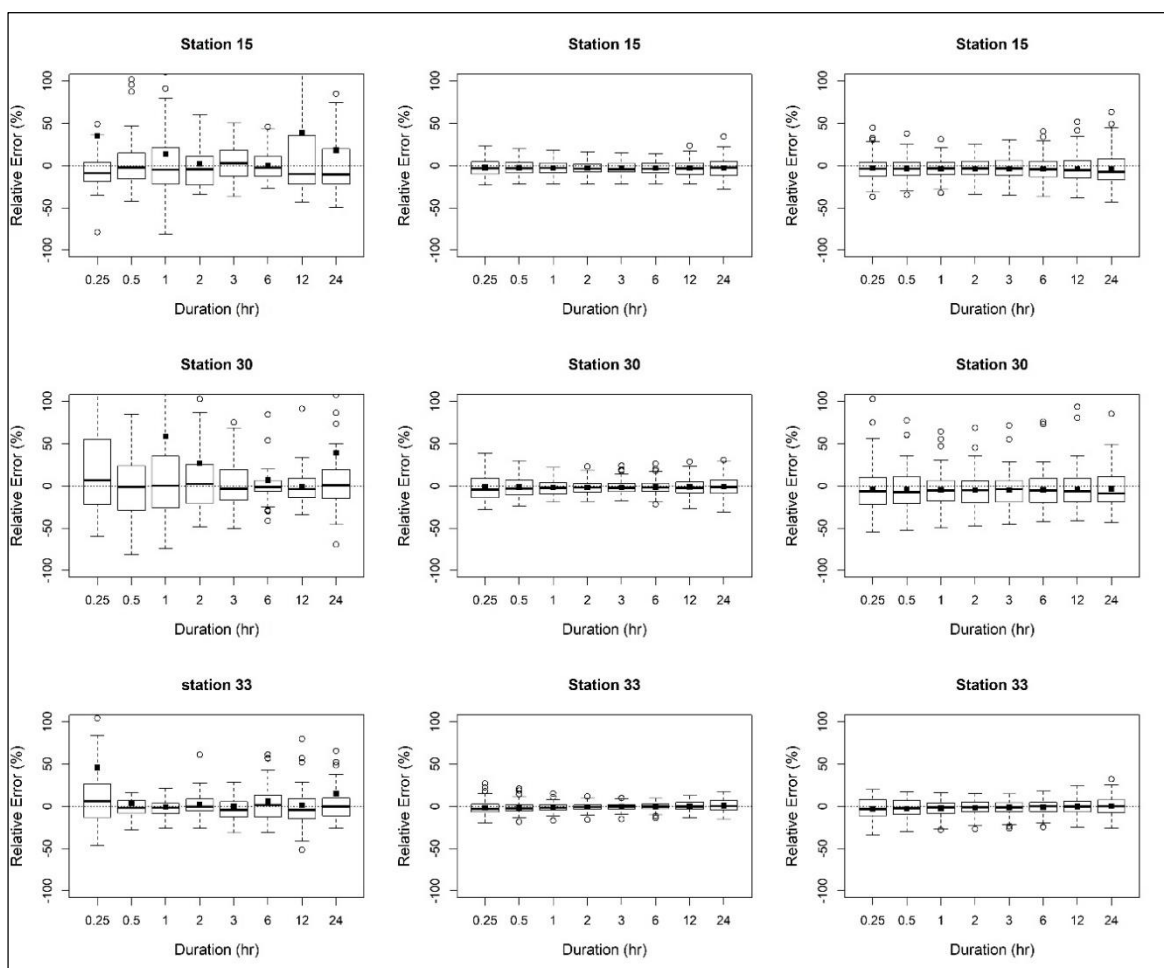


Figure 5.5 Boxplots de l'erreur relative entre quantiles de références ( $T=10$ ans) estimés par les deux modèles M0 et M1 et les quantiles basés sur des séries de 10 années de maxima annuels de pluies tirées aléatoirement. (à gauche) : méthode M0, (au milieu) : méthode M1 cas 2, (à droite) : méthode M1 cas 1.

La figure 5.5 confirme sans aucun doute possible la robustesse de la méthode intégrée M1 comparativement à la méthode M0. La comparaison entre les cas 1 (au milieu) et le cas 2 (à droite) aide à mesurer l'influence de la dépendance entre les cumuls de pluies observés sur différentes durées. Elle est significative, mais la valeur ajoutée de la méthode intégrée

M1 comparativement à la méthode M0 reste aussi très significative même en considérant les dépendances.

## 5.6 Sensibilité des intervalles de crédibilité à la dépendance intra-site des séries temporelles

Les cumuls de pluies maximales annuelles sur diverses durées ne sont pas indépendants tel que montré dans le tableau 5.2. Les calculs présentés dans les sections précédentes sont basés sur une vraisemblance assumant l'indépendance entre les différentes observations, ce qui mène à surestimer l'information contenue dans l'échantillon conjoint et ainsi sous-estimer la largeur de l'intervalle de crédibilité théorique dans le cadre de l'approche intégrée proposée. Il a été proposé dans quelques travaux précédents, d'introduire un modèle multivarié qui tient compte des dépendances dans la formulation de la vraisemblance intégrée (Muller et al., 2008). Cependant, ceci ne fait qu'ajouter de la complexité au modèle intégré proposé et nécessite de plus, le choix d'un modèle multivarié incorporant un nombre beaucoup plus élevé de paramètres qui sont probablement très difficiles à caler et à valider sur un jeu de données.

Tableau 5.2 Coefficient de corrélation moyen entre les séries de cumuls de pluies mesurés sur diverses durées pour 48 stations pluviographiques.

Duration (minutes)	15	30	60	120	180	360	720	1440
15	1	-	-	-	-	-	-	-
30	0.9	1	-	-	-	-	-	-
60	<b>0.74</b>	0.91	1	-	-	-	-	-
120	0.6	0.76	0.9	1	-	-	-	-
180	0.51	0.66	0.81	0.95	1	-	-	-
360	0.39	0.5	0.62	0.77	0.88	1	-	-
720	0.25	0.34	0.43	0.57	0.68	0.87	1	-
1440	<b>0.18</b>	0.25	<b>0.32</b>	0.44	0.53	0.73	0.93	1

Il est proposé d'évaluer la possible influence de la dépendance des résultats avec une approche plus simple. Le même modèle IDF intégré est calé sur la base d'un nombre de durées limité (15, 60 et 1440 minutes), pour lesquelles les séries de cumuls de pluies maximales annuelles observés peuvent être considérées comme relativement indépendantes (Kieffer Weisse, 1998). Le pourcentage de réduction des intervalles de crédibilité résultants comparativement à la méthode M0 sont réduits, mais demeurent significatifs pour presque toutes les stations pluviographiques (voir figure 5.6).

Le gain réel de l'approche intégrée en termes de précision d'estimation des quantiles est impossible à évaluer avec exactitude sans connaître le modèle de dépendance entre les cumuls de pluies maximales annuelles observés sur diverses durées. Mais, l'existence d'une dépendance ne paraît pas changer substantiellement les conclusions des sections précédentes, ceci concorde avec les conclusions de Hosking & Wallis (1988) dans l'étude de l'influence des corrélations inter-site en analyse fréquentielle régionale. Le gain de l'approche intégrée reste substantiel, même en cas de dépendance.



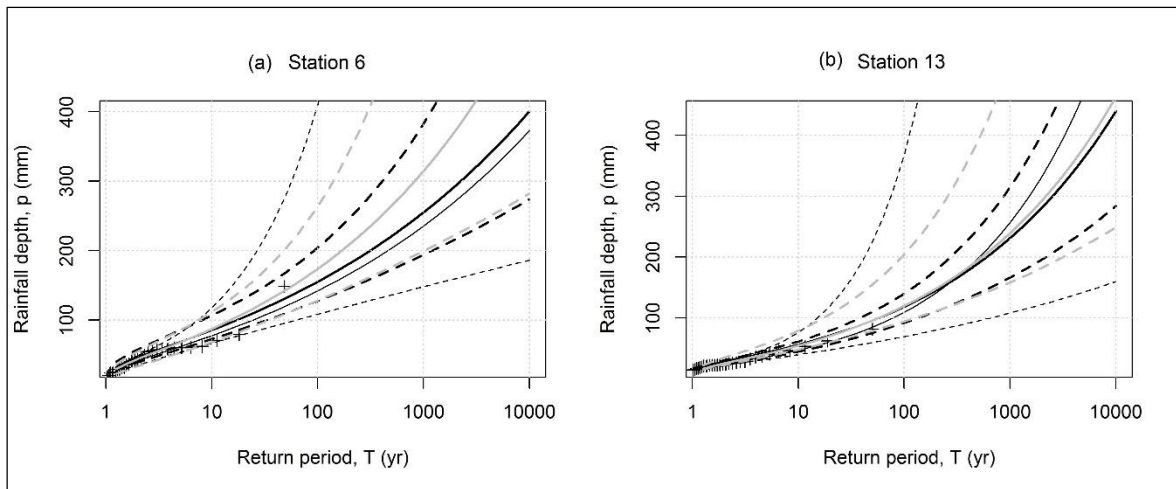


Figure 5.6 Séries des cumuls de pluies maximales annuelles ajustées avec une distribution GEV. Les croix pour les observations. Traits fins : modèle M0. Trait gras : Modèle M1. Trait en tiret : intervalles de crédibilité à 90%. Trait gras en gris montre les résultats obtenus avec la méthode M1 calée sur seulement 3 durées (15min, 60 min. et 24h).

## 5.7 Quelques discussions et conclusion

Ce travail consolide et étend les travaux précédemment cités (Koutsoyiannis et al., 1998; Menabde et al., 1999; Muller et al., 2008) en proposant une approche Bayésienne pour une estimation intégrée des courbes IDF à travers une analyse locale, où les courbes IDF ainsi établies sont théoriquement cohérentes avec les principes d'invariance d'échelle simple et de multifractalité. Mais aussi, il est essentiel de tester empiriquement de telles approches intégrées ainsi que leur valeur ajoutée des échantillons : ici un échantillon relativement large de 48 stations pluviographiques couvrant un territoire de 25 000km<sup>2</sup> au centre Nord de l'Algérie, ce qui est fait dans ce travail de thèse. Le but est de vérifier l'hypothèse d'invariance d'échelle simple, afin d'évaluer les incertitudes sur les quantiles de pluies estimés et de mettre en exergue les bénéfices qu'il est possible d'avoir à travers l'inférence statistique Bayésienne intégrée comparativement aux approches conventionnelles.

La méthode d'estimation Bayésienne intégrée des courbes IDF a été extensivement testée et évaluée sur un jeu relativement large et riche de données de pluies de la région centrale Nord de l'Algérie. Cette méthode permet l'évaluation de tous les paramètres des courbes IDF en une seule étape, puis fournit une estimation complète et cohérente des incertitudes associées sous la forme de distributions a posteriori et d'intervalles de crédibilité pour les paramètres du modèles IDF et les quantiles correspondants, distributions conditionnées des valeurs observées. Néanmoins, il doit être rappelé que les dépendances entre les observations sur diverses durées peuvent difficilement être prises en compte dans la procédure d'inférence, même si des travaux précédents ont apportés une solution partielle (Muller et al., 2008). Les intervalles de crédibilités postérieurs résultants devraient de ce fait être probablement considérés comme sous-estimés.

Excepté le calage cohérent des courbes IDF, l'approche intégrée semble aussi pouvoir significativement réduire les incertitudes reliées à l'inférence statistique comparativement à l'approche traditionnelle consistant à caler les modèles statistiques sur chaque durée séparément. Les résultats présentés confirment cela. Moyennant quelques hypothèses cohérentes avec les observations (i.e. l'invariance d'échelle), les séries maximales annuelles des intensités ou cumuls de pluies sur une durée donnée contiennent une information de valeur qui aide à contraindre les distributions statistiques des intensités ou cumuls de pluies maximum annuels sur d'autres durées.

Les fondamentaux de l'approche proposée ont déjà été présenté dans plusieurs autres travaux, mais c'est la première fois, à notre connaissance, qu'elle est testée et validée sur un échantillon aussi large de données. Des tests et vérifications supplémentaires sur d'autres contextes climatiques seront certainement nécessaires, mais la récente confirmation que l'hypothèse d'invariance d'échelle semble raisonnablement valide sur la grande région qu'est l'Amérique du Nord (Innocenti et al., 2017) indique que de telles approches d'inférence intégrée peuvent être largement utilisées. De plus, cette approche simple peut être relativement directement implémentée dans une version régionale si les homogénéités dans les distributions fréquentielles, aussi bien que dans les valeurs d'exposant d'échelle, sont identifiées entre stations (Halbert et al., 2016). Ceci est illustré dans le chapitre suivant.

Ceci étant dit, toutes les méthodes d'inférence statistique, particulièrement celles visant à créer des échantillons conjoints de données encore plus larges afin de réduire les incertitudes d'inférence, sont basées sur des hypothèses simples si ce n'est simplificatrices, telle que l'hypothèse d'invariance d'échelle simple dans la présente étude de cas. Il est possible de vérifier que ces hypothèses ne sont pas totalement incohérentes avec les observations disponibles, mais cela ne signifie nullement qu'elles sont parfaitement valides. Ces hypothèses contraignent les résultats d'inférence et peuvent introduire des biais, si elles ne sont pas parfaitement concordantes avec les observations, ce qui est pratiquement toujours le cas. Les techniques d'enrichissement des échantillons sont basées sur la balance entre incertitudes et biais. C'est à dire entre la taille de l'échantillon et l'adéquation du modèle calé. De façon intéressante, les résultats obtenus ici, sur une gamme de durées de 15 minutes à 24 heures, indiquent que la balance peut rester positive même dans les cas où il est détecté un écart de l'hypothèse d'invariance d'échelle simple sous-jacente. Ribatet et al. (2007) sont arrivés aux mêmes conclusions en appliquant une analyse fréquentielle régionale sur les écoulements : des estimations plus précises sur les quantiles de débits sont obtenues si le critère d'homogénéité régionale est relaxé dans une certaine mesure, afin de construire de plus larges régions statistiquement homogènes. Au-delà des méthodes, cette question de balance ou d'équilibre est clairement une question centrale s'agissant des méthodes statistiques basées sur l'enrichissement des jeux de données et pourrait faire l'objet d'une attention particulière pour des recherches futures sur cette problématique.



## **Chapitre 6 :**

### **Cartographie des courbes IDF : de l'approche locale vers l'analyse régionale**

## 6.1 Introduction

Tous les avantages et possibilités (cohérence théorique, réduction des incertitudes, robustesse) qu'offre, localement, l'estimation Bayésienne intégrée des courbes IDF proposée dans ce travail ouvrent un large spectre de perspectives pour des combinaisons locales et régionales des mesures conjointes (pooled) avec les objectifs de réduire les incertitudes de l'inférence statistique sur les quantiles de pluie de projet et d'explorer la forme des queues des distributions fréquentielles régionales des extrêmes pluviométriques.

L'utilisation d'une procédure de krigeage pour l'interpolation d'estimation de quantiles, locaux ou régionaux, sur les extrêmes pluviométriques est très fréquente. Cependant, elle présente une incohérence théorique majeure. En effet, le Krigeage suppose un champ de pluie Gaussien, or, les extrêmes pluviométriques ne concordent pas avec la normalité alors introduite par la technique d'interpolation et peut mener à des estimations de faible précision lors d'extrapolation vers des sites non-jaugés. Ainsi, l'objectif visé dans ce chapitre est de prospecter la meilleure solution de régionalisation et de cartographie permettant de combiner les incertitudes issues des estimations des distributions fréquentielles et celles issues des procédures d'interpolation, dans le but d'établir une cartographie des extrêmes pluviométriques limitant les risques de défaillances des ouvrages hydrauliques.

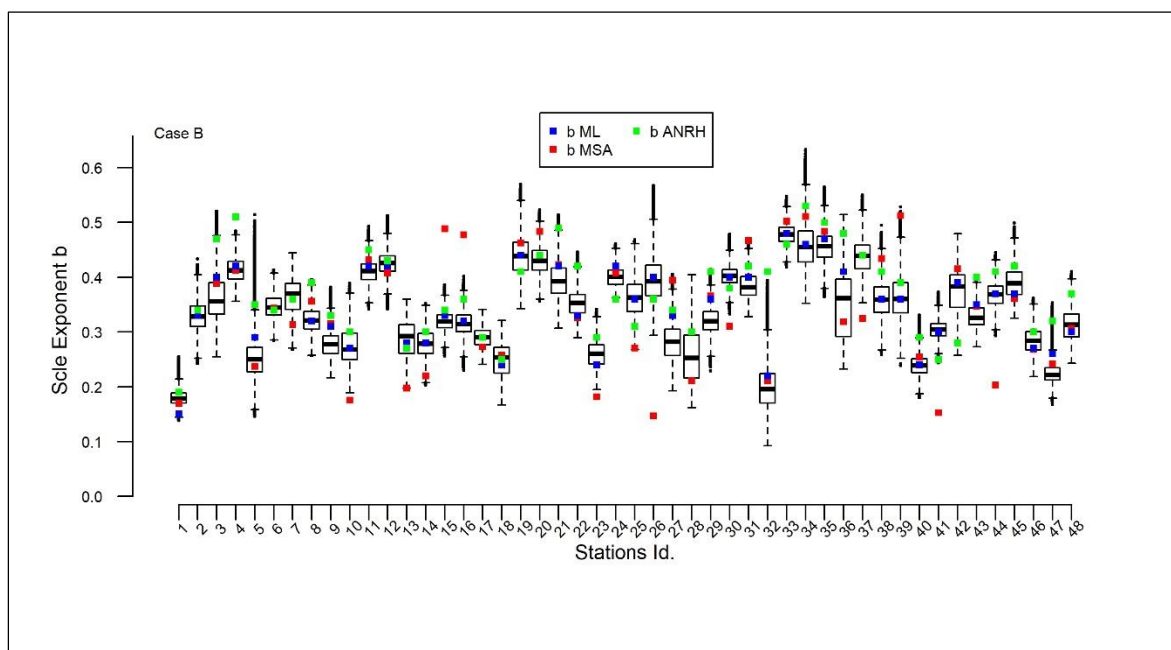


Figure 6.1 Boxplot représentant la distribution postérieure (Bayesian MCMC) sur l'exposant d'échelle  $b$  estimé pour le cas B. Carrés Bleus (Estimation par Maximum de vraisemblance Bayesian MCMC). Carrés rouges (estimation par MSA). Carrés Verts (estimation empirique ANRH – classique en trois étapes)

Certain auteurs (Khelfi et al., 2017; Kieffer-Weisse & Bois, 2001; Zahar & Laborde, 2007) préconisent de spatialiser, en lieu et place des quantiles de pluies extrêmes, quelques caractéristiques des pluies maximales annuelles, tels que le coefficient de Montana, les résidus de régression (linéaire multivariée) entre les quantiles et caractéristiques géographiques ou encore le Gradex. Or dans ce qui a été montré ici l'exposant d'échelle (exemple : coefficient de Montana) est extrêmement sensible à la variabilité d'échantillonnage et sa spatialisation qui

sera entachée d'incertitudes supplémentaires mènera très probablement à la définition d'évènements qui n'existent pas réellement. En effet, la figure 6.1 montre les valeurs possibles (au sens de la distribution à posteriori du Bayesian MCMC) prises par l'exposant d'échelle  $b$  suivant différentes procédures de calcul : Estimation Bayésienne Intégrée proposée dans ce travail, estimation empirique basée sur les observations à travers une analyse d'échelle des moments et enfin l'estimation empirique (fournie par l'ANRH) basée sur le calage classique (en 3 étapes) d'un modèle IDF sur les quantiles de pluies pour diverses durées. Même si globalement les valeurs estimées par les différentes approches sont concordantes, assez souvent les valeurs de  $b$  estimées empiriquement (MSA et ANRH) sont localisées en dehors des intervalles de crédibilités à 95%, ceci est essentiellement dû à la variabilité d'échantillonnage. Sachant qu'en moyenne l'estimation ANRH de  $b$  est différente de 30% (par rapport à l'approche intégrée), et que cet écart peut mener une estimation de quantiles de pluies 50% différente, il est plus que préférable de baser les extrapolations des caractéristiques des pluies sur des estimations plus robustes tels que les quantiles estimés ou mieux encore sur la borne supérieure des intervalles de crédibilités à 70% (Bernier & Veron, 1964; Gaume, 2018), quand ceux-ci sont définis de façon efficiente.

## 6.2 Cartographie des quantiles de pluies à partir d'estimation locale

Etant donné que le modèle local de courbes IDF, proposé ici, est calé, validé et testé, il est étudié la possibilité d'étendre son utilisation dans des analyses d'extrapolation qui permettraient des estimations des pluies horaires sur des zones non-jaugées. Dans un but de comparaison et de prospection de la meilleure approche de cartographie sur la base des modèle M0 et M1 définis au chapitre 5, il est proposé deux procédures différentes :

- a) Modèle M0 qui consiste à interpoler par Krigeage des quantiles de pluies estimés par inférence bayésienne pour chaque durée séparément au niveau de chaque poste PG,
- b) Modèle M1 qui consiste à interpoler par Krigeage des quantiles de pluies estimés par inférence Bayésienne de manière intégrée (intégration temporelle) au niveau de chaque poste PG.

### 6.2.1 Analyse des variogrammes expérimentaux vs. ajustés (M0 et M1)

Tel qu'indiqué au chapitre 2 (section 2.7), la procédure de Krigeage retenue dans ce travail est la plus simple possible. Il s'agit d'un krigeage ordinaire basé sur l'ajustement d'un modèle sphérique au semi-variogramme expérimental. En effet, encore une fois le but de cette cartographie des courbes IDF n'est pas de prospecter la meilleure technique de spatialisation, mais d'illustrer l'incorporation : (a) des incertitudes associées aux estimations de quantiles de pluies et (b) des incertitudes d'interpolation spatiale dans la cartographie des précipitations extrêmes. Dans un souci de clarté, il est présenté seulement les exemples de cartographie sur les quantiles de durées 1440 minutes (une journée) pour une période de retour de 10 ans.

La figure 6.2 montre que les semi-variogrammes pour le deux modèles M0 (fig. 6.2a) et M1 (fig. 6.2b) présentent des valeurs de palier quasiment égales, bien que M0 montre un léger effet de pépite. Toutefois la portée est plus grande de 10 km pour le modèle M1. Ceci indique une meilleure cohérence spatiale des quantiles M1, ce qui était attendu du fait de la robustesse d'estimation apportée par le modèle Bayésien intégré (M1). Aussi, les portées sur les quantiles décennaux sur les 8 durées d'agrégation varient entre 20 et 50 km. Cette faible valeur de portée est assez courante en région Méditerranéenne (Kieffer-Weisse & Bois, 2001; Zahar & Laborde, 2007).

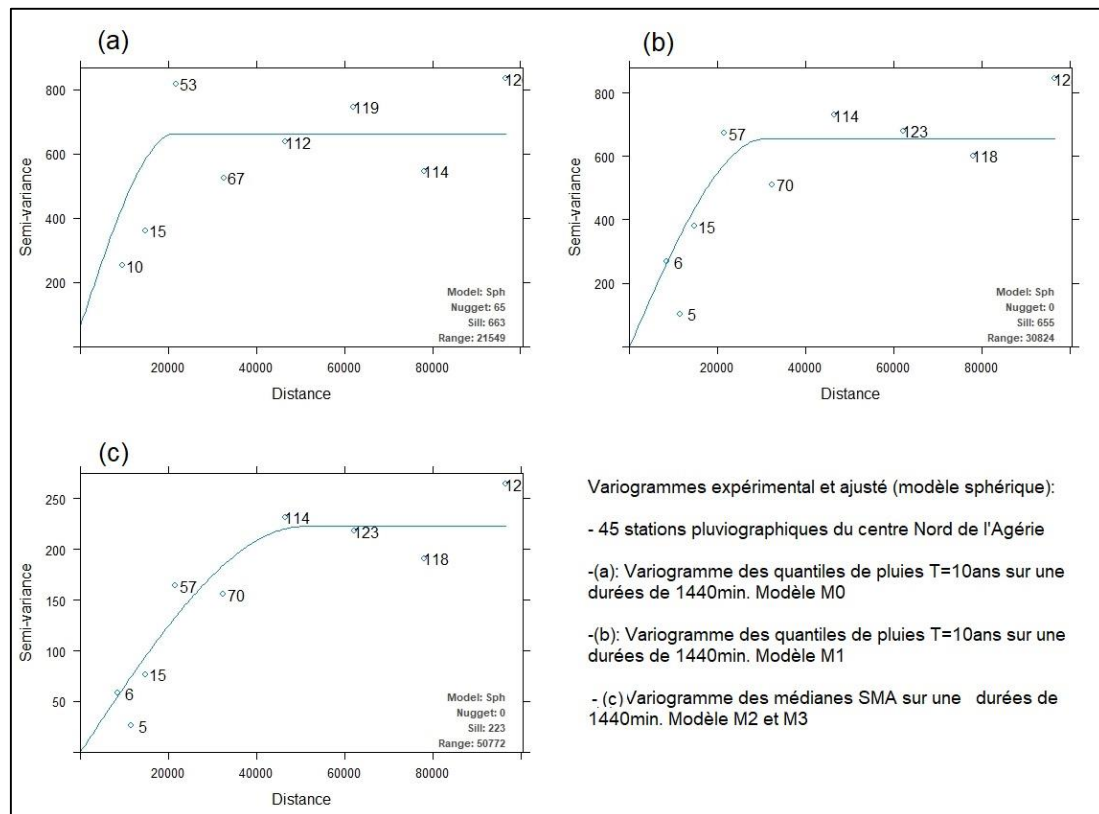


Figure 6.2 Graphiques des variogrammes expérimentaux vs ajustés

### 6.2.2 Interpolation spatiale des quantiles aux sites : Modèle M0 vs. Modèle M1

À une fin purement illustrative, il est retenu les quantiles 10 ans sur une durée de 1440 minutes soit une journée. Pour la cartographie de ces extrêmes pluviométriques, il est considéré une interpolation spatiale par krigeage ordinaire, ce dernier est implémenté à l'aide du package geoR (Ribeiro et al., 2020) sous le logiciel de modélisation statistique R. Les jeux de données observées analysés dans cette étude sont issus de 48 stations pluviographiques. Toutefois, tel que montré au chapitre 5, le modèle M0 donne des estimations de quantiles aberrantes pour les calages sur de courtes séries de mesures. Ainsi, toutes les valeurs de quantiles jugées irréalistes (>200mm pour d= 1440min et T=10ans) ont été retirées des jeux de données.

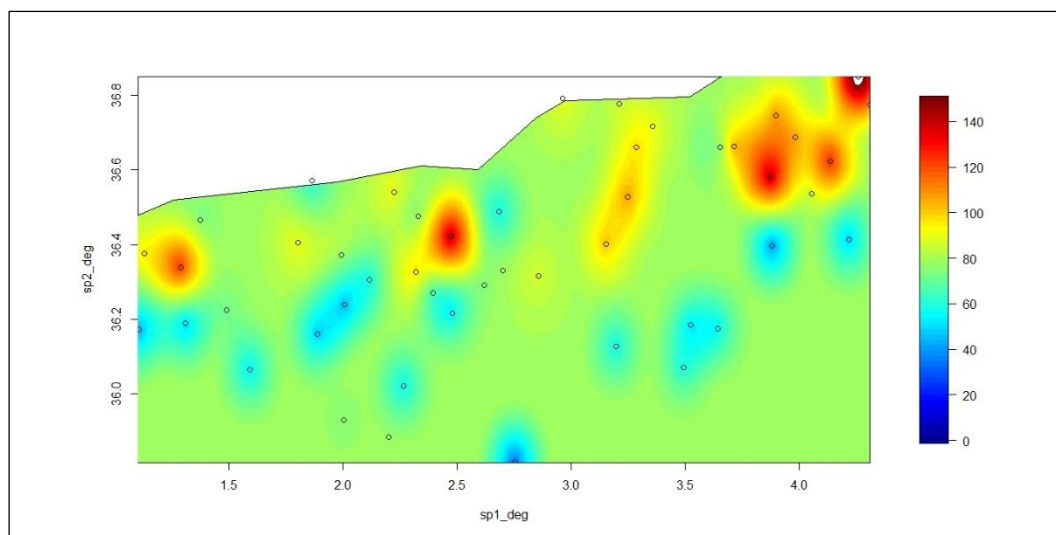


Figure 6.3 Cartographie des quantiles 10 ans sur  $d=1440$  minutes estimés par la méthode M0.

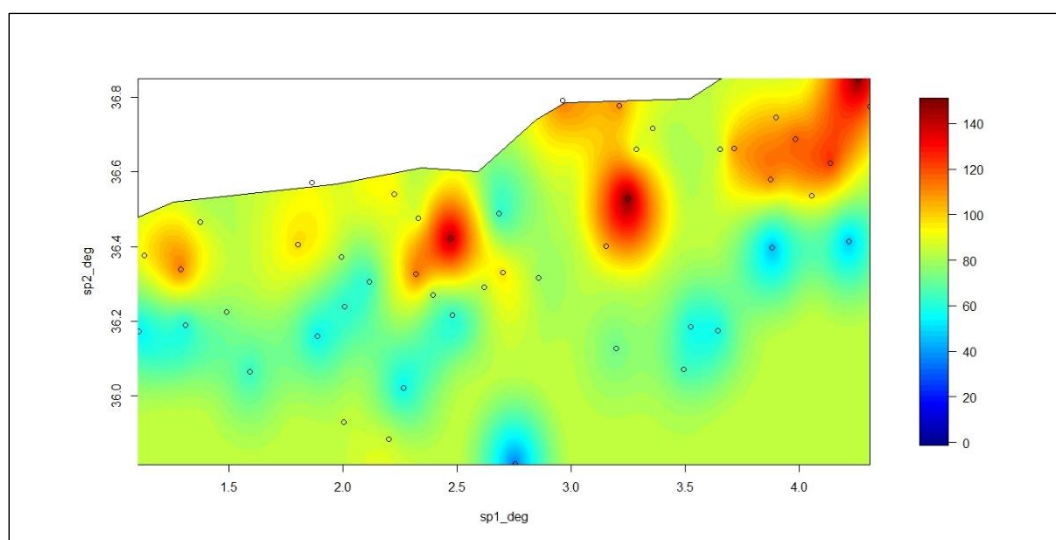


Figure 6.4 Cartographie des quantiles 10 ans sur  $d=1440$  minutes estimés par la méthode M1.

Le comparatif des figures 6.3 et 6.4 n'apporte pas grand-chose, si ce n'est que par endroits la méthode M1 donne des valeurs plus fortes, ce qui s'explique facilement du fait que le modèle M1 incorpore plus de données, en raison de sa robustesse vis-à-vis des séries de courtes durées, et par conséquent une relative meilleure cohérence spatiale. Mais aussi, à cause de certaines stations ayant un comportement particulier, telle que la station 26 des Deux bassins. En effet, cette dernière étant vraisemblablement soumise à un régime pluviométrique différent, à savoir la prédominance des événements convectifs intenses. Etant donné que dans le modèle M1 la distribution sous-jacente est contrainte par les observations sur toutes les durées, il est normal de retrouver au niveau des stations, avec de forts cumuls observés sur les durées de moins de 2h, des estimations de quantiles 10 pour  $d=24h$  plus fortes que les estimations de la méthode M0 sur  $d=24h$ , dont la distribution n'est contrainte en rien par les autres durées.



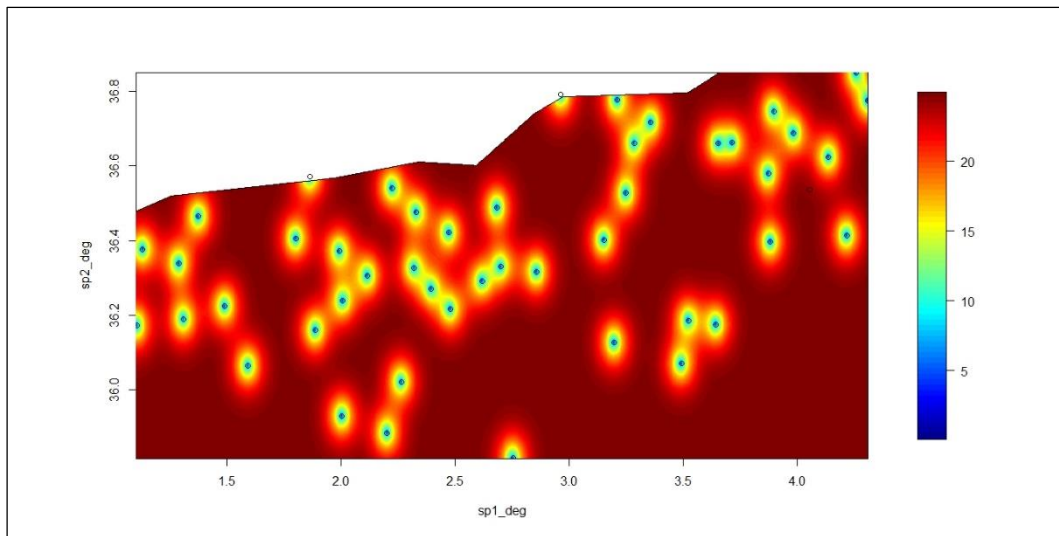


Figure 6.5 Variances d'interpolation sur les quantiles 10 ans pour  $d=1440$  minutes estimés par la méthode M0.

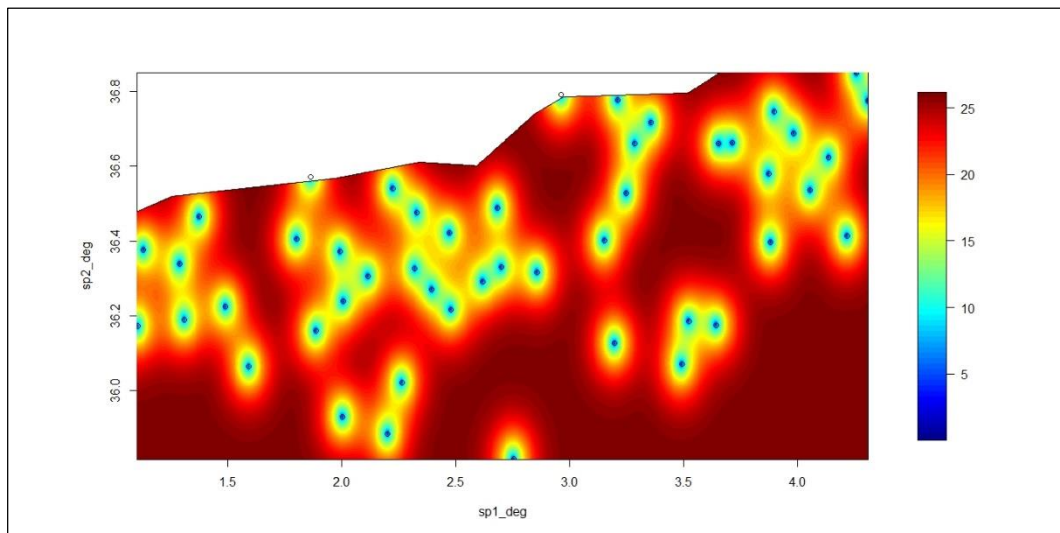


Figure 6.6 Variances d'interpolation sur les quantiles 10 ans pour  $d=1440$  minutes estimés par la méthode M1.

Par contre, l'examen des cartes des variances de Krigeage (figure 6.5 et 6.6), et donc d'incertitude sur l'interpolation spatiale, rend compte du fait que le proche voisinage des stations, où il est possible d'extrapoler avec une incertitude acceptable, est légèrement plus grand dans le modèle M1 que M0. Ceci s'explique par la cohérence spatiale apportée par l'inférence intégrée.

Toutefois, une vue d'ensemble sur les deux cartes (figure 6.5 et 6.6), montre que les incertitudes d'interpolation prennent leurs valeurs maximales (variance des échantillons observés) dès que l'on s'écarte des stations pluviométriques. Ce qui ne laisse aucun doute sur la fiabilité de cette approche d'interpolation spatiale des quantiles aux sites. En effet, une extrapolation avec une bonne précision nécessiterait un plus grand nombre de stations pluviométriques à considérer dans la procédure de Krigeage. Ce qui n'est cependant pas possible vu la faible densité du réseau de mesure pluviométrique de la région centrale du Nord de l'Algérie.

### 6.2.3 Intégration des incertitudes sur les quantiles et sur l'interpolation dans la cartographie des courbes IDF : modèle M0 vs modèle M1.

En l'absence d'autres solutions pour l'estimation des pluies de projets sur des zones non-jaugées, il est très hasardeux de considérer les valeurs de quantiles de pluie ainsi extrapolées. Une alternative sécuritaire consiste à combiner les incertitudes d'estimation (de l'inférence statistique) avec les incertitudes d'interpolation. Cette combinaison est la plus simple possible. Elle consiste à additionner les valeurs des bornes supérieures de l'intervalle de crédibilité à 70% des quantiles de périodes de retour considérées et les incertitudes d'interpolation. En posant une quantité de pluie de 200mm comme étant la valeur au-delà de laquelle les estimations,  $T=10$  ans et  $d=1440$  min., sont considérées comme invraisemblables, les cartes (Figure 6.7 et 6.8) combinant les deux sortes d'incertitudes suscitées pour les deux modèles M0 et M1 montrent des blancs dans l'interpolation spatiale.

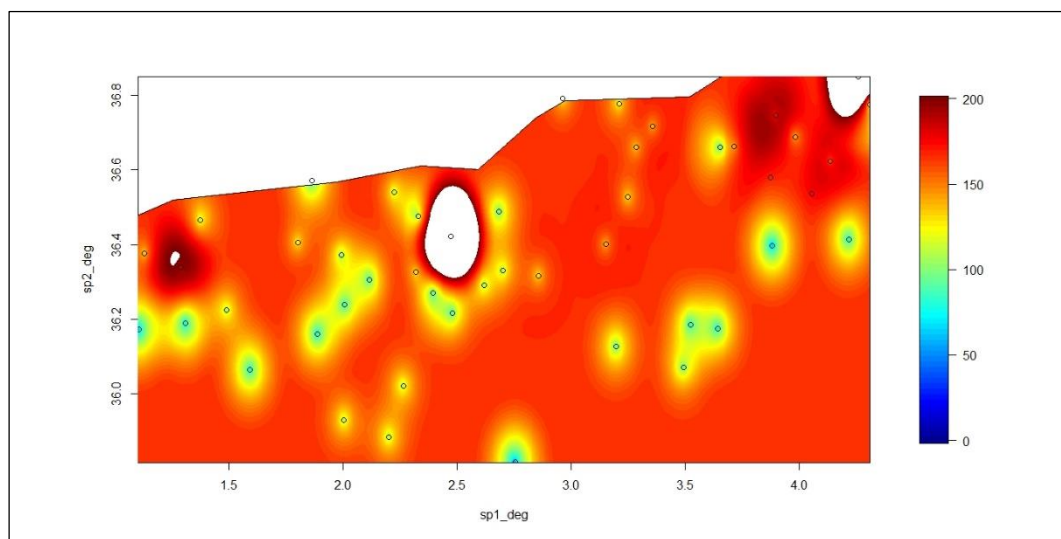


Figure 6.7 Cartographie des bornes supérieures des ICs à 70% combinées aux incertitudes d'interpolation. Période de retour 10ans pour  $d=1440$  minutes estimés par la méthode M0.

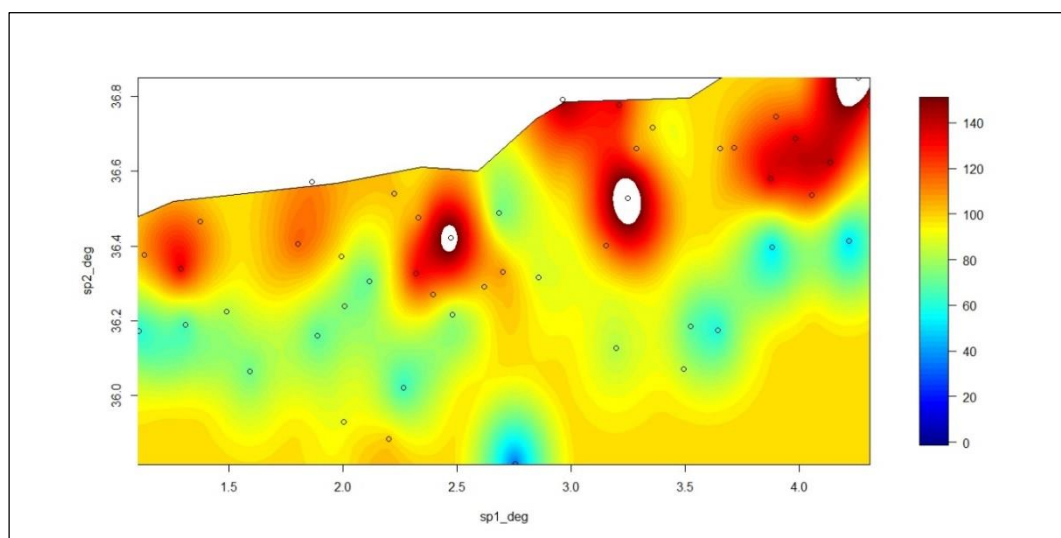


Figure 6.8 Cartographie des bornes supérieures des ICs à 70% combinées aux incertitudes d'interpolation. Période de retour 10ans pour  $d=1440$  minutes estimés par la méthode M1.

Ces vides dans les cartes signifient l'existence de valeurs supérieures à 200 mm, supposées être la quantité de pluie maximale plausible. Cette proposition de combiner les incertitudes bien qu'étant sécuritaire, n'est pas très efficace, particulièrement pour la méthode M0 qui fournit des estimations de quantiles avec des intervalles de crédibilité relativement plus large que la méthode M1. L'utilisation de la méthode M0 pour des interpolations de quantiles au site n'est pas du tout préconisée, tandis que pour le modèle M1, il n'est pas recommandé d'extrapoler sur des distances au-delà de la portée de variogramme des résidus de régression associé au Krigeage (de l'ordre de 30 kilomètres).

### 6.3 Cartographie des quantiles de pluies à partir d'estimation régionale

De nombreuses zones de la région d'étude sont dénuées d'une quelconque station de mesure pluviographique, vu l'usage limité des interpolations spatiales au site et la variabilité du relief du centre Nord de l'Algérie, ceci pose un problème conséquent aux praticiens dans les études d'ingénierie. Toutefois, l'application de procédures d'analyse fréquentielle régionale (Hosking & Wallis, 1997) permettant d'échanger de l'espace contre du temps, est une solution pleinement envisageable. L'analyse fréquentielle régionale traditionnelle sur les précipitations extrêmes consiste à mettre en œuvre une inférence statistique sur un échantillon conjoint composé d'observations (sur une même échelle de temps) provenant de différents sites, sous réserve d'homogénéité spatiale de la région. En second lieu, il s'agit d'interpoler par Krigeage, ou autre techniques, le paramètre de position (typiquement la moyenne ou la médiane) de la série de données de chaque station, et enfin, déduire le quantile de pluie – en multipliant le paramètre de position par le quantile régional adimensionnel – en chaque point de la grille ainsi formée et établir une cartographie des événements pluvieux extrêmes. Au regard des avantages tirés d'une intégration temporelle des observations de pluies maximales annuelles, tel que démontré au chapitre précédent, il est proposé une analyse fréquentielle régionale intégrant des mesures de différents sites mais aussi sur diverses durées. Elle revient donc à contraindre une distribution fréquentielle sur un très large échantillon d'une même variable mesurée dans l'espace et sur le temps. La plausibilité d'application d'une invariance d'échelle simple spatio-temporelle étant permise en assouplissant légèrement les critères d'homogénéité, a été démontré au chapitre 4. Dans un but comparatif, une analyse fréquentielle régionale traditionnelle sur les SMA, sur une seule durée à la fois, est aussi mise en œuvre dans ce qui vient. Les deux procédures régionales sont définies comme suit :

- a) Modèle M2 qui consiste à interpoler par Krigeage le paramètre de position local (de chaque station PG) puis inférer (Bayesian MCMC) les quantiles régionaux pour chaque durée séparément (intégration régionale) et finalement déduire les quantiles, sur la grille où un paramètre de position a été interpolé, de durées  $d$  et de période de retour  $T$ , afin d'établir différents jeux de cartes des précipitations extrêmes ;
- b) Modèle M3 qui consiste à interpoler par Krigeage le paramètre de position local sur une durée d'agrégation de 24h des stations PG, puis d'inférer (Bayesian MCMC) les quantiles régionaux par intégration totale (spatiale et temporelle), et finalement

déduire les quantiles de période de retour  $T$  et de durées  $d$ , sur chaque point de la grille interpolée, afin d'établir différents jeux de cartes des précipitations extrêmes.

### 6.3.1 Modèle M2 : Estimation Bayésienne régionale des courbes IDF.

La cartographie (figure 6.9) des pluies extrêmes sur la base d'estimations régionales des quantiles de période de retour  $T$  et de durée  $d$  est nettement supérieure aux approches basées sur l'interpolation des quantiles au site (figure 6.3 et 6.4). En effet, il y a plus de cohérence spatiale sur les quantiles de pluie, et les 3 régions homogènes sont clairement identifiables (voir chapitre 4). De plus, les incertitudes d'interpolation (figure 6.10) ont nettement été réduites comparativement à M0 et M1, ce qui rend envisageable des extrapolations sur une plus grande portée, et ainsi apporter des estimations relativement acceptables aux sites non-jaugés. A signaler que la spatialisation de la médiane, qui est plus robuste que la moyenne, contribue aussi à la supériorité de la cartographie régionale par rapport à la cartographie locale, car les valeurs de portées sont plus grandes (figure 6.2c)

Toutefois, une inspection plus approfondie de la carte produite sur la base de M2, permet de se rendre compte de quelques incohérences. En effet, la portion centre du littoral Méditerranéen (en région 2) prend des valeurs de quantiles plus proches de celles de la région 1, or il est avéré que cette zone n'est pas soumise au même régime climatique que cette même région 1. La faible densité des stations de mesures, voire leur inexistence, dans cette zone mène à des extrapolations qui ne sont pas très réalistes. De même, la bordure sud de la carte semble avoir un régime pluviométrique plus proche de celui de la région 2 (Méditerranée), là où il devrait être encore plus faible que celui de la région 1 en raison du gradient pluvial Nord-Sud (Méditerranée vers désert). Néanmoins, ce dernier constat s'explique facilement par l'absence de stations de mesures sur la bordure sud de région d'étude.

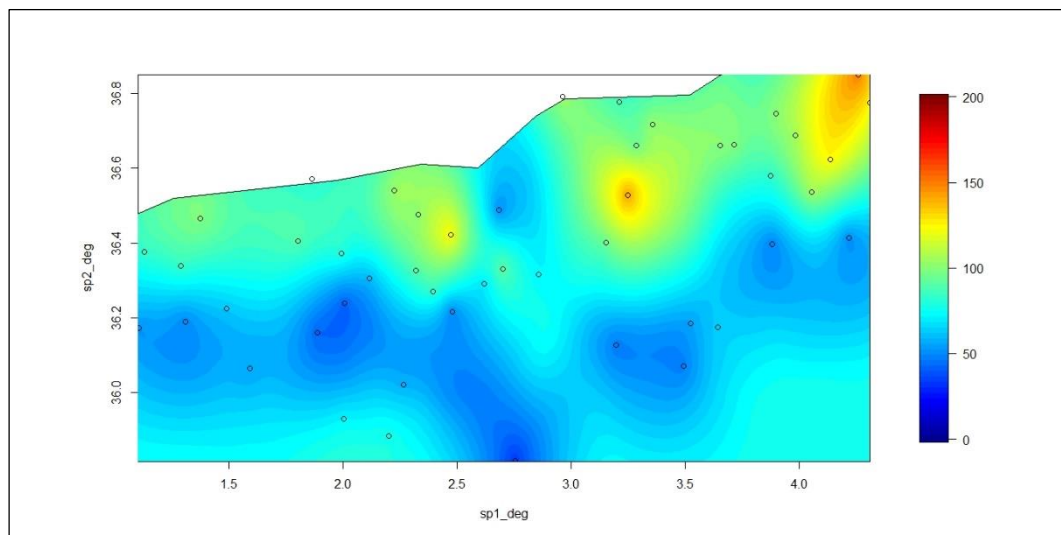


Figure 6.9 Cartographie des quantiles 10 ans sur  $d=1440$  minutes estimés par la méthode M2.

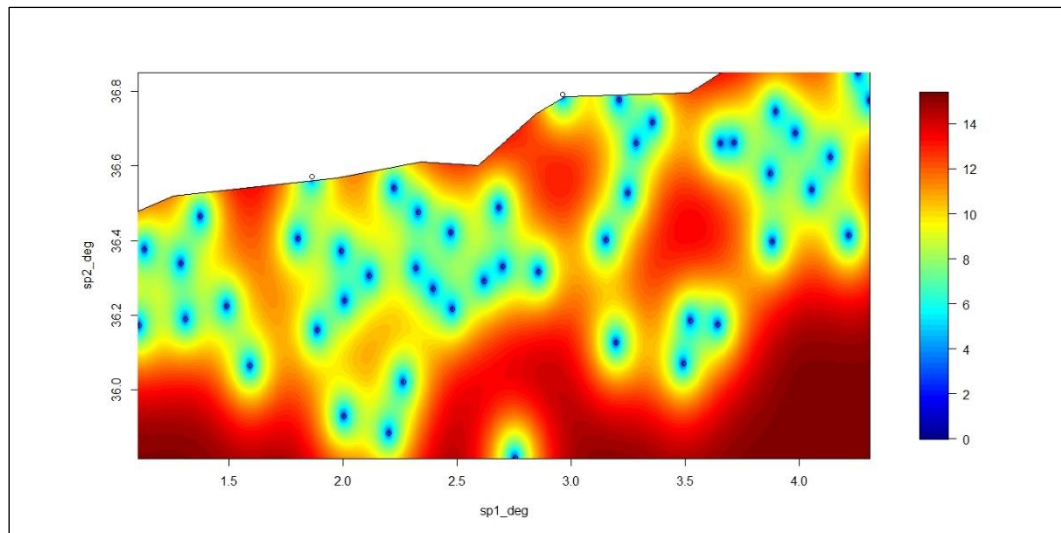


Figure 6.10 Variances d'interpolation sur les quantiles 10 ans pour  $d=1440$  minutes estimés par la méthode M2.

### 6.3.2 Modèle M3 : Estimation Bayésienne intégrée régionale des courbes IDF.

La figure 6.11 ci-après illustre la cartographie des extrêmes pluviométriques sur le modèle IDF Bayésien M3 calé sur les observations des pluies maximales sur diverses durées et pour différentes stations en une seule étape. Cette intégration totale améliore sensiblement les estimations au prix d'une homogénéité spatiale et temporelle définie, par exemple la sous-région A (figure 6.11), sur une superficie de 2000 km<sup>2</sup>. Cela semble assez réduit. Toutefois, les extrapolations à l'intérieur de cette sous-région A sont d'une grande précision au sens des très faibles incertitudes d'interpolation (voir figure 6.12).

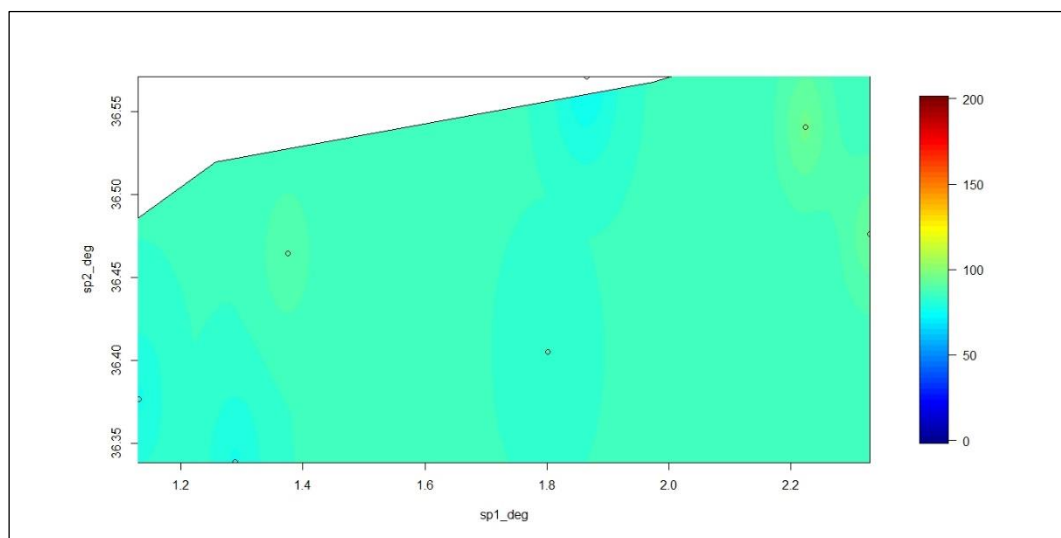


Figure 6.11 Cartographie des quantiles 10 ans sur  $d=1440$  minutes estimés par la méthode M3. Sous-région A (région 2).

Cependant, même si cette région est homogène, il est utile de vérifier de quelle manière la distribution régionale toute intégrée (M3) réoriente les distributions fréquentielles à intégration temporelle au site (modèle M1). Autrement dit, est-il possible de généraliser l'utilisation du

modèle M3, ou bien chaque application pratique nécessite une analyse particulière. Cette dernière question est plus amplement développée en section 6.4.

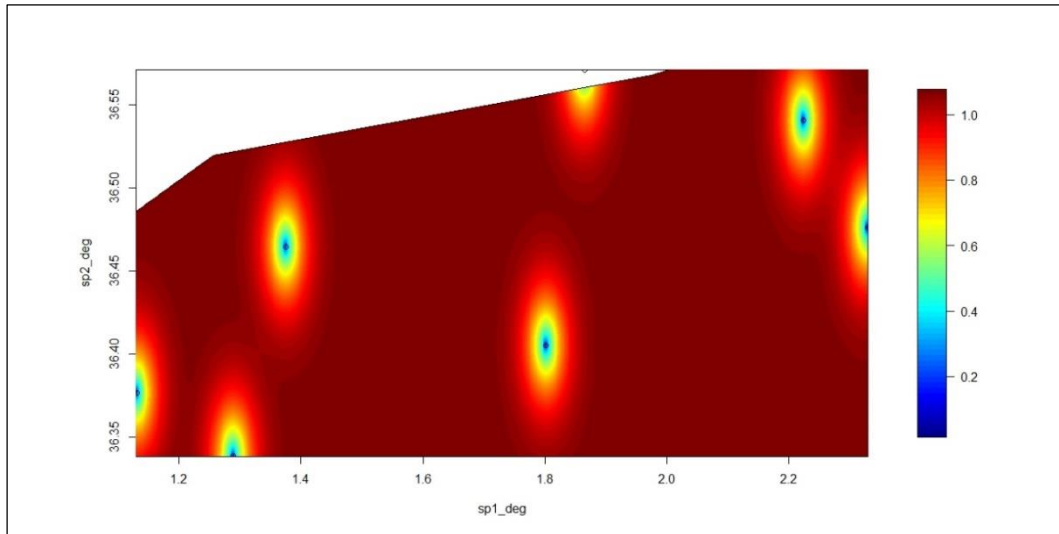


Figure 6.12 Variances d'interpolation sur les quantiles 10 ans pour  $d=1440$  minutes estimés par la méthode M3. Sous-région A (région 2).

### 6.3.3 Intégration des incertitudes sur les quantiles et sur l'interpolation dans la cartographie des courbes IDF : modèle M2 vs modèle M3.

L'inférence fréquentielle Bayésienne pour le modèle M3 offre une distribution à postériori sur un échantillon, très riche en informations spatiale et temporelle, qu'il est possible d'exploiter pour estimer les intervalles de crédibilité sur les quantiles régionaux. De même, l'interpolation spatiale par Krigeage des moyennes des pluies maximales sur 24 heures, fournit en chaque point de la grille interpolée une moyenne et une variance.

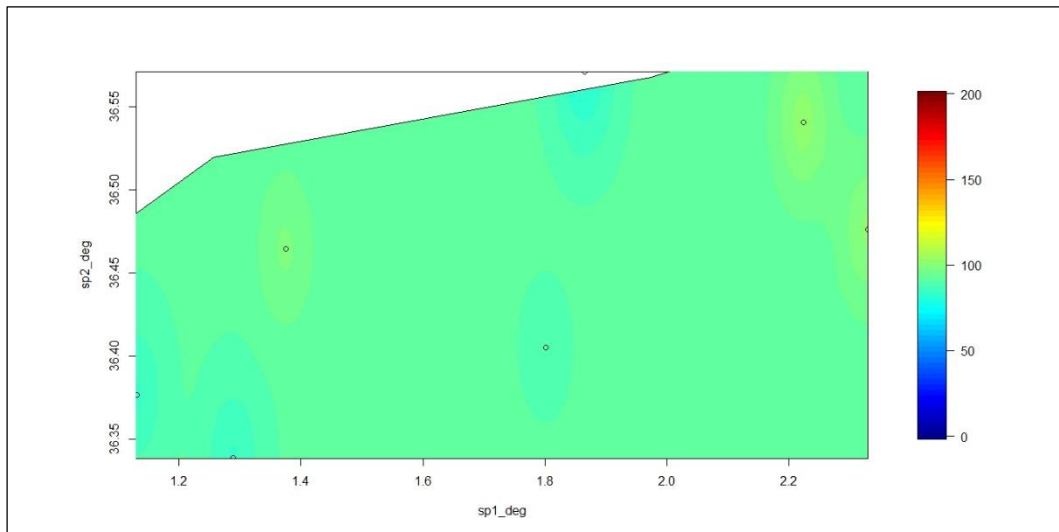


Figure 6.13 Cartographie de la borne supérieure des ICs à 70%, combinant incertitudes d'inférence statistique et d'interpolation spatiale, pour  $T = 10$  ans sur  $d=1440$  minutes estimés par la méthode M3. Sous-région A (région 2).

>Dans ce qui suit il est proposé une combinaison de ces caractéristiques statistiques afin de permettre une estimation efficiente d'intervalles de crédibilité à 70%, rendant compte des incertitudes sur les quantiles et sur l'interpolation, en chaque point de la grille interpolée. Cette combinaison d'incertitudes est obtenue selon les étapes suivantes :

1. Interpolation spatiale par Krigeage de la moyenne sur les séries de pluies maximales journalières (sur 24 heures) dans la région supposée homogène ;
2. En chaque point de la grille interpolée le Krigeage donne une valeur moyenne et la variance correspondante, ce qui constitue une loi normale et permet de faire un tirage aléatoire de 10000 valeurs de pluie moyenne maximale journalière en chacun des points de cette grille ;
3. Récupérer à partir de la distribution à posteriori, du modèle M3, 10 000 (par exemple le dernier 1/3 pour 30 000 itérations) valeurs de quantiles régionaux (adimensionnels) ;
4. Reconstituer un jeu de quantiles au niveau de chaque point de la grille interpolée en multipliant le résultat de l'étape 2 par celui de l'étape 3. Il est ainsi obtenu 10 000 valeurs quantiles en chaque point de la grille interpolée ;
5. Classer les jeux de quantiles ainsi obtenus pour évaluer les intervalles de crédibilité à 70%.

La figure 6.13 illustre l'interpolation spatiale des bornes supérieures d'intervalles de crédibilité combinant entre incertitudes d'estimation et d'interpolation, qui représente la solution sécuritaire, pour des extrapolations vers des sites non-jaugés, la plus efficiente qui soit.

## **6.4 Evaluation et validation des modèles de cartographie des courbes IDF.**

Comme signalé plus haut dans ce chapitre, les distributions fréquentielles régionales ajustées sur des échantillons intégrant des observations, de pluies maximales annuelles, dans l'espace et le temps peuvent fortement réorienter la distribution ajustée sur les observations (sur diverses durées : modèle M1) au site, particulièrement pour les durées horaires et infra-horaires. En effet, la figure 6.14 en est l'illustration parfaite. Ainsi, la distribution fréquentielle M1 de la station 33 de Azazga pour une durée  $d=60$  min n'est pas concordante avec la distribution régionale M3, et de même M2. Ceci s'explique par la présence, dans la même région homogène (sous-région D), de stations ayant observé des événements intenses sur de courtes durées et aboutissant ainsi à fortement réorienter vers le haut la distribution régionale M3. Au regard, des résultats d'ajustements M1 et M3 de la station 12 d'El Anneb, sous-région A, la problématique de la question précédemment posée est relativisée. En effet, la procédure M3 d'ajustement régional sur des observations intégrées dans l'espace et dans le temps peut être performante. Toutefois, son utilisation nécessite d'être évaluée pour chaque cas d'application. Ce qui, en connaissance de la forte variabilité d'échantillonnage et de la faible densité du réseau de mesures pluviographiques de la région d'étude, est tout à fait acceptable.

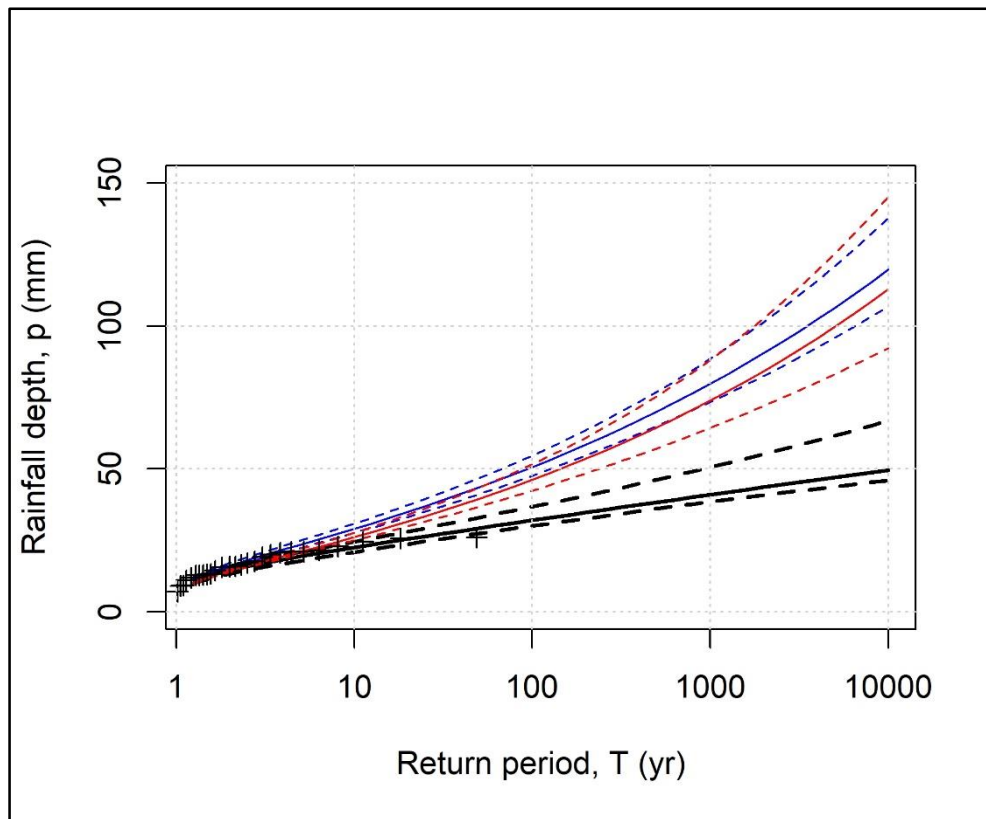


Figure 6.14 Ajustements des SMA pour  $d=60$  min. par le modèle M1 (station 33 : en noir), M2 (région 3, en rouge) et M3 (sous-région D en bleu)

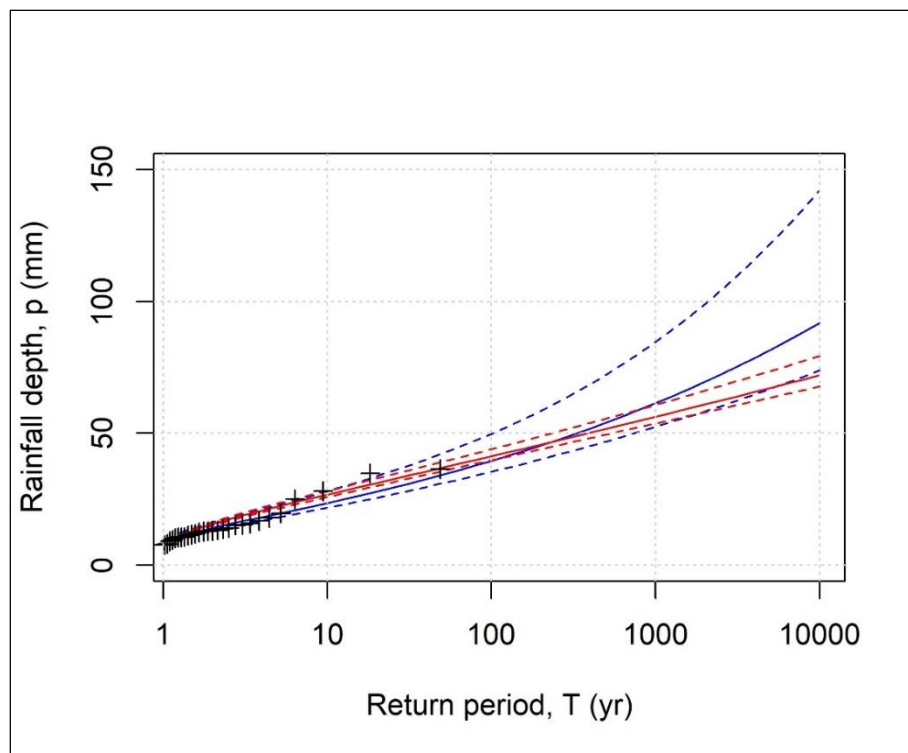


Figure 6.15 Ajustements des SMA pour  $d=60$  min. par le modèle M1 (station 12 : en bleu), et M3 (sous-région A en rouge)



## 6.5 Conclusion du chapitre 6

Le fait que les données observées peuvent présenter une tendance linéaire sur un papier de probabilité approprié, mais ne suivent pas exactement la courbe théorique à ajuster mène à la conviction qu'un évènement de période de retour donnée ne peut être représenté avec une confiance parfaite par la théorie de la probabilité (Chow, 1964). Cette dernière déclaration renseigne plus que tout sur la nécessité absolue de baser les conceptions et les analyses de risque sur les incertitudes d'estimation, ce qui n'est faisable que si elles sont correctement définies (de manière efficiente)

En effet, la meilleure approche de régionalisation, parmi celles étudiées, semble être celle combinant de manière efficiente les incertitudes de l'inférence statistique et celles de l'interpolation spatiale (modèle M3). C'est la borne supérieure de l'intervalle de crédibilité à 70%, résultant de cette fusion d'incertitudes, pour différents quantiles de pluie (i.e. sur diverses durées) qui est cartographiée. En raison des faibles incertitudes sur les estimations locales des pluies et de la grande cohérence spatiale des valeurs estimées, l'approche proposée paraît significativement surpasser l'approche traditionnelle même pour des estimations sur les sites non-jaugés. A noter que les cartographies des extrêmes pluviométriques sur la base du modèle IDF M0 ne sont pas du tout préconisées, et celles basées sur les modèles M1 et M2 ne sont recommandées que pour des extrapolations sur des distances inférieures aux portées des variogrammes sous-jacents.



## **Conclusion générale**

L'interaction entre les sciences climatiques et statistiques, a une portée extraordinaire en termes de recherches scientifiques toutefois, elle nécessite un grand engagement de la part des deux disciplines pour comprendre les nuances résultant des relations complexes entre les données observées et les problèmes du monde réel (Stein, 2019). En effet, ce travail de thèse s'est efforcé, tout au long de ses chapitres, d'identifier la modélisation des relations de courbes IDF permettant d'équilibrer précision, simplicité et robustesse, puis à valider son applicabilité localement et régionalement.

Les courbes intensité-durée-fréquence (IDF) représentent un des outils hydrologiques les plus couramment utilisés dans la gestion des eaux pluviales particulièrement en milieu urbain. Ainsi, il était nécessaire et utile de proposer un modèle autorisant des mises en œuvre sur des séries d'observations pluviographiques de nombre et de longueur limités. Finalement, ce travail a abouti sur une estimation Bayésienne Intégrée des courbes Intensité-Durée-Fréquence locales et régionales. Elle consiste en un calage, en une seule étape, d'un modèle IDF à quatre paramètres (3 pour la distribution GEV et 1 pour la relation d'échelle) sur un échantillon unique composé des maxima annuels de cumuls de pluies sur huit durées d'agrégation (15, 30, 60, 120, 180, 360, 720 et 1440 minutes). C'est l'établissement de la relation de la vraisemblance sur le modèle IDF qui a permis une estimation des paramètres en une seule fois (de façon intégrée), et sa maximisation à travers le calcul inférentiel Bayesian MCMC a permis d'incorporer les incertitudes dans le processus de calcul fournissant ainsi des intervalles de crédibilité les plus fiables possibles sur les paramètres et les quantiles correspondants.

En prélude à la mise en œuvre de l'estimation Bayésienne intégrée, il a fallu valider théoriquement le modèle IDF établi et vérifier sa concordance avec les propriétés des mesures pluviographiques (c'est-à-dire que les observations ne rejettent pas le modèle). Le modèle proposé correspond à une distribution fréquentielle réduite (adimensionnelle) sur les cumuls de pluies maximums annuels, multipliée par un facteur d'échelle. Cette dernière description concorde avec le principe d'invariance d'échelle simple, où la seule définition théoriquement cohérente du facteur d'échelle est : une relation d'exposant d'échelle linéaire dite autosimilaire (c'est-à-dire monofractale passant par l'origine). De même, l'acceptation de l'invariance d'échelle simple et de la structure fractale a été approuvée via l'analyse d'échelles des moments sur les SMA de cumuls de pluie et la vérification de l'homogénéité de l'échantillon conjoint.

Une fois le modèle IDF étudié dans cette thèse, a été calé, validé et testé, il est évalué la possibilité d'étendre son utilisation dans des analyses régionales qui permettraient des estimations des pluies horaires sur des zones non-jaugées. Dans un but de comparaison et de prospection de la meilleure approche régionale, quatre procédures différentes ont été proposées :

- a) Modèle M0 qui consiste à interpoler par Krigeage des quantiles de pluies estimés par inférence bayésienne pour chaque durée séparément au niveau de chaque poste PG,
- b) Modèle M1 qui consiste à interpoler par Krigeage des quantiles de pluies estimés par inférence Bayésienne de manière intégrée (intégration temporelle) au niveau de chaque poste PG ;

- c) Modèle M2 qui consiste à interpoler par Krigeage le paramètre de position local (de chaque station PG) puis inférer (Bayesian MCMC) les quantiles régionaux pour chaque durée séparément (intégration régionale) ;
- d) Modèle M3 qui consiste à interpoler par Krigeage le paramètre de position local sur une durée d'agrégation de 24h des stations PG, puis d'inférer (Bayesian MCMC) les quantiles régionaux par intégration totale (spatiale et temporelle).

## 7.1 Résultats obtenus

Les fondamentaux de la méthodologie proposée dans cette thèse, ont déjà été présentés dans quelques travaux antérieurs (Muller et al., 2008; Van de Vyver, 2015a), toutefois c'est la première fois, à notre connaissance, qu'elle a été validée et testée sur un jeu de données assez large et que son potentiel a été mis en exergue.

Outre la cohérence théorique du modèle IDF proposé, l'approche intégrée présente une nette cohérence avec les ajustements des maxima annuels des pluies, sur une durée à la fois (approche classique avec inférence Bayesian MCMC) pour l'ensemble des stations étudiées. Ce qui dénote d'un équilibre positif entre simplicité du modèle et précision d'estimation (ou statistiquement parlant entre biais et incertitudes). Cependant, il faut garder à l'esprit que les hypothèses d'invariance d'échelle simple et de comportement fractal des observations doivent impérativement être vérifiées pour chaque cas d'application. Toutefois, il est possible d'accepter un léger écart avec le modèle théorique. C'est-à-dire, que ce dernier n'est pas totalement rejeté par les données d'observations. En effet, quelques stations limitées montrent une concordance imparfaite (léger écart) avec l'hypothèse d'invariance d'échelle simple, particulièrement quand les pluies de durées de moins d'une heure sont considérées pour le calage, mais les estimations restent robustes.

En plus de la réduction des incertitudes internes, du fait d'une inférence statistique - sur l'ensemble de la relation IDF- en une étape de calcul (1 étape dans la méthode proposée contre 2 à 3 dans l'approche classique), les incertitudes d'estimations sur les quantiles sont significativement réduites comparativement à un calage du modèle IDF sur chaque durée séparément. Un autre résultat, qui était inattendu, est que cette approche apporte indubitablement de la robustesse, dans le sens où le processus inférentiel intégré appliqué sur des séries d'observation courtes (8 à 12 années) fournit des estimations raisonnables, ce qui représente un atout non négligeable pour des applications d'ingénierie. En plus du constat empirique, cette robustesse vis-à-vis de la petitesse de la taille des enregistrements, a été vérifiée sur quelques milliers de séries simulées (simulations Monte Carlo).

Concernant l'application régionale du modèle IDF Bayesian intégré, il est évident que le modèle M1 (inférence intégrée) est plus performant que le modèle M0 (inférence séparée). Cependant, la faible densité du réseau de mesure montre qu'il est assez hasardeux d'extrapoler, à travers une cartographie des quantiles de pluies pour une période de retour  $T$  et une durée  $d$ , le résultat vers des sites non-jaugés. En effet, l'analyse des cartes des variances, concernant le Krigeage ordinaire implémenté, montre que seul le proche voisinage

des stations autorise une extrapolation avec une incertitude minimale. Pour pallier cette problématique une première proposition de baser les conceptions et analyses du risque en zones non-jaugées, non pas sur les quantiles, mais sur la borne supérieure de l'intervalle de crédibilité à 70% majorée de l'incertitude sur le Krigeage. En dépit de la réduction de l'incertitude grâce à l'estimation Bayésienne intégrée, la valeur proposée reste assez grande, particulièrement pour le modèle M0, et n'est donc pas suffisamment efficace pour des applications d'ingénierie.

Le modèle IDF régional avec intégration spatio-temporelle (M3), c'est à dire sur un échantillon unique d'observations sur différents sites et sur diverses durées d'agrégation, moyennant un assouplissement (dans une certaine mesure) du critère d'homogénéité, semble permettre de surpasser l'approche traditionnelle même pour des estimations sur les sites non-jaugés. En effet, ce dernier fusionne entre les incertitudes d'estimation sur les quantiles et les incertitudes d'interpolation de manière efficace. Ce qui fait qu'en raison des faibles valeurs d'incertitudes sur les quantiles de pluie et de la grande cohérence spatiale des valeurs estimées, le modèle M3 paraît performant pour des applications de conception d'ouvrages hydrauliques et d'analyse du risque d'inondation. Toutefois, avant sa mise en œuvre, la concordance entre les ajustements locaux et l'ajustement régional doit impérativement être vérifiée. En effet, une trop grande variabilité d'échantillonnage (présence d'horsains dans certaines stations de la région supposée homogène) peut mener à fortement réorienter l'ajustement régional comparativement au local, même quand la région est déclarée parfaitement homogène.

## 7.2 Quelques perspectives

Il est impératif de garder à l'esprit que la dépendance entre les observations sur les diverses durées d'agrégation peut être difficilement prise en considération dans la procédure d'inférence. En testant le calage du modèle IDF sur seulement 4–3 durées supposées indépendantes les unes des autres, les intervalles de crédibilité sont effectivement réduits, mais dans une proportion moindre que le calage sur huit durées. Ainsi, pour un calage du modèle IDF sur toutes les durées, les intervalles de crédibilité résultant de la distribution à posteriori, devraient probablement être considérés comme sous-estimés. Muller et al. (2008) ont apportés une solution partielle à ce problème en prenant en considération les corrélations entre les séries de diverses durées dans l'écriture de la vraisemblance de leur modèle IDF, mais en aboutissant à une formulation à 9 paramètres difficilement ajustables, car la convergence de calcul par algorithme MCMC serait affectée par des complications supplémentaires. Aussi, la variance d'estimation augmente avec l'augmentation du nombre de paramètres à inférer. Le règlement de cette question de manière efficace est assez complexe, mais en contrepartie il améliorerait de façon certaine le calcul des intervalles de crédibilité.

Quand la densité du réseau de mesure et la longueur des enregistrements sont réduites, l'efficacité des méthodes statistiques est limitée (Faridzad et al., 2018). Ainsi, en dépit des nombreuses améliorations, mises en avant dans ce travail, dans l'estimation des courbes IDF

sur des séries assez courtes, la faible densité des réseaux de mesures pluviographiques reste quelque peu problématique. En effet, même si le modèle M3 de cartographie, des quantiles de pluies extrêmes issu d'une analyse fréquentielle régionale, est une solution relativement satisfaisante, il reste entaché d'une légère incohérence théorique. Cette dernière réside dans le fait que l'interpolation par Krigeage suppose un champ Gaussien, or les extrêmes pluviométriques s'adaptent mal à une distribution normale. Afin de pallier cette limitation, l'utilisation des processus max-stables semble assez prometteuse.

Aussi, il serait pertinent de mener une analyse comparative, sur l'estimation des débits de projets, sur la base des méthodologies, proposée et classique, dans la quantification des intensités de pluies. Même si les résultats des chapitres 4 et 5 montrent qu'en moyenne l'erreur, sur le coefficient  $b$  de Montana, entre une estimation Bayésienne Intégrée et les valeurs fournies par l'ANRH est de 30% et peut engendrer jusqu'à 50 % sur la valeur de l'intensité de la pluie. Il serait intéressant d'évaluer l'incidence financière en termes de conception d'ouvrages hydrauliques et d'études d'analyse du risque d'inondation.





## Liste des références bibliographiques

- ABH-AHS. (2002). *Le Bassin Algérois : Côtier Algérois, Hodna, Soummam* (Carnet de l'agence No. 1) (p. 37). Alger : Agence des Bassins Hydrographique Algérois, Hodna, Soummam.
- AF Rédaction, L. (2019, September 1). Skikda inondée / Les réseaux sociaux s'enflamment à la vue d'une ville noyée. Retrieved June 2, 2020, from <https://www.algerie-focus.com/2019/09/skikda-inondee-les-reseaux-sociaux-senflamment-a-la-vue-dune-ville-noyee/>
- Agilan, V., & Umamahesh, N. V. (2016). Is the covariate based non-stationary rainfall IDF curve capable of encompassing future rainfall changes? *Journal of Hydrology*, *541*, 1441–1455. <https://doi.org/10.1016/j.jhydrol.2016.08.052>
- Alexander, G. N. (1963). Using the probability of storm transposition for estimating the frequency of rare floods. *Journal of Hydrology*, *1*(1), 46–57. [https://doi.org/10.1016/0022-1694\(63\)90032-5](https://doi.org/10.1016/0022-1694(63)90032-5)
- Alila, Y. (2000). Regional rainfall depth-duration-frequency equations for Canada. *Water Resources Research*, *36*(7), 1767–1778. <https://doi.org/10.1029/2000WR900046>
- Anderson, T. W., & Darling, D. A. (1954). A Test of Goodness of Fit. *Journal of the American Statistical Association*, *49*(268), 765–769. <https://doi.org/10.1080/01621459.1954.10501232>
- Ang, A. H.-S., & Tang, W. H. (2006). *Probability Concepts in Engineering: Emphasis on Applications to Civil and Environmental Engineering* (2 edition). New York: Wiley.
- Arbani, M. (2016, September 30). Intempéries à Laghouat : Évacuation de 66 personnes encerclées par les eaux pluviales (Vidéo)(Actualisé). Retrieved June 2, 2020, from <https://www.algerie1.com/societe/intemperies-a-laghouat-evacuation-de-66-personnes-encerclées-par-les-eaux-pluviales-video-actualise>
- Arnaud, P., Lavabre, J., Sol, B., & Desouches, C. (2008). Régionalisation d'un générateur de pluies horaires sur la France métropolitaine pour la connaissance de l'aléa pluviographique / Regionalization of an hourly rainfall generating model over metropolitan France for flood hazard estimation. *Hydrological Sciences Journal*, *53*(1), 34–47. <https://doi.org/10.1623/hysj.53.1.34>
- Aronica, G. T., & Freni, G. (2005). Estimation of sub-hourly DDF curves using scaling properties of hourly and sub-hourly data at partially gauged site. *Atmospheric Research*, *77*(1–4), 114–123. <https://doi.org/10.1016/j.atmosres.2004.10.025>
- Bairwa, A. K., Khosa, R., & Maheswaran, R. (2016). Developing intensity duration frequency curves based on scaling theory using linear probability weighted moments: A case study from India. *Journal of Hydrology*, *542*, 850–859. <https://doi.org/10.1016/j.jhydrol.2016.09.056>
- Bara, M., Kohnova, S., Gaal, L., Szolgay, J., & Hlavčova, K. (2009). Estimation of IDF curves of extreme rainfall by simple scaling in Slovakia. *Contributions to Geophysics and Geodesy*, *39*(3), 187–206.
- Bárdossy, A., & Pegram, G. (2014). Infilling missing precipitation records – A comparison of a new copula-based method with other techniques. *Journal of Hydrology*, *519*, 1162–1170. <https://doi.org/10.1016/j.jhydrol.2014.08.025>
- Bárdossy, A., & Pegram, G. (2017). Combination of radar and daily precipitation data to estimate meaningful sub-daily point precipitation extremes. *Journal of Hydrology*, *544*, 397–406. <https://doi.org/10.1016/j.jhydrol.2016.11.039>

- Bárdossy, A., & Pegram, G. (2018). Intensity–duration–frequency curves exploiting neighbouring extreme precipitation data. *Hydrological Sciences Journal*, 63(11), 1593–1604. <https://doi.org/10.1080/02626667.2018.1524987>
- Beaulieu, C., Ouarda, T. B. M. J., & Seidou, O. (2007). Synthèse des techniques d’homogénéisation des séries climatiques et analyse d’applicabilité aux séries de précipitations. *Hydrological Sciences Journal*, 52(1), 18–37. <https://doi.org/10.1623/hysj.52.1.18>
- Beck, C., & Cohen, E. G. D. (2003). Superstatistics. *Physica A: Statistical Mechanics and Its Applications*, 322, 267–275. [https://doi.org/10.1016/S0378-4371\(03\)00019-0](https://doi.org/10.1016/S0378-4371(03)00019-0)
- Bendjoudi, H., Hubert, P., Schertzer, D., & Lovejoy, S. (1997). Interprétation multifractale des courbes intensité-durée-fréquence des précipitations. *Comptes Rendus de l’Académie Des Sciences-Series IIA-Earth and Planetary Science*, 325(5), 323–326.
- Benhattab, K., Bouvier, C., & Meddi, M. (2014). Analyse fréquentielle régionale des précipitations journalières maximales annuelles dans le bassin hydrographique - Chécliff, Algérie. *Revue des sciences de l’eau*, 27(3), 189. <https://doi.org/10.7202/1027805ar>
- Ben-Zvi, A. (2009). Rainfall intensity–duration–frequency relationships derived from large partial duration series. *Journal of Hydrology*, 367(1–2), 104–114. <https://doi.org/10.1016/j.jhydrol.2009.01.007>
- Bernard, E., Naveau, P., Vrac, M., & Mestre, O. (2013). Clustering of Maxima: Spatial Dependencies among Heavy Rainfall in France. *Journal of Climate*, 26(20), 7929–7937. <https://doi.org/10.1175/JCLI-D-12-00836.1>
- Bernard, M. M. (1932). Formulas For Rainfall Intensities of Long Duration. *Transactions of the American Society of Civil Engineers*, 96(1), 592–606.
- Bernier, J., & Veron, R. (1964). Sur quelques difficultés rencontrées dans l’estimation d’un débit de crue de probabilité donnée. *Revue de Statistique Appliquée*, 12(1), 25–48.
- Blanchet, J., Ceresetti, D., Molinié, G., & Creutin, J.-D. (2016). A regional GEV scale-invariant framework for Intensity–Duration–Frequency analysis. *Journal of Hydrology*, 540, 82–95. <https://doi.org/10.1016/j.jhydrol.2016.06.007>
- Borga, M., & Vizzaccaro, A. (1997). On the interpolation of hydrologic variables: formal equivalence of multiquadratic surface fitting and kriging. *Journal of Hydrology*, 195(1), 160–171. [https://doi.org/10.1016/S0022-1694\(96\)03250-7](https://doi.org/10.1016/S0022-1694(96)03250-7)
- Borga, M., Vezzani, C., & Fontana, G. D. (2005). Regional Rainfall Depth–Duration–Frequency Equations for an Alpine Region. *Natural Hazards*, 36(1–2), 221–235. <https://doi.org/10.1007/s11069-004-4550-y>
- Bouceffiane, A., & Meddi, M. (2019). Regional growth curves and extreme precipitation events estimation in the steppe area of northwestern Algeria. *Atmósfera*, 32(4), 287–303. <https://doi.org/10.20937/ATM.2019.32.04.03>
- Bougadis, J., & Adamowski, K. (2006). Scaling model of a rainfall intensity-duration-frequency relationship. *Hydrological Processes*, 20(17), 3747–3757. <https://doi.org/10.1002/hyp.6386>
- Boukhelifa, M., Meddi, M., & Gaume, E. (2018). Integrated Bayesian Estimation of Intensity-Duration-Frequency Curves: Consolidation and Extensive Testing of a Method. *Water Resources Research*, 54(10), 7459–7477. <https://doi.org/10.1029/2018WR023366>
- Boukhelifa, M., Touaibia, B., & Hubert, P. (2011). Prévention du risque pluvial par l’élaboration des courbes intensité–durée–fréquence (IDF): application à la ville de Tipasa dans le Nord Ouest Algérien. *Risk in Water Resources Management IAHS PUB*, 347, 36–42.
- Brath, A., Castellarin, A., & Montanari, A. (2003). Assessing the reliability of regional depth-duration-frequency equations for gaged and ungaged sites: ASSESSING THE RELIABILITY OF REGIONAL DDF EQUATIONS. *Water Resources Research*, 39(12). <https://doi.org/10.1029/2003WR002399>
- Buishand, T. A. (1991). Extreme rainfall estimation by combining data from several sites. *Hydrological Sciences Journal*, 36(4), 345–365. <https://doi.org/10.1080/02626669109492519>
- Burboa, A., Vargas, J., & Meier, C. I. (2020). PluvioReader: A software for digitizing weekly siphoning-type pluviograph strip charts. *Computers & Geosciences*, 104463. <https://doi.org/10.1016/j.cageo.2020.104463>
- Burlando, P., & Rosso, R. (1996). Scaling and multiscaling models of depth-duration-frequency curves for storm precipitation. *Journal of Hydrology*, 187(1–2), 45–64.
- Burn, D. H. (1990a). An appraisal of the “region of influence” approach to flood frequency analysis. *Hydrological Sciences Journal*, 35(2), 149–165. <https://doi.org/10.1080/02626669009492415>
- Burn, D. H. (1990b). Evaluation of regional flood frequency analysis with a region of influence approach. *Water Resources Research*, 26(10), 2257–2265. <https://doi.org/10.1029/WR026i010p02257>
- Burn, D. H. (2014). A framework for regional estimation of intensity-duration-frequency (IDF) curves: REGIONAL ESTIMATION OF INTENSITY-DURATION-FREQUENCY (IDF) CURVES. *Hydrological Processes*, 28(14), 4209–4218. <https://doi.org/10.1002/hyp.10231>
- Cannon, A. J. (2018). Non-crossing nonlinear regression quantiles by monotone composite quantile regression neural network, with application to rainfall extremes. *Stochastic Environmental Research and Risk Assessment*, 32(11), 3207–3225. <https://doi.org/10.1007/s00477-018-1573-6>

- Cannon, A. J., & Innocenti, S. (2019). Projected intensification of sub-daily and daily rainfall extremes in convection-permitting climate model simulations over North America: implications for future intensity–duration–frequency curves. *Natural Hazards and Earth System Sciences*, 19(2), 421–440. <https://doi.org/10.5194/nhess-19-421-2019>
- Caporali, E., Cavigli, E., & Petrucci, A. (2008). The index rainfall in the regional frequency analysis of extreme events in Tuscany (Italy). *Environmetrics*, 19(7), 714–724. <https://doi.org/10.1002/env.949>
- Caporali, E., Chiarello, V., & Petrucci, A. (2018). Regional frequency analysis and geosadditive modeling for design storm estimates in the Arno river basin (Italy). *Environmental and Ecological Statistics*, 25(1), 31–52. <https://doi.org/10.1007/s10651-018-0399-1>
- Casas-Castillo, M. C., Llabrés-Brustenga, A., Rius, A., Rodríguez-Solà, R., & Navarro, X. (2018). A single scaling parameter as a first approximation to describe the rainfall pattern of a place: application on Catalonia. *Acta Geophysica*, 66(3), 415–424. <https://doi.org/10.1007/s11600-018-0122-5>
- CERTU. (2003). *La ville et son assainissement* (CEREMA (ex CERTU)). Retrieved from <https://www.eyrolles.com/BTP/Livre/la-ville-et-son-assainissement-9782110940834/>
- Chandra, R., Saha, U., & Mujumdar, P. P. (2015). Model and parameter uncertainty in IDF relationships under climate change. *Advances in Water Resources*, 79, 127–139. <https://doi.org/10.1016/j.advwatres.2015.02.011>
- Chang, K. B., Lai, S. H., & Othman, F. (2016). Comparison of Annual Maximum and Partial Duration Series for Derivation of Rainfall Intensity-Duration-Frequency Relationships in Peninsular Malaysia. *Journal of Hydrologic Engineering*, 21(1), 05015013. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001262](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001262)
- Chebana, F., & Ouarda, T. B. M. J. (2007). Multivariate *L*-moment homogeneity test: MULTIVARIATE L-MOMENT HOMOGENEITY TEST. *Water Resources Research*, 43(8). <https://doi.org/10.1029/2006WR005639>
- Chebana, Fateh, Ouarda, T. B. M. J., Bruneau, P., Barbet, M., El Adlouni, S., & Latraverse, M. (2009). Multivariate homogeneity testing in a northern case study in the province of Quebec, Canada. *Hydrological Processes*, 23(12), 1690–1700. <https://doi.org/10.1002/hyp.7304>
- Cheng, L., & AghaKouchak, A. (2015). Nonstationary Precipitation Intensity-Duration-Frequency Curves for Infrastructure Design in a Changing Climate. *Scientific Reports*, 4(1). <https://doi.org/10.1038/srep07093>
- Chow, V., Maidment, D., & Mays, L. (1988). *Applied Hydrology* (1 edition). New York: McGraw-Hill Science/Engineering/Math.
- Chow, V. T. (1964). *Handbook of applied hydrology: a compendium of water-resources technology*. McGraw-Hill.
- Coles, S., & Pericchi, L. (2003). Anticipating catastrophes through extreme value modelling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 52(4), 405–416. <https://doi.org/10.1111/1467-9876.00413>
- Coles, S., Pericchi, L. R., & Sisson, S. (2003). A fully probabilistic approach to extreme rainfall modeling. *Journal of Hydrology*, 273(1), 35–50.
- D'Agostino, R. B., & Stephens, M. A. (1986). *Goodness-of-Fit-Techniques* (1st ed., Vol. 68). New York: CRC Press.
- Dai, Q., Bray, M., Zhuo, L., Islam, T., & Han, D. (2017). A Scheme for Rain Gauge Network Design Based on Remotely Sensed Rainfall Measurements. *Journal of Hydrometeorology*, 18(2), 363–379. <https://doi.org/10.1175/JHM-D-16-0136.1>
- Dalrymple, T. (1960). *Flood-frequency analyses, Manual of Hydrology: Part 3* (USGS Numbered Series No. 1543- A). U.S. G.P.O., Retrieved from <http://pubs.er.usgs.gov/publication/wsp1543A>
- Das, S. (2019). Extreme rainfall estimation at ungauged sites: Comparison between region-of-influence approach of regional analysis and spatial interpolation technique. *International Journal of Climatology*, 39(1), 407–423. <https://doi.org/10.1002/joc.5819>
- Davison, A. C., Padoan, S. A., & Ribatet, M. (2012). Statistical Modeling of Spatial Extremes. *Statistical Science*, 27(2), 161–186. <https://doi.org/10.1214/11-STS376>
- Davison, A. C., Huser, R., & Thibaud, E. (2013). Geostatistics of Dependent and Asymptotically Independent Extremes. *Mathematical Geosciences*, 45(5), 511–529. <https://doi.org/10.1007/s11004-013-9469-y>
- De Luca, D., & Galasso, L. (2018). Stationary and Non-Stationary Frameworks for Extreme Rainfall Time Series in Southern Italy. *Water*, 10(10), 1477. <https://doi.org/10.3390/w10101477>
- De Michele, C., Zenoni, E., Pecora, S., & Rosso, R. (2011). Analytical derivation of rain intensity–duration–area–frequency relationships from event maxima. *Journal of Hydrology*, 399(3–4), 385–393. <https://doi.org/10.1016/j.jhydrol.2011.01.018>
- De Michele, Carlo, & Avanzi, F. (2018). Superstatistical distribution of daily precipitation extremes: A worldwide assessment. *Scientific Reports*, 8(1), 14204. <https://doi.org/10.1038/s41598-018-31838-z>
- De Michele, Carlo, Kottegoda, N. T., & Rosso, R. (2001). The derivation of areal reduction factor of storm rainfall from its scaling properties. *Water Resources Research*, 37(12), 3247–3252. <https://doi.org/10.1029/2001WR000346>
- Debbarna, N., Choudhury, P., & Roy, P. (2019). Identification of homogeneous rainfall regions using a genetic algorithm involving multi-criteria decision making techniques. *Water Supply*, 19(5), 1491–1499. <https://doi.org/10.2166/ws.2019.018>

- Di Baldassarre, G., Castellarin, A., & Brath, A. (2006). Relationships between statistics of rainfall extremes and mean annual precipitation: an application for design-storm estimation in northern central Italy. *Hydrology and Earth System Sciences*, 10(4), 589–601. <https://doi.org/10.5194/hess-10-589-2006>
- Di Baldassarre, G., Brath, A., & Montanari, A. (2006). Reliability of different depth-duration-frequency equations for estimating short-duration design storms. *Water Resources Research*, 42(12). <https://doi.org/10.1029/2006wr004911>
- Diggle, P., & Ribeiro, P. J. (2007). *Model-based geostatistics*. New York, NY: Springer.
- Dolšak, D., Bezak, N., & Šraj, M. (2016). Temporal characteristics of rainfall events under three climate types in Slovenia. *Journal of Hydrology*, 541, 1395–1405. <https://doi.org/10.1016/j.jhydrol.2016.08.047>
- Durrans, S. R. (2013). Intensity-Duration-Frequency Curves. In *Rainfall: State of the Science* (pp. 159–169). American Geophysical Union (AGU). <https://doi.org/10.1029/2009GM000919>
- Durrans, S. R., & Kirby, J. T. (2004). Regionalization of extreme precipitation estimates for the Alabama rainfall atlas. *Journal of Hydrology*, 295(1–4), 101–107. <https://doi.org/10.1016/j.jhydrol.2004.02.021>
- Efron, B., & Tibshirani, R. (1986). Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy. *Statistical Science*, 1(1), 54–75. <https://doi.org/10.1214/ss/1177013815>
- Efron, Bradley. (2005). *Bayesians, Frequentists, and Scientists* (Technical No. 2005–1B/230). Stanford: Division of Biostatistics, Stanford University.
- El Adlouni, S., & Ouarda, T. B. M. J. (2010). Frequency analysis of extreme rainfall events. In F. Y. Testik & M. Gebremichael (Eds.), *Rainfall: State of the Science* (pp. 171–188). Washington, D. C.: American Geophysical Union. <https://doi.org/10.1029/2010GM000976>
- Endreny, T. A., & Imbeah, N. (2009). Generating robust rainfall intensity–duration–frequency estimates with short-record satellite data. *Journal of Hydrology*, 371(1–4), 182–191. <https://doi.org/10.1016/j.jhydrol.2009.03.027>
- Endreny, T. A., & Pashiardis, S. (2007). The error and bias of supplementing a short, arid climate, rainfall record with regional vs. global frequency analysis. *Journal of Hydrology*, 334(1–2), 174–182. <https://doi.org/10.1016/j.jhydrol.2006.10.005>
- Farizad, M., Yang, T., Hsu, K., Sorooshian, S., & Xiao, C. (2018). Rainfall frequency analysis for ungauged regions using remotely sensed precipitation information. *Journal of Hydrology*, 563, 123–142. <https://doi.org/10.1016/j.jhydrol.2018.05.071>
- Faulkner, D. S., & Prudhomme, C. (1998). Mapping an index of extreme rainfall across the UK. *Hydrology and Earth System Sciences*, 2(2/3), 183–194. <https://doi.org/10.5194/hess-2-183-1998>
- Ferraris, L., Gabellani, S., Rebora, N., & Provenzale, A. (2003). A comparison of stochastic models for spatial rainfall downscaling: RAINFALL DOWNSCALING MODELS. *Water Resources Research*, 39(12). <https://doi.org/10.1029/2003WR002504>
- Fisher, R. A., & Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24(2), 180–190. <https://doi.org/10.1017/S0305004100015681>
- Fisher, Ronald Aymler. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 222(594–604), 309–368. <https://doi.org/10.1098/rsta.1922.0009>
- Fontaine, T. A., & Potter, K. W. (1989). Estimating Probabilities of Extreme Rainfalls. *Journal of Hydraulic Engineering*, 115(11), 1562–1575. [https://doi.org/10.1061/\(ASCE\)0733-9429\(1989\)115:11\(1562\)](https://doi.org/10.1061/(ASCE)0733-9429(1989)115:11(1562))
- François, B., Schlef, K. E., Wi, S., & Brown, C. M. (2019). Design considerations for riverine floods in a changing climate – A review. *Journal of Hydrology*, 574, 557–573. <https://doi.org/10.1016/j.jhydrol.2019.04.068>
- Fuller, W. E. (1914). Flood Flows. *Transactions of the American Society of Civil Engineers*, 77, 567–617.
- Gado, T. A., Hsu, K., & Sorooshian, S. (2017). Rainfall frequency analysis for ungauged sites using satellite precipitation products. *Journal of Hydrology*, 554, 646–655. <https://doi.org/10.1016/j.jhydrol.2017.09.043>
- Ganguli, P., & Coulibaly, P. (2017). Does nonstationarity in rainfall require nonstationary intensity–duration–frequency curves? *Hydrology and Earth System Sciences*, 21(12), 6461–6483. <https://doi.org/10.5194/hess-21-6461-2017>
- García-Marín, A. P., Jiménez-Hornero, F. J., & Ayuso-Muñoz, J. L. (2008). Multifractal analysis as a tool for validating a rainfall model. *Hydrological Processes*, 22(14), 2672–2688. <https://doi.org/10.1002/hyp.6864>
- García-Marín, A. P., Ayuso-Muñoz, J. L., Jiménez-Hornero, F. J., & Estévez, J. (2013). Selecting the best IDF model by using the multifractal approach: IDF MODELS AND MULTIFRACTALITY. *Hydrological Processes*, 27(3), 433–443. <https://doi.org/10.1002/hyp.9272>
- García-Marín, A. P., Estévez, J., Sangüesa-Pool, C., Pizarro-Tapia, R., Ayuso-Muñoz, J. L., & Jimenez-Hornero, F. J. (2015). The use of the exponent  $K(q)$  function to delimit homogeneous regions in regional frequency analysis of extreme annual daily rainfall: MULTIFRACTAL PROPERTIES OF RAINFALL TO FORM HOMOGENEOUS REGIONS IN RFA. *Hydrological Processes*, 29(1), 139–151. <https://doi.org/10.1002/hyp.10284>

- García-Marín, Amanda P., Morbidelli, R., Saltalippi, C., Cifrodelli, M., Estévez, J., & Flammini, A. (2019). On the choice of the optimal frequency analysis of annual extreme rainfall by multifractal approach. *Journal of Hydrology*, 575, 1267–1279. <https://doi.org/10.1016/j.jhydrol.2019.06.013>
- Gargouri-Ellouze, E., & Chebchoub, A. (2008). Modélisation de la structure de dépendance hauteur—durée d'événements pluvieux par la copule de Gumbel. *Hydrological Sciences Journal*, 53(4), 802–817. <https://doi.org/10.1623/hysj.53.4.802>
- Gaume, E., Mouhous, N., & Andrieu, H. (2007). Rainfall stochastic disaggregation models: Calibration and validation of a multiplicative cascade model. *Advances in Water Resources*, 30(5), 1301–1319. <https://doi.org/10.1016/j.advwatres.2006.11.007>
- Gaume, E., Gaál, L., Viglione, A., Szolgay, J., Kohnová, S., & Blöschl, G. (2010). Bayesian MCMC approach to regional flood frequency analyses involving extraordinary flood events at ungauged sites. *Journal of Hydrology*, 394(1–2), 101–117. <https://doi.org/10.1016/j.jhydrol.2010.01.008>
- Gaume, Eric. (2018). Flood frequency analysis: The Bayesian choice. *Wiley Interdisciplinary Reviews: Water*, e1290. <https://doi.org/10.1002/wat2.1290>
- Gaume, Eric, Sivakumar, B., Kolasinski, M., & Hazoumé, L. (2006). Identification of chaos in rainfall temporal disaggregation: Application of the correlation dimension method to 5-minute point rainfall series measured with a tipping bucket and an optical raingage. *Journal of Hydrology*, 328(1–2), 56–64. <https://doi.org/10.1016/j.jhydrol.2005.11.034>
- Gaume, Eric, Bain, V., Bernardara, P., Newinger, O., Barbuc, M., Bateman, A., et al. (2009). A compilation of data on European flash floods. *Journal of Hydrology*, 367(1–2), 70–78. <https://doi.org/10.1016/j.jhydrol.2008.12.028>
- Gaume, Eric, Gaál, L., Viglione, A., Szolgay, J., Kohnová, S., & Blöschl, G. (2010). Bayesian MCMC approach to regional flood frequency analyses involving extraordinary flood events at ungauged sites. *Journal of Hydrology*, 394(1–2), 101–117. <https://doi.org/10.1016/j.jhydrol.2010.01.008>
- Ghanmi, H., Bargaoui, Z., & Mallet, C. (2016). Estimation of intensity-duration-frequency relationships according to the property of scale invariance and regionalization analysis in a Mediterranean coastal area. *Journal of Hydrology*, 541, 38–49. <https://doi.org/10.1016/j.jhydrol.2016.07.002>
- Gnedenko, B. (1943). Sur La Distribution Limite Du Terme Maximum D'Une Serie Aleatoire. *Annals of Mathematics*, 44(3), 423–453. <https://doi.org/10.2307/1968974>
- Goudenhoofd, E., Delobbe, L., & Willems, P. (2017). Regional frequency analysis of extreme rainfall in Belgium based on radar estimates. *Hydrology and Earth System Sciences*, 21(10), 5385–5399. <https://doi.org/10.5194/hess-21-5385-2017>
- Greenwood, J. A., Landwehr, J. M., Matalas, N. C., & Wallis, J. R. (1979). Probability weighted moments: Definition and relation to parameters of several distributions expressible in inverse form. *Water Resources Research*, 15(5), 1049–1054. <https://doi.org/10.1029/WR015i005p01049>
- Grimaldi, S., Kao, S.-C., Castellarin, A., Papalexiou, S. M., Viglione, A., Laio, F., et al. (2011). Statistical Hydrology. In *Treatise on Water Science* (Vol. 2, pp. 479–517). Elsevier. <https://doi.org/10.1016/B978-0-444-53199-5.00046-4>
- Grundmann, J., Hörning, S., & Bárdossy, A. (2019). Stochastic reconstruction of spatio-temporal rainfall patterns by inverse hydrologic modelling. *Hydrology and Earth System Sciences*, 23(1), 225–237. <https://doi.org/10.5194/hess-23-225-2019>
- Gumbel, E. J. (1958). *Statistics of extremes*. New York: Columbia University Press.
- Gupta, V. K., & Waymire, E. (1990). Multiscaling properties of spatial rainfall and river flow distributions. *Journal of Geophysical Research: Atmospheres*, 95(D3), 1999–2009.
- Habibi, B., Meddi, M., & Bouceffiane, A. (2013). Analyse fréquentielle des pluies journalières maximales Cas du Bassin Chott-Chergui. . C, 8.
- Haddad, K., Rahman, A., & Green, J. (2011). Design rainfall estimation in Australia: a case study using L moments and Generalized Least Squares Regression. *Stochastic Environmental Research and Risk Assessment*, 25(6), 815–825. <https://doi.org/10.1007/s00477-010-0443-7>
- Haddad, Khaled, & Rahman, A. (2014). Derivation of short-duration design rainfalls using daily rainfall statistics. *Natural Hazards*, 74(3), 1391–1401. <https://doi.org/10.1007/s11069-014-1248-7>
- Haddad, Khaled, Johnson, F., Rahman, A., Green, J., & Kuczera, G. (2015). Comparing three methods to form regions for design rainfall statistics: Two case studies in Australia. *Journal of Hydrology*, 527, 62–76. <https://doi.org/10.1016/j.jhydrol.2015.04.043>
- Haktanir, T., Cobaner, M., & Kisi, O. (2010). Frequency analyses of annual extreme rainfall series from 5 min to 24 h. *Hydrological Processes*, 24(24), 3574–3588. <https://doi.org/10.1002/hyp.7759>
- Halbert, K., Nguyen, C. C., Payrastra, O., & Gaume, E. (2016). Reducing uncertainty in flood frequency analyses: A comparison of local and regional approaches involving information on extreme historical floods. *Journal of Hydrology*, 541, 90–98. <https://doi.org/10.1016/j.jhydrol.2016.01.017>

- Herath, S. M., Sarukkalgige, P. R., & Nguyen, V. T. V. (2016). A spatial temporal downscaling approach to development of IDF relations for Perth airport region in the context of climate change. *Hydrological Sciences Journal*, 61(11), 2061–2070. <https://doi.org/10.1080/02626667.2015.1083103>
- Hershfield, D. M. (1961). Estimating the Probable Maximum Precipitation. *Journal of the Hydraulics Division*, 87(5), 99–116.
- Hosking, J. R. M. (1990). L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics. *Journal of the Royal Statistical Society. Series B (Methodological)*, 52(1), 105–124.
- Hosking, J. R. M., & Wallis, J. R. (1988). The effect of intersite dependence on regional flood frequency analysis. *Water Resources Research*, 24(4), 588–600. <https://doi.org/10.1029/WR024i004p00588>
- Hosking, J. R. M., & Wallis, J. R. (1993). Some statistics useful in regional frequency analysis. *Water Resources Research*, 29(2), 271–281. <https://doi.org/10.1029/92WR01980>
- Hosking, J. R. M., & Wallis, J. R. (1997). *Regional Frequency analysis: an approach based on L-moments*. New York: Cambridge University Press.
- Hu, G., Bóday, T., & Lucarini, V. (2019). Effects of stochastic parametrization on extreme value statistics: *Chaos*, 29(8). <https://doi.org/doi.org/10.1063/1.5095756>
- Hu, L., Nikolopoulos, E. I., Marra, F., & Anagnostou, E. N. (2020). Sensitivity of flood frequency analysis to data record, statistical model, and parameter estimation methods: An evaluation over the contiguous United States. *Journal of Flood Risk Management*, 13(1). <https://doi.org/10.1111/jfr3.12580>
- Huard, D., Mailhot, A., & Duchesne, S. (2010). Bayesian estimation of intensity–duration–frequency curves and of the return period associated to a given rainfall event. *Stochastic Environmental Research and Risk Assessment*, 24(3), 337–347. <https://doi.org/10.1007/s00477-009-0323-1>
- Huser, R., Dombry, C., Ribatet, M., & Genton, M. G. (2019). Full likelihood inference for max-stable data. *Stat*, 8(1). <https://doi.org/10.1002/sta4.218>
- Iliopoulou, T., & Koutsoyiannis, D. (2020). Projecting the future of rainfall extremes: better classic than trendy. *Journal of Hydrology*, 125005. <https://doi.org/10.1016/j.jhydrol.2020.125005>
- Iliopoulou, T., Koutsoyiannis, D., & Montanari, A. (2018). Characterizing and Modeling Seasonality in Extreme Rainfall. *Water Resources Research*, 54(9), 6242–6258. <https://doi.org/10.1029/2018WR023360>
- Ilorme, F., & Griffis, V. W. (2013). A novel procedure for delineation of hydrologically homogeneous regions and the classification of ungauged sites for design flood estimation. *Journal of Hydrology*, 492, 151–162. <https://doi.org/10.1016/j.jhydrol.2013.03.045>
- Innocenti, S., Mailhot, A., & Frigon, A. (2017). Simple scaling of extreme precipitation in North America. *Hydrology and Earth System Sciences*, 21(11), 5823–5846. <https://doi.org/10.5194/hess-21-5823-2017>
- Institute of Hydrology. (1999). *Flood estimation Handbook* (Vol. Vol. 1-5). Wallingford: UK Centre for Ecology & Hydrology.
- ISL-BRGM. (2005). *Etude de réduction de la vulnérabilité du massif de Bouzaréah aux catastrophes naturelles. Rapport de tâche 2. Volume 2 - Risque d'inondation* (Technique No. ISL R3114) (p. 36). Alger.
- Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, 81(348), 158–171. <https://doi.org/10.1002/qj.49708134804>
- Johnson, F., Hutchinson, M. F., The, C., Beesley, C., & Green, J. (2016). Topographic relationships for design rainfalls over Australia. *Journal of Hydrology*, 533, 439–451. <https://doi.org/10.1016/j.jhydrol.2015.12.035>
- Johnson, R. A., & Wichern, D. W. (2007). *Applied multivariate statistical analysis* (6th ed). Upper Saddle River, N.J: Pearson Prentice Hall.
- Kang, B., Kim, E., Kim, J., & Moon, S. (2019). Comparative Study on Spatiotemporal Characteristics of Fixed-Area and Storm-Centered ARFs. *Journal of Hydrologic Engineering*, 24(10), 04019044. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001839](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001839)
- Katz, R. W., Parlange, M. B., & Naveau, P. (2002). Statistics of extremes in hydrology. *Advances in Water Resources*, 25(8–12), 1287–1304. [https://doi.org/10.1016/S0309-1708\(02\)00056-8](https://doi.org/10.1016/S0309-1708(02)00056-8)
- Kebaili Bargaoui, Z., & Chebbi, A. (2009). Comparison of two kriging interpolation methods applied to spatiotemporal rainfall. *Journal of Hydrology*, 365(1), 56–73. <https://doi.org/10.1016/j.jhydrol.2008.11.025>
- Keifer, j. c., & Chu, h. h. (1957). synthetic storm patterns for drainage design. *Journal of Hydraulic Division*, 89(4).
- Khelfi, M. E. A., Touaibia, B., & Guastaldi, E. (2017). Regionalisation of the “intensity-duration-frequency” curves in Northern Algeria. *Arabian Journal of Geosciences*, 10(20), 441. <https://doi.org/10.1007/s12517-017-3214-7>
- Kieffer Weisse, A. (1998). *Etude des précipitations exceptionnelles de pas de temps court en relief accidenté (Alpes Françaises): Méthode de cartographie des précipitations extrêmes* (Ph. D. de mécanique des milieux géophysiques et Environnement). Institut National Polytechnique de Grenoble, Grenoble.

- Kieffer-Weisse, A., & Bois, P. (2001). Topographic Effects on Statistical Characteristics of Heavy Rainfall and Mapping in the French Alps. *JOURNAL OF APPLIED METEOROLOGY*, 40, 21.
- Kim, T.-J., Kwon, H.-H., & Lima, C. (2018). A Bayesian partial pooling approach to mean field bias correction of weather radar rainfall estimates: Application to Osungsan weather radar in South Korea. *Journal of Hydrology*, 565, 14–26. <https://doi.org/10.1016/j.jhydrol.2018.07.082>
- Kingumbi, A., & Mailhot, A. (2010). Courbes Intensité–Durée–Fréquence (IDF): comparaison des estimateurs des durées partielles et des maximums annuels. *Hydrological Sciences Journal*, 55(2), 162–176. <https://doi.org/10.1080/02626660903545995>
- Kirby, W. (1974). Algebraic boundedness of sample statistics. *Water Resources Research*, 10(2), 220–222. <https://doi.org/10.1029/WR010i002p00220>
- Kitanidis, P. K. (1997). *Introduction to Geostatistics: Applications in Hydrogeology*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511626166>
- Kite, G. W. (1975). Confidence limits for design events. *Water Resources Research*, 11(1), 48–53. <https://doi.org/10.1029/WR011i001p00048>
- Klemes, V. (1993). Probability of extreme hydrometeorological events - a different approach. In Z. W. Kundzewicz, D. Rosbjerg, S. P. Simonovic, & K. Takeuchi (Eds.), *Extreme Hydrological Events: Precipitation, Floods and Droug* (Vol. 213, pp. 167–176). Yokohama, Japan: IAHS Publ.
- Kondratieva, T., & Amarchi, H. (2015). Régionalisation des précipitations journalières extrêmes : cas de la région située au Nord-Est de l'Algérie. *Hydrological Sciences Journal*, 60(3), 498–507. <https://doi.org/10.1080/02626667.2014.988154>
- Koutsoyiannis, D. (1999). A probabilistic view of hershfield's method for estimating probable maximum precipitation. *Water Resources Research*, 35(4), 1313–1322. <https://doi.org/10.1029/1999WR900002>
- Koutsoyiannis, D. (2004a). Statistics of extremes and estimation of extreme rainfall: I. Theoretical investigation / Statistiques de valeurs extrêmes et estimation de précipitations extrêmes: I. Recherche théorique. *Hydrological Sciences Journal*, 49(4). <https://doi.org/10.1623/hysj.49.4.575.54430>
- Koutsoyiannis, D. (2004b). Statistics of extremes and estimation of extreme rainfall: II. Empirical investigation of long rainfall records / Statistiques de valeurs extrêmes et estimation de précipitations extrêmes: II. Recherche empirique sur de longues séries de précipitations. *Hydrological Sciences Journal*, 49(4). <https://doi.org/10.1623/hysj.49.4.591.54424>
- Koutsoyiannis, D. (2019). Knowable moments for high-order stochastic characterization and modelling of hydrological processes. *Hydrological Sciences Journal*, 64(1), 19–33. <https://doi.org/10.1080/02626667.2018.1556794>
- Koutsoyiannis, D., & Langousis, A. (2011). Precipitation. In P. Wilderer (Ed.), *Treatise on Water Science* (Vol. 2, pp. 27–77). Oxford: Elsevier. <https://doi.org/10.1016/B978-0-444-53199-5.00027-0>
- Koutsoyiannis, D., Kozonis, D., & Manetas, A. (1998). A mathematical framework for studying rainfall intensity-duration-frequency relationships. *Journal of Hydrology*, 206(1–2), 118–135.
- Koutsoyiannis, D., Dimitriadis, P., Lombardo, F., & Stevens, S. (2018). From Fractals to Stochastics: Seeking Theoretical Consistency in Analysis of Geophysical Data. In A. A. Tsonis (Ed.), *Advances in Nonlinear Geosciences* (pp. 237–278). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-319-58895-7\\_14](https://doi.org/10.1007/978-3-319-58895-7_14)
- Kuczera, G. (1999). Comprehensive at-site flood frequency analysis using Monte Carlo Bayesian inference. *Water Resources Research*, 35(5), 1551–1557. <https://doi.org/10.1029/1999WR900012>
- Kuichling, E. (1889). The relation between the rainfall and the discharge of sewers in populous district. *Transactions of the American Society of Civil Engineers*, 20, 1–56.
- Kyselý, J., Pícek, J., & Huth, R. (2007a). Formation of homogeneous regions for regional frequency analysis of extreme precipitation events in the Czech Republic. *Studia Geophysica et Geodaetica*, 51(2), 327–344. <https://doi.org/10.1007/s11200-007-0018-3>
- Kyselý, J., Gaál, L., & Pícek, J. (2011). Comparison of regional and at-site approaches to modelling probabilities of heavy precipitation. *International Journal of Climatology*, 31(10), 1457–1472. <https://doi.org/10.1002/joc.2182>
- Laborde, J. P. (1986). Pour une approche géostatistique des coefficients d'abattement des précipitations. *La Houille Blanche*, (3), 221–228. <https://doi.org/10.1051/lhb/1986021>
- Laio, F. (2004). Cramer-von Mises and Anderson-Darling goodness of fit tests for extreme value distributions with unknown parameters: Goodness of fit tests for extreme value distributions. *Water Resources Research*, 40(9). <https://doi.org/10.1029/2004WR003204>
- Laio, F., Di Baldassarre, G., & Montanari, A. (2009). Model selection techniques for the frequency analysis of hydrological extremes: MODEL SELECTION CRITERIA. *Water Resources Research*, 45(7). <https://doi.org/10.1029/2007WR006666>

- Landwehr, J. M., Matalas, N. C., & Wallis, J. R. (1979). Probability weighted moments compared with some traditional techniques in estimating Gumbel Parameters and quantiles. *Water Resources Research*, *15*(5), 1055–1064. <https://doi.org/10.1029/WR015i005p01055>
- Lang, M., Ouarda, T. B. M. J., & Bobée, B. (1999). Towards operational guidelines for over-threshold modeling. *Journal of Hydrology*, *225*(3–4), 103–117. [https://doi.org/10.1016/S0022-1694\(99\)00167-5](https://doi.org/10.1016/S0022-1694(99)00167-5)
- Langousis, A., & Veneziano, D. (2007). Intensity-duration-frequency curves from scaling representations of rainfall: SCALING REPRESENTATIONS OF RAINFALL. *Water Resources Research*, *43*(2). <https://doi.org/10.1029/2006WR005245>
- Langousis, A., Veneziano, D., Furcolo, P., & Lepore, C. (2009). Multifractal rainfall extremes: Theoretical analysis and practical estimation. *Chaos, Solitons & Fractals*, *39*(3), 1182–1194. <https://doi.org/10.1016/j.chaos.2007.06.004>
- Libertino, A., Allamano, P., Laio, F., & Claps, P. (2018). Regional-scale analysis of extreme precipitation from short and fragmented records. *Advances in Water Resources*, *112*, 147–159. <https://doi.org/10.1016/j.advwatres.2017.12.015>
- Lima, C. H. R., Kwon, H.-H., & Kim, J.-Y. (2016). A Bayesian beta distribution model for estimating rainfall IDF curves in a changing climate. *Journal of Hydrology*, *540*, 744–756. <https://doi.org/10.1016/j.jhydrol.2016.06.062>
- Lima, C. H. R., Kwon, H.-H., & Kim, Y.-T. (2018). A local-regional scaling-invariant Bayesian GEV model for estimating rainfall IDF curves in a future climate. *Journal of Hydrology*, *566*, 73–88. <https://doi.org/10.1016/j.jhydrol.2018.08.075>
- de Lima, M. I. P., & de Lima, J. L. M. P. (2009). Investigating the multifractality of point precipitation in the Madeira archipelago. *Nonlinear Processes in Geophysics*, *16*(2), 299–311. <https://doi.org/10.5194/npg-16-299-2009>
- Lloyd-Davies, D. E. (1906). The elimination of storm water from sewerage systems. In *Proceedings of Institution of Civil Engineers* (Vol. 164, pp. 41–67).
- Lombardo, F., Volpi, E., Koutsoyiannis, D., & Papalexiou, S. M. (2014). Just two moments! A cautionary note against use of high-order moments in multifractal models in hydrology. *Hydrology and Earth System Sciences*, *18*(1), 243–255. <https://doi.org/10.5194/hess-18-243-2014>
- Lombardo, F., Volpi, E., Koutsoyiannis, D., & Serinaldi, F. (2017). A theoretically consistent stochastic cascade for temporal disaggregation of intermittent rainfall: DISAGGREGATION OF INTERMITTENT RAINFALL. *Water Resources Research*, *53*(6), 4586–4605. <https://doi.org/10.1002/2017WR020529>
- Lovejoy, S., & Schertzer, D. (2013). *The Weather and Climate: Emergent Laws and Multifractal Cascades*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139093811>
- Lu, L.-H., & Stedinger, J. R. (1992). Sampling variance of normalized GEV/PWM quantile estimators and a regional homogeneity test. *Journal of Hydrology*, *138*(1–2), 223–245. [https://doi.org/10.1016/0022-1694\(92\)90166-S](https://doi.org/10.1016/0022-1694(92)90166-S)
- Lubes, H., & Masson, J. M. (1991). Méthode des moments de probabilité pondérés : application à la loi de Jenkinson. *Hydrologie Continentale*, *6*(1), 67–84.
- Ly, S., Charles, C., & Degré, A. (2011). Geostatistical interpolation of daily rainfall at catchment scale: the use of several variogram models in the Ourthe and Ambleve catchments, Belgium. *Hydrology and Earth System Sciences*, *15*(7), 2259–2274. <https://doi.org/10.5194/hess-15-2259-2011>
- Madsen, H., Rosbjerg, D., & HARREMOËS, P. (1993). Application of the partial duration series approach in the analysis of extreme rainfalls. In *Extreme Hydrological Events: Precipitation, Floods and Droughts* (Vol. 213, pp. 257–266). Yokohama, Japan: IAHS Publ.
- Madsen, H., Rasmussen, P. F., & Rosbjerg, D. (1997). Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events: 1. At-site modeling. *Water Resources Research*, *33*(4), 747–757. <https://doi.org/10.1029/96WR03848>
- Madsen, H., Pearson, C. P., & Rosbjerg, D. (1997). Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events: 2. Regional modeling. *Water Resources Research*, *33*(4), 759–769. <https://doi.org/10.1029/96WR03849>
- Madsen, H., Arnbjerg-Nielsen, K., & Mikkelsen, P. S. (2009). Update of regional intensity–duration–frequency curves in Denmark: Tendency towards increased storm intensities. *Atmospheric Research*, *92*(3), 343–349. <https://doi.org/10.1016/j.atmosres.2009.01.013>
- Mailhot, A., Duchesne, S., Caya, D., & Talbot, G. (2007). Assessment of future change in intensity–duration–frequency (IDF) curves for Southern Quebec using the Canadian Regional Climate Model (CRCM). *Journal of Hydrology*, *347*(1–2), 197–210. <https://doi.org/10.1016/j.jhydrol.2007.09.019>
- Mandelbrot, B. B. (1974). Intermittent turbulence in self-similar cascades: divergence of high moments and dimension of the carrier. *Journal of Fluid Mechanics*, *62*(2), 331–358. <https://doi.org/10.1017/S0022112074000711>
- Marani, M. (2003). On the correlation structure of continuous and discrete point rainfall: ON THE CORRELATION STRUCTURE OF POINT RAINFALL. *Water Resources Research*, *39*(5). <https://doi.org/10.1029/2002WR001456>



- Marani, M., & Ignaccolo, M. (2015). A metastatistical approach to rainfall extremes. *Advances in Water Resources*, 79, 121–126. <https://doi.org/10.1016/j.advwatres.2015.03.001>
- Marquínez, J., Lastra, J., & García, P. (2003). Estimation models for precipitation in mountainous regions: the use of GIS and multivariate analysis. *Journal of Hydrology*, 270(1–2), 1–11. [https://doi.org/10.1016/S0022-1694\(02\)00110-5](https://doi.org/10.1016/S0022-1694(02)00110-5)
- Marra, F., & Morin, E. (2015). Use of radar QPE for the derivation of Intensity–Duration–Frequency curves in a range of climatic regimes. *Journal of Hydrology*, 531, 427–440. <https://doi.org/10.1016/j.jhydrol.2015.08.064>
- Marra, F., Morin, E., Peleg, N., Mei, Y., & Anagnostou, E. N. (2017). Intensity–duration–frequency curves from remote sensing rainfall estimates: comparing satellite and weather radar over the eastern Mediterranean. *Hydrology and Earth System Sciences*, 21(5), 2389–2404. <https://doi.org/10.5194/hess-21-2389-2017>
- Marra, F., Nikolopoulos, E. I., Anagnostou, E. N., & Morin, E. (2018). Metastatistical Extreme Value analysis of hourly rainfall from short records: Estimation of high quantiles and impact of measurement errors. *Advances in Water Resources*, 117, 27–39. <https://doi.org/10.1016/j.advwatres.2018.05.001>
- Martins, E. S., & Stedinger, J. R. (2000). Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resources Research*, 36(3), 737–744. <https://doi.org/10.1029/1999WR900330>
- Matheron, G. (1971). *The Theory of Regionalized Variables and Its Application* (Vol. 5). Paris (France): Centre de morphologie mathématique, Ecole Nationale Supérieure des Mines de Paris. Retrieved from <https://books.google.dz/books?id=1ygBMwEACAAJ>
- McCuen, R. H. (2001). Generalized Flood Skew: Map versus Watershed Skew. *Journal of Hydrologic Engineering*, 6(4), 293–299. [https://doi.org/10.1061/\(ASCE\)1084-0699\(2001\)6:4\(293\)](https://doi.org/10.1061/(ASCE)1084-0699(2001)6:4(293))
- McCuen, R. H. (2004). *Hydrologic Analysis and Design* (3rd ed.). Upper Saddle River, N.J: Prentice Hall.
- McGraw, D., Nikolopoulos, E. I., Marra, F., & Anagnostou, E. N. (2019). Precipitation frequency analyses based on radar estimates: An evaluation over the contiguous United States. *Journal of Hydrology*, 573, 299–310. <https://doi.org/10.1016/j.jhydrol.2019.03.032>
- Mebarki, A. (2003). CARTOGRAPHIE AUTOMATIQUE DES PRECIPITATIONS : APPLICATION A L'EST ALGERIEN. *Sciences & Technologie B*, 20, 100–107.
- Meddi, M., & Toumi, S. (2015a). Spatial variability and cartography of maximum annual daily rainfall under different return periods in Northern Algeria. *Journal of Mountain Science*, 12(6), 1403–1421.
- Meddi, M., Toumi, S., & Assani, A. A. (2017). Application of the L-moments approach to the analysis of regional flood frequency in Northern Algeria. *International Journal of Hydrology Science and Technology*, 7(1), 26.
- Meersmans, J., Van Weverberg, K., De Baets, S., De Ridder, F., Palmer, S. J., van Wesemael, B., & Quine, T. A. (2016). Mapping mean total annual precipitation in Belgium, by investigating the scale of topographic control at the regional scale. *Journal of Hydrology*, 540, 96–105. <https://doi.org/10.1016/j.jhydrol.2016.06.013>
- Mélèse, V., Blanchet, J., & Molinié, G. (2018). Uncertainty estimation of Intensity–Duration–Frequency relationships: A regional analysis. *Journal of Hydrology*, 558, 579–591. <https://doi.org/10.1016/j.jhydrol.2017.07.054>
- Mélèse, V., Blanchet, J., & Creutin, J. (2019). A Regional Scale-Invariant Extreme Value Model of Rainfall Intensity-Duration-Area-Frequency Relationships. *Water Resources Research*, 2018WR024368. <https://doi.org/10.1029/2018WR024368>
- Menabde, M., Seed, A., & Pegram, G. (1999). A simple scaling model for extreme rainfall. *Water Resources Research*, 35(1), 355–339.
- Météo France. (2020). COMPRENDRE LE CLIMAT par Météo-France. Retrieved August 16, 2020, from <http://météofrance.com/comprendre-climat>
- Meylan, P., & Musy, A. (1999). *Hydrologie fréquentielle* (Première). Bucarest: \*H\*C\*A\*.
- Mohyont, B., Demarée, G. R., & Faka, D. N. (2004). Establishment of IDF-curves for precipitation in the tropical area of Central Africa - comparison of techniques and results. *Natural Hazards and Earth System Sciences*, 4(3), 375–387. <https://doi.org/10.5194/nhess-4-375-2004>
- Mohyont, Bernard, & Demarée, G. R. (2006). Courbes intensité—durée—fréquence des précipitations à Yangambi, Congo, au moyen de différents modèles de type Montana. *Hydrological Sciences Journal*, 51(2), 239–253.
- Molnar, P., & Burlando, P. (2005). Preservation of rainfall properties in stochastic disaggregation by a simple random cascade model. *Atmospheric Research*, 77(1–4), 137–151. <https://doi.org/10.1016/j.atmosres.2004.10.024>
- Mondal, A., & Daniel, D. (2019). Return Levels under Nonstationarity: The Need to Update Infrastructure Design Strategies. *Journal of Hydrologic Engineering*, 24(1), 04018060. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001738](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001738)
- Mondal, A., & Mujumdar, P. P. (2015). Modeling non-stationarity in intensity, duration and frequency of extreme rainfall over India. *Journal of Hydrology*, 521, 217–231. <https://doi.org/10.1016/j.jhydrol.2014.11.071>

- Montanari, A. (2011). Uncertainty of Hydrological Predictions. In *Treatise on Water Science* (Vol. 2, pp. 459–478). Elsevier. <https://doi.org/10.1016/B978-0-444-53199-5.00045-2>
- Morbidelli, R., Saltalippi, C., Flammini, A., Cifrodelli, M., Picciafuoco, T., Corradini, C., et al. (2017). Effect of temporal aggregation on the estimate of annual maximum rainfall depths for the design of hydraulic infrastructure systems. *Journal of Hydrology*, 554, 710–720. <https://doi.org/10.1016/j.jhydrol.2017.09.050>
- Muller, Aurélie, Bacro, J.-N., & Lang, M. (2008). Bayesian comparison of different rainfall depth–duration–frequency relationships. *Stochastic Environmental Research and Risk Assessment*, 22(1), 33–46. <https://doi.org/10.1007/s00477-006-0095-9>
- Muller, Aurelie, Arnaud, P., Lang, M., & Lavabre, J. (2009). Uncertainties of extreme rainfall quantiles estimated by a stochastic rainfall model and by a generalized Pareto distribution / Incertitudes des quantiles extrêmes de pluie estimés par un modèle stochastique d’averses et par une loi de Pareto généralisée. *Hydrological Sciences Journal*, 54(3), 417–429. <https://doi.org/10.1623/hysj.54.3.417>
- Mulvaney, T. J. (1850). On the use of self-registering rain and flood gauges in making observations on the relation of rainfall and flood discharges in a given catchment. *Transactions of Institution of Civil Engineers*, 4(2)(18).
- Nerantzaki, S. D., & Papalexiou, S. M. (2019). Tails of extremes: Advancing a graphical method and harnessing big data to assess precipitation extremes. *Advances in Water Resources*, 134, 103448. <https://doi.org/10.1016/j.advwatres.2019.103448>
- Ng, J. L., Abd Aziz, S., Huang, Y. F., Mirzaei, M., Wayayok, A., & Rowshon, M. K. (2019). Uncertainty analysis of rainfall depth duration frequency curves using the bootstrap resampling technique. *Journal of Earth System Science*, 128(5), 113. <https://doi.org/10.1007/s12040-019-1154-1>
- Nguyen, C. C., Gaume, E., & Payraastre, O. (2014). Regional flood frequency analyses involving extraordinary flood events at ungauged sites: further developments and validations. *Journal of Hydrology*, 508, 385–396. <https://doi.org/10.1016/j.jhydrol.2013.09.058>
- Nhat, L. M., Tachikawa, Y., Sayama, T., & Takara, K. (2007). A simple scaling characteristics of rainfall in time and space to derive intensity duration frequency relationships. *PROCEEDINGS OF HYDRAULIC ENGINEERING*, 51, 73–78.
- ObservAlgerie.com. (2018, December 14). Inondation : Oued Chiffa déborde sur la route nationale N1 à Blida [Réseau social]. Retrieved from <https://www.facebook.com/observalgerie/posts/2106836416005336>
- Olsson, J., & Berndtsson, R. (1998). Temporal rainfall disaggregation based on scaling properties. *Water Science and Technology*, 37(11), 73–79. <https://doi.org/10.2166/wst.1998.0439>
- Olsson, Jonas, & Burlando, P. (2002). Reproduction of temporal scaling by a rectangular pulses rainfall model. *Hydrological Processes*, 16(3), 611–630. <https://doi.org/10.1002/hyp.307>
- Olsson, Jonas, Södling, J., Berg, P., Wern, L., & Eronn, A. (2019). Short-duration rainfall extremes in Sweden: a regional analysis. *Hydrology Research*, 50(3), 945–960. <https://doi.org/10.2166/nh.2019.073>
- Ombadi, M., Nguyen, P., Sorooshian, S., & Hsu, K. (2018). Developing Intensity-Duration-Frequency (IDF) Curves From Satellite-Based Precipitation: Methodology and Evaluation. *Water Resources Research*, 54(10), 7752–7766. <https://doi.org/10.1029/2018WR022929>
- OMM. (1983). *Rapport annuel de l’Organisation météorologique mondiale 1982*. Genève: OMM.
- Ouali, D., & Cannon, A. J. (2018). Estimation of rainfall intensity–duration–frequency curves at ungauged locations using quantile regression methods. *Stochastic Environmental Research and Risk Assessment*, 32(10), 2821–2836. <https://doi.org/10.1007/s00477-018-1564-7>
- Ouali, D., Chebana, F., & Ouarda, T. B. M. J. (2016). Quantile Regression in Regional Frequency Analysis: A Better Exploitation of the Available Information. *Journal of Hydrometeorology*, 17(6), 1869–1883. <https://doi.org/10.1175/JHM-D-15-0187.1>
- Ouarda, T. B. M. J., Yousef, L. A., & Charron, C. (2019). Non-stationary intensity-duration-frequency curves integrating information concerning teleconnections and climate change. *International Journal of Climatology*, 39(4), 2306–2323. <https://doi.org/10.1002/joc.5953>
- Overeem, A., Buishand, T. A., & Holleman, I. (2009). Extreme rainfall analysis and estimation of depth-duration-frequency curves using weather radar: RAINFALL ANALYSIS, ESTIMATION OF DDF CURVES. *Water Resources Research*, 45(10). <https://doi.org/10.1029/2009WR007869>
- Overeem, Aart, Buishand, A., & Holleman, I. (2008). Rainfall depth-duration-frequency curves and their uncertainties. *Journal of Hydrology*, 348(1–2), 124–134. <https://doi.org/10.1016/j.jhydrol.2007.09.044>
- Padoan, S. A., Ribatet, M., & Sisson, S. A. (2010). Likelihood-Based Inference for Max-Stable Processes. *Journal of the American Statistical Association*, 105(489), 263–277. <https://doi.org/10.1198/jasa.2009.tm08577>
- Pagliero, L., Bouraoui, F., Diels, J., Willems, P., & McIntyre, N. (2019). Investigating regionalization techniques for large-scale hydrological modelling. *Journal of Hydrology*, 570, 220–235. <https://doi.org/10.1016/j.jhydrol.2018.12.071>

- Paixao, E., Mirza, M. M. Q., Shephard, M. W., Auld, H., Klaassen, J., & Smith, G. (2015). An integrated approach for identifying homogeneous regions of extreme rainfall events and estimating IDF curves in Southern Ontario, Canada: Incorporating radar observations. *Journal of Hydrology*, 528, 734–750. <https://doi.org/10.1016/j.jhydrol.2015.06.015>
- Palynchuk, B., & Guo, Y. (2008). Threshold analysis of rainstorm depth and duration statistics at Toronto, Canada. *Journal of Hydrology*, 348(3), 535–545. <https://doi.org/10.1016/j.jhydrol.2007.10.023>
- Pan, C., Wang, X., Liu, L., Huang, H., & Wang, D. (2017). Improvement to the Huff Curve for Design Storms and Urban Flooding Simulations in Guangzhou, China. *Water*, 9(6), 411. <https://doi.org/10.3390/w9060411>
- Panthou, G., Vischel, T., Lebel, T., Blanchet, J., Quantin, G., & Ali, A. (2012). Extreme rainfall in West Africa: A regional modeling: EXTREME RAINFALL MAPPING WEST AFRICA. *Water Resources Research*, 48(8). <https://doi.org/10.1029/2012WR012052>
- Panthou, G., Vischel, T., Lebel, T., Quantin, G., & Molinié, G. (2014). Characterising the space–time structure of rainfall in the Sahel with a view to estimating IDAF curves. *Hydrology and Earth System Sciences*, 18(12), 5093–5107. <https://doi.org/10.5194/hess-18-5093-2014>
- Panthou, Gérémy. (2013). *Analyse des extrêmes pluviométriques en Afrique de l'Ouest et de leur évolution au cours des 60 dernières années* (Thèse de doctorat). Université de Grenoble, Grenoble.
- Papacharalampous, G., Tyralis, H., & Koutsoyiannis, D. (2019). Comparison of stochastic and machine learning methods for multi-step ahead forecasting of hydrological processes. *Stochastic Environmental Research and Risk Assessment*, 33(2), 481–514. <https://doi.org/10.1007/s00477-018-1638-6>
- Papacharalampous, G., Tyralis, H., Langousis, A., Jayawardena, A. W., Sivakumar, B., Mamassis, N., et al. (2019). Probabilistic Hydrological Post-Processing at Scale: Why and How to Apply Machine-Learning Quantile Regression Algorithms. *Water*, 11(10), 2126. <https://doi.org/10.3390/w11102126>
- Papalexioi, S. M. (2018). Unified theory for stochastic modelling of hydroclimatic processes: Preserving marginal distributions, correlation structures, and intermittency. *Advances in Water Resources*, 115, 234–252. <https://doi.org/10.1016/j.advwatres.2018.02.013>
- Papalexioi, S. M., & Koutsoyiannis, D. (2013). Battle of extreme value distributions: A global survey on extreme daily rainfall: SURVEY ON EXTREME DAILY RAINFALL. *Water Resources Research*, 49(1), 187–201. <https://doi.org/10.1029/2012WR012557>
- Papalexioi, S. M., & Serinaldi, F. (2020). Random Fields Simplified: Preserving Marginal Distributions, Correlations, and Intermittency, With Applications From Rainfall to Humidity. *Water Resources Research*, 56(2). <https://doi.org/10.1029/2019WR026331>
- Papalexioi, S. M., Koutsoyiannis, D., & Makropoulos, C. (2013). How extreme is extreme? An assessment of daily rainfall distribution tails. *Hydrology and Earth System Sciences*, 17(2), 851–862. <https://doi.org/10.5194/hess-17-851-2013>
- Papalexioi, S. M., Markonis, Y., Lombardo, F., AghaKouchak, A., & Foufoula-Georgiou, E. (2018). Precise Temporal Disaggregation Preserving Marginals and Correlations (DiPMaC) for Stationary and Nonstationary Processes. *Water Resources Research*, 54(10), 7435–7458. <https://doi.org/10.1029/2018WR022726>
- Pappas, C., Papalexioi, S. M., & Koutsoyiannis, D. (2014). A quick gap filling of missing hydrometeorological data. *Journal of Geophysical Research: Atmospheres*, 119(15), 9290–9300. <https://doi.org/10.1002/2014JD021633>
- Park, J., Onof, C., & Kim, D. (2019). A hybrid stochastic rainfall model that reproduces some important rainfall characteristics at hourly to yearly timescales. *Hydrology and Earth System Sciences*, 23(2), 989–1014. <https://doi.org/10.5194/hess-23-989-2019>
- Pavlovic, S., Perica, S., St Laurent, M., & Mejía, A. (2016). Intercomparison of selected fixed-area areal reduction factor methods. *Journal of Hydrology*, 537, 419–430. <https://doi.org/10.1016/j.jhydrol.2016.03.027>
- Payrastre, O., Gaume, E., & Andrieu, H. (2011). Usefulness of historical information for flood frequency analyses: Developments based on a case study: HISTORICAL INFORMATION FOR FLOOD FREQUENCY ANALYSES. *Water Resources Research*, 47(8). <https://doi.org/10.1029/2010WR009812>
- Payrastre, Olivier. (2005). *Faisabilité et utilité du recueil de données historiques pour l'étude des crues extrêmes de petits cours d'eau: Etude du cas de quatre bassins versants affluents de l'Aude* (Thèse de doctorat). Ecole Nationale des Ponts et Chaussées, Paris (France).
- Payrastre, Olivier, Gaume, E., & Andrieu, H. (2013). Information historique et étude statistique des crues extrêmes: quelles caractéristiques souhaitables pour les inventaires de crues historiques? *La Houille Blanche*, (3), 5–11. <https://doi.org/10.1051/lhb/2013019>
- Peck, A., Prodanovic, P., & Simonovic, S. P. P. (2012). Rainfall Intensity Duration Frequency Curves Under Climate Change: City of London, Ontario, Canada. *Canadian Water Resources Journal*, 37(3), 177–189. <https://doi.org/10.4296/cwrj2011-935>
- Peleg, N., Marra, F., Faticchi, S., Paschalis, A., Molnar, P., & Burlando, P. (2018). Spatial variability of extreme rainfall at radar subpixel scale. *Journal of Hydrology*, 556, 922–933. <https://doi.org/10.1016/j.jhydrol.2016.05.033>
- Pilgrim, D. H., & Cordery, I. (1975). Rainfall Temporal Patterns for Design Floods. *Journal of the Hydraulics Division*, 101(1), 81–95.

- Pizarro, R., Valdés, R., Abarza, A., & Garcia-Chevesich, P. (2015). A simplified storm index method to extrapolate intensity-duration-frequency (IDF) curves for ungauged stations in central Chile: STORM INDEX METHOD TO EXTRAPOLATE IDF CURVES. *Hydrological Processes*, 29(5), 641–652. <https://doi.org/10.1002/hyp.10187>
- Prudhomme, C. (1999). Mapping a statistic of extreme rainfall in a mountainous region. *Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere*, 24(1), 79–84. [https://doi.org/10.1016/S1464-1909\(98\)00015-X](https://doi.org/10.1016/S1464-1909(98)00015-X)
- Ragno, E., AghaKouchak, A., Love, C. A., Cheng, L., Vahedifard, F., & Lima, C. H. R. (2018). Quantifying Changes in Future Intensity-Duration-Frequency Curves Using Multimodel Ensemble Simulations. *Water Resources Research*, 54(3), 1751–1764. <https://doi.org/10.1002/2017WR021975>
- Read, L. K., & Vogel, R. M. (2015). Reliability, return periods, and risk under nonstationarity. *Water Resources Research*, 51(8), 6381–6398. <https://doi.org/10.1002/2015WR017089>
- Reed, D. W., Faulkner, D. S., & Stewart, E. J. (1999). The FORGEX method of rainfall growth estimation II: Description. *Hydrology and Earth System Sciences*, 3(2), 197–203. <https://doi.org/10.5194/hess-3-197-1999>
- Reis, D. S., & Stedinger, J. R. (2005). Bayesian MCMC flood frequency analysis with historical information. *Journal of Hydrology*, 313(1–2), 97–116. <https://doi.org/10.1016/j.jhydrol.2005.02.028>
- Remenieras, G. (1986). *L'hydrologie de l'ingénieur* (2ème). Paris (France): Librairie Eyrolles. Retrieved from <https://www.eyrolles.com/BTP/Livre/l-hydrologie-de-l-ingenieur-9782212052206/>
- Renard, B. (2011). A Bayesian hierarchical approach to regional frequency analysis: HIERARCHICAL REGIONAL ESTIMATION. *Water Resources Research*, 47(11). <https://doi.org/10.1029/2010WR010089>
- Requena, A. I., Chebana, F., & Ouarda, T. B. M. J. (2017). Heterogeneity measures in hydrological frequency analysis: review and new developments. *Hydrology and Earth System Sciences*, 21(3), 1651–1668. <https://doi.org/10.5194/hess-21-1651-2017>
- Ribatet, M., & Sedki, M. (2013). Extreme value copulas and max-stable processes. *Journal de La Société Française de Statistique*, 154(1), 138–150.
- Ribatet, M., Sauquet, E., Grésillon, J.-M., & Ouarda, T. B. M. J. (2007). A regional Bayesian POT model for flood frequency analysis. *Stochastic Environmental Research and Risk Assessment*, 21(4), 327–339. <https://doi.org/10.1007/s00477-006-0068-z>
- Ribeiro, P. J., Diggle, P., Schlather, M., Bivand, R., & Ripley, B. (2020). geoR: Analysis of Geostatistical Data (Version 1.8-1). Retrieved from <https://CRAN.R-project.org/package=geoR>
- Ricciardelli, E., Cimini, D., Di Paola, F., Romano, F., & Viggiano, M. (2014). A statistical approach for rain intensity differentiation using Meteosat Second Generation–Spinning Enhanced Visible and InfraRed Imager observations. *Hydrology and Earth System Sciences*, 18(7), 2559–2576. <https://doi.org/10.5194/hess-18-2559-2014>
- Rios Gaona, M. F., Overeem, A., Leijnse, H., & Uijlenhoet, R. (2015). Measurement and interpolation uncertainties in rainfall maps from cellular communication networks. *Hydrology and Earth System Sciences*, 19(8), 3571–3584. <https://doi.org/10.5194/hess-19-3571-2015>
- Ritschel, C., Ulbrich, U., Névir, P., & Rust, H. W. (2017). Precipitation extremes on multiple timescales – Bartlett–Lewis rectangular pulse model and intensity–duration–frequency curves. *Hydrology and Earth System Sciences*, 21(12), 6501–6517. <https://doi.org/10.5194/hess-21-6501-2017>
- Robert, C., & Casella, G. (2004). *Monte Carlo Statistical Methods* (2nd ed.). New York: Springer-Verlag. <https://doi.org/10.1007/978-1-4757-4145-2>
- Roche, P.-A., Miquel, J., Gaume, F., & Marsily, G. de. (2012). *Hydrologie quantitative : Processus, modèles et aide à la décision* (1 Pap/DVD). Paris; Berlin; Heidelberg: Springer Verlag France.
- Rosbjerg, D., & Madsen, H. (1996). The role of regional information in estimation of extreme point rainfalls. *Atmospheric Research*, 42(1), 113–122. [https://doi.org/10.1016/0169-8095\(95\)00057-7](https://doi.org/10.1016/0169-8095(95)00057-7)
- Rosbjerg, D., & Madsen, H. (2004). Advanced approaches in PDS/POT modelling of extreme hydrological events. In *Hydrology. Science and Practice for the 21st Century. Proceedings* (pp. 217–220). Londres: British Hydrological Society. Retrieved from <https://orbit.dtu.dk/en/publications/advanced-approaches-in-pdspot-modelling-of-extreme-hydrological-e>
- Salas, L. de, & Fernández, J. A. (2007). “In-site” regionalization to estimate an intensity-duration-frequency law: a solution to scarce spatial data in Spain. *Hydrological Processes*, 21(25), 3507–3513. <https://doi.org/10.1002/hyp.6551>
- Sane, Y., Panthou, G., Bodian, A., Vischel, T., Lebel, T., Dacosta, H., et al. (2018). Intensity–duration–frequency (IDF) rainfall curves in Senegal. *Natural Hazards and Earth System Sciences*, 18(7), 1849–1866. <https://doi.org/10.5194/nhess-18-1849-2018>
- Schertzer, D., & Lovejoy, S. (1987). Physical modeling and analysis of rain and clouds by anisotropic scaling multiplicative processes. *Journal of Geophysical Research*, 92(D8), 9693. <https://doi.org/10.1029/JD092iD08p09693>
- Schiavo Bernardi, E., Allasia, D., Basso, R., Freitas Ferreira, P., & Tassi, R. (2015). TRMM rainfall estimative coupled with Bell (1969) methodology for extreme rainfall characterization. *Proceedings of the International Association of Hydrological Sciences*, 369, 163–168. <https://doi.org/10.5194/piahs-369-163-2015>

- Sebillé, Q., Fougères, A.-L., & Mercadier, C. (2017). Modeling extreme rainfall A comparative study of spatial extreme value models. *Spatial Statistics*, 21, 187–208. <https://doi.org/10.1016/j.spasta.2017.06.009>
- Serinaldi, F. (2010). Multifractality, imperfect scaling and hydrological properties of rainfall time series simulated by continuous universal multifractal and discrete random cascade models. *Nonlinear Processes in Geophysics*, 17(6), 697–714. <https://doi.org/10.5194/npg-17-697-2010>
- Serinaldi, Francesco, & Kilsby, C. G. (2015). Stationarity is undead: Uncertainty dominates the distribution of extremes. *Advances in Water Resources*, 77, 17–36. <https://doi.org/10.1016/j.advwatres.2014.12.013>
- Shahabul Alam, Md., & Elshorbagy, A. (2015). Quantification of the climate change-induced variations in Intensity–Duration–Frequency curves in the Canadian Prairies. *Journal of Hydrology*, 527, 990–1005. <https://doi.org/10.1016/j.jhydrol.2015.05.059>
- Sivakumar, B. (2001). Is a chaotic multi-fractal approach for rainfall possible? *Hydrological Processes*, 15(6), 943–955. <https://doi.org/10.1002/hyp.260>
- Sivapalan, M., & Blöschl, G. (1998). Transformation of point rainfall to areal rainfall: Intensity-duration-frequency curves. *Journal of Hydrology*, 204(1–4), 150–167. [https://doi.org/10.1016/S0022-1694\(97\)00117-0](https://doi.org/10.1016/S0022-1694(97)00117-0)
- Smith, A., Sampson, C., & Bates, P. (2015). Regional flood frequency analysis at the global scale. *Water Resources Research*, 51(1), 539–553. <https://doi.org/10.1002/2014WR015814>
- Smith, R. L. (1985). Maximum Likelihood Estimation in a Class of Nonregular Cases. *Biometrika*, 72(1), 67–90. <https://doi.org/10.2307/2336336>
- Soltani, S., Helfi, R., Almasi, P., & Modarres, R. (2017). Regionalization of Rainfall Intensity-Duration-Frequency using a Simple Scaling Model. *Water Resources Management*, 31(13), 4253–4273. <https://doi.org/10.1007/s11269-017-1744-0>
- Song, J. J., Kwon, S., & Lee, G. (2015). Incorporation of parameter uncertainty into spatial interpolation using Bayesian trans-Gaussian kriging. *Advances in Atmospheric Sciences*, 32(3), 413–423. <https://doi.org/10.1007/s00376-014-4040-4>
- Stedinger, J. R. (1983a). Confidence Intervals for Design Events. *Journal of Hydraulic Engineering*, 109(1), 13–27. [https://doi.org/10.1061/\(ASCE\)0733-9429\(1983\)109:1\(13\)](https://doi.org/10.1061/(ASCE)0733-9429(1983)109:1(13))
- Stedinger, J. R. (1983b). Design events with specified flood risk. *Water Resources Research*, 19(2), 511–522. <https://doi.org/10.1029/WR019i002p00511>
- Stedinger, J. R., Vogel, R. M., & Foufoula-Georgiou. (1993). Frequency analysis of extreme events (Version McGraw-Hill). In *Handbook of hydrology* (McGraw-Hill). New York: Maiment, D.R.
- Stein, M. L. (2019). Some Statistical Issues in Climate Science, 17.
- Stephens, M., A. (1976). *Asymptotic power of EDF statistics for exponentiality against gamma and Weibull alternatives* (Technical Report No. 297). Stanford, CA.: Stanford University.
- Stephens, M. A. (1977). Goodness of Fit for the Extreme Value Distribution. *Biometrika*, 64(3), 583. <https://doi.org/10.2307/2345336>
- Stewart, E. J., Reed, D. W., Faulkner, D. S., & Reynard, N. S. (1999). The FORGEX method of rainfall growth estimation I: Review of requirement. *Hydrology and Earth System Sciences*, 3(2), 187–195. <https://doi.org/10.5194/hess-3-187-1999>
- St-Hilaire, A., Ouarda, T. B. M. J., Lachance, M., Bobée, B., Barbet, M., & Bruneau, P. (2003). La régionalisation des précipitations : une revue bibliographique des développements récents. *Revue des sciences de l'eau*, 16(1), 27. <https://doi.org/10.7202/705497ar>
- Svensson, C., & Jones, D. A. (2010). Review of rainfall frequency estimation methods: Review of rainfall frequency estimation methods. *Journal of Flood Risk Management*, 3(4), 296–313. <https://doi.org/10.1111/j.1753-318X.2010.01079.x>
- Svensson, Cecilia, Clarke, R. T., & Jones, D. A. (2007). An experimental comparison of methods for estimating rainfall intensity-duration-frequency relations from fragmentary records. *Journal of Hydrology*, 341(1–2), 79–89. <https://doi.org/10.1016/j.jhydrol.2007.05.002>
- Szolgay, J., Parajka, J., Kohnová, S., & Hlavčová, K. (2009). Comparison of mapping approaches of design annual maximum daily precipitation. *Atmospheric Research*, 92(3), 289–307. <https://doi.org/10.1016/j.atmosres.2009.01.009>
- Taibi, S., Meddi, M., Mahé, G., & Assani, A. (2017). Relationships between atmospheric circulation indices and rainfall in Northern Algeria and comparison of observed and RCM-generated rainfall. *Theoretical and Applied Climatology*, 127(1–2), 241–257. <https://doi.org/10.1007/s00704-015-1626-4>
- Tanner, M. A. (1996). *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions* (3rd ed.). New York: Springer-Verlag. <https://doi.org/10.1007/978-1-4612-4024-2>
- Tawn, J., Shooter, R., Towe, R., & Lamb, R. (2018). Modelling spatial extreme events with environmental applications. *Spatial Statistics*, 28, 39–58. <https://doi.org/10.1016/j.spasta.2018.04.007>
- Teegavarapu, R. S. V., Salas, J. D., & Stedinger, J. R. (Eds.). (2019). *Statistical Analysis of Hydrologic Variables: Methods and Applications*. Reston, VA: American Society of Civil Engineers. <https://doi.org/10.1061/9780784415177>

- Thorndahl, S., Einfalt, T., Willems, P., Nielsen, J. E., ten Veldhuis, M.-C., Arnbjerg-Nielsen, K., et al. (2017). Weather radar rainfall data in urban hydrology. *Hydrology and Earth System Sciences*, 21(3), 1359–1380. <https://doi.org/10.5194/hess-21-1359-2017>
- Todorovic, P. (1978). Stochastic models of floods. *Water Resources Research*, 14(2), 345–356. <https://doi.org/10.1029/WR014i002p00345>
- Totaro, V., Gioia, A., & Iacobellis, V. (2020). Numerical investigation on the power of parametric and nonparametric tests for trend detection in annual maximum series. *Hydrology and Earth System Sciences*, 24(1), 473–488. <https://doi.org/10.5194/hess-24-473-2020>
- Tsoukalas, I., Efstratiadis, A., & Makropoulos, C. (2019). Building a puzzle to solve a riddle: A multi-scale disaggregation approach for multivariate stochastic processes with any marginal distribution and correlation structure. *Journal of Hydrology*, 575, 354–380. <https://doi.org/10.1016/j.jhydrol.2019.05.017>
- Tye, M. R., & Cooley, D. (2015). A spatial model to examine rainfall extremes in Colorado's Front Range. *Journal of Hydrology*, 530, 15–23. <https://doi.org/10.1016/j.jhydrol.2015.09.023>
- Tyralis, H., & Langousis, A. (2019). Estimation of intensity–duration–frequency curves using max-stable processes. *Stochastic Environmental Research and Risk Assessment*, 33(1), 239–252. <https://doi.org/10.1007/s00477-018-1577-2>
- Tyralis, H., Papacharalampous, G., & Tantane, S. (2019). How to explain and predict the shape parameter of the generalized extreme value distribution of streamflow extremes using a big dataset. *Journal of Hydrology*, 574, 628–645. <https://doi.org/10.1016/j.jhydrol.2019.04.070>
- Uboldi, F., Sulis, A. N., Lussana, C., Cislighi, M., & Russo, M. (2014). A spatial bootstrap technique for parameter estimation of rainfall annual maxima distribution. *Hydrology and Earth System Sciences*, 18(3), 981–995. <https://doi.org/10.5194/hess-18-981-2014>
- Van de Vyver, H. (2015a). Bayesian estimation of rainfall intensity–duration–frequency relationships. *Journal of Hydrology*, 529, 1451–1463. <https://doi.org/10.1016/j.jhydrol.2015.08.036>
- Van de Vyver, H. (2015b). On the estimation of continuous 24-h precipitation maxima. *Stochastic Environmental Research and Risk Assessment*, 29(3), 653–663. <https://doi.org/10.1007/s00477-014-0912-5>
- Van de Vyver, Hans. (2018). A multiscaling-based intensity-duration-frequency model for extreme precipitation. *Hydrological Processes*, 32(11), 1635–1647. <https://doi.org/10.1002/hyp.11516>
- Van de Vyver, Hans, & Demarée, G. R. (2010). Construction of Intensity–Duration–Frequency (IDF) curves for precipitation at Lubumbashi, Congo, under the hypothesis of inadequate data. *Hydrological Sciences Journal*, 55(4), 555–564. <https://doi.org/10.1080/02626661003747390>
- Veneziano, D., & Furcolo, P. (2002). Multifractality of rainfall and scaling of intensity-duration-frequency curves: MULTIFRACTALITY OF RAINFALL AND IDF CURVES. *Water Resources Research*, 38(12), 42-1-42–12. <https://doi.org/10.1029/2001WR000372>
- Veneziano, D., & Langousis, A. (2010). SCALING AND FRACTALS IN HYDROLOGY. In B. Sivakumar & R. Berndtsson, *Advances in Data-Based Approaches for Hydrologic Modeling and Forecasting* (pp. 107–243). WORLD SCIENTIFIC. [https://doi.org/10.1142/9789814307987\\_0004](https://doi.org/10.1142/9789814307987_0004)
- Veneziano, D., Langousis, A., & Furcolo, P. (2006). Multifractality and rainfall extremes: A review: MULTIFRACTALITY AND RAINFALL EXTREMES. *Water Resources Research*, 42(6). <https://doi.org/10.1029/2005WR004716>
- Veneziano, D., Lepore, C., Langousis, A., & Furcolo, P. (2007). Marginal methods of intensity-duration-frequency estimation in scaling and nonscaling rainfall. *Water Resources Research*, 43(10), W10418. <https://doi.org/10.1029/2007WR006040>
- Viglione, A., Laio, F., & Claps, P. (2007). A comparison of homogeneity tests for regional frequency analysis: HOMOGENEITY TESTS FOR REGIONAL FREQUENCY ANALYSIS. *Water Resources Research*, 43(3). <https://doi.org/10.1029/2006WR005095>
- Viglione, Alberto, Merz, R., Salinas, J. L., & Blöschl, G. (2013). Flood frequency hydrology: 3. A Bayesian analysis. *Water Resources Research*, 49(2), 675–692. <https://doi.org/10.1029/2011WR010782>
- Viglione, Alberto, Hosking, J. R. M., Laio, F., Miller, A., Gaume, E., Payrastre, O., et al. (2014). nsRFA: Non-supervised Regional Frequency Analysis (Version 0.7-12). Retrieved from <https://cran.r-project.org/web/packages/nsRFA/index.html>
- Viglione, Alberto, Hosking, J. R. M., Laio, F., Miller, A., Gaume, E., Payrastre, O., et al. (2020). nsRFA: Non-Supervised Regional Frequency Analysis (Version 0.7-15). Retrieved from <https://CRAN.R-project.org/package=nsRFA>
- Vogel, R. M., & Fennessey, N. M. (1993). L moment diagrams should replace product moment diagrams. *Water Resources Research*, 29(6), 1745–1752. <https://doi.org/10.1029/93WR00341>
- Vogel, R. M., & McMartin, D. E. (1991). Probability Plot Goodness-of-Fit and Skewness Estimation Procedures for the Pearson Type 3 Distribution. *Water Resources Research*, 27(12), 3149–3158. <https://doi.org/10.1029/91WR02116>
- Volpi, E., Fiori, A., Grimaldi, S., Lombardo, F., & Koutsoyiannis, D. (2015). One hundred years of return period: Strengths and limitations. *Water Resources Research*, 51(10), 8570–8585. <https://doi.org/10.1002/2015WR017820>

- Volpi, Elena. (2019). On return period and probability of failure in hydrology. *Wiley Interdisciplinary Reviews: Water*, 6(3), e1340. <https://doi.org/10.1002/wat2.1340>
- Volpi, Elena, Fiori, A., Grimaldi, S., Lombardo, F., & Koutsoyiannis, D. (2019). Save hydrological observations! Return period estimation without data decimation. *Journal of Hydrology*, 571, 782–792. <https://doi.org/10.1016/j.jhydrol.2019.02.017>
- Vorstermans, V. (2019, September 30). *Return level estimation: a theoretical comparison between the peak-over-threshold approach and the block maxima method* (Master). ERASMUS UNIVERSITY ROTTERDAM, ROTTERDAM. Retrieved from <https://thesis.eur.nl/pub/50405>
- Wallis, J. R., Matalas, N. C., & Slack, J. R. (1974). Just a moment! *Water Resources Research*, 10(2), 211–219. <https://doi.org/10.1029/WR010i002p00211>
- Wallis, J. R., Schaefer, M. G., Barker, B. L., & Taylor, G. H. (2007). Regional precipitation-frequency analysis and spatial mapping for 24-hour and 2-hour durations for Washington State. *Hydrology and Earth System Sciences*, 11(1), 415–442. <https://doi.org/10.5194/hess-11-415-2007>
- Wasserman, L. (2004). *All of statistics: A concise course in statistical inference*. (1st ed.). New York: Springer Science+Business Media.
- Weiss, L. L. (1964). Ratio of True to Fixed-Interval Maximum Rainfall. *Journal of the Hydraulics Division*, 90(1), 77–82.
- Western, A. W., Anderson, B., Siriwardena, L., Chiew, F. H. S., Seed, A., & Blöschl, G. (2011). The within-day behaviour of 6 minute rainfall intensity in Australia. *Hydrology and Earth System Sciences*, 15(8), 2561–2579. <https://doi.org/10.5194/hess-15-2561-2011>
- Westra, S., Fowler, H. J., Evans, J. P., Alexander, L. V., Berg, P., Johnson, F., et al. (2014). Future changes to the intensity and frequency of short-duration extreme rainfall: FUTURE INTENSITY OF SUB-DAILY RAINFALL. *Reviews of Geophysics*, 52(3), 522–555. <https://doi.org/10.1002/2014RG000464>
- Willems, P. (2000). Compound intensity/duration/frequency-relationships of extreme precipitation for two seasons and two storm types. *Journal of Hydrology*, 233(1), 189–205.
- Willems, P. (2001). A spatial rainfall generator for small spatial scales. *Journal of Hydrology*, 252(1–4), 126–144. [https://doi.org/10.1016/S0022-1694\(01\)00446-2](https://doi.org/10.1016/S0022-1694(01)00446-2)
- Willems, P. (2012). Impacts of Climate Change on Rainfall Extremes and Urban Drainage Systems. *Water Intelligence Online*, 11. <https://doi.org/10.2166/9781780401263>
- Willems, P., Olsson, J., Arnbjerg-Nielsen, K., Beecham, S., Pathirana, A., Bülow Gregersen, I., et al. (2012). *Impacts of Climate Change on Rainfall Extremes and Urban Drainage Systems* (Vol. 11). London: IWA Publishing. Retrieved from <https://iwaponline.com/ebooks/book/631/>
- Wright, D. B., Yu, G., & England, J. F. (2020). Six decades of rainfall and flood frequency analysis using stochastic storm transposition: Review, progress, and prospects. *Journal of Hydrology*, 585, 124816. <https://doi.org/10.1016/j.jhydrol.2020.124816>
- Xu, Y.-P., & Tung, Y.-K. (2009). Constrained scaling approach for design rainfall estimation. *Stochastic Environmental Research and Risk Assessment*, 23(6), 697–705. <https://doi.org/10.1007/s00477-008-0250-6>
- Yilmaz, A. G., Hossain, I., & Perera, B. J. C. (2014). Effect of climate change and variability on extreme rainfall intensity–frequency–duration relationships: a case study of Melbourne. *Hydrology and Earth System Sciences*, 18(10), 4065–4076. <https://doi.org/10.5194/hess-18-4065-2014>
- Yoon, S., Kumphon, B., & Park, J.-S. (2015). Spatial Modelling of Extreme Rainfall in Northeast Thailand. *Procedia Environmental Sciences*, 26, 45–48. <https://doi.org/10.1016/j.proenv.2015.05.021>
- Yu, P.-S., & Chen, C.-J. (1997). POTENTIAL OF EXTENDING THE RAINFALL INTENSITY-DURATION-FREQUENCY RELATIONSHIP TO NON-RECORDING RAIN GAUGES. *Hydrological Processes*, 11, 377–390. [https://doi.org/doi.org/10.1002/\(SICI\)1099-1085\(19970330\)11:4<377::AID-HYP437>3.0.CO;2-4](https://doi.org/doi.org/10.1002/(SICI)1099-1085(19970330)11:4<377::AID-HYP437>3.0.CO;2-4)
- Yu, P.-S., Yang, T.-C., & Lin, C.-S. (2004). Regional rainfall intensity formulas based on scaling property of rainfall. *Journal of Hydrology*, 295(1–4), 108–123. <https://doi.org/10.1016/j.jhydrol.2004.03.003>
- Zahar, Y., & Laborde, J.-P. (2007). Modélisation statistique et synthèse cartographique des précipitations journalières extrêmes de Tunisie. *Revue des sciences de l'eau / Journal of Water Science*, 20(4), 409–424. <https://doi.org/10.7202/016914ar>
- Zeng, Q., Chen, H., Xu, C.-Y., Jie, M.-X., Chen, J., Guo, S.-L., & Liu, J. (2018). The effect of rain gauge density and distribution on runoff simulation using a lumped hydrological modelling approach. *Journal of Hydrology*, 563, 106–122. <https://doi.org/10.1016/j.jhydrol.2018.05.058>
- Zeroual, A., Assani, A. A., & Meddi, M. (2017). Combined analysis of temperature and rainfall variability as they relate to climate indices in Northern Algeria over the 1972–2013 period. *Hydrology Research*, 48(2), 584–595. <https://doi.org/10.2166/nh.2016.244>
- Zeroual, A., Assani, A. A., Meddi, M., & Alkama, R. (2019). Assessment of climate change in Algeria from 1951 to 2098 using the Köppen–Geiger climate classification scheme. *Climate Dynamics*, 52(1–2), 227–243. <https://doi.org/10.1007/s00382-018-4128-0>

- Zhu, Z., Wright, D. B., & Yu, G. (2018). The Impact of Rainfall Space-Time Structure in Flood Frequency Analysis. *Water Resources Research*, 54(11), 8983–8998. <https://doi.org/10.1029/2018WR023550>
- Zorzetto, E., Botter, G., & Marani, M. (2016). On the emergence of rainfall extremes from ordinary events: EXTREMES EMERGE FROM ORDINARY EVENTS. *Geophysical Research Letters*, 43(15), 8076–8082. <https://doi.org/10.1002/2016GL069445>





**Annexes**

# Liste des annexes

1	Annexe 1 - Outils statistiques .....	II
1.1	Test d'adéquation d'Anderson-Darling (AD) .....	II
1.2	Test de discordance de Hosking et Wallis (D) .....	II
1.3	Test d'homogénéité de Hosking et Wallis (H).....	III
2	Annexe 2 - Caractéristiques statistiques de SMA de pluies.....	V
2.1	Liste de codes ANRH des stations Pluviographiques .....	V
2.2	Evolution des totaux pluviométriques pour 49 postes PG .....	VI
2.3	Diagramme de dispersions des cumuls de pluies maximales de 49 stations PG pour 8 durées .....	IX
3	Annexe 3 - Résultats de l'analyse d'échelle des moments de 48 SMA .....	X
4	Annexe 4 - Séries simulées : médianes et incertitudes des cumuls de pluies vs. durées.....	XI
4.1	Gamme de durées de 15 à 1440 minutes (8 durées).....	XI
4.2	Gamme de durées de 60 à 1440 minutes (6 durées).....	XI
5	Annexe 5 - Graphiques d'ajustements GEV par BayMCMC : M0 vs M1.....	XII
5.1	Durées d'agrégation de d = 15 à 1440 minutes .....	XII
5.2	Durées d'agrégation de d= 60 à 1440 minutes .....	XII
5.3	Durées d'agrégation de =15, 60 et 1440 minutes.....	XIII
6	Annexe 6 - Publication dans la revue scientifique Water Resources Research .....	XIV
7	Annexe 7 - communications scientifiques .....	XXXIII
7.1	EVAN 2019 .....	XXXIII
7.2	EGU General Assembly 2018.....	XXXIV

## Annexe 1 - Outils statistiques

### 1.1 Test d'adéquation d'Anderson-Darling (AD)

La statistique  $A^2$  du test d'adéquation d'Anderson-Darling est donnée par l'expression suivante (D'Agostino & Stephens, 1986; Stephens, 1977) :

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^n \{(2i - 1) \ln F(y_i) + (2n - 2i + 1) \ln(1 - F(y_i))\} \quad (A1.1)$$

Avec  $F(y_i)$  la fréquence théorique au non dépassement de la distribution de probabilité sous-jacente pour chaque valeur de la série.  $n$  étant la taille de la série classée suivant un ordre croissant. Les seuils de rejets de l'hypothèse nulle  $H_0$  du test d'Anderson-Darling pour les distributions fréquentielles **des valeurs extrêmes** sont données par (Stephens, 1977) :

Tableau A 1.1: Seuils de rejet du test AD pour des distributions des valeurs extrêmes

Statistique (AD)	Cas p	Modification tenant compte de la taille de l'effectif	Seuil de signification $\alpha$				
			0.75	0.90	0.95	0.975	0.99
$A^2$	<b>Cas 0</b> : les paramètres de position et d'échelle sont connus	Aucune	-	1.933	2.492	3.070	3.857
	<b>Cas 1</b> : le paramètre de position est inconnu (à estimer) et celui d'échelle est connu	$A^2(1 + 0.3/n)$	0.736	1.062	1.321	1.591	1.959
	<b>Cas 2</b> : le paramètre de position est connu et celui d'échelle est à estimer (inconnu)	Aucune	1.060	1.725	2.277	2.854	3.640
	<b>Cas 3</b> : les paramètres de position et d'échelle sont inconnus et doivent être estimés	$A^2(1 + 0.2/\sqrt{n})$	0.474	0.637	0.757	0.877	1.038

### 1.2 Test de discordance de Hosking et Wallis (D)

Le test de discordance de Hosking et Wallis permet de vérifier si un site (ou une station de mesure) appartient à une groupe supposé homogène. C'est-à-dire que la mesure de discordance permet d'identifier les sites qui sont grossièrement discordants avec le groupe dans son ensemble. La discordance est mesurée en termes de L-moments ou plutôt de ratios de L-moments. La discordance s'exprime comme suit (Hosking & Wallis, 1997) :

En supposant qu'il y a  $N$  sites dans le groupe. Soit  $\bar{u} = [t^i, t_3^i, t_4^i]^T$  un vecteur contenant les valeurs de  $t$ ,  $t_3$  et  $t_4$  pour un site  $i$  : l'indice  $T$  dénote d'un vecteur ou d'une matrice. Soit :

$$\bar{u} = N^{-1} \sum_{i=1}^N u_i \quad (A1.2)$$

Avec  $\bar{u}$  la moyenne (non pondérée) du groupe. Définissons la matrice des sommes des carrés et des produits croisés :

$$A = N^{-1} \sum_{i=1}^N (u_i - \bar{u})(u_i - \bar{u})^T \quad (A1.3)$$

La mesure de la discordance pour un site  $i$  s'écrit alors :

$$D_i = \frac{1}{3} N (u_i - \bar{u})^T A^{-1} (u_i - \bar{u}) \quad (A1.4)$$

Le site  $i$  est déclaré comme discordant si  $D_i$  prend une grande valeur. Toutefois, la définition de grand dépend du nombre de sites dans le groupe. Hosking & Wallis (1997) suggèrent de considérer un site comme discordant s'il y a dépassement des valeurs critiques données dans le tableau A 1.2.

Tableau A 1.2 : Valeurs critiques de mesure de discordance  $D_i$ .

Nombre de sites dans une région	Valeur critique	Nombre de sites dans une région	Valeur critique
5	1.333	10	2.491
6	1.648	11	2.632
7	1.917	12	2.757
8	2.140	13	2.869
9	2.329	14	2.971
		≥15	3

### 1.3 Test d'homogénéité de Hosking et Wallis (H)

Supposant que la région proposée compte  $N$  sites, où chaque site  $i$  a une longueur d'enregistrement de  $n_i$  et des ratios de L-moments sur les échantillons  $t^i, t_3^i, t_4^i$ . Les moyennes régionales de L-Cv, L-Cs et LCk sont respectivement :  $t^R, t_3^R, t_4^R$ . Ces derniers sont pondérés proportionnellement à la longueur de l'échantillon, par exemple :

$$t^R = \sum_{i=1}^N n_i t^i / \sum_{i=1}^N n_i \quad (A1.5)$$

L'écart-type pondéré L-CVs au site peut être calculé par :

$$V = \left\{ \sum_{i=1}^N n_i (t^i - t^R)^2 / \sum_{i=1}^N n_i \right\}^{1/2} \quad (A1.6)$$

Il s'agit ensuite d'ajuster une distribution kappa (équation A1.7) aux ratios des L-moments moyens régionaux 1,  $t^R, t_3^R, t_4^R$ .

$$x(F) = \xi + \frac{\alpha}{k} \left\{ 1 - \left( \frac{1 - F^h}{h} \right)^k \right\} \quad (A1.7)$$

La distribution kappa comporte quatre paramètres :  $\xi$  (position),  $\alpha$  (échelle),  $k$ (forme),  $h$ .

Une fois la loi kappa ajustée, il s'agit de simuler un grand nombre  $N_{sim}$  de réalisations d'une région avec  $N$  sites, chacun ayant cette loi kappa comme distribution de fréquence. Les régions simulées sont homogènes et ne comportent pas de corrélations sérielles ou de corrélations croisées. Tous les sites ont la même longueur d'enregistrement que leurs contreparties réelles. Ensuite, il suffit de calculer  $V$  pour chaque région simulée, puis à partir des simulations de calculer la moyenne  $\mu_V$  et l'écart-type  $\sigma_V$  de  $N_{sim}$  valeurs de  $V$ . la mesure de l'hétérogénéité  $H$  est donné par :

$$H = \frac{V - \mu_V}{\sigma_V} \quad (A1.8)$$

Hosking & Wallis (1997) suggèrent de déclarer une région hétérogène si  $H$  est suffisamment grand. Si  $H < 1$  la région est acceptablement homogène, si  $1 \leq H < 2$  la région est possiblement homogène, si  $H > 2$  la région est certainement homogène.

## Annexe 2 - Caractéristiques statistiques de SMA de pluies

### 2.1 Liste de codes ANRH des stations Pluviographiques

Code ANRH	Nom Station	Identification (Id.) station PG	Code ANRH	Nom Station	Identification (Id.) station PG
011208	Boughzoul	1	020509	Birmandreis	25
011405	Ghrib (barrage)	2	020601	Deux Bassins	26
011501	Tamezguida	3	020602	Hamiz	27
011509	Medea	4	020642	Rouiba	28
011514	Djenane Ben Ouadhah	5	020646	Keddara	29
011603	Bordj Amir AEK	6	021014	Oued Zeboudj	30
011605	Theniet el Had	7	021112	Ouzra	31
011609	Deurdeur	8	021132	Mouzaia	32
011717	Khemis Miliana	9	021504	Azzazga	33
011718	Harraza	10	021601	Aghrib	34
011801	Arib Ebda	11	021712	Beni Yenn	35
011806	El Anneb	12	021717	Ait Djema	36
011901	El Touaibia	13	021801	Bouassem	37
012108	Oued Fodda	14	021915	Argoub	38
012203	Oum Drou (Ponteba)	15	021924	Ain Zaouia	39
012205	Benaria	16	050103	Souk el Djemaa	40
012218	Domaine Si tayeb	17	090104	Souagui	41
012219	Chlef	18	090202	Mahterre	42
020101	Berkhissa	19	090314	Dechmia	43
020201	Abou el Hassene	20	090506	Tizi Ghenif	44
020224	Bissa	21	090512	ChabetElAmeur	45
020306	Meurad	22	150101	Sour el Ghozlène	46
020315	Gouraya	23	150208	Bouira	47
020329	Boukerdane	24	150301	M'Chedallah	48
020502	Ouled Fayet	Eliminer	021234	Soumaa	49

## 2.2 Evolution des totaux pluviométriques pour 49 postes PG

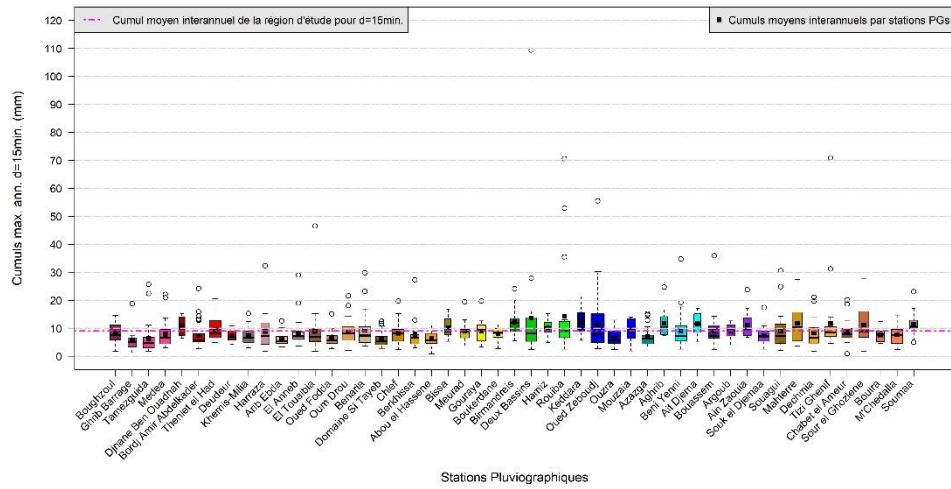


Figure A 2-1 Evolution des totaux pluviométriques pour d= 15 minutes.

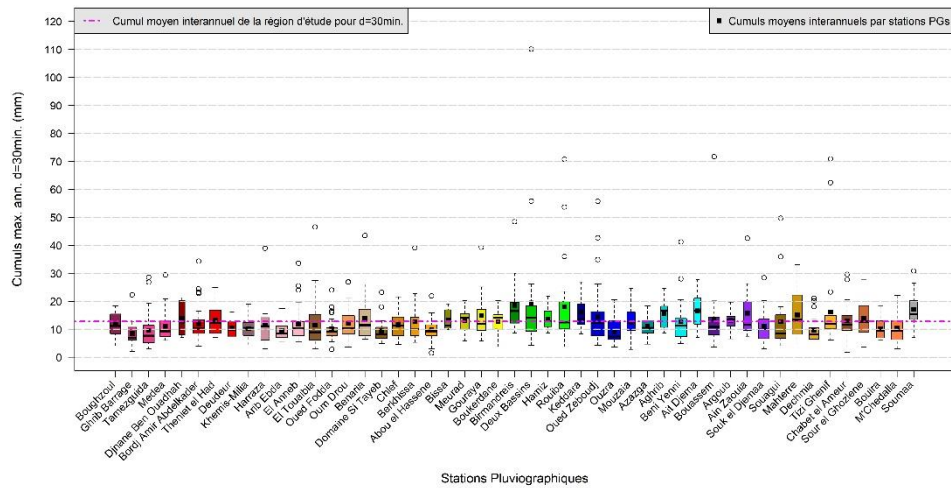


Figure A 2-2 Evolution des totaux pluviométriques pour d= 30 minutes.



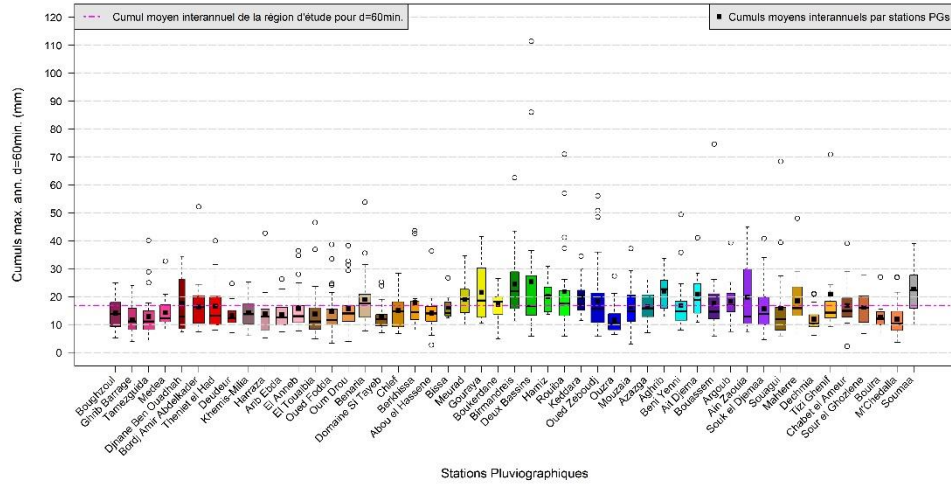


Figure A 2-3 Evolution des totaux pluviométriques pour d= 60 minutes.

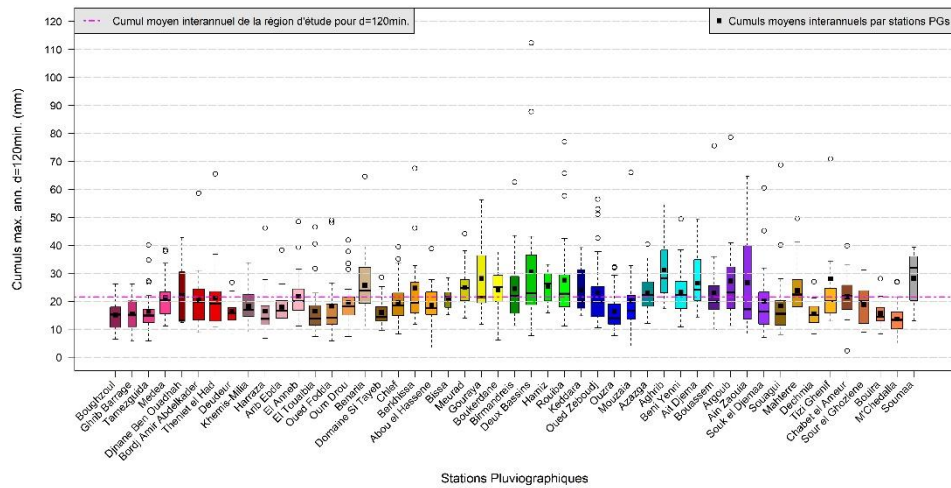


Figure A 2-4 Evolution des totaux pluviométriques pour d= 120 minutes.

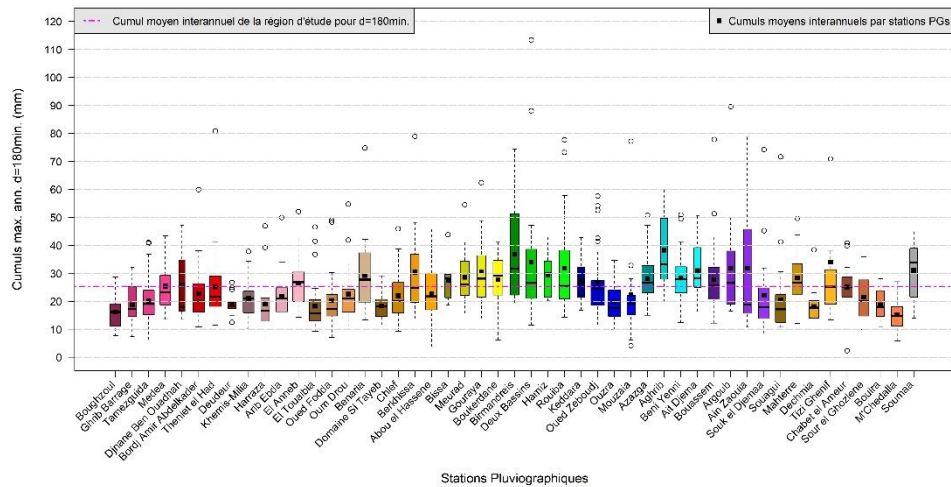


Figure A 2-5 Evolution des totaux pluviométriques pour d= 180 minutes.

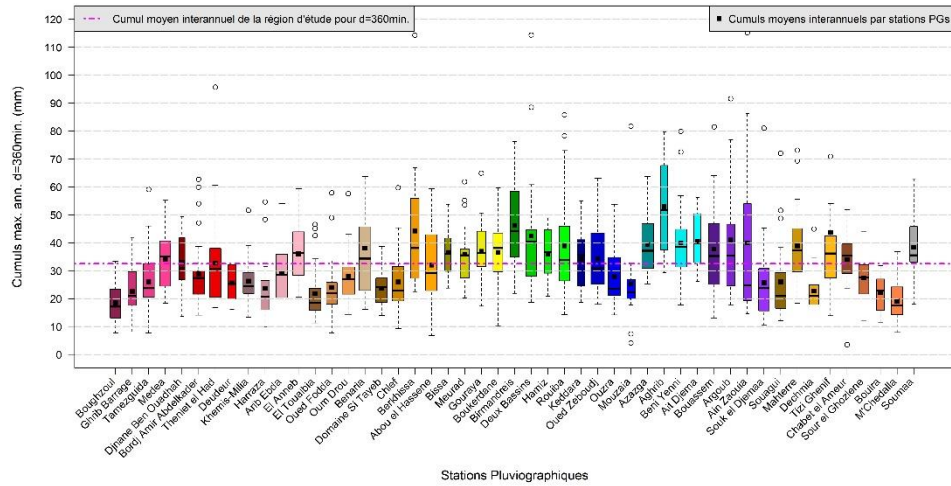


Figure A 2-6 Evolution des totaux pluviométriques pour d= 360 minutes.

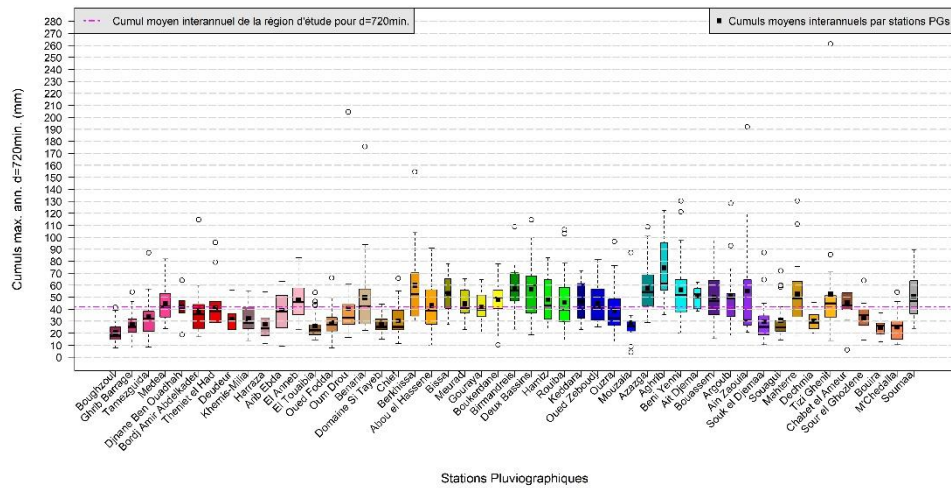


Figure A 2-7 Evolution des totaux pluviométriques pour d= 720 minutes.

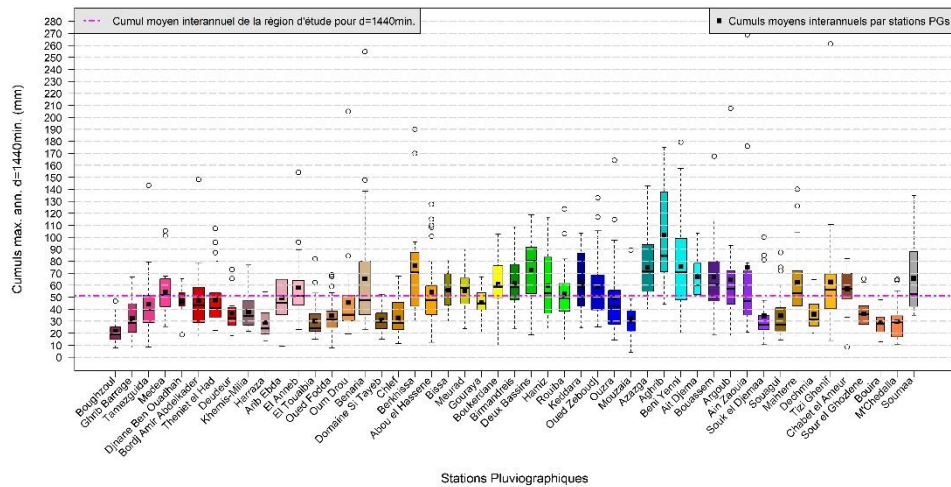


Figure A 2-8 Evolution des totaux pluviométriques pour d= 1440 minutes.

### 2.3 Diagramme de dispersions des cumuls de pluies maximales de 49 stations PG pour 8 durées

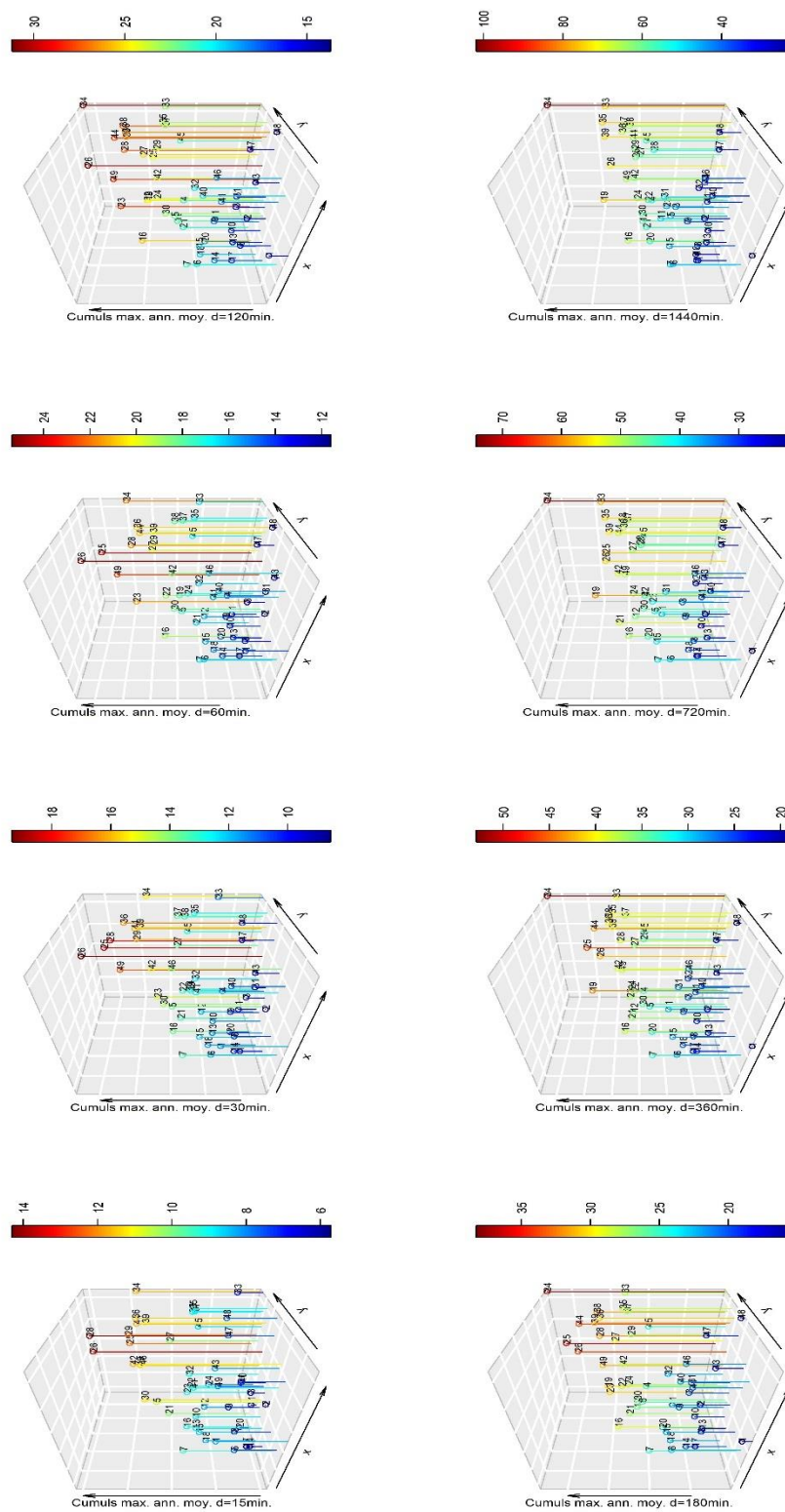
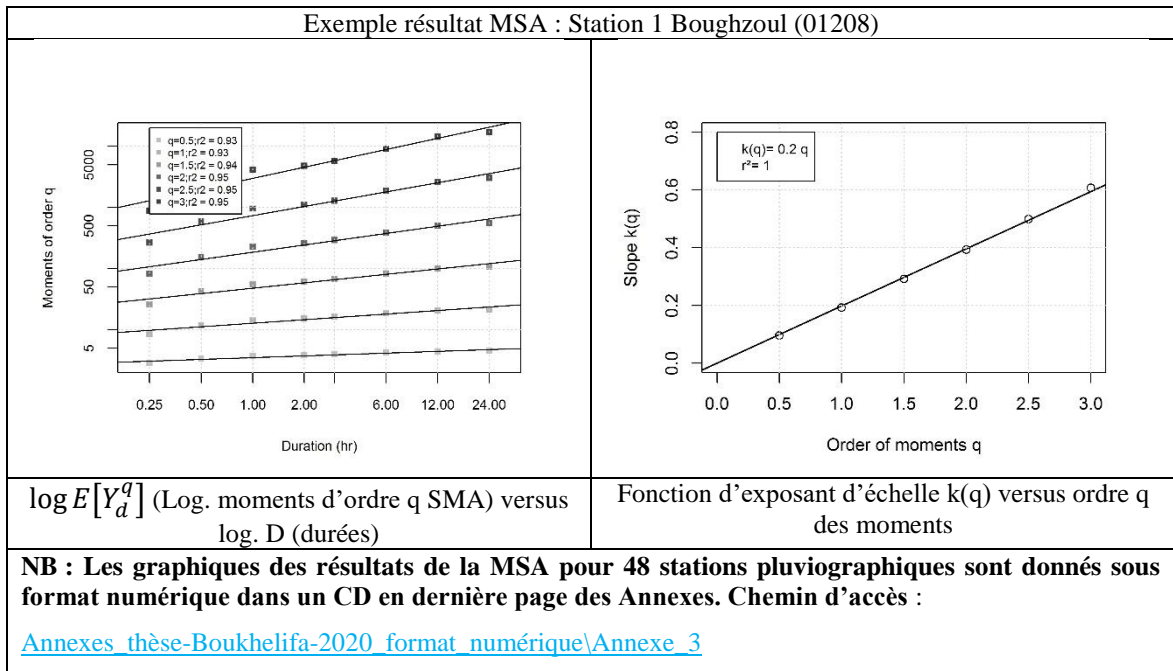


Figure A 2-9 Scatter plot des cumuls de pluies maximales de 49 stations PG

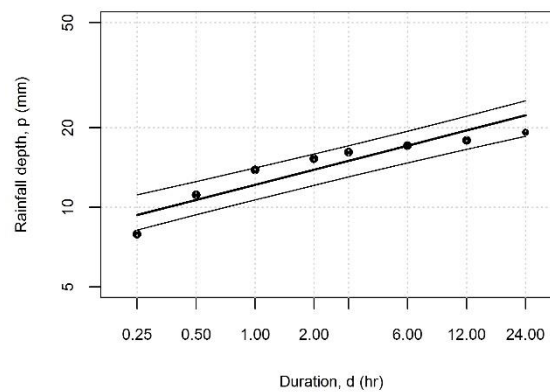
## Annexe 3 - Résultats de l'analyse d'échelle des moments de 48 SMA



## Annexe 4 - Séries simulées : médianes et incertitudes des cumuls de pluies vs. durées.

### 4.1 Gamme de durées de 15 à 1440 minutes (8 durées)

Exemple résultat médianes simulées et empiriques des SMA de pluies : Station 1 Boughzoul (011208)

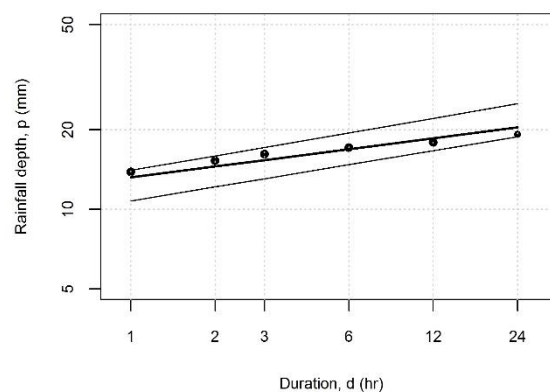


Graphique des médianes empiriques (points), médianes théoriques BayMCMC, Intervalles de crédibilités à 80% BayMCMC pour 8 durées d'agrégation

**NB : Graphiques Annexe\_4-1 sont donnés sous format numérique dans un CD en dernière page des Annexes. Chemin d'accès : [Annexes thèse-Boukhelifa-2020 format numérique\Annexe 4\Annexe 4-1](#)**

### 4.2 Gamme de durées de 60 à 1440 minutes (6 durées)

Exemple résultat médianes simulées et empiriques des SMA de pluies : Station 1 Boughzoul (011208)

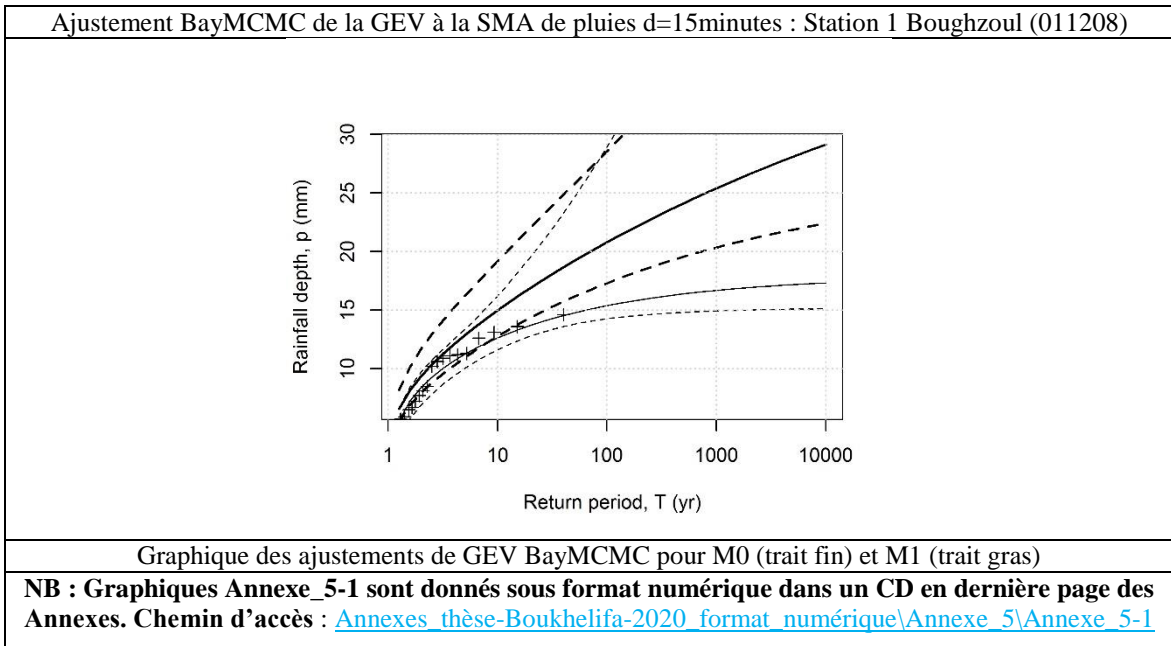


Graphique des médianes empiriques (points), médianes théoriques BayMCMC, Intervalles de crédibilités à 80% BayMCMC pour 6 durées d'agrégation

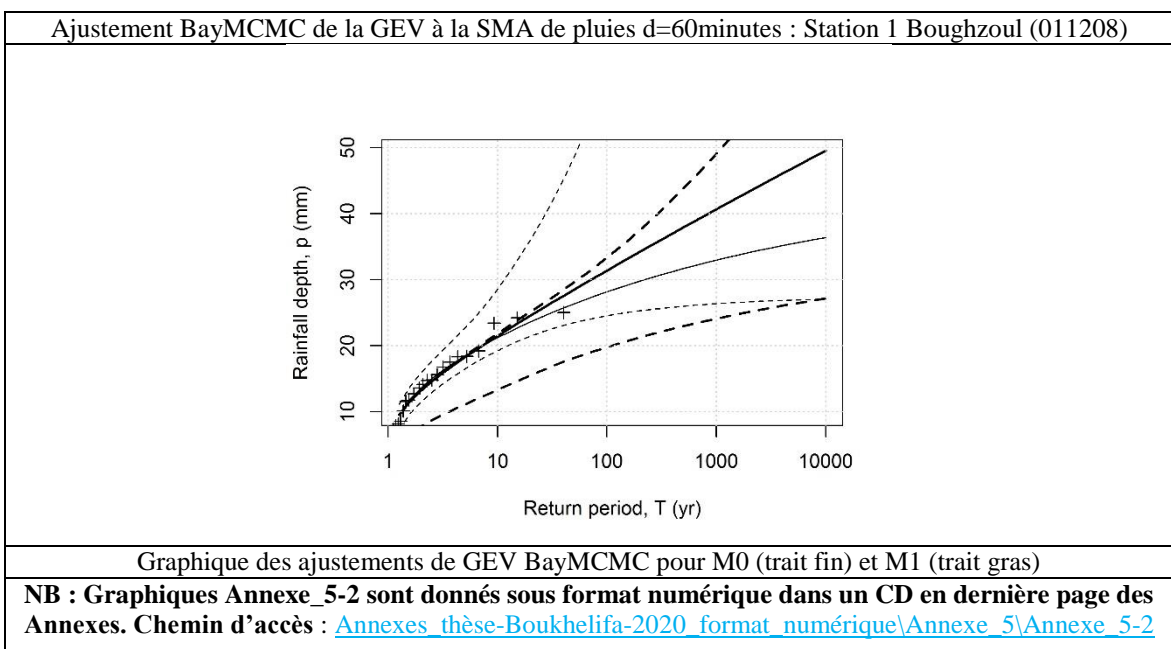
**NB : Graphiques Annexe\_4-2 sont donnés sous format numérique dans un CD en dernière page des Annexes. Chemin d'accès : [Annexes thèse-Boukhelifa-2020 format numérique\Annexe 4\Annexe 4-2](#)**

## Annexe 5 - Graphiques d'ajustements GEV par BayMCMC : M0 vs M1

### 5.1 Durées d'agrégation de $d = 15$ à 1440 minutes

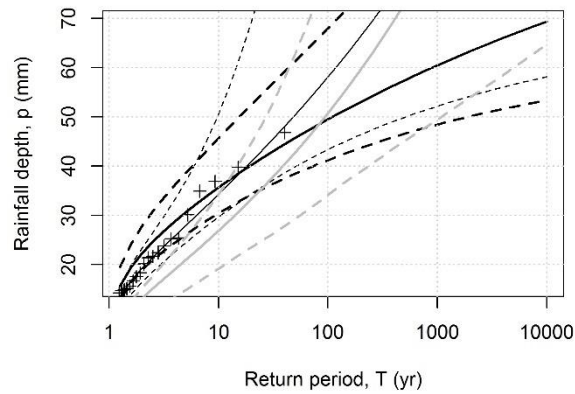


### 5.2 Durées d'agrégation de $d= 60$ à 1440 minutes



### 5.3 Durées d'agrégation de =15, 60 et 1440 minutes

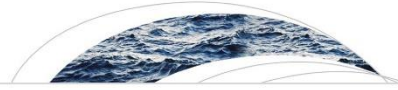
Ajustement BayMCMC de la GEV à la SMA de pluies d=1440minutes : Station 1 Boughzoul (011208)



Graphique des ajustements de GEV BayMCMC pour M0 (trait fin) et M1 (trait gras) calé sur 3 durées (15, 60 et 1440 minutes)

**NB : Graphiques Annexe\_5-3 sont donnés sous format numérique dans un CD en dernière page des Annexes. Chemin d'accès : [Annexes thèse-Boukhelifa-2020 format numérique\Annexe\\_5\Annexe\\_5-3](#)**

# Annexe 6 - Publication dans la revue scientifique Water Resources Research



## Water Resources Research

### RESEARCH ARTICLE

10.1029/2018WR023366

#### Key Points:

- The integrated Bayesian approach allows to reduce the uncertainties in rainfall quantiles estimates
- Robustness in the estimation of the IDF curves for short observed series
- More reliable and consistent estimation of the scaling factor than through traditional moment scaling analyses

#### Correspondence to:

M. Boukhelifa,  
mustapha.boukhelifa@gmail.com;  
m.boukhelifa@ensh.dz

#### Citation:

Boukhelifa, M., Meddi, M., & Gaume, E. (2018). Integrated Bayesian estimation of intensity-duration-frequency curves: Consolidation and extensive testing of a method. *Water Resources Research*, 54. <https://doi.org/10.1029/2018WR023366>

Received 23 MAY 2018

Accepted 10 SEP 2018

Accepted article online 19 SEP 2018

## Integrated Bayesian Estimation of Intensity–Duration–Frequency Curves: Consolidation and Extensive Testing of a Method

M. Boukhelifa<sup>1</sup>, M. Meddi<sup>1</sup>, and E. Gaume<sup>2</sup>

<sup>1</sup>Ecole Nationale Supérieure d'Hydraulique, Laboratoire GEE, Blida, Algeria, <sup>2</sup>IFSTTAR-GERS, Nantes, France

**Abstract** Intensity-duration-frequency (IDF) curves are one of the most common rainfall statistical models used in hydrologic design and analysis projects. The uncertainties related to the elaboration of these IDF curves have nevertheless seldom been evaluated in the past. The article will recall the existing link between the IDF formulation and some properties of the rainfall series such as simple scaling and multifractal structure. Assuming that these properties are valid, the IDF curves formulation is then the product of a dimensionless (i.e., reduced) distribution function for the annual maximum rainfall intensities/depths and a duration-dependent scaling factor. Its parameters can be evaluated in an integrated way (i.e., based on a unique pooled sample of peak intensities over a range of durations: from 15 min to 24 hr). The use of likelihood-based Bayesian Markov chain Monte Carlo statistical inference methods for this evaluation provides consistent uncertainties for all the parameters of the IDF relation and for the corresponding rainfall quantiles. This methodology has been tested, via a local analysis, on a large data set of 48 rain gauge records, spread over the north central part of Algeria (25,000 km<sup>2</sup>), under various climatic regimes. The integrated approach is undoubtedly consistent with estimates from annual maximum rainfall fitted to single durations. Furthermore, credibility intervals are significantly reduced. Also, this integrated approach appears to be robust: Unlike the traditional method based single durations, it generally provides rational quantile estimates, even when short observed series are available. This is a significant advantage for engineering applications.

### 1. Introduction

The rational method developed from the middle of the nineteenth century (Kuichling, 1889; Lloyd-Davies, 1906; Mulvaney, 1850) relies on the fact that the peak discharge at a watershed outlet, considering a constant runoff coefficient, is proportional to the maximum average rainfall intensity over a period equal to the time of concentration of this watershed. The rational method has since established itself as a reference method for the design of stormwater structures, particularly in urban areas. Its uses require the knowledge of the frequency of occurrence of rainfall intensities over different durations. This information is synthesized as intensity-duration-frequency (IDF) curves (Chow et al., 1988). Beyond design, these last years, IDF curves have also been in the issue of a large number of research works, especially for analyzing the impact of climate change on extreme precipitations (Agilan & Umamahesh, 2016; Chandra et al., 2015; Cheng & AghaKouchak, 2015; Herath et al., 2016; Lima et al., 2016; Mailhot et al., 2007; Mondal & Mujumdar, 2015; Peck et al., 2012; Shahabul Alam & Elshorbagy, 2015).

Traditionally, the IDF curves are established by fitting a statistical distribution, separately, on annual maximum series of rainfall intensities/depths over several durations. The parameters of the IDF relation are calculated empirically in a second step. Therefore, the estimation of the IDF curves proceeds in two or three calculation steps. Uncertainties are associated with each of these estimation steps (Mohyont & Demarée, 2006; Overeem et al., 2008; Van de Vyver & Demarée, 2010). In particular, they are related to the limited size of the measured data sets available for IDF curve calibration (Koutsoyiannis et al., 1998). These estimation uncertainties are difficult to evaluate in the classical estimation procedure of step-by-step IDF curves and are often ignored. However, these estimation uncertainties are an integral part of the statistical information derived from the analysis of observations and must be taken into account in risk assessment or structural design (Coles et al., 2003; Gaume, 2018).

The methodology proposed and tested here has two objectives: (i) to make the best use of the data and their properties for the calibration of the IDF curves and (ii) to evaluate the uncertainties related to the estimated parameters of the IDF models and corresponding quantiles.





Regarding the first objective, many research works have highlighted the link between the formulation of IDF relationships and the scaling properties of rainfall series (Bendjoudi et al., 1997; Burlando & Rosso, 1996; Menabde et al., 1999). In particular, in case of simple scaling, the probability distributions of rainfall intensities/depths are identical over a certain range of time scales (durations) except for specific scale factor depending on the considered duration. The distributions are then the product of a dimensionless reduced distribution and of a duration-dependent scaling factor. As in the case of regional statistical analysis (Hosking & Wallis, 1997), this property makes it possible to calibrate the reduced probability distribution function from a pooled sample of annual maximum series of intensities/depths observed over various duration. The fitted sample being larger, it is hoped by this approach to obtain more accurate estimates. Since the early works of Koutsoyiannis et al. (1998) and Menabde et al. (1999), many research works have used this simple scaling property for IDF curve calibration from pooled samples (Aronica & Freni, 2005; Bara et al., 2009; Blanchet et al., 2016; Ghanmi et al., 2016; Innocenti et al., 2017; Menabde et al., 1999; Muller et al., 2008; Nhat et al., 2007; Panthou et al., 2014; Xu & Tung, 2009; Yu et al., 2004). Most of these works proceed in two steps, first calibrating the relation between scaling factor and duration, then in a second step fitting the reduced distribution to the sample of normalized rainfall intensities/depths (i.e., divided by the scaling factor). This approach does not allow to easily evaluate the uncertainties associated with the whole calibration procedure of the IDF curves. Only a few studies have proposed integrated approaches allowing a combined calibration of the parameters of the scaling function and the reduced probability distribution parameters, opening the possibility of an assessment of the associated uncertainties (Blanchet et al., 2016; Muller et al., 2008).

Regarding the second objective of assessing uncertainties, the Bayesian approaches, considered as the most suited (Coles & Pericchi, 2003; Reis & Stedinger, 2005), have progressively developed in all the fields of statistical hydrology and also for the estimation of IDF curves (Chandra et al., 2015; Cheng & AghaKouchak, 2015; Huard et al., 2010; Rosbjerg & Madsen, 1996; Van de Vyver, 2015). The possibility of a combination of Bayesian statistical inference and the integrated approach for estimating IDF curves has been illustrated by some works (Muller et al., 2008; Van de Vyver, 2015). These publications were, however, limited to a few application sites, chosen for illustrative purposes. But the integrated approach is based on a hypothesis. The hypothesis that the possible biases introduced by the constraints set by the integrated calibration and a necessarily simplified mathematical expression of the IDF curves will be largely compensated by the increase of the size of the samples on which statistical inferences are conducted. The relevance of this hypothesis can be tested via numerical simulations: cf., for instance, chapter 7 of Hosking and Wallis (1997) on regional statistical analyzes or Halbert et al. (2016). But it is also essential to empirically test such integrated approaches and their added value over a large sample of data.

This paper consolidates and extends the previously cited work (Koutsoyiannis et al., 1998; Menabde et al., 1999; Muller et al., 2008) to propose a Bayesian approach for integrated estimation of IDF curves through a local analysis, where the established IDF curves relationships are theoretically consistent with the principle of simple scaling and multifractality. But it is also essential to empirically test such integrated approaches and their added value over a large sample of data: 48 recording rain gauges stations covering a territory of 25,000 km<sup>2</sup> in the central north of Algeria, which is done in this paper. It aims at verifying the validity of the hypothesis of the simple scaling, to assess uncertainties on rainfall quantile estimates and to highlight the benefits that can be gained from integrated Bayesian statistical inference if compared to more conventional approaches.

The article is structured in five sections. After the introduction, the second section justifies the integrated formulation of the IDF relationship and presents the principles of Bayesian inference. The third section presents and evaluates the results of applying the integrated approach to the data sample by comparing the results obtained with those of the traditional approach of fitting separately a probability distribution to each duration. In the fourth section, the assumptions underlying the integrated approach and its apparent qualities are questioned. Finally, the conclusions and perspectives are presented in the fifth section.

## 2. Methods

Some notations must be recalled before presenting the details of the methods. Let  $Y$  be a random variable. The function  $F(y) = P(Y \leq y)$  is called cumulative probability density function of the variable  $Y$ . If  $Y$  corresponds

to a yearly maximum value, the return period is defined as  $T(\mathbf{y}) = 1/[1 - F(\mathbf{y})]$ . The quantile  $y_T$  is the value  $y$ , which image through the function  $T(\mathbf{y})$  is equal to  $T$ :  $y_T = F^{-1}(1 - 1/T)$ .

### 2.1. Mathematical Formulation of the IDF Relationship

Several formulas have been proposed to relate a given rainfall intensity or depth quantile  $y_{d,T}$  to the corresponding duration  $d$  and return period  $T$ . Koutsoyiannis et al. (1998) and several authors after them suggested the use of the following formula (equation (1)):

$$y_{d,T} = \frac{a(T)}{(d + \theta)^\eta} \quad (1)$$

But other formulas may be proposed as the one in use in France or Italy (Chocat, 1997):

$$y_{d,T} = a(T)d^{b(T)} \quad (2)$$

where  $\eta$ ,  $\theta$ , and  $b$  are numerical coefficient depending on site location and climatic conditions. Even if it appears convenient to separate the functional dependence on  $d$  and  $T$  as in equation (1), this hypothesis may not be the best suited to the observed data set. This section aims at showing the existing relation between the possible form of the IDF relationship and some commonly observed statistical properties of rainfall series, namely, multifractal structure and simple scaling. Let us start from a more general formulation of the IDF relationship (equation (3)):

$$y_{d,T} = a(T)g(d) \quad (3)$$

In case of a simple scaling, the statistical distributions of rainfall intensities or depth for the various durations, at a given location, are identical apart from a duration-dependent scaling factor:

$$y_{d,T} = a(T)g(d) \quad \text{or} \quad \frac{y_{d,T}}{g(d)} = a(T) \quad (4)$$

In other words, in case of simple scaling, the function  $g$  does not depend on the return period. Note that the IDF formulation proposed by Koutsoyiannis et al. (1998; equation (1)) is a particular form of equation (4) with  $g(d) = (d + \theta)^{-b}$ . Its use entails an implicit hypothesis of simple scaling. In equation (4),  $g(d)$  is the scaling factor, depending on the duration  $d$  and  $a(T)$  is the reduced statistical distribution, which can be selected among any possible distribution type.

Furthermore, it has also been reported that annual maximum series of rainfall intensity/depths series display multifractal scaling properties. Let us note  $Y_d$  the random variable of observed annual maximum intensities/depths for duration  $d$  and  $E[Y_d^q]$  the moment of order  $q$  of this random variable. Multifractal scaling implies the following relation (equation (5)) with  $k(q)$  the so-called scaling exponent function (Burlando & Rosso, 1996; Menabde et al., 1999; Veneziano & Furcolo, 2002).

$$E[Y_d^q] = \alpha_q d^{k(q)} \quad (5)$$

If equation (4) holds, the distribution of  $Y_d/g(d)$  and its moments do not depend on  $d$ .

$$E\left[\left(\frac{Y_d}{g(d)}\right)^q\right] = \alpha_q d^{k(q)} g(d)^{-q} \quad (6)$$

This implies that the term  $d^{k(q)} g(d)^{-q}$  does not depend on  $d$ .

$$d^{-k(q)} g(d)^q = h(q) \Rightarrow g(d) = h(q)^{1/q} d^{k(q)/q} \quad (7)$$

But  $g(d)$  does by definition not depend on  $q$ , hence  $h(q)^{1/q} = cst$  and most importantly  $k(q)/q$  is constant and equal to  $k(1)$ . Two conclusions can be drawn from the combination of both hypotheses concerning the statistical properties of rainfall intensities (simple scaling and multifractality).

1. First, the function  $g(d)$  is necessarily of the form  $g(d) = \beta d^b$  with  $b = k(1)$  and  $\beta = h(q)^{1/q}$ . Hence, the IDF equation is of the following form:

$$y_{d,T} = a'(T)d^b \quad \text{with} \quad a'(T) = \beta a(T) \quad (8)$$

2. Second, the multifractal scaling exponent function should be linear:  $k(q) = q k(1)$

As a conclusion, equation (8) is the natural consequence of the combination of the two hypotheses concerning the statistical properties of rainfall intensity/depths series: simple scaling and multifractal structure. A similar conclusion had already been drawn based on a slightly different demonstration by Menabde et al. (1999). Equation (8) corresponds exactly to equation (1) when the parameter  $\theta$  is equal to 0. Several applications of that previous model came to the empirical conclusion that the parameter  $\theta$  was often close to 0 (Blanchet et al., 2016; Koutsoyiannis et al., 1998; Menabde et al., 1999; Van de Vyver, 2015). The choice of equation (8) is thus consistent with the outcome of numerous previous works.

Based on this theoretical result, it is suggested to proceed in three steps for the local calibration of the IDF curves (equation (8)):

1. Evaluate if the two hypotheses (simple scaling and multifractal structure) are reasonably consistent with the available data sets, which will justify the use of equation (8) as a model for the IDF curves. By the way, it can be verified that a linear scaling exponent function fits to the observations, which is an additional indication of the adequacy of the simple scaling hypothesis.
2. Calibrate the IDF curve with an integrated approach that is described in the next section.
3. Evaluate the consistency and added value of the integrated Bayesian approach when compared to separated Bayesian approaches (i.e., applying inferential process on each duration separately; see section 2.3).

## 2.2. Bayesian Inference Procedure

The Bayesian inference procedure used herein is very similar to the one proposed in Nguyen et al. (2014) and Halbert et al. (2016) for regional flood frequency analyses, the area of the watersheds being replaced by the duration  $d$ . This approach is included in the nsRFA package of the R statistical software (Viglione et al., 2014) and corresponds to the functions BayesianMCMC and BayesianMCMCreg. The principles are shortly presented here.

Let  $Y_d$  be a sample of  $n$  observed annual maximum rainfall intensities or depths over the duration  $d$  ( $Y_d = \{y_{d_1}^1, y_{d_2}^1, \dots, y_{d_n}^1\}$ ) and  $Y$  the corresponding sample of  $n$  annual maximum values for  $l$  different considered durations  $\{d_1, d_2, \dots, d_l\}$ .

$$\mathbf{Y} = \begin{Bmatrix} y_{d_1}^1 & \dots & y_{d_1}^n \\ \vdots & \ddots & \vdots \\ y_{d_l}^1 & \dots & y_{d_l}^n \end{Bmatrix}$$

The shape of the IDF relationship being defined (equation (8)), the parameters of the corresponding statistical model have to be estimated: that is, the parameter  $b$  as well as the parameters of the selected statistical distribution for the reduced intensities or depths  $F_{\theta}(y_d/d^b)$ . For sake of generality, a three-parameter generalized extreme value (GEV) distribution is selected in the applications presented hereafter. This distribution is often considered as well suited to describe the occurrence of hydrological variables (Koutsoyiannis, 2004a, 2004b) and has already been tested with success in the considered test region (Meddi & Toumi, 2015). Its cumulative density function has the following form, with the vector  $\theta$  of position, scale and shape parameters:  $\theta = \{\xi, \alpha, k\}$ .

$$F_{\theta}(y) = \exp \left[ - \left( 1 - \frac{k(y - \xi)}{\alpha} \right)^{1/k} \right]_{\alpha > 0} \quad (9)$$

The corresponding density function has the following expression:

$$f_{\theta}(y) = \frac{\partial F_{\theta}(y)}{\partial y} = \frac{1}{\alpha} \left( 1 - \frac{k(y - \xi)}{\alpha} \right)^{1/k-1} \exp \left[ - \left( 1 - \frac{k(y - \xi)}{\alpha} \right)^{1/k} \right]_{y > 0} \quad (10)$$

The distribution type being selected, the likelihood of the samples  $Y_d$  or  $Y$  given the values of the parameters  $\xi$ ,  $\alpha$ ,  $k$ , and  $b$  can be computed if the various observed rainfall intensities or depths are considered as independent (equation (11)):

$$\ell(Y_d|\theta) = \prod_{i=1}^n f_{\theta}(y_d^i) \quad \text{and} \quad \ell(Y|\theta, b) = \prod_{i=1}^n \prod_{j=1}^l f_{\theta} \left( \frac{y_{d_j}^i}{d^b} \right) \quad (11)$$

The consequences of the simplifying hypothesis of independence between the maximum yearly intensities/depths computed for various durations  $d$  will be discussed later on section 4.2.

The likelihood expression being defined, Bayesian Markov chain Monte Carlo (MCMC) algorithms can be used to derive the posterior probability density functions of the parameters  $\xi$ ,  $\alpha$ ,  $k$ , and  $b$  and of the corresponding rainfall intensity or depth quantiles  $(y_{d,T} = d^b F_{\theta}^{-1}(1 - 1/T))$ , according to the observed sample  $Y$ . These algorithms have become relatively common in hydrology and will not be described herein. Interested readers may refer to previous publications for more details (Gaume et al., 2010; Reis & Stedinger, 2005; Viglione et al., 2013).

The proposed procedure has two major advantages if compared to numerous previous works on IDF curves. First, the parameters of the IDF curves are not estimated in several successive steps but in one single integrated step. Second, related estimation uncertainties, in the form of posterior distributions, are evaluated.

### 2.3. Consistency and Efficiency Evaluation

Two main aspects are considered to evaluate the proposed integrated IDF calibration method and compare the results obtained with the proposed integrated approach (method M1) and the traditional approach consisting in calibrating a statistical distribution for each duration separately (method M0). First, it is verified that the calibrated distributions resulting from the proposed integrated approach, that is, constrained by the observed values at all durations, are consistent with the observed samples at the various considered durations. The Anderson-Darling (AD) statistical test will be used to compare the observed samples for each duration and the maximum likelihood calibrated integrated distributions. It is known to be a powerful test (Stephens, 1976), placing more weight than other tests in the tail of the distributions, which are the main focus of interest in IDF analyses (Laio, 2004). It has therefore often been used in previous studies on IDF curves (Ben-Zvi, 2009; Chang et al., 2016; Gargouri-Ellouze & Chebchoub, 2008; Panthou et al., 2014; Van de Vyver & Demarée, 2010). The lower the value of the AD statistics  $A^2$ , the better the fit (Ben-Zvi, 2009). According to Stephens (1977), the critical value of  $A^2$  for a significance level of 10% is equal to 1.933. This critical value should nevertheless be considered as a guideline rather than an accurate reference for the comparison for two reasons. First, it has been established for Gumbel distributions when GEV distributions, especially Fréchet-type distributions with larger variability and skewness, are considered herein. The critical value to be considered should therefore probably be larger than this reference value. Second, since the maximum likelihood distributions are calibrated against the observed data, but not only, since the other time steps are considered, the critical value would have to be adjusted, that is, reduced, for a fair comparison (Stephens, 1977). In the absence of any reference work on  $A^2$  values in situation comparable to the one presented in this paper and considering to the complexity of the question, it has been decided to stick to the standard reference critical value of  $A^2$ .

Two other criteria will be considered to evaluate the consistency and efficiency of method M1 if compared to method M0, both based on the Bayesian MCMC computed credibility intervals. The proportion of M1 model 90% credibility interval included in the M0 model 90% credibility interval (percentage of overlap) is computed to further test the consistency of the integrated approach. Finally, the ratio between the width of the two 90% credibility intervals of the M1 and M0 models (percentage of reduction) is the selected criterion to evaluate the added value or efficiency (i.e., increase of estimation accuracy) of the proposed integrated approach. The influence of the possible dependencies between the observed series at various durations on the computed M1 credibility intervals will be discussed in section 4.2.

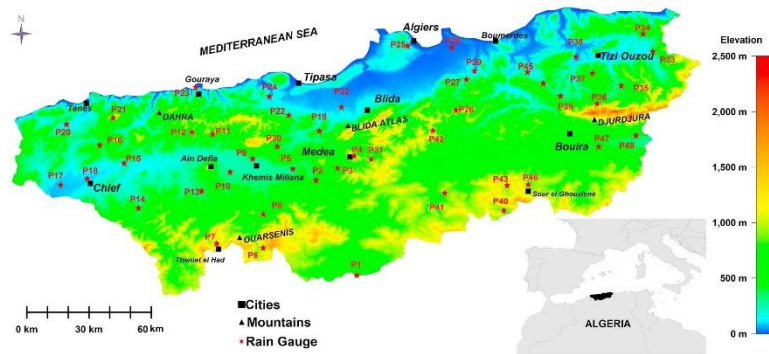


Figure 1. Study area: topography and location of the rain gauges.

### 3. Implementation and Testing

#### 3.1. Study Area and Data Set

The test case covers the north central part of Algeria (Figure 1). It is a 25,000-km<sup>2</sup> area comprised between the Mediterranean Sea in the north and the Tellien Atlas Mountains in the south. The relief is marked, with several mountain ranges and a culminating point at Mont Lalla Khadidja (2,308 m) in the Djurdjura Mountains. The climate is of Mediterranean type. The mean annual rainfall amounts ranging from 300 to 1,000 mm, with a decreasing gradient from north to south and east to west (Taibi et al., 2017; Zeroual et al., 2017) and a variability related to the relief (Meddi & Toumi, 2015). The region is frequently affected by severe flash floods as for instance the 2001 Bab el Oued catastrophic flood in Algiers, which killed more than 300 people. Improved rain and flood hazard assessment is of crucial importance.

The existing observation network entails 48 automatic rain gauges (Table 1 and Figure 1) with recording durations ranging from 8 to 38 years (22 years on average). The data analyzed hereafter have been provided by the Agence Nationale des Ressources Hydrauliques (in French; [www.anrh.dz](http://www.anrh.dz)) and consist of annual maximum series of rainfall depths cumulated over various durations for each rain gauge: 15 to 1,440 min. The observed maximum daily rainfall depth ranges from 50 to 200 mm, which is in the magnitude to the 10-year daily rainfall depths observed for instance in the South Mediterranean part of France (Arnaud et al., 2008), one the most exposed regions to flash floods in Europe and around the Mediterranean (Gaume, 2016; Gaume et al., 2009).

#### 3.2. Verification of the Simple Scaling and Multifractal Structure of the Series

Recalling section 2, the proposed simple IDF relation (equation (8)) is related to two properties of the rainfall series, namely, multifractal structure and simple scaling. Prior to the calibration of the proposed integrated IDF model, it should be verified if the data sets reasonably match these two properties. Two complementary approaches are used:

1. Verification of the log-log linearity of the relation between the moments of order  $q$  and the aggregation duration  $d$  (Burlando & Rosso, 1996; Gupta & Waymire, 1990), as well as of the linearity of the resulting so-called scaling exponent function  $k(q)$ , sign of a possible simple scaling. According to the proposal of Innocenti et al. (2017). This analysis will be called hereafter moment scaling analysis (MSA). If linear, the slope of the scaling exponent function may be considered as a first estimation of the parameter  $b$  in equation (8).
2. Implementation of the homogeneity test proposed by Hosking and Wallis (1997) to verify that the distributions of rainfall depths at various aggregation durations  $d$  can be considered as similar if properly rescaled (simple scaling). Note that this test is usually used in regional frequency analyses to test the statistical homogeneity between various sites. But it is basically a homogeneity test and is used here to compare distributions at various aggregation durations for each single site.

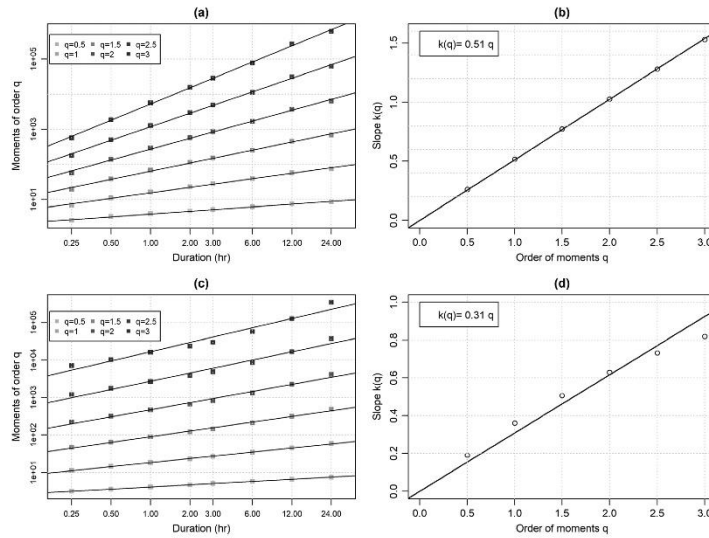


**Table 1**  
Coordinates, Altitude, and Recording Lengths of the Rain Gauges in the Study Area

Id	Gauge	Latitude (deg)	Longitude (deg)	Altitude (m)	Length (year)	Id	Gauge	Latitude (deg)	Longitude (deg)	Altitude (m)	Length (year)
1	Boughzoul	35.82	2.75	643	24	25	Birmandreis	36.79	2.96	140	16
2	Ghrib Barrage	36.22	2.48	460	15	26	Deux Bassins	36.53	3.25	890	24
3	Tamezguida	36.29	2.62	490	26	27	Hamiz	36.66	3.29	590	12
4	Medea	36.33	2.70	935	24	28	Rouiba	36.78	3.21	20	24
5	Djnane Ben Oudhah	36.27	2.40	320	8	29	Keddara	36.72	3.36	150	13
6	Bordj Amir Abdelkader	35.89	2.20	1,080	29	30	Oued Zeboudj	36.33	2.32	400	38
7	Theniet EL Had	35.93	2.00	1,160	27	31	Ouzra	36.32	2.86	930	24
8	Deurdeur	36.02	2.26	575	14	32	Mouzaia	36.49	2.68	140	14
9	Khemis Milliana	36.31	2.12	300	25	33	Azazga	36.77	4.31	500	29
10	Harraza	36.24	2.01	312	16	34	Aghrib	36.85	4.26	680	11
11	Arib Ebda	36.37	1.99	280	28	35	Beni Yenni	36.62	4.14	850	29
12	El Anneb	36.41	1.80	1,075	29	36	Ait Djema	36.54	4.06	940	10
13	El Touaibia	36.16	1.89	335	30	37	Bouassem	36.69	3.98	600	26
14	Oued Fodda	36.06	1.59	500	25	38	Argoub	36.75	3.90	100	16
15	Oum Drou	36.22	1.49	153	32	39	Ain Zaouia	36.58	3.87	260	12
16	Benaria	36.34	1.29	390	30	40	Souk el Djemaa	36.07	3.49	829	24
17	Domaine Si Tayeb	36.17	1.10	107	29	41	Souagui	36.13	3.20	850	24
18	Chlef	36.19	1.31	130	28	42	Mahterre	36.40	3.15	875	22
19	Berkhissa	36.42	2.47	400	15	43	Dechmia	36.19	3.52	1,144	20
20	Abou el Hassene	36.38	1.13	165	28	44	TiziGhenif	36.66	3.71	415	28
21	Bissa	36.46	1.38	1,100	8	45	ChabetElAmeur	36.66	3.66	260	24
22	Meurad	36.48	2.33	300	17	46	Sour El Ghozlene	36.18	3.64	882	23
23	Gouraya	36.57	1.86	34	17	47	Bouira	36.40	3.88	510	14
24	Boukerdane	36.54	2.22	110	20	48	M'Chedalla	36.41	4.22	450	29

Figure 2 illustrates the results of the MSA for two representative rain gauges. The computed determination coefficients  $R^2$  between  $\log E[Y_d^q]$  and  $\log d$  appear to be, for all gauges and tested moment order values, with some very few exceptions, are larger than 0.9. Likewise, the  $R^2$  between  $k(q)$  and  $q$  are all larger than 0.94 (see Table 2), except for rain gauge 26. A departure from the linear model (curvature) of the scaling exponent function  $k(q)$  appears nevertheless in some cases such as for the gauging station 30 (Figure 2d). It is difficult to state, due to the limited size of the data samples, if this reveals a clear departure from the theoretical simple scaling model, or if this can be attributed to the sensitivity of the larger-order moments, and hence the associated  $k(q)$  values, to sampling variability. It seems nevertheless that the rainfall events generating the largest observed daily total amount at some gauges are convective events. The total rainfall amount is then concentrated within 1 or 2 hr (daily maximum almost equal to the 2-hr maximum). Such situations reduce the slope of the  $\log E[Y_d^q]$  versus  $\log d$  relation (i.e., the value  $k(q)$ ), particularly for the higher-order moments. The predominance of such events in the annual maximum series is particularly marked at gauge 26. As a first partial conclusion, multifractality and simple scaling appear globally consistent with the available data set based on the MSA. Innocenti et al. (2017) came to the same conclusion based on the analysis of a large data set of 2,700 rain gauge records in North America: The simple scaling hypothesis appears consistent with records, especially for aggregation durations larger than 1 hr. It is nevertheless important to keep in mind the possible limits just mentioned. If the tails of the statistical distributions are controlled by convective rainfall events with specific scaling properties, observations may depart significantly from the simple scaling hypothesis. Calibrating a common reduced statistical distribution, based on a pooled sample of rescaled rainfall depths, would then introduce biases. More precisely, it could lead to underestimate low-frequency quantiles for the smallest durations  $d$  and overestimate them for the largest aggregation duration. This will have to be checked in the analysis of the implementation results of the proposed integrated model.

The homogeneity test proposed by Hosking and Wallis (1997) gives another point of view on the plausibility of the simple scaling hypothesis (Table 2). Theoretically, the  $H_1$  value should be a normal  $N(0,1)$  variable. If simple scaling holds for the rain gauge data, its average value should be around 0 for the 48 rain gauge and exceed about 10% of the time the 90% quantile of the  $N(0,1)$  distribution: 1.28. The  $H_1$  values



**Figure 2.** Log-log-linear regression between moments of order  $q$  and the aggregation duration  $d$ . (a and c) gauges 33 and 30 and adjusted linear regression to the corresponding scaling exponent function  $k(q)$ . (b and d) gauges 33 and 30.

**Table 2**

*Hosking and Wallis Homogeneity Test, H1 Values, and Determination Coefficients Between  $k(q)$  and  $q$ : Case A (15 min to 24 h) and Case B (1 hr to 24 h)*

Id	Location name	A H1	B H1	A R <sup>2</sup>	B R <sup>2</sup>	Id	Location name	A H1	B H1	A R <sup>2</sup>	B R <sup>2</sup>
1	Boughzoul	-1.06	-0.59	1.00	1.00	25	Birmandreis	-1.09	-0.48	1.00	0.99
2	Ghrib Barrage	-1.17	-0.96	1.00	1.00	26	Deux Bassins	<b>1.30</b>	0.81	0.70	0.87
3	Tamezguida	0.49	-1.29	1.00	1.00	27	Hamiz	<b>1.76</b>	<b>2.57</b>	1.00	<b>1.00</b>
4	Medea	<b>2.52</b>	-2.28	1.00	1.00	28	Rouiba	0.18	-1.21	0.96	0.99
5	Djnane Ben Oudhah	1.14	0.58	1.00	0.99	29	Keddara	-0.43	-1.18	1.00	1.00
6	Borj Amir Abdelkader	-0.71	-1.60	1.00	1.00	30	Oued Zeboudj	<b>5.10</b>	<b>1.58</b>	0.99	1.00
7	Theniet EL Had	-1.97	-1.70	1.00	1.00	31	Ouzra	<b>3.93</b>	<b>1.71</b>	1.00	1.00
8	Deurdeur	-0.05	0.06	1.00	1.00	32	Mouzaia	-1.66	-1.78	1.00	1.00
9	Khemis Miliana	0.02	-0.76	1.00	1.00	33	Azazga	0.80	-0.43	1.00	1.00
10	Harraza	-0.31	-1.54	0.98	0.98	34	Aghrib	-0.94	-0.67	1.00	1.00
11	Arib Ebda	-0.12	0.48	1.00	1.00	35	Beni Yenni	0.60	0.14	1.00	1.00
12	El Anneb	0.63	0.75	1.00	1.00	36	Ait Djema	<b>2.03</b>	<b>2.44</b>	1.00	1.00
13	El Touaibia	-0.03	-0.93	0.96	0.99	37	Bouassem	-0.16	-0.80	0.99	0.99
14	Oued Fodda	-1.31	-1.21	1.00	1.00	38	Argoub	0.27	-1.19	1.00	1.00
15	Oum Drou	-0.08	0.02	1.00	0.99	39	Ain Zaouia	-0.51	-1.83	1.00	1.00
16	Benaria	0.51	0.42	1.00	1.00	40	Souk el Djemaa	-1.48	-2.02	1.00	1.00
17	Domaine Si Tayeb	<b>3.10</b>	0.14	1.00	1.00	41	Souagui	-0.75	-0.95	0.99	0.95
18	Chlef	-2.29	-2.33	1.00	1.00	42	Mahterre	0.82	-0.40	1.00	1.00
19	Berkhissa	-2.19	-2.13	1.00	1.00	43	Dechmia	<b>2.75</b>	-1.14	1.00	1.00
20	Abou el Hassene	0.27	0.75	1.00	1.00	44	TiziGhenif	NAN	NAN	1.00	0.97
21	Bissa	-1.09	-1.38	1.00	1.00	45	ChabetElAmeur	<b>1.43</b>	0.21	1.00	1.00
22	Meurad	-1.90	-2.30	1.00	1.00	46	Sour El Ghozlene	<b>2.95</b>	-1.29	1.00	1.00
23	Gouraya	<b>1.45</b>	<b>1.69</b>	1.00	1.00	47	Bouira	-1.80	-1.92	1.00	1.00
24	Boukerdane	-2.13	-1.87	1.00	1.00	48	M'Chedalla	0.88	0.16	1.00	1.00

Note. The H1 values exceeding the threshold rejection, of homogeneity, at significance level 10% are shown in bold.



**Table 3**  
AD Statistics, Consistency, and Efficiency Evaluation Criteria

id	N	PR/d = 1 hr	PR/d = 24 hr	OP/d = 1 hr	OP/d = 24 hr	A <sup>2</sup> /d = 1 hr	A <sup>2</sup> /d = 24 hr	id	N	PR/d = 1 hr	PR/d = 24 hr	OP/d = 1 hr	OP/d = 24 hr	A <sup>2</sup> /d = 1 hr	A <sup>2</sup> /d = 24 hr
1	24	60 (62)	73 (79)	86 (65)	92 (39)	1.02 (0.3)	1.21 (0.47)	25	16	96 (96)	11 (6)	82 (73)	75 (87)	0.38 (0.29)	0.78 (0.61)
2	15	70 (65)	47 (48)	100 (86)	100 (86)	0.28 (0.35)	0.94 (0.73)	26	24	86 (81)	-438 (-540)	85 (73)	0 (0)	0.5 (0.51)	1.3 (1.26)
3	26	82 (71)	42 (23)	99 (85)	100 (92)	0.47 (0.39)	0.71 (0.38)	27	12	83 (51)	97 (93)	100 (93)	57 (64)	0.74 (0.29)	1.7 (1.33)
4	24	85 (78)	48 (33)	100 (100)	100 (100)	0.42 (1.1)	<b>2.03</b> (0.7)	28	25	79 (58)	35 (-15)	100 (93)	100 (100)	0.31 (0.6)	0.83 (0.44)
5	8	100 (100)	-13 (-9)	78 (69)	66 (99)	1.41 (1.33)	0.51 (0.45)	29	13	96 (96)	66 (61)	100 (77)	100 (81)	0.23 (0.31)	0.6 (0.65)
6	29	93 (89)	75 (68)	87 (90)	100 (94)	1.66 (0.96)	1.25 (0.79)	30	38	78 (74)	70 (69)	100 (74)	100 (100)	0.33 (0.75)	1.23 (0.66)
7	28	96 (95)	69 (67)	100 (98)	100 (100)	0.77 (0.67)	0.56 (0.6)	31	32	- (-)	86 (73)	0 (22)	81 (76)	0.9 (0.63)	0.62 (1.5)
8	14	90 (76)	96 (90)	76 (82)	69 (71)	0.28 (0.3)	1.69 (1.17)	32	14	71 (62)	70 (66)	100 (98)	100 (93)	0.36 (0.22)	0.47 (0.38)
9	25	64 (42)	97 (96)	93 (100)	100 (82)	0.75 (0.59)	1.53 (0.27)	33	29	27 (12)	74 (72)	100 (100)	100 (88)	0.93 (0.28)	1.48 (1.01)
10	16	91 (89)	74 (71)	100 (87)	100 (100)	0.31 (0.33)	0.94 (0.47)	34	11	91 (90)	83 (82)	100 (99)	100 (89)	0.3 (0.29)	0.71 (0.72)
11	28	84 (77)	46 (31)	78 (95)	100 (96)	0.64 (0.24)	1.52 (0.93)	35	29	85 (79)	69 (58)	100 (88)	100 (91)	0.29 (0.5)	0.98 (1.19)
12	29	89 (86)	64 (58)	90 (74)	100 (93)	0.63 (1.93)	<b>2.09</b> (1.31)	36	10	100 (99)	100 (100)	72 (85)	0 (59)	0.5 (0.63)	0.56 (0.59)
13	30	86 (78)	84 (77)	100 (85)	100 (100)	0.37 (0.91)	1.86 (1.07)	37	26	80 (76)	42 (32)	91 (81)	100 (100)	0.4 (0.48)	0.36 (0.28)
14	25	73 (64)	38 (34)	74 (86)	100 (100)	1.13 (0.75)	0.94 (0.36)	38	16	71 (-)	92 (75)	100 (82)	65 (98)	0.55 (0.72)	0.66 (0.42)
15	32	52 (32)	88 (85)	100 (100)	97 (83)	0.82 (0.77)	1.85 (1.03)	39	12	99 (99)	97 (95)	100 (100)	100 (100)	0.49 (0.29)	0.59 (0.33)
16	30	56 (21)	92 (89)	100 (100)	85 (75)	0.92 (0.64)	1.8 (0.73)	40	24	80 (67)	77 (69)	100 (100)	100 (100)	0.57 (0.36)	1.45 (0.65)
17	29	90 (86)	49 (44)	63 (51)	100 (97)	1.13 (1.33)	<b>3.41</b> (1.04)	41	28	96 (93)	83 (74)	82 (95)	100 (100)	0.28 (0.76)	1.93 (0.81)
18	28	70 (33)	59 (26)	100 (100)	100 (98)	0.79 (0.94)	<b>2.14</b> (0.9)	42	22	77 (70)	70 (60)	100 (89)	100 (95)	0.32 (0.39)	0.47 (0.71)
19	15	96 (91)	97 (95)	100 (100)	100 (100)	0.43 (0.44)	1.1 (0.66)	43	20	77 (68)	84 (80)	100 (100)	100 (100)	0.32 (0.54)	0.85 (0.98)
20	28	37 (-52)	83 (63)	83 (66)	81 (82)	1.34 (1.04)	<b>2.5</b> (1.76)	44	28	93 (90)	48 (34)	100 (98)	100 (100)	1.14 (1.54)	0.66 (0.58)
21	8	100 (100)	74 (60)	0 (20)	100 (100)	0.76 (0.66)	0.65 (0.81)	45	24	62 (33)	-139 (-305)	65 (69)	0 (14)	0.85 (1.14)	0.81 (0.74)
22	17	90 (81)	40 (9)	82 (86)	100 (94)	0.71 (0.33)	0.55 (0.28)	46	23	68 (54)	68 (50)	100 (76)	100 (80)	0.25 (0.47)	0.31 (0.62)
23	17	98 (98)	-91 (-47)	67 (50)	0 (58)	1.59 ( <b>1.99</b> )	<b>2.31</b> (0.69)	47	14	93 (90)	75 (65)	76 (77)	100 (90)	0.35 (0.74)	0.7 (0.37)
24	20	19 (-19)	46 (26)	84 (84)	99 (85)	0.46 (0.34)	1.02 (0.76)	48	29	77 (73)	80 (76)	99 (68)	100 (77)	0.26 (0.46)	0.62 (0.87)

Note. N = series length. Value between brackets correspond to the case B. A<sup>2</sup> value in bold exceeds A<sup>2</sup> critical value (1.933) for a significance level of 10%. PR = percentage of reduction of the range of the 90% credibility interval for T = 100 years. OP = overlap percentage between credibility intervals computed for each separate duration and with the integrated model.

exceeding this threshold are shown in bold in Table 2. In case B, where aggregation duration between 1 and 24 hr are considered, the threshold is exceeded at five gauges only (i.e., about 10% of the samples). The scale range 1–24 hr is often the range over which simple scaling has been considered to be valid in several previous works (Blanchet et al., 2016; Burlando & Rosso, 1996; Panthou et al., 2014; Yu et al., 2004). In case A (durations considered ranging from 15 min to 24 hr), the average H1 value (0.20) remains small but 11 computed values (almost ¼ of the sample) exceed 1.28. The validity of simple scaling over this larger range of durations is doubtful. This also is in accordance with numerous previous works. It is important to mention that the test does not account for the dependencies existing between the rainfall depths annual maximum series available at different time steps. The conclusions of the test may therefore be considered as a little optimistic. Even if less consistent with the hypotheses of the proposed integrated IDF model, case A will also be considered for sake of comparison, in the next sections.

**3.3. Results of the Statistical Inference**

According to section 2.2, four parameters have to be estimated through the statistical Bayesian inference procedure: the three parameters of the reduced GEV distribution ( $\xi$ ,  $\alpha$ , and  $\kappa$ ) and the scaling parameter  $b$ . The inference procedure consists basically in a random sampling of a series of probable parameter sets, typically some thousands, according to the observed data. The rainfall depth quantiles corresponding to these parameter sets can then be computed and credibility intervals estimated based on these thousands of sampled quantile values. It is pointed out that the Bayesian MCMC inference process is conducted with *noninformative* priors. Although the use of informative prior for instance, by bounding the possible values of the shape parameter  $\kappa$  is feasible and avoids, eventually, unrealistic estimates. However, it was not considered necessary to introduce an additional level of complexity. Indeed, the estimates (noninformative prior) of the shape parameters are consistent with L-moment estimates in the same study region (Meddi & Toumi, 2015; Papalexiou & Koutsoyiannis, 2013). Table 3 provides the results obtained for the consistency and efficiency criterions presented in section 2.3 for all gauging stations and the two cases A and B. The 100-year quantile has been selected for the computation of the criterions in Table 3, but the results are consistent for all quantiles. For sake of clarity only the results for the durations 1 and 24 hr are presented below. The results for the other aggregation durations are similar. Figure 4 illustrates the results for gauge 33, a station that passed perfectly the homogeneity tests, and gauge 30 that exhibited doubtful homogeneity and simple scaling.



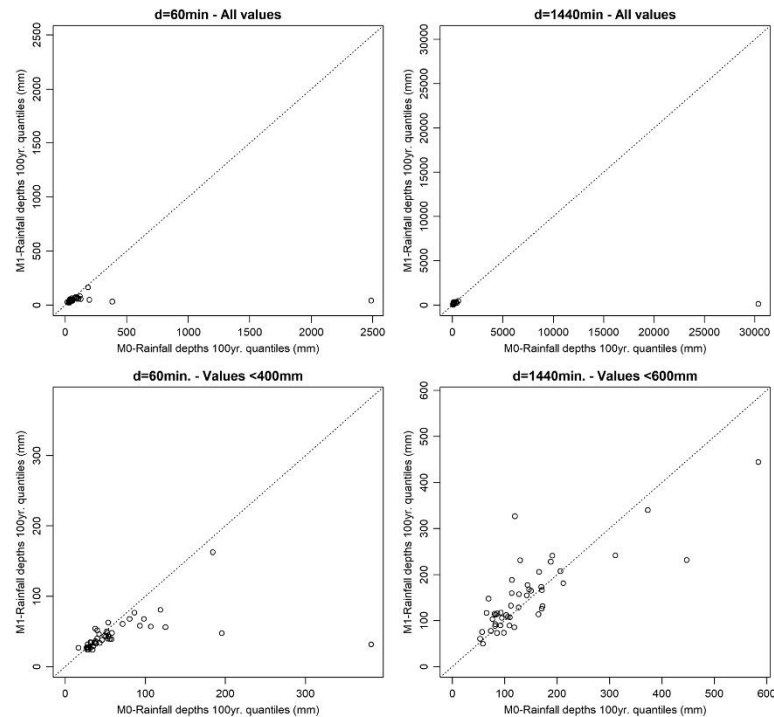


Except for some rare cases, the tested integrated approach (M1) appears to provide smaller credibility intervals than the inferences based on separate durations (M0): positive percentage of reduction. The percentage of reduction of the 100-year quantile between methods M0 and M1 is on average equal to 70% for the various considered rain gauges and durations (Table 3). This is of course not totally surprising. The credibility intervals computed through the Bayesian inference procedure reflect the information content of the analyzed data sets and hence are related to their sizes. This also explains why the percentage of reduction is generally larger for case A if compared to case B. But this general trend—that is, reduction of the credibility intervals with the amount of data—could nevertheless be counterbalanced by an inadequacy between the data set and the calibrated statistical model, for instance, if the pooled data sets were statistically heterogeneous. The results show that even if the proposed statistical simple scaling model is not perfectly suited to the data sets, the trade-off between the amount of data and the model adequacy seems positive according to the computed credibility intervals. This is particularly well revealed by case A, which shows on average better performances than case B, despite a lower consistency with the simple scaling hypothesis. It must nevertheless be kept in mind that possible dependency between the observed annual maximum rainfall depths at various durations as well as heterogeneities in the distributions have not been accounted for. The likelihood formulation used in the integrated method assumes independence and homogeneity. The credibility intervals computed with method M1 are therefore certainly underestimated: see discussion in the next section or results based on simulated heterogeneous samples provided in previous publications (Halbert et al., 2016).

More importantly, the distributions calibrated through the integrate method M1 appear consistent with the distributions calibrated on separate durations (generally good overlap between the credibility intervals) and with the measured data sets as illustrated by the AD test results (Table 3). The overlap often reaches 100%. That is, the integrated approach improves the accuracy of the estimated quantile if compared to the reference method M0. It is significantly lower than 50%, sometimes equal to 0%, for a reduced proportion of rain gauges: less than 10%. Note that the comparisons are based on computed 90% credibility intervals. The unknown effective distribution, that the statistical inference procedure is seeking to evaluate, has theoretically 10% chance to lie outside the computed intervals. It is therefore not surprising that the integrated approach, providing additional information, may sometimes reorient significantly the adjusted distribution. This, combined with the possible underestimation of the credibility intervals for the method M1, may lead to no overlap in some cases. It has been moreover been verified that 100-year quantiles estimated with the integrated method do make sense if compared to the values estimated at the nearby gauges. For some gauges with short recorded series, the traditional statistical inference method M0 may lead to clearly absurd quantile estimations (see example of gauges 5 and 36, respectively for a duration of 1 and 24 hr in Figure 3). The integrated method M1 does not provide such absurd estimates. Further tests in the next section will confirm the relative robustness of the integrated method. The AD test comparing the maximum likelihood distributions obtained through the integrated method M1 with the measured data sets also provides extremely satisfactory results (Table 3). Despite the important constraints set to the integrated IDF model, the adjusted distributions remain consistent with the observed data sets at all durations. The number of values exceeding the 10% significance level is a little high for the 24-hr rainfall depths when subhourly data are accounted for (case A): six values in bold in Table 3. This may be interpreted as an indication that the simple scaling hypothesis over the whole range of considered durations in case A (15 to 1,440 min) is not perfectly suited to the observed data sets.

On the other hand, the integrated method does not show a clear tendency to reduce the estimated 1-hr quantiles and increase the 24-hr quantiles as illustrated in Figure 4. This is also confirmed for the gauges where the AD threshold is exceeded. The possible bias due to a departure from the simple scaling model related to a growing proportion of convective events for high return periods may exist but is not detectable in the results and has probably a limited influence for the estimation of 10- to 100-year quantiles.

Finally, Figure 5 illustrates the posterior distributions for the calibrated scaling factor  $b$  compared with the slopes resulting from the MSA. Both estimates appear overall consistent but the results of the MSA are much more variable: that is, sensitive to sampling variability and estimation method. The MSA estimates sometimes lie outside the range of possible values according to the integrated method: That is, they are inconsistent with the GEV distribution and the data sets. This illustrates that the integrated method should certainly be preferred to a two-step method such as the one recently proposed by Innocenti et al. (2017) consisting in

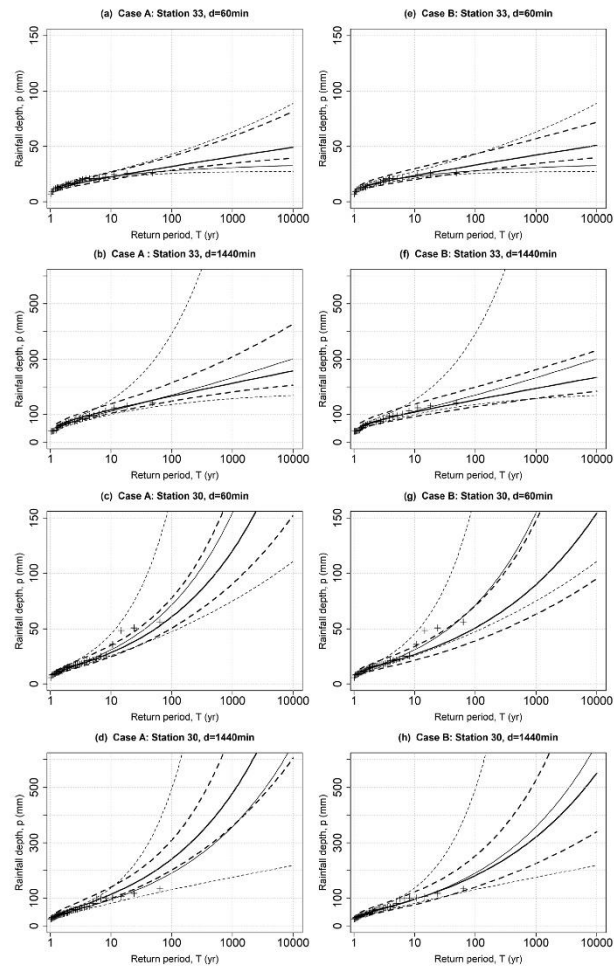


**Figure 3.** Rainfall depths 100-year quantiles: M0 model versus M1 model.

calibrating first the scaling factor  $b$  based on the MSA and then the reduced statistical distribution. It will provide more consistent and robust outcomes. The calibrated scaling factor values are difficult to compare due to estimations uncertainties. Most of these values seem to be comprised between 0.3 and 0.4. Gauge 1 is outlying with a significantly lower estimated value, sign that maximum annual values may be due to short-duration convective events. This is consistent with its location in the south subdesert part of the area. Likewise, the rain gauge with the apparent largest estimated  $b$  values (longer lasting large rainfall events) appears to be close to the Mediterranean Sea. This spatial consistence suggests that the proposed integrated model could be regionalized in the future.

#### 4. Discussion

The previously presented figures, based on a relatively large but yet finite data set, leave some open questions. First, the adequacy of the scaling factor expression  $d^b$  in equation (8) has only been indirectly evaluated through the MSA and some theoretical derivations. Departure from this log-linear relation could be a major source of heterogeneity and may limit the added value of the integrated approach. Likewise, the likelihood formulation is based on an assumption of independence, which is probably not met for the pooled samples of rescaled maximum rainfall depths observed at various durations. To which extent is then the theoretically computed credibility intervals for the proposed integrated approach undervalued? Finally, the integrated approach seems to provide robust quantile estimates (i.e., estimates less sensitive to sampling variability). This seems particularly well illustrated for gauges with limited data samples such as gauges 36 (Figures 8a

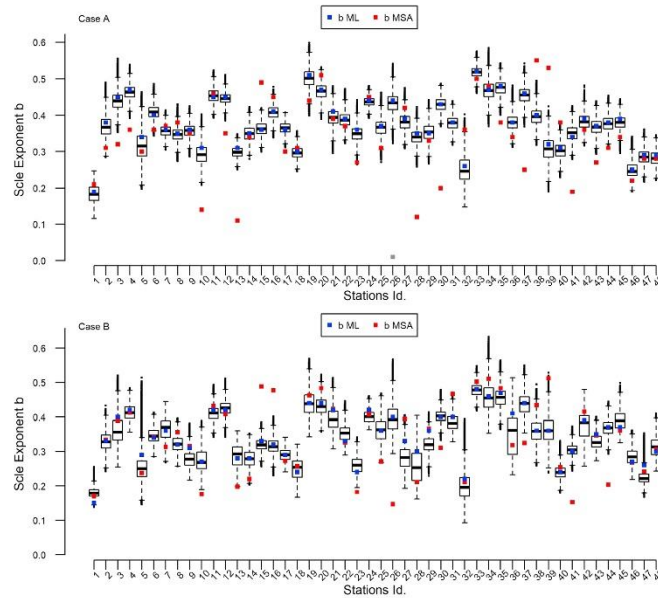


**Figure 4.** Rainfall depths annual maximums series fitted to a generalized extreme value distribution. (a–d) For the Case A (15 min to 24 h) and (e–h) for Case B (60 min to 24 h). Crosses for observed values. Thin lines: M0 model. Thick lines: M1 model. Dashed lines: 90% credibility intervals.

and 8b) and by the lower proportion of outlying estimated quantiles (Figure 3). Is it possible to confirm this property in a more systematic way? These three questions are further discussed hereafter.

#### 4.1. Adequacy of the Relation Between Duration and Scaling Factor

To confirm that the calibrated scaling factor relation  $d^b$  is reasonably suited to the observed data sets, the medians of the observed samples are compared to the medians corresponding to the calibrated IDF distributions. The median has been selected rather than the mean value or any other quantile value because of its limited sensitivity to sampling variability. A Monte Carlo simulation is conducted to evaluate the 80%

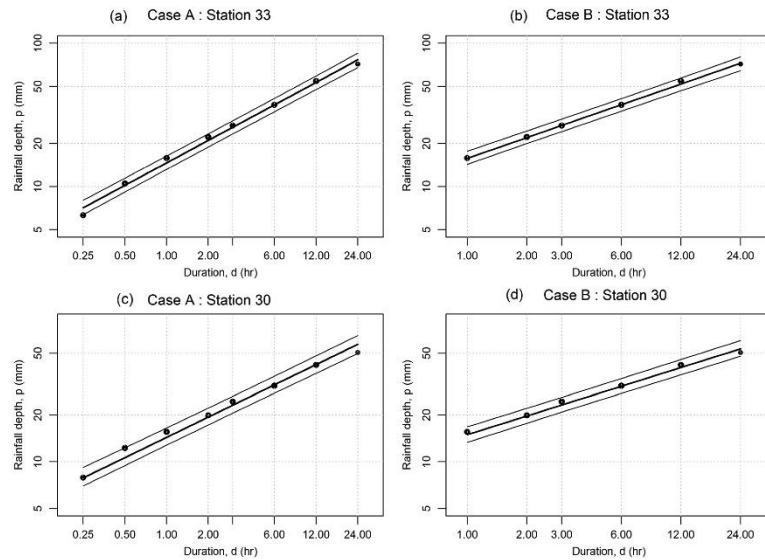


**Figure 5.** Posterior distributions of the estimated scaling exponents  $b$  for cases A and B (Bayesian Markov chain Monte Carlo procedure). Blue squares (maximum likelihood, ML), red square (estimation based on the moment scaling analysis, MSA).

confidence intervals for the medians corresponding to the calibrated distributions (Figure 6). For each rain gauge and duration, a series of depths of the same length as the observed samples is randomly drawn for each of the parameter sets sampled through the Bayesian MCMC calibration procedure. The confidence intervals hence account for both: uncertainties in the calibrated parameter values of the IDF model and sampling variability. Note that this analysis is very similar to Figure 2 for the first-order moment:  $b = k(1)$ . But the confidence intervals help measure the distance between observed and modeled values. As illustrated in Figure 6, a large proportion of the observed sample median values is comprised in the 80% estimated confidence intervals, for both cases A and B. Moreover, the case A and B calibrated scaling factor relations are very similar. This is confirmed for all rain gauges. This, along with the limited width of the 80% confidence intervals, confirms the adequacy of the proposed scaling relation and does not point the need for introducing more complex relations with a larger number of parameters such as in previous works (Koutsoyiannis et al., 1998; Muller et al., 2008).

#### 4.2. Sensitivity of the Computed Credibility Intervals to Dependencies Between Observations

Annual maximum rainfall depths for various durations are not independent as illustrated in Table 4. The computations presented in the previous section are based on a likelihood formulation assuming independence between the various observations thus overrate the information content of the pooled samples and underestimates the widths of the theoretically computed credibility intervals for the proposed integrated approach. It has been proposed in some previous works to introduce multivariate models to account for dependencies in the formulation of an integrated likelihood (Muller et al., 2008). This nevertheless adds complexity to the proposed integrated model and necessitates the choice of a multivariate model with an increased number of parameters that can probably hardly be calibrated and validated against observed data sets. It is proposed here to evaluate the possible influence of dependence on the results with a simpler approach. The same integrated IDF model is calibrated but based on a much more limited number of



**Figure 6.** (a–d) Evolutions of the median rainfall depths with duration (log-log scales). Estimates based on the measured samples (black dots) and values corresponding to the Bayesian inference results for the integrated intensity-duration-frequency method: expected value (bold line) and 80% credibility intervals (thin lines).

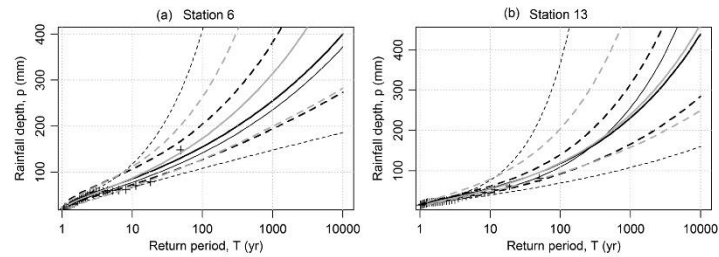
durations (15, 60, and 1,440 min), for which annual maximum observed rainfall depths series can be considered as relatively independent (Kieffer Weisse, 1998). The resulting percentages of reduction of the credibility intervals if compared to method M0 are diminished but remain significant for almost all the rain gauges (see Figures 7a and 7b for an illustration). The real gain of the integrated approach in terms of quantile estimation accuracy is impossible to evaluate precisely without knowing precisely the dependence model between maximum rainfall depths observed at various durations. But the existence of dependence does not appear to change substantially the conclusions of the previous sections. The gain of the integrated approach is substantial.

#### 4.3. Robustness of the Integrated Approach

The estimated quantiles as well as the credibility intervals are dependent on the samples. The integrated IDF calibration approaches sometimes lead to a significant reorientation of the calibrated distributions especially for short-duration series with a resulting limited overlap between the computed M0 and M1 credibility intervals (gauges 5, 21, or 36, for instance, in Table 3 and illustrations in Figures 8a–8d). Calibrations based on single durations (M0) sometimes lead to quantile estimates that are absurd (much too high) or inconsistent with the values estimated at the nearby gauges. With exception of the daily quantile estimate at gauge 26, there is always an exception, the integrated approach M1 appears to provide reasonable estimates at all sites even when the observed series is of limited size (see Figures 8a–8d: gauges 36 or 5). This apparent robustness is an additional property that seems to confirm what has been illustrated through the computed credibility intervals: The maximum depths measured at a given duration provide useful information for conditioning the statistical distributions of maximum depths at other durations. Monte

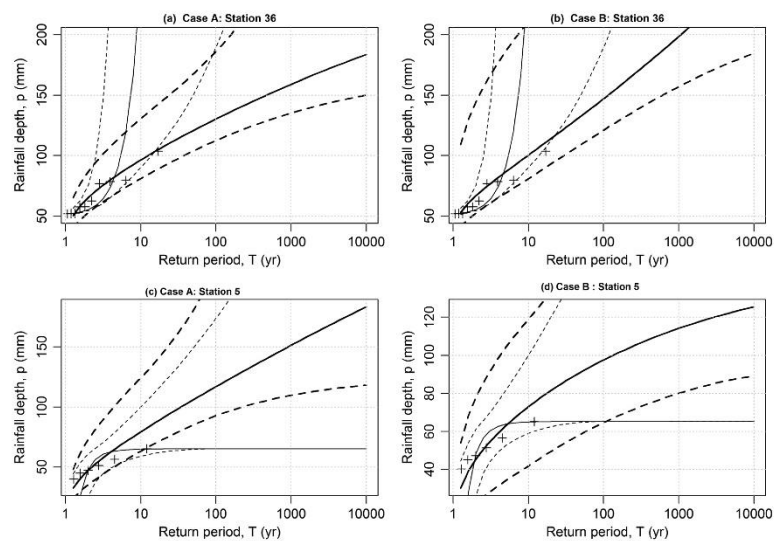
**Table 4**  
Average Coefficients of Correlation Between the Series of Rainfall Depths Measured at Various Durations for the 48 Rain Gauges

Duration (min)	15	30	60	120	180	360	720	1,440
15	1	—	—	—	—	—	—	—
30	0.9	1	—	—	—	—	—	—
60	0.74	0.91	1	—	—	—	—	—
120	0.6	0.76	0.9	1	—	—	—	—
180	0.51	0.66	0.81	0.95	1	—	—	—
360	0.39	0.5	0.62	0.77	0.88	1	—	—
720	0.25	0.34	0.43	0.57	0.68	0.87	1	—
1,440	0.18	0.25	0.32	0.44	0.53	0.73	0.93	1

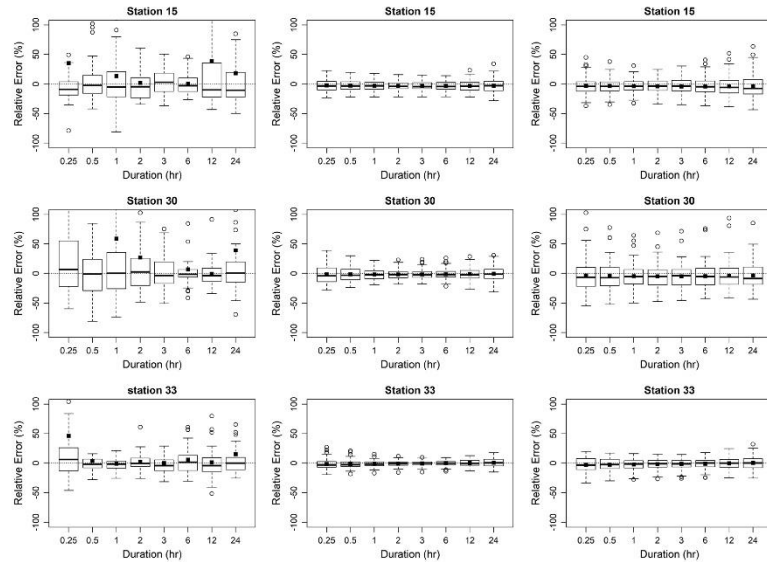


**Figure 7.** (a, b) Rainfall depths annual maximums series fitted to a generalized extreme value distribution. Crosses for observed values, thin lines for method M0, bold lines for method M1, and dashed lines for 90% credibility intervals. The bold gray lines illustrate the results obtained with the method M1 calibrated with three durations only (15 min, 60 min, and 24 h).

Carlo simulations have been conducted to confirm this apparent robustness. Three stations with relatively long observed series from various sub-regions have been selected: stations 15, 30 and 33 (see Figure 1 for location), with respective recording durations of 32, 38, and 29 years. Reference maximum likelihood 100-year quantiles for the various durations have been estimated with methods M0 and M1 based on the complete observed series. For each station, 100 series of 10 years have been drawn randomly with replacement from the observed series. The various durations were drawn independently (case 1) or all duration from a given year drawn at the same time to maintain the possible dependence between the rainfall depths at various durations (case 2). The maximum likelihood 100-year quantiles were then estimated again based on these randomly drawn 10-year series and the corresponding relative error if compared to the reference estimate computed.



**Figure 8.** (a–d) Rainfall depths annual maximums series fitted to a generalized extreme value distribution. Crosses for observed values, thin lines for method M0, bold lines for method M1, and dashed lines for 90% credibility intervals.



**Figure 9.** Boxplot of the relative errors between reference quantiles ( $T = 10$  years) estimated for both separated (M0) and integrated (M1) duration Bayesian intensity-duration-frequency model and quantiles based on 10-year randomly drawn annual maximums series. Left column: M0 method, central column M1 method case 1, and the right column M1 method case 2.

Figure 9 confirms without doubts the robustness of the integrated method M1 if compared to method M0. The comparison between the central (case 1) and the right (case 2) columns of Figure 9 helps measure the influence of the dependence between rainfall depth observed at various time steps. It is significant, but the added value of the integrated method M1 if compared to method M0 remains very significant even if this dependence is considered.

## 5. Conclusion

A method for an integrated Bayesian estimation of the IDF curves has been thoroughly tested and evaluated against a relatively large and rich rainfall data set from the north central region of Algeria. The method enables the evaluation of all parameters of IDF curves in one single step and provides a comprehensive and consistent estimation of the associated uncertainties in the form of posterior distributions and credibility intervals for the IDF model parameters and the corresponding quantiles, given the observed values. It must nevertheless be recalled that dependence between observations at various time steps can hardly be accounted for in the inference procedure, even if some few previous works introduced some partial solutions to this question (Muller et al., 2008). The resulting posterior credibility intervals should therefore probably considered as underestimated.

Apart from the consistent calibration of IDF curves, the integrated approach also appears to significantly reduce uncertainties related to statistical inference if compared to the traditional approach consisting in calibrating statistical models for each duration separately. The presented results confirm that, provided some assumptions consistent with the observations (i.e., simple scaling), annual maximum series of rainfall intensities or depths measured at a given duration contain valuable information that help constraining the statistical distributions of annual maximum rainfall intensities or depths at other time steps.



The basics of the proposed approach have already been presented in several previous works, but it is the first time, to our knowledge, that it is tested and validated on such a large data set. Further tests and validation in other climatic contexts will certainly be needed, but the recent confirmation that the simple scaling assumption seems to be reasonably valid over large areas such as North America (Innocenti et al., 2017) indicate that such integrated inference approaches could be widely used. Moreover, the proposed simple approach can relatively straightforwardly be implemented in a regional version if homogeneities in the statistical distributions as well as in the values of the scaling exponent are identified between gauges (Halbert et al., 2016). This clearly opens a large spectrum of perspectives for local and regional combinations of data pooling with the objective (1) to reduce the statistical inference uncertainties to estimated rainfall quantiles for design purposes and risk assessment and (2) to try to explore the shape of tails of rainfall intensity statistical distributions.

This being said, all statistical inference methods, especially methods aiming at creating larger pooled data samples to reduce inference uncertainties, are based on simple if not simplistic hypotheses, such as the simple scaling hypothesis in the present case study. It can be verified that these hypotheses are not totally inconsistent with the available observations, but this does not mean that they are perfectly valid. These hypotheses constrain the inference results and may therefore introduce biases if not perfectly suited to the observations, which is probably always the case. Pooling techniques are thus based on a trade-off between uncertainty and bias: that is, between the size of the data sample and the adequacy of the calibrated model. Interestingly, the results obtained here for the time step range 15 min to 1 day indicate that the trade-off may remain positive even in case of a detectable departure for the underlying simple scaling hypothesis. Ribatet et al. (2007) came to the same conclusions applying regional flood frequency analyses: More accurate discharge quantile estimates were obtained if the regional homogeneity criteria were relaxed to some extent, to build larger statistically homogeneous regions. Beyond the methods, this question of trade-off or balance is clearly a central question for statistical methods based on data pooling and could be a matter of particular attention for future research on this issue.

#### Acknowledgments

L'Agence Nationale des ressources Hydraulique (ANRH) is gratefully acknowledge for providing the rainfall records of 48 automatic rain gauges from the central north part of Algeria. These data are available from ANRH under a reasoned request (contact at [www.anrh.dz](http://www.anrh.dz)). The authors acknowledge the constructive comments and suggestions of the Associate Editor and two reviewers. The first author thanks all the members of the laboratory Eau et Environnement at IFSTTAR (Bouguenais, Nantes, France) for hosting him for a year. He also thanks the laboratory GEE at ENSH (Blida, Algeria) for facilitating and supervising the preparation of this paper as part of his *doctorat ès science* thesis. The authors declare no conflict of interest.

#### References

- Agilan, V., & Umamahesh, N. V. (2016). Is the covariate based non-stationary rainfall IDF curve capable of encompassing future rainfall changes? *Journal of Hydrology*, 541, 1441–1455. <https://doi.org/10.1016/j.jhydrol.2016.08.052>
- Arnaud, P., Lavabre, J., Sol, B., & Desouches, C. (2008). Régionalisation d'un générateur de pluies horaires sur la France métropolitaine pour la connaissance de l'aléa pluviographique/Regionalization of an hourly rainfall generating model over metropolitan France for flood hazard estimation. *Hydrological Sciences Journal*, 53(1), 34–47. <https://doi.org/10.1623/hysj.53.1.34>
- Aronica, G. T., & Freni, G. (2005). Estimation of sub-hourly DDF curves using scaling properties of hourly and sub-hourly data at partially gauged site. *Atmospheric Research*, 77(1–4), 114–123. <https://doi.org/10.1016/j.atmosres.2004.10.025>
- Bara, M., Kohnova, S., Gaal, L., Szolgay, J., & Hlav'cova, K. (2009). Estimation of IDF curves of extreme rainfall by simple scaling in Slovakia. *Contributions to Geophysics and Geodesy*, 39(3), 187–206. Retrieved from [http://www.mis.sav.sk/journals/uploads/11270958Bara-et-al\\_CGG-39-3.pdf](http://www.mis.sav.sk/journals/uploads/11270958Bara-et-al_CGG-39-3.pdf)
- Bendjoudi, H., Hubert, P., Schertzer, D., & Lovejoy, S. (1997). Interprétation multifractale des courbes intensité-durée-fréquence des précipitations. *Comptes Rendus de l'Académie Des Sciences-Series IIA-Earth and Planetary Science*, 325(S), 323–326. Retrieved from <http://www.sciencedirect.com/science/article/pii/S1251805097813791>
- Ben-Zvi, A. (2009). Rainfall intensity–duration–frequency relationships derived from large partial duration series. *Journal of Hydrology*, 367(1–2), 104–114. <https://doi.org/10.1016/j.jhydrol.2009.01.007>
- Blanchet, J., Ceresetti, D., Molinié, G., & Creutin, J.-D. (2016). A regional GEV scale-invariant framework for intensity–duration–frequency analysis. *Journal of Hydrology*, 540, 82–95. <https://doi.org/10.1016/j.jhydrol.2016.06.007>
- Burlando, P., & Rosso, R. (1996). Scaling and multiscaling models of depth-duration-frequency curves for storm precipitation. *Journal of Hydrology*, 187(1–2), 45–64. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0022169496030867>. [https://doi.org/10.1016/S0022-1694\(96\)03086-7](https://doi.org/10.1016/S0022-1694(96)03086-7)
- Chandra, R., Saha, U., & Mujumdar, P. P. (2015). Model and parameter uncertainty in IDF relationships under climate change. *Advances in Water Resources*, 79, 127–139. <https://doi.org/10.1016/j.advwatres.2015.02.011>
- Chang, K. B., Lai, S. H., & Othman, F. (2016). Comparison of annual maximum and partial duration series for derivation of rainfall intensity-duration-frequency relationships in peninsular Malaysia. *Journal of Hydrologic Engineering*, 21(1), 05015013. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001262](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001262)
- Cheng, L., & AghaKouchak, A. (2015). Nonstationary precipitation intensity-duration-frequency curves for infrastructure Design in a Changing Climate. *Scientific Reports*, 4(1). <https://doi.org/10.1038/srep07093>
- Chocat, B. (1997). *Encyclopédie de l'hydrologie urbaine et de l'assainissement* (4th ed.). Paris: Hermès - Lavoisier. Retrieved from [BTP/Livre/encyclopedie-de-l-hydrologie-urbaine-et-de-l-assainissement-9782743001261](http://BTP/Livre/encyclopedie-de-l-hydrologie-urbaine-et-de-l-assainissement-9782743001261)
- Chow, V., Maidment, D., & Mays, L. (1988). *Applied hydrology* (1st ed.). New York: McGraw-Hill Science/Engineering/Math.
- Coles, S., & Pericchi, L. (2003). Anticipating catastrophes through extreme value modelling. *Journal of the Royal Statistical Society: Series C: Applied Statistics*, 52(4), 405–416. <https://doi.org/10.1111/1467-9876.00413>
- Coles, S., Pericchi, L. R., & Sisson, S. (2003). A fully probabilistic approach to extreme rainfall modeling. *Journal of Hydrology*, 273(1–4), 35–50. [https://doi.org/10.1016/S0022-1694\(02\)00353-0](https://doi.org/10.1016/S0022-1694(02)00353-0)





- Gargouri-Ellouze, E., & Chebchoub, A. (2008). Modélisation de la structure de dépendance hauteur—durée d'événements pluvieux par la copule de Gumbel. *Hydrological Sciences Journal*, 53(4), 802–817. <https://doi.org/10.1623/hysj.53.4.802>
- Gaume, E. (2016). In Alliance nationale de recherche pour l'environnement (France) (Ed.), *The Mediterranean region under climate change: A scientific update*. Marseille: IRD éditions.
- Gaume, E. (2018). Flood frequency analysis: The Bayesian choice. *Wiley Interdisciplinary Reviews Water*, 5(4), e1290. <https://doi.org/10.1002/wat.2.1290>
- Gaume, E., Bain, V., Bernardara, P., Newinger, O., Barbuc, M., Bateman, A., et al. (2009). A compilation of data on European flash floods. *Journal of Hydrology*, 367(1–2), 70–78. <https://doi.org/10.1016/j.jhydrol.2008.12.028>
- Gaume, E., Gaál, L., Viglione, A., Szolgay, J., Kohnová, S., & Blöschl, G. (2010). Bayesian MCMC approach to regional flood frequency analyses involving extraordinary flood events at ungauged sites. *Journal of Hydrology*, 394(1–2), 101–117. <https://doi.org/10.1016/j.jhydrol.2010.01.008>
- Ghanmi, H., Bargaoui, Z., & Mallet, C. (2016). Estimation of intensity-duration-frequency relationships according to the property of scale invariance and regionalization analysis in a Mediterranean coastal area. *Journal of Hydrology*, 541, 38–49. <https://doi.org/10.1016/j.jhydrol.2016.07.002>
- Gupta, V. K., & Waymire, E. (1990). Multiscaling properties of spatial rainfall and river flow distributions. *Journal of Geophysical Research*, 95(D3), 1999–2009. <https://doi.org/10.1029/JD095iD03p01999/full>
- Halbert, K., Nguyen, C. C., Payrastré, O., & Gaume, E. (2016). Reducing uncertainty in flood frequency analyses: A comparison of local and regional approaches involving information on extreme historical floods. *Journal of Hydrology*, 541, 90–98. <https://doi.org/10.1016/j.jhydrol.2016.01.017>
- Herath, S. M., Sarukkalgige, P. R., & Nguyen, V. T. V. (2016). A spatial temporal downscaling approach to development of IDF relations for Perth airport region in the context of climate change. *Hydrological Sciences Journal*, 61(11), 2061–2070. <https://doi.org/10.1080/02626667.2015.1083103>
- Hosking, J. R. M., & Wallis, J. R. (1997). *Regional frequency analysis: An approach based on L-moments*. New York: Cambridge University Press. <https://doi.org/10.1017/CBO9780511529443>
- Huard, D., Mailhot, A., & Duchesne, S. (2010). Bayesian estimation of intensity–duration–frequency curves and of the return period associated to a given rainfall event. *Stochastic Environmental Research and Risk Assessment*, 24(3), 337–347. <https://doi.org/10.1007/s00477-009-0323-1>
- Innocenti, S., Mailhot, A., & Frigon, A. (2017). Simple scaling of extreme precipitation in North America. *Hydrology and Earth System Sciences*, 21(11), 5823–5846. <https://doi.org/10.5194/hess-21-5823-2017>
- Kieffer Weisse, A. (1998). *Etude des précipitations exceptionnelles de pas de temps court en relief accidenté (Alpes Françaises): Méthode de cartographie des précipitations extrêmes* (Ph. D. de mécanique des milieux géophysiques et Environnement). Institut National Polytechnique de Grenoble, Grenoble.
- Koutsyiannis, D. (2004a). Statistics of extremes and estimation of extreme rainfall: I. Theoretical investigation/Statistiques de valeurs extrêmes et estimation de précipitations extrêmes: I. Recherche théorique. *Hydrological Sciences Journal*, 49(4), 575–590. <https://doi.org/10.1623/hysj.49.4.575.54430>
- Koutsyiannis, D. (2004b). Statistics of extremes and estimation of extreme rainfall: II. Empirical investigation of long rainfall records/Statistiques de valeurs extrêmes et estimation de précipitations extrêmes: II. Recherche empirique sur de longues séries de précipitations. *Hydrological Sciences Journal*, 49(4), 591–610. <https://doi.org/10.1623/hysj.49.4.591.54424>
- Koutsyiannis, D., Kozonis, D., & Manetas, A. (1998). A mathematical framework for studying rainfall intensity-duration-frequency relationships. *Journal of Hydrology*, 206(1–2), 118–135. [https://doi.org/10.1016/S0022-1694\(98\)00097-3](https://doi.org/10.1016/S0022-1694(98)00097-3)
- Kuichling, E. (1889). The relation between the rainfall and the discharge of sewers in populous district. *Transactions of the American Society of Civil Engineers*, 20, 1–56.
- Laio, F. (2004). Cramer-von Mises and Anderson-Darling goodness of fit tests for extreme value distributions with unknown parameters: Goodness of fit tests for extreme value distributions. *Water Resources Research*, 40, W09308. <https://doi.org/10.1029/2004WR003204>
- Lima, C. H. R., Kwon, H.-H., & Kim, J.-Y. (2016). A Bayesian beta distribution model for estimating rainfall IDF curves in a changing climate. *Journal of Hydrology*, 540, 744–756. <https://doi.org/10.1016/j.jhydrol.2016.06.062>
- Lloyd-Davies, D. E. (1906). The elimination of storm water from sewerage systems. (Including appendix and plate). *Minutes of the Proceedings of the Institution of Civil Engineers*, 164(1906), 41–67. <https://doi.org/10.1680/imotp.1906.16637>
- Mailhot, A., Duchesne, S., Caya, D., & Talbot, G. (2007). Assessment of future change in intensity–duration–frequency (IDF) curves for southern Quebec using the Canadian Regional Climate Model (CRCM). *Journal of Hydrology*, 347(1–2), 197–210. <https://doi.org/10.1016/j.jhydrol.2007.09.019>
- Meddi, M., & Toumi, S. (2015). Spatial variability and cartography of maximum annual daily rainfall under different return periods in Northern Algeria. *Journal of Mountain Science*, 12(6), 1403–1421. <https://doi.org/10.1007/s11629-014-3084-3>
- Menabde, M., Seed, A., & Pegram, G. (1999). A simple scaling model for extreme rainfall. *Water Resources Research*, 35(1), 355–339.
- Mohyont, B., & Demarée, G. R. (2006). Courbes intensité—durée—fréquence des précipitations à Yangambi, Congo, au moyen de différents modèles de type Montana. *Hydrological Sciences Journal*, 51(2), 239–253. <https://doi.org/10.1623/hysj.51.2.239>
- Mondal, A., & Mujumdar, P. P. (2015). Modeling non-stationarity in intensity, duration and frequency of extreme rainfall over India. *Journal of Hydrology*, 521, 217–231. <https://doi.org/10.1016/j.jhydrol.2014.11.071>
- Muller, A., Bacro, J.-N., & Lang, M. (2008). Bayesian comparison of different rainfall depth–duration–frequency relationships. *Stochastic Environmental Research and Risk Assessment*, 22(1), 33–46. <https://doi.org/10.1007/s00477-006-0095-9>
- Mulvaney, T. J. (1850). On the use of self-registering rain and flood gauges in making observations on the relation of rainfall and flood discharges in a given catchment. *Transactions of Institution of Civil Engineers*, 4(2), 18.
- Nguyen, C. C., Gaume, E., & Payrastré, O. (2014). Regional flood frequency analyses involving extraordinary flood events at ungauged sites: Further developments and validations. *Journal of Hydrology*, 508, 385–396. <https://doi.org/10.1016/j.jhydrol.2013.09.058>
- Nhat, L. M., Tachikawa, Y., Sayama, T., & Takara, K. (2007). A simple scaling characteristics of rainfall in time and space to derive intensity duration frequency relationships. *Proceedings of Hydraulic Engineering*, 51, 73–78. <https://doi.org/10.2208/prohe.51.73>
- Overeem, A., Buisshand, A., & Holleman, I. (2008). Rainfall depth-duration-frequency curves and their uncertainties. *Journal of Hydrology*, 348(1–2), 124–134. <https://doi.org/10.1016/j.jhydrol.2007.09.044>
- Panthou, G., Vischel, T., Lebel, T., Quantin, G., & Molinié, G. (2014). Characterising the space–time structure of rainfall in the Sahel with a view to estimating IDAF curves. *Hydrology and Earth System Sciences*, 18(12), 5093–5107. <https://doi.org/10.5194/hess-18-5093-2014>
- Papalexiou, S. M., & Koutsyiannis, D. (2013). Battle of extreme value distributions: A global survey on extreme daily rainfall: Survey on extreme daily rainfall. *Water Resources Research*, 49, 187–201. <https://doi.org/10.1029/2012WR021257>



- Peck, A., Prodanovic, P., & Simonovic, S. P. P. (2012). Rainfall intensity duration frequency curves under climate change: City of London, Ontario, Canada. *Canadian Water Resources Journal*, 37(3), 177–189. <https://doi.org/10.4296/cwrj2011-935>
- Reis, D. S., & Stedinger, J. R. (2005). Bayesian MCMC flood frequency analysis with historical information. *Journal of Hydrology*, 313(1–2), 97–116. <https://doi.org/10.1016/j.jhydrol.2005.02.028>
- Ribatet, M., Sauquet, E., Grésillon, J.-M., & Ouarda, T. B. M. J. (2007). A regional Bayesian POT model for flood frequency analysis. *Stochastic Environmental Research and Risk Assessment*, 21(4), 327–339. <https://doi.org/10.1007/s00477-006-0068-z>
- Rosbjerg, D., & Madsen, H. (1996). The role of regional information in estimation of extreme point rainfalls. *Atmospheric Research*, 42(1–4), 113–122. [https://doi.org/10.1016/0169-8095\(95\)00057-7](https://doi.org/10.1016/0169-8095(95)00057-7)
- Shahabul Alam, M., & Elshorbagy, A. (2015). Quantification of the climate change-induced variations in intensity–duration–frequency curves in the Canadian prairies. *Journal of Hydrology*, 527, 990–1005. <https://doi.org/10.1016/j.jhydrol.2015.05.059>
- Stephens, M. A. (1976). *Asymptotic power of EDF statistics for exponentiality against gamma and Weibull alternatives (technical report no. 297)*. Stanford, CA: Stanford University.
- Stephens, M. A. (1977). Goodness of fit for the extreme value distribution. *Biometrika*, 64(3), 583. <https://doi.org/10.2307/2345336>
- Talbi, S., Meddi, M., Mahé, G., & Assani, A. (2017). Relationships between atmospheric circulation indices and rainfall in Northern Algeria and comparison of observed and RCM-generated rainfall. *Theoretical and Applied Climatology*, 127(1–2), 241–257. <https://doi.org/10.1007/s00704-015-1626-4>
- Van de Vyver, H. (2015). Bayesian estimation of rainfall intensity–duration–frequency relationships. *Journal of Hydrology*, 529, 1451–1463. <https://doi.org/10.1016/j.jhydrol.2015.08.036>
- Van de Vyver, H., & Demarée, G. R. (2010). Construction of intensity–duration–frequency (IDF) curves for precipitation at Lubumbashi, Congo, under the hypothesis of inadequate data. *Hydrological Sciences Journal*, 55(4), 555–564. <https://doi.org/10.1080/02626661003747390>
- Veneziano, D., & Furcolo, P. (2002). Multifractality of rainfall and scaling of intensity-duration-frequency curves: Multifractality of rainfall and IDF curves. *Water Resources Research*, 38(12), 1306. <https://doi.org/10.1029/2001WR000372>
- Vigione, A., Hosking, J. R. M., Laio, F., Miller, A., Gaume, E., Payrastre, O., et al. (2014). nsRFA: Non-supervised regional frequency analysis (version 0.7-12). Retrieved from <https://cran.r-project.org/web/packages/nsRFA/index.html>
- Vigione, A., Merz, R., Salinas, J. L., & Blöschl, G. (2013). Flood frequency hydrology: 3. A Bayesian analysis. *Water Resources Research*, 49, 675–692. <https://doi.org/10.1029/2011WR010782>
- Xu, Y.-P., & Tung, Y.-K. (2009). Constrained scaling approach for design rainfall estimation. *Stochastic Environmental Research and Risk Assessment*, 23(6), 697–705. <https://doi.org/10.1007/s00477-008-0250-6>
- Yu, P.-S., Yang, T.-C., & Lin, C.-S. (2004). Regional rainfall intensity formulas based on scaling property of rainfall. *Journal of Hydrology*, 295(1–4), 108–123. <https://doi.org/10.1016/j.jhydrol.2004.03.003>
- Zeroual, A., Assani, A. A., & Meddi, M. (2017). Combined analysis of temperature and rainfall variability as they relate to climate indices in Northern Algeria over the 1972–2013 period. *Hydrology Research*, 48(2), 584–595. <https://doi.org/10.2166/nh.2016.244>

## Annexe 7 - communications scientifiques

### 7.1 EVAN 2019

*Fourth international conference on Advances in Extreme Value Analysis and Application to Natural Hazard (EVAN), 17 to 19 September 2019, Chatou, France*

#### **Robust estimation and mapping of extreme rainfall quantiles**

**Mustapha Boukhelifa<sup>1</sup>, Eric Gaume<sup>2</sup>, Mohamed Meddi<sup>1</sup>**

<sup>1</sup>Ecole Nationale Supérieure d'Hydraulique (ENSH), Laboratoire GEE, Route de Soumaa, BP31, Blida, 09000, Algeria. E-mail : [m.boukhelifa@ensh.dz](mailto:m.boukhelifa@ensh.dz)

<sup>2</sup>IFSTTAR-GERS, Route de Bouaye, CS 5004, Bouguenais – Nantes, 44344, France.

Drainage network design and urban flood mapping etc., remains a challenging topic in many countries, mainly when only short record lengths and low-density raingauge network are available. To overcome these limitations, a methodology that allows a robust regional extreme rainfall return level estimation is proposed. This methodology combines local and regional merging and analysis methods for rainrate series. Regional rainfall intensities quantile estimates over various time steps (from 15 min to 24 hour) are obtained from an integrated (in space and time) Bayesian estimation of Intensity-Duration-frequency curves. This integrated framework relies on the assumption that annual maximum series of rainfall intensities satisfy to simple scaling hypothesis and show multifractal behavior. Both properties are verified on the available series prior to the analysis. The proposed integrated approach appears to be robust (i.e. providing consistent estimation even when implemented on short observed series) and reduces significantly the uncertainties on point rainfall quantiles estimates if compared to the standard local Intensity-Duration-Frequency analyses. In a second step, the impact of the proposed procedure on the uncertainties of spatially interpolated rainfall quantiles based on kriging is evaluated. Upper bounds of the 70% credibility intervals for various rainrate quantiles are mapped, combining local estimation and spatial interpolation uncertainties. Due to lower point rainfall estimation uncertainties and higher spatial consistency of the estimated values, the proposed approach appears to significantly outperform the traditional approach also if implemented to provide estimates at ungauged sites. The implementation and tests have been conducted on a large data sets of 48 automatic raingauge records from the North central part of Algeria.

## 7.2 EGU General Assembly 2018

Geophysical Research Abstracts  
Vol. 20, EGU2018-9563, 2018  
EGU General Assembly 2018  
© Author(s) 2018. CC Attribution 4.0 license.



### **Integrated bayesian estimation of IDF curves: consolidation and extensive testing of a method**

Mustapha Boukhelifa (1), Eric Gaume (2), and Mohamed Meddi (1)

(1) Ecole Nationale Supérieure d'Hydraulique, Laboratoire GEE, Blida, Algeria., (2) IFSTTAR, GERS, EE, F-44344 Bouguenais, France

Intensity-duration-frequency (IDF) curves are one of the most common rainfall model used for instance for the design of urban drainage systems. The uncertainties related to the elaboration of these IDF curves have nevertheless seldom been evaluated in the past. The presentation will recall the existing link between the IDF relationship and some properties of the rainfall series that are often observed such as simple scaling. The resulting IDF relationship is the product of a dimensionless (i.e. reduced) distribution function for the annual maximum rainfall intensities and a duration-dependent scaling factor. Its parameters can be evaluated in an integrated way (i.e. based on a unique mixed sample of peak intensities over a range of durations: 15 minutes to 24 hours). The use of likelihood based Bayesian Monte-Carlo Markov Chain (MCMC) statistical inference methods for this evaluation provides consistent uncertainties for all the parameters of the IDF relationships and the corresponding rainfall quantiles. This methodology has been tested on a large dataset of 48 rain-gauge records, spread over the North Central part of Algeria (25000 km<sup>2</sup>), under various climatic regimes. The integrated approach is undoubtedly consistent with estimates from annual maximum rainfall fitted to a single duration. Furthermore, credibility intervals are significantly reduced. Also, this approach provides robustness in the sense where the integrated inferential process applied on short observation series (8 to 12 years) provides rational estimates, which represents a significant advantage for engineering applications.

