

Titre: Estimation statistique de la précipitation maximale probable à l'aide
Title: de la loi de Pearson de type I

Auteur: Anne Martin
Author:

Date: 2024

Type: Mémoire ou thèse / Dissertation or Thesis

Référence: Martin, A. (2024). Estimation statistique de la précipitation maximale probable à
Citation: l'aide de la loi de Pearson de type I [Mémoire de maîtrise, Polytechnique
Montréal]. PolyPublie. <https://publications.polymtl.ca/59163/>

 **Document en libre accès dans PolyPublie**
Open Access document in PolyPublie

URL de PolyPublie: <https://publications.polymtl.ca/59163/>
PolyPublie URL:

**Directeurs de
recherche:** Jonathan Jalbert, & Élyse Fournier
Advisors:

Programme: Maîtrise recherche en mathématiques appliquées
Program:

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

**Estimation statistique de la précipitation maximale probable à l'aide de la loi
de Pearson de type I**

ANNE MARTIN

Département de mathématiques et de génie industriel

Mémoire présenté en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*
Mathématiques appliquées

Juillet 2024

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Ce mémoire intitulé :

**Estimation statistique de la précipitation maximale probable à l'aide de la loi
de Pearson de type I**

présenté par **Anne MARTIN**

en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*
a été dûment accepté par le jury d'examen constitué de :

Julie CARREAU, présidente

Jonathan JALBERT, membre et directeur de recherche

Élyse FOURNIER, membre et codirectrice de recherche

Jean-Luc MARTEL, membre

DÉDICACE

*Pour Ông ngoại et Grand-mère,
en échange de quelques bánh xèo
et peut-être un carré au coconut*

REMERCIEMENTS

Je voudrais remercier l'ensemble des personnes m'ayant permis de près ou de loin de réaliser ce projet.

D'abord, merci à Jonathan Jalbert, mon énergique directeur de recherche, pour ton partage, ta pédagogie et ton ouverture. Merci à Élyse Fournier, mon (autant sinon plus) énergique co-directrice pour ta disponibilité et tes encouragements.

Merci à Gabriel Gobeil de m'avoir non seulement mise au courant de l'existence du département de mathématiques de Polytechnique, mais également aidé à toutes les étapes de ma maîtrise, littéralement. Merci à Duy Anh Alexandre et à Auguste Paoli pour votre soutien moral et votre amitié.

Merci à Mitacs, au projet Arrimé ainsi qu'à Hydro-Québec pour le soutien financier.

Finalement, merci à Cyril pour ta douceur, mon papa et ma maman pour tout le reste.

RÉSUMÉ

Les ingénieurs civils effectuent la conception des infrastructures importantes exposées aux aléas hydrométéorologiques tels que les barrages hydroélectriques à l'aide de l'estimation des précipitations maximales probables (PMP). L'Organisation Mondiale Météorologique (OMM) définit la PMP comme *la quantité maximale d'eau qui peut physiquement s'accumuler sur une période de temps et une région donnée, selon la saison et sans tenir compte des tendances climatiques à long terme*. Ainsi, une surestimation de la PMP peut entraîner des coûts inutiles lors de l'aménagement des ouvrages, et une sous-estimation peut avoir des conséquences dramatiques sur les populations avoisinantes ainsi que sur la production d'énergie locale. Les méthodes actuelles du calcul de la PMP présentent de nombreuses failles : certaines variables utilisées ne sont pas directement observables et requièrent une suite d'approximations afin d'être utilisées ; l'incertitude n'est pas toujours prise en compte et peut parfois être complexe à déterminer ; les changements climatiques (CC), qui exacerbent les événements de précipitations extrêmes, sont difficiles à intégrer aux calculs.

L'objectif de ce travail est de proposer une méthode d'estimation statistique et objective de la PMP répondant à la définition donnée par l'OMM et pour laquelle l'estimation de l'incertitude et l'intégration des effets des CC sont possibles. Cette approche inédite met à profit la distribution de Pearson de type I, une généralisation de la loi bêta dont les bornes sont données par le domaine de définition des précipitations. Ainsi, cette nouvelle méthode présume que chaque averse observée correspond à une fraction de la PMP, supposée comme la borne supérieure de la distribution de probabilité modélisant les pluies.

À la suite d'une série de tests, trois méthodes d'ajustement de la distribution sont suggérées : la méthode des moments, une méthode mixte mélangeant l'approche précédente à l'estimateur du maximum de vraisemblance ainsi qu'une méthode usant de l'inférence bayésienne. Cette méthodologie est appliquée à deux stations hydrométéorologiques de la région métropolitaine de Montréal, soit : l'aéroport international Pierre-Elliott-Trudeau de Montréal et St-Hubert. L'étude de cas indique que l'approche bayésienne performe mieux que les estimateurs fréquentistes et donne des valeurs s'apparentant à celles obtenues à l'aide de méthodes d'estimation usuelles de la PMP, avec comme avantages additionnels une simplicité de calcul ainsi que la possibilité de générer des graphiques diagnostiques sinon inaccessibles. La méthode est limitée par le nombre d'observations disponibles aux stations, mais pourrait être appliquée dans le cas de la disponibilité de nouveaux produits climatiques d'observation.

ABSTRACT

Civil engineers carry out the design of infrastructures exposed to hydrometeorological hazards, like hydroelectric dams, with the estimation of probable maximum precipitation (PMP). The World Meteorological Organization (WMO) defines PMP as *the greatest depth of precipitation for a given duration meteorologically possible for a design watershed or a given storm area at a particular location at a particular time of year, with no allowance made for long-term climatic trends*. Thus, an overestimation of the PMP can lead to unnecessary expenses when building those infrastructures, and an underestimation can have dramatic consequences on neighbouring populations and local energy production. Current PMP estimation methods pose multiple flaws: some variables used aren't directly observable and require a series of approximations to be estimated; uncertainties aren't always taken into account and can sometimes be complex to approximate; climate change (CC), which intensify extreme precipitation events, are difficult to integrate to computations.

This work aims to suggest a statistical and objective method of estimation of the PMP, following the WMO definition and for which the uncertainties calculation and the CC integration are possible. This novel approach takes advantage of the Pearson type I distribution, a generalization of the beta distribution for which the limits are given by the definition domain of the precipitations. Thus, this new method assumes that every observed rain event corresponds to a fraction of the PMP, presumed as the upper limit of the probability distribution modelling the rains.

Following a series of tests, three distribution adjustment methods are suggested: the method of moments, a mixed method blending the precedent approach and the maximum likelihood estimation as well a Bayesian inference approach. This methodology is applied to two hydrometeorological stations in the Montréal metropolitan region: the international Montréal Pierre-Elliott-Trudeau airport and St-Hubert. The study case indicates that the Bayesian approach performs better than the frequentist estimators, and returns similar values to those obtained with the usual PMP estimation methods, with the additional advantages of simplified computation as well as the possibility to generate diagnostic graphics inaccessible otherwise. The technique is limited by the number of available observations but could be applied in the case of the disponibility of new climatic observation products.

TABLE DES MATIÈRES

DÉDICACE	iii
REMERCIEMENTS	iv
RÉSUMÉ	v
ABSTRACT	vi
TABLE DES MATIÈRES	vii
LISTE DES TABLEAUX	ix
LISTE DES FIGURES	x
LISTE DES SIGLES ET ABRÉVIATIONS	xii
LISTE DES ANNEXES	xiii
CHAPITRE 1 INTRODUCTION	1
1.1 Éléments de la problématique	2
1.2 Objectifs de recherche	3
1.3 Plan du mémoire	3
CHAPITRE 2 REVUE DE LITTÉRATURE	5
2.1 Définitions théorique et opérationnelle de la PMP	5
2.2 Estimation de la PMP théorique : méthodes hydrométéorologiques	5
2.2.1 Transposition des tempêtes	5
2.2.2 Maximisation de l'humidité	6
2.2.3 Processus de sélection des tempêtes	7
2.2.4 Utilisation de simulations climatiques	8
2.3 Estimation de la PMP opérationnelle	8
2.3.1 Méthode empirique d'Hershfield	8
2.3.2 Analyse fréquentielle	10
2.4 PMP et changements climatiques	11
2.5 Calcul des incertitudes	12

CHAPITRE 3 THÉORIE	14
3.1 Méthodes d'estimation des paramètres	15
3.1.1 Méthode des moments	15
3.1.2 Maximum de la vraisemblance	16
3.1.3 Inférence bayésienne	17
3.2 Méthodes Monte-Carlo par chaîne de Markov	18
3.2.1 Algorithme de Metropolis-Hasting	18
3.2.2 Échantillonnage de Gibbs	19
3.2.3 Monte-Carlo hybride	20
CHAPITRE 4 MÉTHODOLOGIE	23
4.1 Pearson de type I à quatre paramètres	25
4.1.1 Corrélation entre les paramètres	27
4.2 Pearson de type I à trois paramètres	29
4.2.1 Estimation bayésienne des paramètres	34
4.3 Pearson de type I à trois paramètres reparamétrisée	36
4.4 Méthodologie choisie	37
4.4.1 Données	37
4.4.2 Méthode d'estimation de la PMP bêta	38
4.4.3 Méthodes usuelles d'estimation de la PMP	41
CHAPITRE 5 RÉSULTATS	43
5.1 Station de l'aéroport international Montréal-Trudeau	43
5.2 Station de St-Hubert	48
5.3 Discussion	52
CHAPITRE 6 DISCUSSION ET CONCLUSION	55
6.1 Synthèse des travaux	55
6.2 Limitations de la solution proposée et améliorations futures	55
RÉFÉRENCES	58
ANNEXES	63

LISTE DES TABLEAUX

Tableau 4.1	Valeurs théoriques des paramètres des distributions échantillonnées pour le modèle de Pearson de type I	25
Tableau 4.2	Moyennes des 100 estimations par maximisation de la vraisemblance des paramètres ainsi que les quantiles 2.5% et 97.5% des estimations de b pour la distribution C	27
Tableau 5.1	Paramètres estimés pour différents modèles et méthodes d'estimations, à partir des données observées de précipitation journalières entre 1953 et 2024 à la station de Montréal	47
Tableau 5.2	PMP estimées à l'aide de méthodes usuelles pour la station de Montréal	47
Tableau 5.3	Paramètres estimés pour différents modèles et méthodes d'estimations, à partir des données observées de précipitation journalières entre 1949 et 2024 à la station de Saint-Hubert	52
Tableau 5.4	PMP estimées à l'aide de méthodes usuelles pour St-Hubert	52
Tableau A.1	Coefficients de régression entre $\hat{a}$ et $\hat{\alpha}$ pour chacune des distributions et des méthodes d'estimations	63
Tableau A.2	Coefficients de régression entre $\hat{b}$ et $\hat{\beta}$ pour chacune des distributions et des méthodes d'estimations	64

LISTE DES FIGURES

Figure 3.1	Graphiques de densités de probabilité et de fonctions de répartition possibles pour la distribution de Pearson de type I	14
Figure 3.2	Courbes de niveaux associées à différentes valeurs de l'énergie du modèle en fonction du couple (θ, ϕ)	22
Figure 4.1	Précipitations journalières observées à la station Montréal-Trudeau du 1er mai 1953 au 28 mai 2024	24
Figure 4.2	Fonctions de probabilité des distributions échantillonnées pour le modèle de Pearson de type I	26
Figure 4.3	Estimation de b en fonction de la taille de l'échantillon, pour les distributions de Pearson de type I à quatre paramètres C et D	28
Figure 4.4	Illustration de la flexibilité de la distribution de Pearson de type I . .	29
Figure 4.5	Ajustement des échantillons de plus grandes tailles tirés des distribution C et D (Pearson de type I à quatre paramètres)	30
Figure 4.6	Corrélation entre les estimations des paramètres (a, α) et (b, β) pour la distribution A	30
Figure 4.7	Corrélation entre les estimations des paramètres (a, α) et (b, β) pour la distribution B	31
Figure 4.8	Corrélation entre les estimations des paramètres (a, α) et (b, β) pour la distribution C	31
Figure 4.9	Corrélation entre les estimations des paramètres (a, α) et (b, β) pour la distribution D	32
Figure 4.10	Estimation de b en fonction de la taille de l'échantillon, pour chaque distribution de Pearson de type I à trois paramètres échantillonnée .	33
Figure 4.11	Ajustement des échantillons de plus grandes tailles tirés des distribution C et D (Pearson de type I à trois paramètres)	34
Figure 4.12	Estimation bayésienne de b en fonction de la taille de l'échantillon, pour les distribution A et B	36
Figure 4.13	Boîtes à moustaches des moyennes des chaînes de b proposées par l'échantillonneur de Gibbs à partir d'échantillons de taille 5000 et 8000	37
Figure 4.14	Corrélation entre les estimations des paramètres (b, μ) et (b, ν) pour les distributions A et B	38
Figure 4.15	Corrélation entre les estimations des paramètres (b, μ) et (b, ν) pour les distributions C et D	39

Figure 4.16	Estimation de b en fonction de la taille de l'échantillon, pour la distribution de Pearson de type I à trois paramètres (reparamétrisation) A	39
Figure 4.17	Estimation de b en fonction de la taille de l'échantillon, pour les distributions de Pearson de type I à trois paramètres (reparamétrisation) échantillonnée B, C et D	40
Figure 4.18	Précipitations journalières observées à la station de Saint-Hubert du 4 juillet 1949 au 29 mai 2024	41
Figure 5.1	Diagrammes quantile-quantile et probabilité-probabilité pour le modèle bêta avec a fixé à 0mm et ajustée sur les données de Montréal à l'aide de trois méthodes différentes (moments, mixte et Gibbs)	43
Figure 5.2	Traces et histogrammes de la chaîne de Markov générée par Gibbs sur les données de Montréal pour le modèle bêta avec $a = 0$	44
Figure 5.3	Diagrammes quantile-quantile et probabilité-probabilité pour le modèle bêta à quatre paramètres avec $a = 0.2$ fixé, ajusté sur les données de Montréal à l'aide des méthodes choisies (moments, mixte et Gibbs)	45
Figure 5.4	Traces et histogrammes de la chaîne de Markov générée par l'échantillonneur de Gibbs sur les données de Montréal pour le modèle bêta avec a fixé à 0.2	46
Figure 5.5	Diagrammes quantile-quantile et probabilité-probabilité pour la distribution Pearson de type I à trois paramètres ajustée sur les données de St-Hubert à l'aide de trois méthodes différentes (moments, mixte et Gibbs)	48
Figure 5.6	Traces et histogrammes de la chaîne de Markov générée par l'échantillonneur de Gibbs sur les données de St-Hubert pour le modèle bêta avec a fixé à 0	49
Figure 5.7	Diagrammes quantile-quantile et probabilité-probabilité pour la distribution Pearson de type I à quatre paramètres avec $a = 0.2$ fixé, ajustée ajustée sur les données de St-Hubert à l'aide de trois méthodes différentes (moments, mixte et Gibbs)	50
Figure 5.8	Traces et histogrammes de la chaîne de Markov générée par l'échantillonneur de Gibbs sur les données de St-Hubert pour le modèle bêta avec a fixé à 0.2	51

LISTE DES SIGLES ET ABRÉVIATIONS

CMP	Crue maximale probable
PMP	Précipitation maximale probable
OMM	Organisation Météorologique Mondiale
CC	Changements climatiques
GEV	Extremum généralisée
MCMC	Méthodes de Monte-Carlo par chaînes de Markov

LISTE DES ANNEXES

Annexe A	Corrélation des paramètres	63
----------	--------------------------------------	----

CHAPITRE 1 INTRODUCTION

Le Répertoire des barrages québécois recense plus de 6000 ouvrages de plus d'un mètre de hauteur dispersés sur l'ensemble du territoire (CEHQ, 2023). Hydro-Québec, la société d'État responsable de la production, du transport et de la distribution d'électricité au niveau de la province, possède à elle seule plus de 700 barrages et autres ouvrages de retenue qui permettent le bon fonctionnement de ses centrales électriques. Le surdimensionnement de ces infrastructures lors de leur construction ou de leur rénovation peut entraîner des coûts inutiles. Leur sous-dimensionnement engendre quant à lui des risques au niveau de la sécurité des barrages, de l'environnement, des populations avoisinantes et peut ainsi également occasionner des coûts excessifs. Selon le règlement qui découle de la loi sur la sécurité des barrages, les ingénieurs civils doivent dimensionner ceux-ci à l'aide d'une estimation de crue extrême locale, soit la crue maximale probable (CMP), la crue décennale ou la crue millénaire selon le risque en cas de rupture (Québec, 2023). Ces deux dernières sont les crues dont les probabilités de survenue à chaque année sont estimées à 0.0001 et 0.001 respectivement. La CMP, quant à elle, est la plus grande crue théoriquement possible d'un bassin spécifique, résultant de la combinaison de conditions météorologiques et hydrologiques extrêmes (LaRocque, 2013; OMM, 2009).

La valeur de cette crue est calculée à partir entre autres de la précipitation maximale probable (PMP), qui est définie selon l'Organisation Météorologique Mondiale (OMM, 2009) comme la quantité maximale d'eau qui peut physiquement s'accumuler sur une période de temps et une région donnée, selon la saison. La dernière édition du *Manuel sur l'Estimation de la Précipitation Maximale Probable (PMP)* publié par l'OMM en 2009 définit un éventail de méthodes associées au calcul de la PMP. Certaines méthodes, comme la transposition des tempêtes, sont utilisées afin d'augmenter le nombre d'événements à considérer lors du calcul de la PMP d'une zone précise. D'autres, comme la maximisation de l'humidité ou l'estimation empirique d'Hershfield, permettent d'obtenir une valeur numérique de la PMP.

Au Québec, un ensemble de méthodes hydrométéorologiques incluant la maximisation de l'humidité et la transposition des tempêtes est utilisé afin de déterminer la PMP d'une région. La maximisation de l'humidité considère les événements passés de pluie et projette la hauteur des précipitations dans le cas où cette tempête se serait produite au moment où l'humidité spécifique de l'air était maximale. Pour ce faire, on doit recenser la hauteur des précipitations des événements de pluie extrêmes ainsi que l'humidité de l'air aux mêmes moments, en plus d'estimer la quantité maximale d'eau pouvant possiblement se trouver dans

la colonne d'air au-dessus de la région, souvent définie comme l'eau précipitable maximale. Chaque événement de pluie extrême est maximisé, et la PMP est choisie comme la plus grande hauteur de précipitation de l'ensemble de tempêtes maximisées.

1.1 Éléments de la problématique

La définition actuelle ainsi que les méthodes d'estimation de la PMP présentent plusieurs limitations. D'abord, la définition d'une précipitation maximale *probable* laisse place à l'interprétation. En effet, mathématiquement, la plus haute précipitation possible sur une certaine région pourrait être très grande par rapport aux données observées aux stations. La distribution statistique des pluies pourrait même être non bornée supérieurement, ce qui impliquerait une PMP infinie. Un tel résultat serait bien sûr inutile d'un point de vue pratique. Certains auteurs (Ben Alaya *et al.*, 2018; Salas *et al.*, 2014; Visser *et al.*, 2022) font la distinction entre les PMP dites théorique et opérationnelle, concepts qui seront abordés dans le second chapitre. Néanmoins, ce manque de rigueur de la définition implique la subjectivité de l'utilisateur en omettant de préciser la signification des termes *maximale* et *probable*.

La définition donnée par l'OMM et mentionnée à la section précédente fait également abstraction de la réalité climatique. Effectivement, une liste grandissante d'articles souligne la relation entre réchauffement climatique et augmentation des valeurs estimées de PMP (Beauchamp *et al.*, 2013; Ben Alaya *et al.*, 2020b; Clavet-Gaumont *et al.*, 2017; Mahoney *et al.*, 2018; Rouhani et Leconte, 2018, 2020; Rousseau *et al.*, 2014; Visser *et al.*, 2022). Puisqu'un barrage est une infrastructure ayant une longue durée de vie et que la PMP utilisée lors de la construction de celui-ci doit refléter les précipitations les plus extrêmes sur cet intervalle de temps, la prise en compte des tendances climatiques est nécessaire.

D'un côté strictement fonctionnel, le début de l'observation des données nécessaires au calcul de la PMP est généralement relativement récent, ce qui rend le nombre d'événements de pluie extrêmes insuffisant malgré les algorithmes de transposition des tempêtes. En effet, un échantillon enregistré sur vingt ans n'aura pas le même profil que celui recueilli sur cent ans et les valeurs des quantiles extrêmes ne seront probablement pas du tout les mêmes. Un petit échantillon implique généralement une grande variance sur les estimations, ce qui n'est pas toujours pris en compte par les méthodes de l'OMM. Il est également nécessaire de noter qu'au Québec, les données enregistrées de pluie et de neige sont fréquemment indiscernables, ce qui peut possiblement restreindre le nombre de tempêtes utilisables.

En se basant sur le guide de l'OMM, on peut remarquer que les méthodologies utilisées spécifiquement pour l'estimation de la PMP, excluant les algorithmes de transposition des

tempêtes, ont des lacunes mathématiques et physiques. Ces méthodologies ont été développées au courant de la seconde moitié du dernier siècle et étaient des solutions d'ingénierie raisonnables à l'époque vu le nombre de données climatiques limité, la capacité de calcul et la nécessité de construire des barrages sécuritaires. Ces méthodes de calcul de la PMP n'ont peu ou pas évoluées depuis leur conception. De plus, certaines modifications sont parfois apportées sans justifications physiques ou mathématiques, comme l'adoption d'une limite supérieure pour le ratio humidité maximale sur humidité observée lors de la maximisation de l'humidité (Rousseau *et al.*, 2014).

Relativement à la maximisation de l'humidité et spécifiquement pour le Québec, il est à noter que l'humidité spécifique n'est pas une valeur mesurée aux stations. Il est donc nécessaire de convertir le point de rosée en humidité, ce qui peut potentiellement introduire un manque de précision supplémentaire aux estimations. De plus, les observations de point de rosée ne sont pas disponibles à toutes les stations du Québec, ce qui force les utilisateurs à transposer les données de stations voisines. Finalement, le processus de sélection des tempêtes extrêmes utilisées pour le calcul de la PMP manque actuellement d'objectivité. Les événements choisis dépendent de l'utilisateur et cette variation peut affecter grandement la valeur finale de la PMP.

1.2 Objectifs de recherche

Le premier objectif de ce travail est de développer une méthodologie de calcul de la PMP théorique statistiquement rigoureuse et développée à partir de la définition actuelle de l'OMM. Cette nouvelle méthodologie se doit également d'être facilement ajustable aux changements climatiques (CC) et son incertitude doit être aisément quantifiable.

Le second objectif consiste à tester la méthode sur des échantillons de données réels. Ces observations proviennent de deux stations témoins du Québec : l'aéroport international Pierre-Elliott-Trudeau de Montréal et Saint-Hubert. Ces estimations de la PMP sont par la suite comparées avec celles obtenues à l'aide de méthodes usuelles de calcul de la PMP, soit la maximisation de l'humidité et la méthode empirique d'Hershfield.

1.3 Plan du mémoire

Ce mémoire est divisé de la manière suivante : le second chapitre passe en revue la littérature existante sur la PMP. Le troisième chapitre met en lumière la théorie nécessaire à la compréhension de la méthode développée par ce projet. Cette méthodologie est décrite dans le quatrième chapitre et le cinquième chapitre rapporte les résultats associés aux diverses

méthodes d'estimation de la PMP pour les stations choisies. Une définition actuelle plus sensible aux réalités climatiques et hydrologiques propres au Québec sera finalement mise en valeur dans la conclusion du mémoire, aux côtés d'une discussion sur les limites du projet et les avenues possibles suite à celui-ci.

CHAPITRE 2 REVUE DE LITTÉRATURE

Ce chapitre recense les différentes définitions et méthodes d'estimation de la PMP dans la littérature, avec une attention particulière portée sur l'extension des définitions en contexte de changements climatiques et l'utilisation de la théorie statistique.

2.1 Définitions théorique et opérationnelle de la PMP

La définition de la PMP donnée par l'OMM (2009) indique que celle-ci correspond à la quantité *maximale* de pluie pouvant s'abattre sur une zone déterminée. Cependant, certains auteurs (Ben Alaya *et al.*, 2018; Salas *et al.*, 2014; Visser *et al.*, 2022) différencient la PMP *opérationnelle* de celle *théorique*. Ceux-ci soutiennent que la PMP estimée est une valeur purement pratique qui permet le dimensionnement des infrastructures et ne devrait pas être confondue avec la PMP théorique, qui serait elle la vraie limite supérieure physique des précipitations (finie ou infinie). La PMP opérationnelle, soulignent Ben Alaya *et al.* (2018), n'est pas nécessairement un concept scientifique, mais bien une solution rationnelle qui permet la conception d'ouvrages exposés aux aléas hydrométéorologiques lorsque les connaissances scientifiques sont insuffisantes.

Dans le présent travail, les approches qui tentent à l'aide de méthodes statistiques ou hydro-météorologiques d'estimer la borne supérieure des précipitations pour un lieu donné seront qualifiées de méthodes d'estimation théorique de la PMP, alors que celles qui calculent plutôt un quantile extrême sans être la limite supérieure des pluies sont classées comme méthodes d'estimation opérationnelle.

2.2 Estimation de la PMP théorique : méthodes hydrométéorologiques

2.2.1 Transposition des tempêtes

On sait que le nombre d'événements de pluie extrêmes est limité par définition. Il arrive également que la zone pour laquelle la PMP est requise enregistre des données depuis un nombre insuffisant d'années. Afin d'augmenter le nombre d'événements de pluie extrêmes qui seront utilisées lors du calcul de la PMP, il est possible d'utiliser les données de tempêtes de zones avoisinantes, donc probables sur la région d'intérêt. Plusieurs techniques prenant en compte l'orographie et autres caractéristiques des différentes régions ont été développées au cours des dernières décennies afin de transposer les tempêtes de manière réaliste (OMM,

2009). Cette transposition des tempêtes peut être incorporée dans le processus de sélection de tempêtes des méthodes d'estimation de la PMP. Comme mentionné au chapitre 1, le mélange transposition des tempêtes et maximisation de l'humidité est celui le plus couramment utilisé pour le calcul de la PMP au Québec.

2.2.2 Maximisation de l'humidité

La maximisation de l'humidité estime la PMP en utilisant la relation entre quantité d'eau précipitée et humidité de l'air. On estime à l'aide de n événements de précipitation la PMP de la façon suivante :

$$\begin{aligned} PMP &= \max_{i \in \{1, \dots, n\}} \left\{ P_i \times \frac{EP_{\max}}{EP_i} \right\} \\ &= \max_{i \in \{1, \dots, n\}} \left\{ \frac{P_i}{EP_i} \right\} \times EP_{\max} \end{aligned} \quad (2.1)$$

avec P_i et EP_i l'eau précipitée et précipitable d'un événement i et $EP_{\max}$ l'eau précipitable maximale de la zone d'intérêt. L'eau précipitable maximale $EP_{\max}$ peut être déterminée comme une valeur de retour exceptionnelle de la variable, généralement la valeur de retour 100 ans (Clavet-Gaumont *et al.*, 2017; Rouhani et Leconte, 2018, 2020; Rousseau *et al.*, 2014), ou encore comme la valeur maximale observée si le jeu de données est suffisamment grand (Ben Alaya *et al.*, 2018; Salas *et al.*, 2014). L'eau précipitable n'est pas une donnée observable et doit être approximée par une fonction de l'humidité spécifique de la colonne d'air au-dessus de la région (OMM, 2009). Dans la majorité des stations météorologiques du Québec et ailleurs dans le monde, cette humidité spécifique n'est pas non plus enregistrée. On doit alors passer par une estimation de la quantité d'eau dans la colonne d'air en fonction du point de rosée au sol. Cette approximation recommandée par l'OMM (2009) s'appuie sur des tests effectués à partir de données d'Oakland en Californie (U.S. Weather United States Weather Bureau, 1960) et Fairbanks en Alaska (Miller, 1963) et requiert des conditions pseudo-adiabatiques. Selon Viswanadham (1981), la corrélation entre point de rosée au sol et eau précipitable est plus grande (>0.60) dans les régions de latitudes élevée ($>25^\circ$) et moins grande (<0.50) dans les régions de basse latitudes ($<20^\circ$). Ces approximations entraînent une incertitude qui n'est généralement pas prise en compte dans le calcul de la PMP. Entre autres, une étude menée par Chen et Bradley (2006) dans la région de Chicago indique que l'hypothèse de conditions pseudo-adiabatiques pourrait mener à la surestimation de l'eau précipitable. Rouhani et Leconte (2020) estiment pour leur part l'eau précipitable journalière EP à l'aide des données aériennes (*upper-air data*) ainsi qu'à partir du point de rosée, puis comparent les valeurs estimées pour trois bassins versants québécois pour des conditions

de climats actuels et futurs. Les auteurs notent que les valeurs de PMP estimées varient grandement selon l'eau précipitable utilisée, en particulier pour les bassins plus au sud.

Les quotients EP_{max}/EP_i et P_i/EP_i sont respectivement appelés ratio de maximisation et efficacité de précipitation. Ces fractions permettent deux interprétations de la maximisation de l'humidité. Le premier ratio donne son nom à la maximisation de l'humidité, puisqu'il maximise l'eau présente dans la colonne d'air lors de l'événement i à l'aide de EP_{max} local. Ainsi, on peut voir le calcul de la PMP comme la précipitation qu'aurait eue l'événement de pluie i si l'humidité de l'air avait été maximale. La PMP est choisie comme la plus grande valeur obtenue parmi toutes les tempêtes maximisées et correspond à la première ligne de l'équation 2.1. Ce ratio de maximisation est parfois borné arbitrairement par une valeur numérique entre 1.5 et 2.5 pour éviter la surestimation de la PMP (Hansen, 1988; OMM, 2009; Rousseau *et al.*, 2014; Schreiner et Riedel, 1978). Cette borne est subjective en raison du manque de justifications physiques ou mathématiques derrière son utilisation. Rouhani et Leconte (2016) redéfinissent le ratio de maximisation afin qu'il ne soit pas nécessaire d'imposer une limite supérieure potentiellement arbitraire.

L'efficacité de précipitation permet une seconde interprétation de la PMP : celle-ci est définie comme le produit entre le maximum observé du ratio P_i/EP_i et l'eau précipitable maximale EP_{max} (seconde ligne de l'équation 2.1). Ben Alaya *et al.* (2018) exploitent cette définition afin de modéliser la dépendance réelle entre les valeurs extrêmes de l'efficacité de précipitation et de l'eau précipitable. L'article démontre ainsi que la comonotonie imposée par l'équation 2.1 induit une surestimation de la PMP en Amérique du Nord.

2.2.3 Processus de sélection des tempêtes

Généralement, les méthodes hydrométéorologiques impliquent une étape de sélection des tempêtes extrêmes qui sont par la suite transposées et maximisées. Certains articles (Beauchamp *et al.*, 2013) utilisent toutes les tempêtes pour lesquelles la quantité de précipitation dépasse un certain seuil, alors que d'autres exploitent l'ensemble des précipitations observées sur un intervalle de temps (Ben Alaya *et al.*, 2018). Dans l'industrie, ce processus de sélection est effectué par des météorologues et dépend de facteurs physiques (CEHQ et SNC-Lavalin, 2004; DTN et MGS Engineering, 2020; EWGR, 2020; OMM, 2009). Puisque le nombre de tempêtes sélectionnées est petit et varie d'un calcul à l'autre, ce processus d'élimination induit une grande variabilité dans l'estimation de la PMP.

2.2.4 Utilisation de simulations climatiques

Les méthodes d'estimation de la PMP actuellement présentées par l'OMM ne considèrent que les données de précipitations observées aux stations hydrométéorologiques. Des méthodologies adaptées à certaines simulations régionales ont précédemment été élaborées pour des régions du Canada (Beauchamp *et al.*, 2013; Clavet-Gaumont *et al.*, 2017; Rouhani et Leconte, 2018, 2020; Rousseau *et al.*, 2014), de l'Amérique du Nord (Ben Alaya *et al.*, 2018, 2020a,b; Hiraga *et al.*, 2021; Kunkel *et al.*, 2013) ou d'autres régions du monde (Sarkar et Maity, 2020; Visser *et al.*, 2022). L'utilisation de ces simulations climatiques, en plus de permettre la considération d'un plus grand nombre de tempêtes extrêmes lors du calcul de la PMP, rendent possible l'estimation de PMP futures. À l'exception de Sarkar et Maity (2020), qui emploient la méthode empirique d'Hershfield, la grande majorité des articles cités précédemment étudient l'évolution de la PMP estimée à l'aide de méthodes hydrométéorologiques.

L'utilisation de données simulées du climat est étroitement liée à la modélisation de l'évolution de la PMP due aux changements climatiques, sujet qui sera abordé à la section 2.4. Néanmoins, ces simulations peuvent également être employées afin de pallier au manque de données observées (Rouhani et Leconte, 2016) ou pour estimer des PMP spécifiques à certains climats précis. Par exemple, Hiraga *et al.* (2021) utilisent la version 3.9.1 du *Weather Research and Forecasting Model* afin de modéliser le climat des hivers du nord-ouest des États-Unis.

Beauchamp *et al.* (2013) mentionnent l'importance de rester critique quant à l'interprétation de résultats provenant de simulations climatiques, puisque ceux-ci ajoutent une couche supplémentaire d'incertitude dans les estimations. Particulièrement, l'utilisation de simulations de climats projetés devrait faire l'objet d'une attention particulière puisque les données produites sont elles-mêmes des estimations dont on ne peut pas connaître l'exactitude.

2.3 Estimation de la PMP opérationnelle

2.3.1 Méthode empirique d'Hershfield

Initialement développée au courant des années 1960 par Hershfield (1961a), cette méthode d'estimation de la PMP est l'une des plus populaires par sa simplicité et sa rapidité d'utilisation. Elle a également l'avantage de ne pas exiger un grand nombre de données hydrométéorologiques : en effet, seuls les maxima annuels des hauteurs de précipitations sur une durée déterminée ainsi que le nombre d'années de la série sont nécessaires au calcul. Elle est généralement qualifiée comme une technique statistique, mais sera plutôt considérée ici comme empirique puisqu'elle utilise les moments empiriques de séries de précipitations observées. Il

est important de remarquer que cette méthode est développée spécifiquement pour le calcul de la PMP de petits bassins versants (OMM, 2009).

La méthode se base sur l'équation d'analyse fréquentielle de Chow (1951), qui souligne que chacun des quantiles d'une distribution de probabilité peut être exprimé comme une fonction de la moyenne et de la distance de ce quantile avec la moyenne. Cet écart est le produit de l'écart-type et d'un facteur de fréquence K , qui doit être déterminé selon le quantile d'intérêt. Ainsi, la PMP peut être estimée selon

$$PMP = \bar{X}_n + K_m S_n, \quad (2.2)$$

avec $\bar{X}_n$ et S_n la moyenne et l'écart-type de la série de n maxima annuels de précipitations journalières et K_m le nombre d'écarts-types à ajouter à la moyenne afin d'obtenir le quantile associé à la PMP. L'indice m est associé à l'événement de pluie le plus important de la série de maxima annuels.

Avant de calculer la PMP à l'aide de l'équation 2.2, il est nécessaire d'ajuster $\bar{X}_n$ et S_n selon l'événement maximal observé, selon le nombre d'années disponibles n de la série et si les données sont enregistrées sur des intervalles de temps fixes (Hershfield, 1961a,b). Il est également nécessaire d'ajuster le facteur de fréquence selon le lieu d'intérêt. Au départ, la valeur maximale de K_m est estimée à 15 par Hershfield (1961a) à partir de 2645 stations sur le territoire des États-Unis et est considérée comme invariante. Cependant, Hershfield (1965) souligne que cette valeur enveloppante K_m est inversement proportionnelle à la moyenne annuelle de la station d'intérêt. Ainsi, la PMP d'une région sèche se trouvera à un plus grand nombre d'écarts-types de la moyenne que celle d'une région humide. Hershfield (1965) précise également la dépendance entre la durée des précipitations et la valeur de K_m .

Cette méthode a été incrémentalement améliorée depuis sa conception afin de prendre en compte entre autres les réalités de différents milieux et types de données (OMM, 2009). Notamment, Singh *et al.* (2018) et Salas *et al.* (2014) s'intéressent au calcul de l'incertitude des composantes de l'équation 2.2 ou des PMP estimées, alors que Sarkar et Maity (2020) modélisent les variations temporelles de la PMP afin de prendre en compte les changements climatiques dans le calcul.

Il est à noter que les durées comparées par Hershfield (1965) et disponibles dans le manuel de l'OMM (2009) sont toutes plus petites ou égales à 24h, ce qui ne peut pas servir dans le cas de calcul de PMP de plus longues durées. De plus, les ajustements sur les moments empiriques proposés par l'OMM sont uniquement disponibles pour le cas de PMP 24h. Finalement, une grande partie des calculs sur lesquels se base la théorie est effectuée uniquement pour les

48 états des États-Unis contigus. On peut supposer que ce territoire n'est pas représentatif des autres régions du monde. Les améliorations mentionnées précédemment sont également spécifiques à certaines régions et ne peuvent pas toujours être généralisées.

On peut également remettre en question la légitimité de la procédure : en effet, l'équation 2.2 définit simplement la PMP comme un quantile extrême de la distribution des maxima annuels de précipitation et non pas comme la borne supérieure de celle-ci. D'ailleurs, Koutsoyiannis (1999) démontre que les valeurs estimées par la méthode d'Hershfield correspondent en moyenne à des valeurs de retour de 60 000 ans d'une distribution de loi d'extremum généralisée (GEV) ajustée aux séries de maxima annuels de pluie. Ainsi, il semble important de souligner que les PMP estimées par la méthode d'Hershfield sont des PMP opérationnelles plutôt que théoriques.

2.3.2 Analyse fréquentielle

Bien que non mentionnée dans le guide de l'OMM (2009), l'analyse fréquentielle est une méthode d'analyse des extrêmes de pluie qui pourrait être explorée comme méthode d'estimation de la PMP. Comme la méthode d'Hershfield, cette approche détermine la PMP comme un quantile extrême d'une distribution ajustée aux données de précipitation : on estime donc une PMP opérationnelle. Cette approche, qui n'utilise que les données extrêmes de précipitations, peut être séparée en deux catégories de modèles : les modèles de maxima par bloc et ceux de dépassement de seuil.

Sans mentionner spécifiquement la PMP, Martins et Stedinger (2000) s'intéressent à l'utilisation de la théorie des valeurs extrêmes, plus spécifiquement à la distribution GEV, dans l'estimation de quantiles extrêmes pour des données hydrologiques. La fonction de répartition de la loi GEV est donnée par

$$F(x) = \exp \left\{ - \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right)^{-1/\xi} \right] \right\}, \quad \{x : 1 + \xi \left(\frac{x - \mu}{\sigma} \right) > 0\}, \quad (2.3)$$

avec $\mu \in \mathbb{R}$ un paramètre de position, $\sigma \in \mathbb{R}^+$ un paramètre d'échelle et $\xi \in \mathbb{R}$ un paramètre de forme, qui contrôle le poids des queues de la distribution (Coles, 2001).

En particulier, Martins et Stedinger (2000) soulignent que malgré la performance insatisfaisante de la maximisation de la vraisemblance sur certains ensembles de données hydrologiques de petites tailles, la théorie des valeurs extrêmes ne devrait pas pour autant être ignorée. En effet, l'utilisation de l'inférence bayésienne, entre autres par la définition d'une distribution à priori informative pour le paramètre de forme, permet une amélioration substantielle de l'estimation des quantiles extrêmes. Papalexiou et Koutsoyiannis (2013) concluent quant à

eux que les distributions des maxima annuels de précipitation sur 24 heures devraient suivre une distribution de Fréchet, c'est-à-dire une loi GEV pour laquelle la queue de droite est lourde et non bornée. Ainsi, selon un tel modèle, une précipitation maximale probable dans le sens d'une borne supérieure de la distribution des pluies n'existerait pas. Afin de calculer une PMP théorique plutôt qu'opérationnelle, Eliasson (1997) modifie la queue de droite d'une distribution Gumbel, une loi GEV pour laquelle les queues sont légères et infinies, afin que celle-ci soit bornée supérieurement. Ainsi, cette borne supérieure devient l'estimation de la PMP. Cette méthode soulève cependant le problème de la sous-estimation de la PMP, puisqu'on force une limite supérieure sur une distribution non-bornée.

En utilisant un modèle de dépassement de seuil plutôt que de blocs maxima, c'est-à-dire en ajustant une distribution de Pareto généralisée à toutes les données de précipitations journalières excédant un certain seuil élevé, Serinaldi et Kilsby (2014) concluent eux aussi que la distribution présente une queue lourde et non bornée. N'utilisant pas des distributions propres à la théorie des valeurs extrêmes, mais toujours en adoptant des modèles de dépassement de seuil, Papalexiou *et al.* (2013) ont la même conclusion en ajustant différents types de queues de distribution.

Théoriquement, l'analyse fréquentielle de valeurs extrêmes devrait passer par la théorie des valeurs extrêmes. Cependant, le nombre de données disponibles peut souvent être un frein à son utilisation. En effet, les distributions GEV et Pareto généralisée sont des lois limites, c'est-à-dire que la convergence vers ces distributions est atteinte en supposant que les maxima sont ceux d'échantillons de taille infinie (GEV) ou que le seuil tend vers l'infini (Pareto généralisée), ce qui est impossible en pratique. Par exemple, on utilise les maxima d'échantillons de taille $n = 365$ (c'est-à-dire le maxima annuel) pour l'ajustement de la distribution GEV à des données journalières de précipitation. Un ensemble de données échantillonné sur 20 ans ne permet donc l'utilisation que de 20 maxima.

Cependant, l'analyse fréquentielle, par l'utilisation de variables explicatives, peut potentiellement bien s'adapter au contexte non stationnaire induit par le dérèglement climatique. De plus, plusieurs méthodes de calcul de l'incertitude des valeurs estimées existent et sont disponibles selon l'approche et la méthode d'estimation.

2.4 PMP et changements climatiques

Dans son guide de 2009, l'OMM détermine que les PMP sont des valeurs stationnaires. Ainsi, si l'on suit les indications de l'organisme, les changements climatiques ne sont pas pris en compte lors du calcul de la PMP. Or, il est admis que ceux-ci ont un impact direct sur

les précipitations extrêmes et devraient donc être considérés lors de l'estimation. À l'aide de simulations de climats projetés, Kunkel *et al.* (2013) démontrent une croissance de la concentration de vapeur d'eau dans l'air au niveau mondial, sans évolution suffisante des valeurs du mouvement vertical ascendant ou de la vitesse du vent horizontal, phénomènes permettant de compenser la hausse de l'humidité de l'air. Cette augmentation implique des valeurs futures de EP_{max} plus grandes que dans le passé et ainsi une croissance de la PMP.

Beauchamp (2010) et Beauchamp *et al.* (2013) initient une méthodologie usant de simulations climatiques afin d'estimer des PMP actuelles et futures spécifiques au territoire québécois et la testent sur le bassin versant Manic-5. Rousseau *et al.* (2014) actualisent et appliquent la méthode de Beauchamp *et al.* (2013) sur les bassins versants Kénogami et Yamaska. Les résultats de Beauchamp *et al.* (2013) et Rousseau *et al.* (2014) appuient l'hypothèse d'une PMP en augmentation sous l'effet des changements climatiques pour des bassins spécifiques du territoire québécois en prenant en compte l'augmentation théorique de l'humidité de l'air liée aux dérèglements climatiques démontrée par Kunkel *et al.* (2013). Utilisant la méthode développée par Rousseau *et al.* (2014), Clavet-Gaumont *et al.* (2017) observe également une augmentation probable de la PMP estimée sur la période 2041-2070 par rapport à celle de la période 1971-2000 pour trois bassins (Kénogami, Saguenay et Manic-5) sur les cinq étudiés, avec une absence de consensus pour les deux derniers (Nelson et Mattagami).

Ben Alaya *et al.* (2020b) estime une augmentation de la PMP d'environ 4% pour 1°C de réchauffement sur le territoire du Canada et des États-Unis contigus à l'aide du modèle extrême bivarié de Ben Alaya *et al.* (2018) et de l'utilisation de la température comme covariable. Ailleurs dans le monde, Sarkar et Maity (2020) (Inde) et Visser *et al.* (2022) (Australie) estiment également des augmentations de la PMP reliées au dérèglement climatique pour divers scénarios de réchauffement.

2.5 Calcul des incertitudes

Les méthodes présentées jusqu'à présent ne retournent qu'une seule valeur, celle de la PMP. Cependant, il peut être utile d'estimer l'incertitude sur la valeur estimée puisque celle-ci ne peut être considérée que comme une approximation de la précipitation maximale probable réelle (OMM, 2009). En effet, le caractère extrême de la PMP et l'absence de précipitation observée d'ampleur similaire rend son estimation hautement incertaine. À l'exception des méthodes fréquentielles d'estimation, l'incertitude est relativement complexe à intégrer aux calculs. Certains chercheurs se sont néanmoins penché sur le problème, en particulier pour la maximisation de l'humidité et la méthode empirique d'Hershfield.

Pour ce qui est de la maximisation de l'humidité, Ben Alaya *et al.* (2018) supposent que les maxima annuels d'efficacité de précipitation et d'eau précipitable suivent une distribution GEV. Les auteurs construisent ainsi un modèle extrême bivarié à l'aide de différentes copules. L'article démontre qu'il est possible de retrouver la PMP calculée de façon déterministe à l'aide d'une méthodologie raisonnable sur le plan statistique. L'utilisation d'un tel modèle probabiliste permet de déterminer un intervalle de valeurs de PMP à l'aide du calcul des percentiles 10%, 50% et 90% pour une grande région des États-Unis et du Canada. Micovic *et al.* (2015) identifient les sources d'incertitudes hydrométéorologiques lors de l'estimation de la PMP et étudient les effets qu'ont les différentes variables sur la valeur de celle-ci. Dans un même ordre d'idée, d'autres chercheurs estiment l'incertitude sur la PMP en testant plusieurs méthodes d'approximation de l'eau précipitable *EP*, à partir de données observées (Chen et Bradley, 2006) et de simulations climatiques (Clavet-Gaumont *et al.*, 2017; Rouhani et Leconte, 2020). Salas *et al.* (2014) se sont pour leur part attardés à l'étude de l'incertitude sur les estimations de la PMP estimée à l'aide de la méthode empirique d'Hershfield. Ceux-ci s'intéressent particulièrement à l'incertitude causée par le calcul des moments empiriques d'ordre 1 et 2 utilisés par cette méthode.

CHAPITRE 3 THÉORIE

La loi bêta est une loi de probabilité continue et souple dont la fonction de densité de probabilité est définie par

$$f_{(X|\alpha,\beta)}(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}, \quad (3.1)$$

définie sur l'intervalle unitaire avec $\alpha, \beta > 0$ des paramètres de forme et Γ la fonction gamma. Cette distribution est versatile et fréquemment utilisée afin de modéliser les variables aléatoires continues restreintes à l'intervalle $[0, 1]$. Il est possible de généraliser cette distribution vers tout intervalle continu et fini des réels à l'aide de deux paramètres de borne $a \in \mathbb{R}$ et $b \in \mathbb{R}$, où $a < b$, à l'aide de la transformation $Y = X(b-a) + a$, pour $X \sim \text{Beta}(\alpha, \beta)$ et $Y \sim \text{Beta}(a, b, \alpha, \beta)$. La fonction de densité de cette nouvelle distribution, qu'on appellera bêta à quatre paramètres ou distribution de Pearson de type I (Johnson *et al.*, 1995), est donnée par

$$f_{(Y|a,b,\alpha,\beta)}(y) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{(y-a)^{\alpha-1} (b-y)^{\beta-1}}{(b-a)^{\alpha+\beta-1}}, \quad a < y < b. \quad (3.2)$$

La forme de la distribution dépend des paramètres (α, β) . Le cas $\alpha = \beta$ implique que la densité de probabilité est symétrique. Lorsque $\alpha < \beta$, le poids de la distribution se déplace vers la gauche et inversement lorsque $\alpha > \beta$. Plus l'écart entre les deux paramètres est grand, plus la densité de probabilité est asymétrique. La fonction de densité est convexe lorsqu'un ou deux paramètres ≤ 1 et concave lorsque les deux paramètres sont > 1 . Quelques exemples de distributions possibles sont tracées sur les graphiques de la figure 3.1.

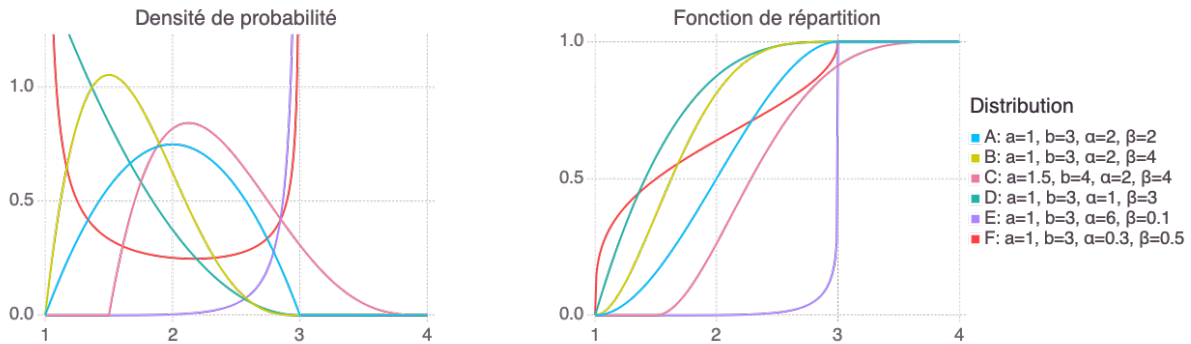


FIGURE 3.1 Graphiques de densités de probabilité (à gauche) et de fonctions de répartition (à droite) possibles pour la distribution de Pearson de type I

Il est possible de ré-exprimer la distribution de Pearson de type I en utilisant les paramètres de formes donnés par

$$\mu = \frac{\alpha}{\alpha + \beta}, \quad \nu = \alpha + \beta$$

avec $\alpha = \mu\nu$ et $\beta = \nu(1 - \mu)$ (Zhou et Huang, 2022). Ainsi, on obtient la fonction de densité

$$f_{(Y|a,b,\mu,\nu)}(y) = \frac{\Gamma(\mu)}{\Gamma(\mu\nu)\Gamma(\nu(1 - \mu))} \frac{(y - a)^{\mu\nu-1}(b - y)^{\nu(1-\mu)-1}}{(b - a)^{\nu-1}}, \quad a < y < b. \quad (3.3)$$

On remarque que le nouveau paramètre μ correspond à la moyenne de la distribution transposée sur l'intervalle $[0, 1]$, c'est-à-dire la moyenne de la distribution bêta standard associée. Les quatre premiers moments (moyenne, variance, asymétrie et kurtosis) peuvent être exprimés comme :

$$m = a + \mu(b - a) \quad (3.4)$$

$$v = \frac{\mu(1 - \mu)}{\nu + 1}(b - a)^2 \quad (3.5)$$

$$s = \frac{2(1 - 2\mu)\sqrt{\nu + 1}}{(\nu + 2)\sqrt{\mu(1 - \mu)}} \quad (3.6)$$

$$k = \frac{6(\mu - 1\nu + 2\mu^2 - \mu\nu - 2\mu + \nu + 1)}{\mu(1 - \mu)(\nu + 2)(\nu + 3)} + 3. \quad (3.7)$$

3.1 Méthodes d'estimation des paramètres

3.1.1 Méthode des moments

Les quatre premiers moments de la paramétrisation standard sont donnés par les expressions suivantes :

$$m = a + \frac{\alpha(b - a)}{\alpha + \beta} \quad (3.8)$$

$$v = \frac{\alpha\beta(b - a)^2}{(\alpha + \beta)^2(\alpha + \beta + 1)} \quad (3.9)$$

$$s = \frac{2(\beta - \alpha)\sqrt{\alpha + \beta + 1}}{(\alpha + \beta + 2)\sqrt{\alpha\beta}} \quad (3.10)$$

$$k = \frac{6(\alpha^3 - \alpha^2(2\beta - 1) + \beta^2(\beta + 1) - 2\alpha\beta(\beta + 2))}{\alpha\beta(\alpha + \beta + 2)(\alpha + \beta + 3)} + 3 \quad (3.11)$$

On peut remarquer que les moments d'ordre 3 et 4, données par les équations 3.10 et 3.11, ne dépendent que des paramètres de forme α et β . À partir de ces quatre expressions, il est possible de dériver les quatre paramètres de la loi de Pearson de type I (Johnson *et al.*, 1995). Ainsi, on a

$$\alpha = -\frac{s(k+3) + \left(\frac{-s(k+3) - \sqrt{s^2(k+3)^2 - 4(2k-3s^2-6)(4k-3s^2)}}{2(2k-3s^2-6)} \right) (10k - 12s^2 - 18)}{\sqrt{(s(k+3))^2 - 4(2k-3s^2-6)(4k-3s^2)}} + 1 \quad (3.12)$$

$$\beta = -\frac{s(k+3) - \left(\frac{-s(k+3) + \sqrt{s^2(k+3)^2 - 4(2k-3s^2-6)(4k-3s^2)}}{2(2k-3s^2-6)} \right) (10k - 12s^2 - 18)}{\sqrt{(s(k+3))^2 - 4(2k-3s^2-6)(4k-3s^2)}} + 1 \quad (3.13)$$

$$a = m - \frac{\alpha\sqrt{v}}{\alpha + \beta} \left(\frac{\sqrt{s^2(k+3)^2 - 4(2k-3s^2-6)(4k-3s^2)}}{(2k-3s^2-6)} \right) \quad (3.14)$$

$$b = a + \sqrt{v} \left(\frac{\sqrt{s^2(k+3)^2 - 4(2k-3s^2-6)(4k-3s^2)}}{(2k-3s^2-6)} \right) \quad (3.15)$$

Afin d'estimer les paramètres de la distribution à partir d'un échantillon fini, il suffit de calculer les moments empiriques des données, puis de remplacer ceux-ci dans les équations 3.12 à 3.15.

3.1.2 Maximum de la vraisemblance

La maximisation de la vraisemblance est une méthode qui, comme son nom l'indique, cherche à trouver les paramètres telle que la vraisemblance du modèle échantillonnal est maximisée. Afin de simplifier les calculs, on utilise généralement la log-vraisemblance. Dans le cas où $Y_i \sim \text{Beta}(a, b, \alpha, \beta)$, dont la densité de probabilité est donnée par l'équation 3.2, la log-vraisemblance d'un échantillon de taille n correspond à

$$\begin{aligned} \mathcal{L}(a, b, \alpha, \beta; \mathbf{y}) &= \ln \left\{ \prod_{i=1}^n \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{(y_i - a)^{\alpha-1} (b - y_i)^{\beta-1}}{(b - a)^{\alpha+\beta-1}} \right\} \\ &= -n \ln \left\{ \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \right\} - n(\alpha + \beta - 1) \ln(b - a) \\ &\quad + (\alpha - 1) \sum_{i=1}^n \ln(y_i - a) + (\beta - 1) \sum_{i=1}^n \ln(b - y_i) \end{aligned}$$

Cette fonction est optimisée numériquement par rapport aux paramètres afin d'obtenir les estimations $(\hat{a}, \hat{b}, \hat{\alpha}, \hat{\beta})$. En effet, la fonction gamma, qui apparaît à trois reprises dans la densité de probabilité, ainsi que la fonction digamma, sa dérivée, ne permettent pas une solution analytique. Lorsque les paramètres de bornes sont connus, il est possible d'estimer

itérativement les paramètres de forme (Johnson *et al.*, 1995). Lorsque ceux-ci sont assez grands (généralement $a > 2$ et $b > 2$), Johnson *et al.* (1995) propose une approximation analytique comme point de départ à partir de l'approximation $\psi(t) \approx \ln(t - 1/2)$, avec ψ la fonction digamma.

3.1.3 Inférence bayésienne

Le cadre bayésien diffère de celui fréquentiste (méthode des moments, maximisation de la vraisemblance) par son interprétation des paramètres du modèle échantillonnal. En effet, pour une distribution $f_{(\mathbf{Y}|\boldsymbol{\theta})}(\mathbf{y})$, on considère non seulement $\mathbf{Y}$ mais également $\boldsymbol{\theta}$ comme une variable aléatoire. Ainsi, on peut octroyer une distribution $f_{\boldsymbol{\theta}}(\boldsymbol{\theta})$, qualifiée *a priori*, à cette seconde variable aléatoire. Cette distribution permet de modéliser les informations connues de $\boldsymbol{\theta}$. Dans l'éventualité où aucune ou peu d'information est disponible au préalable, il est possible d'utiliser des distributions *a priori* vagues ou impropres, qui ont une moins grande influence sur l'estimation finale de $\boldsymbol{\theta}$. Cette interprétation des paramètres nous permet l'utilisation du théorème de Bayes, qui indique que

$$f_{(\boldsymbol{\theta}|\mathbf{Y}=\mathbf{y})}(\boldsymbol{\theta}) = \frac{f_{(\mathbf{Y}|\boldsymbol{\theta})}(\mathbf{y}) \times f_{\boldsymbol{\theta}}(\boldsymbol{\theta})}{\int_{\boldsymbol{\theta}} f_{(\mathbf{Y}|\boldsymbol{\theta})}(\mathbf{y}) \times f_{\boldsymbol{\theta}}(\boldsymbol{\theta}) d\boldsymbol{\theta}}, \quad (3.16)$$

avec $\boldsymbol{\theta} \in \Theta$ et $f_{(\boldsymbol{\theta}|\mathbf{Y})}(\boldsymbol{\theta})$ la distribution *a posteriori* de $\boldsymbol{\theta}$, c'est-à-dire la distribution estimée du paramètre $\boldsymbol{\theta}$ sachant les données observées $\mathbf{y}$. Ainsi, plutôt que d'obtenir une estimation ponctuelle des paramètres du modèle échantillonnal, on obtient une distribution de probabilité associée à ceux-ci.

Peu de modèles échantillonnaux n'ont qu'un seul paramètre. Par exemple, dans le cas de la distribution de Pearson de type I, $\boldsymbol{\theta} = [a, b, \alpha, \beta]$. La distribution *a posteriori* $f_{(\boldsymbol{\theta}|\mathbf{Y})}(\boldsymbol{\theta}) = f_{(a,b,\alpha,\beta|\mathbf{Y})}(a, b, \alpha, \beta)$ est donc une distribution conjointe de tous les paramètres. Généralement, certains paramètres sont d'intérêt alors que d'autres non : ceux-ci sont appelés paramètres de nuisance. Afin d'obtenir la loi *a posteriori* marginale d'un paramètre d'intérêt, par exemple la distribution $f_{(a|\mathbf{Y})}(a)$ pour a , on doit d'abord estimer la loi conjointe pour tous les paramètres $f_{(a,b,\alpha,\beta|\mathbf{Y})}(a, b, \alpha, \beta)$, puis intégrer cette distribution par rapport aux paramètres de nuisance b, α et β :

$$f_{(a|\mathbf{Y})}(a) = \int_{\Theta_2 \times \Theta_3 \times \Theta_4} f_{(a,b,\alpha,\beta|\mathbf{Y})}(a, b, \alpha, \beta) db d\alpha d\beta,$$

où $b \in \Theta_2$, $\alpha \in \Theta_3$ et $\beta \in \Theta_4$.

Dans le cas des méthodes d'échantillonnage, introduites à la section suivante, il n'est pas nécessaire d'intégrer la distribution *a posteriori* conjointe afin d'obtenir les lois *a posteriori* marginales d'intérêt, on peut simplement faire abstraction des observations liées aux para-

mètres de nuisance.

3.2 Méthodes Monte-Carlo par chaîne de Markov

Dans un contexte bayésien, afin de trouver la distribution *a posteriori*, on utilise des méthodes numériques d'échantillonnage. En effet, les calculs menant à l'estimation de la loi *a posteriori* demandent de grandes ressources computationnelles, particulièrement celui de l'intégrale au dénominateur de l'équation 3.16. L'objectif d'un algorithme d'échantillonnage est de créer un échantillon provenant de la distribution *a posteriori* sans avoir à calculer explicitement celle-ci. Les méthodes de Monte-Carlo par chaînes de Markov (MCMC) sont une des classes d'échantillonneurs répandues, utilisées pour leur mise en œuvre relativement facile en dépit de la complexité des densités à échantillonner. Ces méthodes sont généralement fiables, bien qu'il puisse être difficile d'en déterminer leur précision et d'évaluer leur convergence.

Les chaînes de Markov sont un processus stochastique qui sont, dans le cas présent, sans mémoire. Ainsi, seul l'état actuel influence le choix de l'état subséquent. Ainsi, les MCMC ont pour but de créer une chaîne de Markov dont la distribution stationnaire correspond à celle d'intérêt, $f_{(\theta|\mathbf{Y})}(\theta)$. Pour ce faire, des candidats sont générés à partir d'une distribution dite instrumentale, puis ajoutés ou non à la chaîne selon une probabilité d'acceptation donnée. Il existe un grand nombre d'algorithmes MCMC, puisqu'il est relativement simple d'ajuster ceux-ci selon le problème d'intérêt en personnalisant la distribution instrumentale, la probabilité d'acceptation ou encore l'approche utilisée afin de générer les candidats.

3.2.1 Algorithme de Metropolis-Hasting

L'algorithme de Metropolis-Hasting est une classe de méthodes MCMC qui permet d'échantillonner la distribution *a posteriori* connue à une constante de normalisation près. Dans le cas de la distribution de l'équation 3.16, la constante de normalisation correspond à l'intégrale au dénominateur. On note $\tilde{f}_{(\theta|\mathbf{Y})}(\theta) = f_{(\mathbf{Y}|\theta)}(\mathbf{y}) \times f_{\theta}(\theta)$ cette fonction proportionnelle à la distribution *a posteriori*, et θ_0 l'état initial du paramètre choisi arbitrairement. L'échantillonnage se déroule comme suit : pour $i \in \{0, \dots, n-1\}$,

1. On génère un candidat θ_{i+1}^* de la distribution instrumentale $q_{(\theta^*|\theta_i)}(\theta^*)$;
2. On calcule le taux d'acceptation

$$e = \frac{\tilde{f}_{(\theta_{i+1}^*|\mathbf{Y})}(\theta_{i+1}^*) q_{(\theta_i|\theta_{i+1}^*)}(\theta_i)}{\tilde{f}_{(\theta_i|\mathbf{Y})}(\theta_i) q_{(\theta_{i+1}^*|\theta_i)}(\theta_{i+1}^*)};$$

3. On génère une valeur u d'une Uniforme(0, 1)
 - si $u \leq e$, on accepte le candidat et on pose $\theta_{i+1} = \theta_{i+1}^*$

— sinon, on rejette le candidat et on pose $\theta_{i+1} = \theta_i$.

On note que le premier point de l'étape 3 implique que si le taux d'acceptation ≥ 1 , alors on accepte nécessairement le candidat. De plus, on peut remarquer que

$$\begin{aligned} e &= \frac{\tilde{f}_{(\theta_{i+1}^*|\mathbf{Y})}(\theta_{i+1}^*) q_{(\theta_i|\theta_{i+1}^*)}(\theta_i)}{\tilde{f}_{(\theta_i|\mathbf{Y})}(\theta_i) q_{(\theta_{i+1}^*|\theta_i)}(\theta_{i+1}^*)} \\ &= \frac{\tilde{f}_{(\theta_{i+1}^*|\mathbf{Y})}(\theta_{i+1}^*) q_{(\theta_i|\theta_{i+1}^*)}(\theta_i)}{\tilde{f}_{(\theta_i|\mathbf{Y})}(\theta_i) q_{(\theta_{i+1}^*|\theta_i)}(\theta_{i+1}^*)} \times \frac{\int_{\Theta} f_{(\mathbf{Y}|\theta)}(\mathbf{y}) f_{\theta}(\theta) d\theta}{\int_{\Theta} f_{(\mathbf{Y}|\theta)}(\mathbf{y}) f_{\theta}(\theta) d\theta} \\ &= \frac{f_{(\theta_{i+1}^*|\mathbf{Y})}(\theta_{i+1}^*) q_{(\theta_i|\theta_{i+1}^*)}(\theta_i)}{f_{(\theta_i|\mathbf{Y})}(\theta_i) q_{(\theta_{i+1}^*|\theta_i)}(\theta_{i+1}^*)}. \end{aligned}$$

Bien que l'algorithme de Metropolis-Hasting garantisse d'échantillonner tôt ou tard la loi cible, l'échantillon obtenu peut être très auto-corrélé. Ceci s'explique par une exploration assez locale des états possibles à chacune des itérations, causée par l'amplitude limitée du pas généré par la distribution instrumentale. Ces méthodes demandent donc un grand nombre d'itérations avant de bien représenter la distribution d'intérêt.

D'ordinaire, la chaîne produite par l'algorithme n'est pas utilisée dans son entièreté. En effet, on coupe les premières itérations générées lors l'étape appelée «période de chauffe». Au cours de cette phase, la dispersion de la distribution instrumentale est ajustée afin d'obtenir un taux d'acceptation des candidats. Afin de diminuer la corrélation entre les états, il est également possible d'amincir la chaîne en ne gardant qu'un maillon par groupe de k états générés.

3.2.2 Échantillonnage de Gibbs

Lorsque le modèle comporte plusieurs paramètres, comme dans le cas de la Pearson de type I où $\theta = [a, b, \alpha, \beta]$, l'algorithme de Metropolis-Hasting standard génère, puis accepte ou refuse θ en bloc. Il est cependant possible de ne générer qu'un seul élément du vecteur paramètre à la fois. Cette catégorie d'algorithmes de Metropolis-Hasting est appelé échantillonneur de Gibbs.

Généralement, celui-ci prend avantage des distributions *a posteriori* dites conditionnelles complètes, qui sont associées à un cas spécial de lois *a priori*, conjuguées au modèle échantillonnal. Cette astuce permet d'éviter le calcul du taux d'acceptation des étapes 2 et 3 de l'algorithme de Metropolis-Hasting. Cependant, les distributions *a priori* conjuguées ne sont pas toujours aisées à déterminer et certaines lois de probabilités n'en n'ont tout simplement pas. Il est alors toujours possible d'utiliser l'algorithme de Gibbs mais les étapes 2 et 3 ne peuvent être écartées. Un à un, chaque élément du vecteur paramètre est ainsi tiré, accepté ou refusé selon le taux d'acceptation. Ceux-ci sont mis à jour préalablement au tirage de

l'élément suivant.

3.2.3 Monte-Carlo hybride

On appelle Monte Carlo hybride, ou *Hamiltonian Monte Carlo*, un algorithme de Metropolis-Hasting qui utilise la théorie développée en dynamique hamiltonienne afin de se déplacer dans l'espace des états. Cette méthode permet une exploration de l'espace plus rapide qu'un algorithme de Metropolis-Hasting régulier puisqu'elle propose des états successifs éloignés avec une grande probabilité d'acceptation. Cette distance entre les états successifs réduit également leur corrélation, ce qui implique que le nombre d'échantillons nécessaires à l'algorithme afin de représenter la distribution *a posteriori* est moins grand.

On pose ϕ , une variable auxiliaire indépendante de même dimension que θ telle que $\phi \sim \mathcal{N}(0, M)$ avec M généralement une matrice diagonale. Cette nouvelle variable est appelée quantité de mouvement, ou *momentum*, en parallèle avec la dynamique hamiltonienne. Comme en physique, on a $\phi = M \frac{d\theta}{dt}$. Un système d'équations différentielles relie les deux variables θ et ϕ de la façon suivante :

$$\frac{d\theta}{dt} = \frac{\partial H}{\partial \phi} \quad (3.17)$$

$$\frac{d\phi}{dt} = -\frac{\partial H}{\partial \theta}. \quad (3.18)$$

Le terme à droite de l'équation 3.18 correspond au gradient de la distribution *a posteriori*. En effet, H se réfère à l'hamiltonien, défini comme

$$H(\theta, \phi) = U(\theta) + K(\phi),$$

un opérateur représentant l'énergie totale d'un système. Ainsi, U se rapporte à l'énergie potentielle et K à l'énergie cinétique. Si on augmente la densité *a posteriori* de ϕ afin d'obtenir la densité conjointe $f_{(\theta, \phi|\mathbf{Y})}(\theta, \phi) = f_{(\theta|\mathbf{Y})}(\theta)f_{\phi}(\phi)$, on a donc

$$-U(\theta) = \ln f_{(\theta|\mathbf{Y})}(\theta) \propto \ln f_{(\mathbf{Y}|\theta)}(\mathbf{y}) + \ln f_{\theta}(\theta) \quad (3.19)$$

$$-K(\phi) = \ln f_{\phi}(\phi) \quad (3.20)$$

$$\Rightarrow -H(\theta, \phi) = \ln f_{(\theta, \phi|\mathbf{Y})}(\theta, \phi) \quad (3.21)$$

Finalement, on note que la variable t ne correspond pas réellement au temps mais fait plutôt référence à l'évolution de la chaîne au cours de la progression de l'algorithme.

Avant d'expliciter celui-ci, il semble important de définir la méthode saute-mouton, ou *leap-frog*, qui est utilisée afin de résoudre numériquement le système d'équations différentielles.

Un saut est divisé en trois étapes : pour l'état i ,

1. D'abord, ϕ_i effectue un demi-saut. On utilise 3.18 tel que

$$\phi_i \left(t + \frac{\Delta t}{2} \right) = \phi_i(t) - \frac{\Delta t}{2} \frac{dU(\boldsymbol{\theta})}{d\boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_i(t)}$$

Puisque l'on effectue généralement $L > 1$ sauts de suite, L et Δt sont choisis tels que $L \cdot \Delta t = 1$.

2. On met ensuite $\boldsymbol{\theta}_i$ à jour à l'aide de 3.17

$$\begin{aligned} \boldsymbol{\theta}_i(t + \Delta t) &= \boldsymbol{\theta}_i(t) - \Delta t \frac{dK(\phi)}{d\phi} \Big|_{\phi=\phi_i(t+\Delta t/2)} \\ &= \boldsymbol{\theta}_t + \Delta t M^{-1} \phi_i \left(t + \frac{\Delta t}{2} \right) \end{aligned}$$

3. Finalement, ϕ_i complète son saut

$$\phi_i(t + \Delta t) = \phi_i \left(t + \frac{\Delta t}{2} \right) - \frac{\Delta t}{2} \frac{dU(\boldsymbol{\theta})}{d\boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_i(t+\Delta)}$$

Concrètement, l'amplitude du déplacement des variables $\boldsymbol{\theta}$ et ϕ est déterminé par le niveau d'énergie du système : un niveau élevé implique un déplacement important alors qu'une quantité d'énergie faible implique un déplacement moindre.

L'algorithme Monte Carlo hybride peut ainsi être formulé de la manière suivante : pour $i \in \{0, \dots, n-1\}$

1. On sélectionne aléatoirement le niveau d'énergie de l'itération en tirant ϕ_i de la distribution normale $\mathcal{N}(0, M)$. Si $i = 0$, l'état initial de $\boldsymbol{\theta}$ doit également être choisi arbitrairement ;
2. $(\boldsymbol{\theta}_i, \phi_i)$ effectuent L sauts afin de générer un candidat $(\boldsymbol{\theta}^*, \phi^*)$;
3. On calcule le taux d'acceptation

$$e = \frac{f_{(\boldsymbol{\theta}^*, \phi^* | \mathbf{Y})}(\boldsymbol{\theta}^*, \phi^*)}{f_{(\boldsymbol{\theta}_i, \phi_i | \mathbf{Y})}(\boldsymbol{\theta}_i, \phi_i)},$$

4. Afin d'accepter ou de refuser le candidat, on exécute la troisième étape de l'algorithme de Metropolis-Hasting général défini au début de la sous-section. Le paramètre ϕ_{i+1} n'est pas enregistré puisqu'il n'est pas d'intérêt et sera sélectionné aléatoirement au début de l'itération suivante.

On prend par exemple $\phi \sim \mathcal{N}(0, 1)$, $\theta | \mathbf{Y} \sim \text{Gamma}(2, 3)$ et $L = 10$. L'énergie H du modèle en fonction de ces deux paramètres est illustré par la figure 3.2. Pour une certaine itération i , on a $\theta_i(t)$ et $\phi_i(t)$ tels que $H(\theta_i(t), \phi_i(t)) = 4.0$. On utilise la conservation de l'énergie du système (équations 3.17 et 3.18) afin d'explorer ce niveau, c'est-à-dire que les L sauts de mouton de l'itération i suivent la courbe $H = 4.0$. Ainsi, le chemin emprunté par la méthode saute-mouton est tracé en rouge et est borné à droite par le candidat (θ^*, ϕ^*) . Celui-ci sera accepté ou refusé selon le taux d'acceptation calculé.

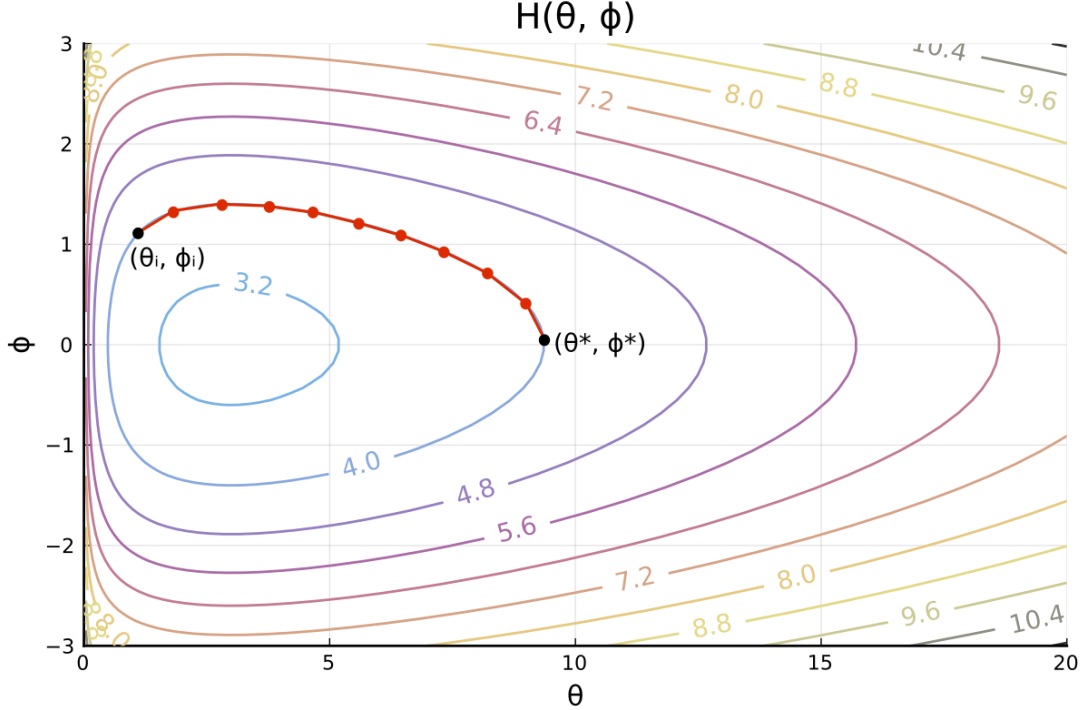


FIGURE 3.2 Courbes de niveaux associées à différentes valeurs de l'énergie du modèle en fonction du couple (θ, ϕ) . On remarque que les zones associée à de fortes probabilités d'occurrence pour $f_{(\theta, \phi | \mathbf{Y})}(\theta, \phi)$ (au centre à gauche) sont reliée à des niveaux d'énergie faibles, comme l'équation 3.21 l'indique. Les droites et points rouges représentent $L = 10$ sauts de la méthode saute-mouton, avec le point initial de l'itération $(\theta_i(t), \phi_i(t))$ et le candidat proposé (θ^*, ϕ^*) .

Le nombre optimal de sauts L peut être estimé à chaque itération afin de maximiser la distance entre deux états consécutifs et diminuer la corrélation entre ceux-ci. Cette variante de l'algorithme de Monte Carlo hybride est appelée *No U-Turn Sampler*, ou NUTS.

CHAPITRE 4 MÉTHODOLOGIE

Selon la définition de l'OMM de la PMP, chaque événement de pluie correspond à une fraction de la valeur de la PMP. Il est ainsi possible de relier cette fraction à une variable aléatoire suivant une distribution bêta définie sur l'intervalle $[0, 1]$. De même, il est envisageable de modéliser la hauteur des précipitations plutôt que la fraction de la PMP réalisée. Ainsi, chaque événement de pluie est une réalisation de la distribution de Pearson de type I avec la PMP comme borne supérieure. Ceci signifie que l'on associe chaque événement de pluie à la variable y et la PMP à la borne b de l'équation 3.2. Par conséquent, il suffit d'ajuster le modèle à l'ensemble des précipitations observées d'un lieu donné afin d'estimer la PMP de cet endroit. La PMP ainsi estimée grâce à cette méthode est une PMP théorique plutôt qu'opérationnelle.

Cette approche permet d'utiliser l'entièreté des données de précipitations plutôt qu'un sous-ensemble de celles-ci, comme dans le cas de l'analyse fréquentielle qui n'utilise que les maxima annuels (GEV) ou les dépassements de seuil (Pareto généralisée), ou encore de la maximisation de l'humidité qui n'emploie qu'une sélection de tempêtes extrêmes. Ainsi, la subjectivité liée à la sélection des événements inclus dans le calcul de la PMP est évitée. Cette approche suit également la définition de l'OMM qui décrit la PMP comme «the theoretical maximum precipitation for a given duration under modern meteorological conditions» OMM (2009). La distribution choisie semble idéale par sa flexibilité et l'existence de limites inférieure et supérieure. De plus, cette méthode peut potentiellement prendre en compte des covariables comme l'année ou la température, ce qui permettrait de considérer les CC dans l'estimation de la PMP, résolvant ainsi facilement le problème de non-stationnarité du climat. La PMP étant l'un des paramètres de la distribution, une estimation de l'incertitude est également accessible.

Généralement, les échantillons de précipitations journalières observées ont une forme très asymétrique, avec un poids important pour les petites précipitations. On peut le voir à la figure 4.1 qui représente les pluies journalières observées à Montréal. On s'attend donc à ce que la valeur du paramètre de forme α soit beaucoup plus petite que celle de β , avec $\alpha \leq 1$.

Différentes approches sont étudiées afin d'ajuster la distribution aux données de précipitations, soit : la méthode des moments (MM), la maximisation de la vraisemblance (MV), l'échantillonneur de Gibbs, l'algorithme NUTS ainsi qu'une procédure associant les deux méthodes fréquentistes, appelée ici méthode mixte. Cette dernière use de la méthode des moments afin d'estimer les paramètres de forme de la distribution (α, β) , puis emploie le

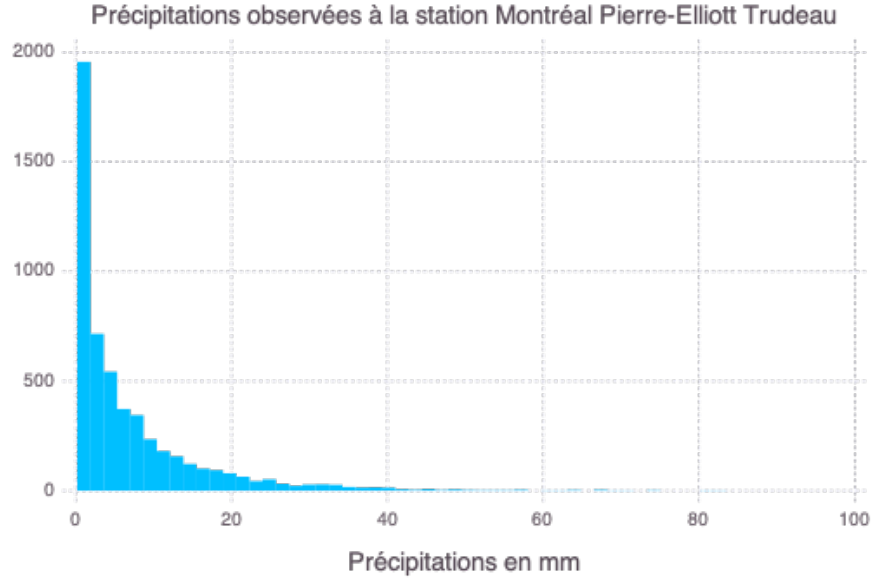


FIGURE 4.1 Précipitations journalières observées à la station Montréal-Trudeau du 1er mai 1953 au 28 mai 2024

maximum de vraisemblance afin d'estimer les paramètres de bornes (a, b) . Cette méthode cherche à tirer parti des propriétés des deux approches fréquentistes étudiées. C'est également la méthode qui permet d'estimer les valeurs initiales des estimations par maximum de vraisemblance.

L'estimation des paramètres par méthode des moments, présentée à la section 3.1.1, a l'avantage d'être relativement simple et rapide dans son exécution. Comme mentionné précédemment, il est possible de résoudre analytiquement le système d'équations 3.8 à 3.11 pour obtenir les expressions de chacun des paramètres. Cependant, si la taille de l'échantillon n'est pas suffisamment grande, il est possible que les moments empiriques ne soient pas une bonne approximation des moments théoriques. Les estimateurs obtenus à l'aide des équations 3.12 à 3.15 sont ainsi potentiellement incorrects et s'écartent parfois du support des paramètres.

La section 3.1.2 introduit l'estimation par maximum de vraisemblance et évoque la méthode itérative proposée par Johnson *et al.* (1995) lorsque les bornes de la distribution sont connues. Dans le cas présent, il est possible de fixer le paramètre de borne inférieure a , ce qui pourrait faciliter l'estimation par maximum de vraisemblance. Cependant, b est bien sûr indéterminé *a priori* et l'approche simplifiée ne peut pas être employée. De plus, un des paramètres de forme, α , est nécessairement ≤ 1 : il est donc impossible d'approximer la fonction digamma. Bien que la maximisation de la vraisemblance ne soit pas impossible, celle-ci ne peut pas être

facilitée par la méthode susmentionnée. Néanmoins, lorsque la taille de l'échantillon tend vers l'infini, l'estimateur par maximum de vraisemblance converge généralement en probabilité vers les valeurs réelles des paramètres. Ainsi, cette approche donne usuellement de meilleures estimations que la méthode des moments, lorsqu'elle converge.

De son côté, l'inférence bayésienne est moins sensible à l'identifiabilité des paramètres et des maxima locaux que les estimateurs fréquentistes énumérés plus tôt. Cette caractéristique est d'autant plus importante dans le cas présent au vu de la grande flexibilité de la loi Pearson de type I. Cette approche permet également une caractérisation plus précise de l'incertitude sur l'estimation des paramètres, attribut précieux dans l'optique de l'évaluation de celle-ci sur la PMP estimée.

4.1 Pearson de type I à quatre paramètres

On veut d'abord déterminer si la méthode d'estimation de la PMP à l'aide de la distribution de Pearson de type I est justifiable, c'est-à-dire si l'approche estime les bons paramètres pour des échantillons synthétiques. Pour ce faire, on crée plusieurs ensembles de m échantillons de taille variable $n \in \{100, 500, 1000, 2000, 5000, 8000\}$ et de paramètres (a, b, α, β) fixés. Puis, à partir de ces échantillons, on ajuste différents modèles de Pearson de type I. Les valeurs exactes des paramètres des quatre distributions échantillonnées sont données dans le tableau 4.1 et les distributions de probabilités sont tracées à la figure 4.2.

TABLEAU 4.1 Valeurs théoriques des paramètres des distributions échantillonnées pour le modèle de Pearson de type I

Distribution	a	b	α	β
A	0	50	2	2
B	0	50	1	3
C	0	50	0.1	6
D	0	50	0.1	15

Pour chaque taille n , on tire $m = 100$ échantillons pour lesquels on ajuste à chaque fois le modèle par méthode des moments, par maximisation de la vraisemblance et par la méthode mixte. La figure 4.3 tracent les moyennes des valeurs estimées pour le paramètre b , qui correspondrait à la PMP dans un contexte de pluie. On peut voir que lorsque les valeurs réelles des paramètres de forme α et β sont rapprochées, c'est-à-dire pour les distributions A et B, les trois méthodes d'estimation performant bien. En général, on peut voir que la méthode des moments est plus stable : même si les intervalles de confiance sont plus larges, la moyenne se trouve plus proche de la valeur réelle de b . Une moyenne plus éloignée couplée d'un intervalle

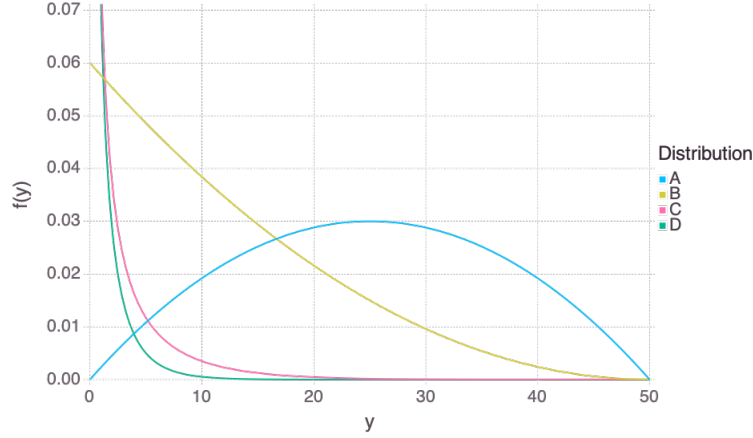


FIGURE 4.2 Fonctions de probabilité des distributions échantillonnées pour le modèle de Pearson de type I

de confiance plus étroit, comme dans le cas de la maximisation de la vraisemblance, implique qu'un petit nombre d'estimations se trouve très loin de la valeur réelle.

Lorsque les valeurs réelles des paramètres de forme s'éloignent, c'est-à-dire pour les distributions C et D, on peut voir dans la figure 4.3 qu'aucune des méthodes d'estimation ne fonctionne très bien. En général, la méthode mixte est stable, mais sous-estime fortement le paramètre de borne supérieure. Les résultats obtenus par maximisation de la vraisemblance, eux, sont très peu convaincants. On peut voir que la moyenne estimée de la maximisation de la vraisemblance pour la distribution C dépasse de loin les quantiles 97.5%, particulièrement pour les échantillons de taille 500, 1000 et 2000. Ces résultats sont dus à l'instabilité de la méthode, qui surestime de beaucoup la borne supérieure pour un petit nombre d'échantillons, ce qui déplace la moyenne vers le haut. Le tableau 4.2 affiche ce résultat dans les colonnes associées aux moyennes des estimations de b . Ainsi, on peut noter que malgré une mauvaise estimation moyenne de b , les quantiles 2.5% et 97.5% restent raisonnables (même s'ils sont inférieurs à la valeur réelle de b). Dans ce même tableau, on peut également remarquer que la performance de l'estimation moyenne des paramètres a et α est corrélée. Similairement, la performance du couple $(\hat{\bar{b}}, \hat{\bar{\beta}})$ semble elle aussi reliée.

TABLEAU 4.2 Moyennes des 100 estimations par maximisation de la vraisemblance des paramètres ainsi que les quantiles 2.5% et 97.5% des estimations de b pour la distribution C

Taille	$\bar{a}$	$\bar{b}$	$\bar{\alpha}$	$\bar{\beta}$	$\bar{q}_{b, 2.5\%}$	$\bar{q}_{b, 97.5\%}$
100	8.7×10^{-15}	15.1	0.1	1.6	6.8	28.6
500	-4.6×10^4	2678.7	1.1×10^9	7.0×10^7	14.1	31.9
1000	-6.2×10^4	2088.3	1.2×10^9	4.0×10^7	17.2	37.8
2000	-3.4×10^4	2844.4	8.6×10^8	7.0×10^7	20.7	39.6
5000	1.7×10^{-31}	30.9	0.1	7.3	24.8	39.0
8000	2.9×10^{-33}	31.9	0.1	5.8	26.8	40.0

Les écarts entre les valeurs estimées et théoriques observés à la figure 4.3 peuvent être expliqués par la grande flexibilité de la distribution. En effet, on peut imaginer que pour un échantillon donné de taille finie, les paramètres peuvent potentiellement prendre un très grand nombre de valeurs différentes qui permettent de bien ajuster la distribution. La figure 4.4 illustre ce manque d'identifiabilité des paramètres : en effet, alors même que les trois distributions représentées n'ont pas des paramètres similaires, les courbes tracées sur le graphique sont presque identiques. On peut d'ailleurs noter que la distribution dont les paramètres sont ajustés par maximisation de la vraisemblance, représentée par la courbe rose sur la figure, a une borne supérieure $b = 8.1 \times 10^7$, très éloignée de la valeur réelle $b = 1$ de la distribution échantillonnée.

Pour résoudre ce problème, on peut augmenter la taille des échantillons des distributions C et D. On teste chacune des trois méthodes d'estimation pour les tailles échantillonnales $\{50\ 000, 75\ 000, 100\ 000, 200\ 000, 300\ 000, 500\ 000\}$. Les résultats sont affichés dans les graphiques de la figure 4.5. Pour la méthode des moments, on remarque que l'algorithme se stabilise relativement rapidement, tendance observable déjà sur la figure 4.3. Autant la méthode mixte que la maximisation de la vraisemblance donnent des résultats améliorés par rapport aux estimations associées à des tailles échantillonnales plus petites, mais toujours pas probants. En général, ces deux techniques sous-estiment fortement la borne supérieure b .

4.1.1 Corrélation entre les paramètres

On a remarqué plus tôt, dans le tableau 4.2, que la performance des estimations des couples $(\hat{a}, \hat{\alpha})$ et $(\hat{b}, \hat{\beta})$ semblait corrélées. On peut facilement vérifier ces relations en traçant les nuages de points des estimations de ces couples de paramètres. Les figures 4.6, à 4.9 confirment la relation positive entre le paramètre de borne supérieure et β ainsi que celle négative entre le paramètre de borne inférieure et α pour les quatre distributions. On peut également trouver les coefficients de régression linéaire entre $(\hat{a}, \hat{\alpha})$ et $(\hat{b}, \hat{\beta})$ pour chacune des distributions et

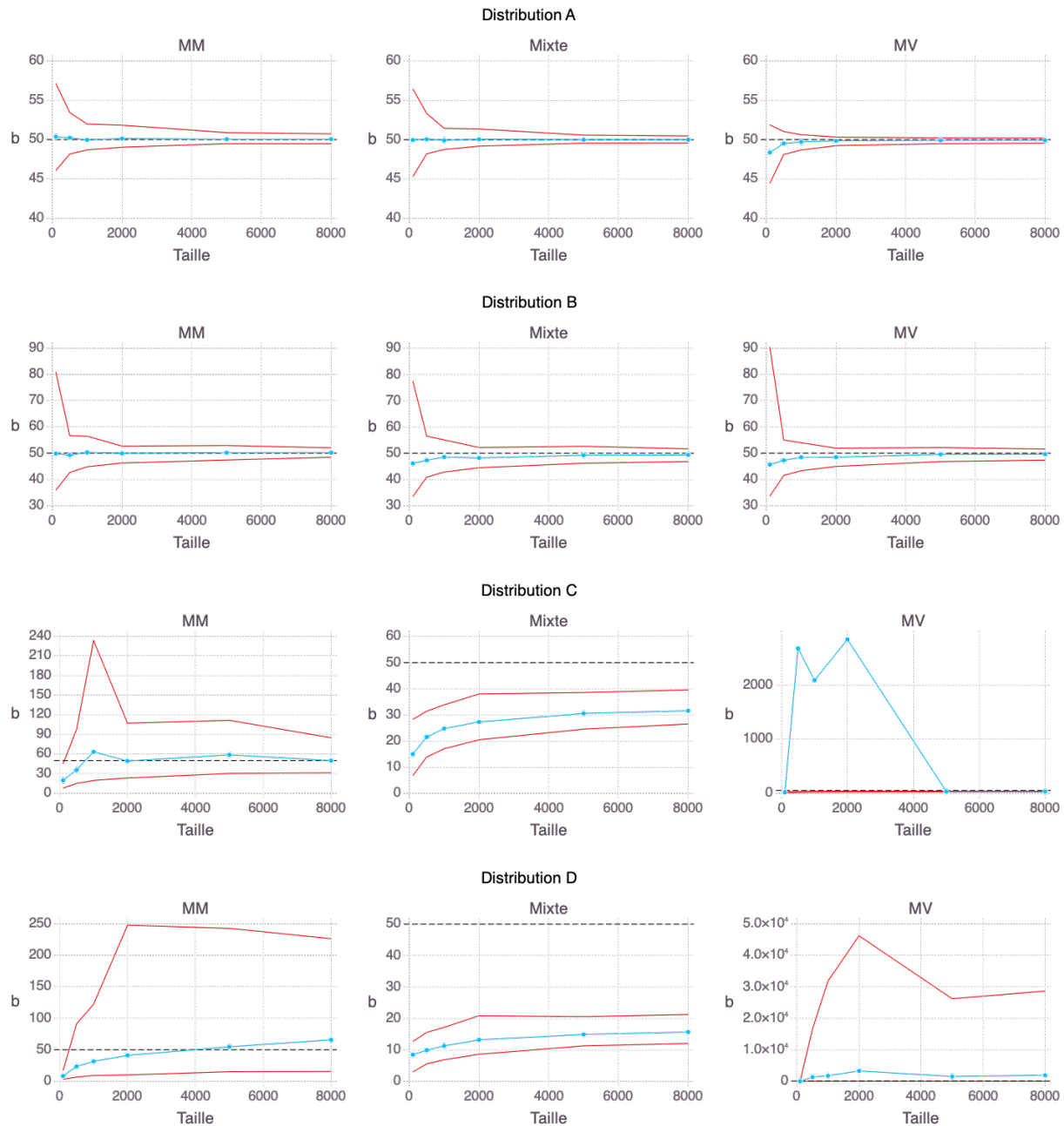


FIGURE 4.3 Estimation de b en fonction de la taille de l'échantillon, pour les distributions C et D du modèle bêta à quatre paramètres. De gauche à droite : méthode des moments, estimation mixte, maximisation de la vraisemblance. Les points bleus correspondent à la moyenne des estimations de b pour les 100 échantillons de même taille, tandis que les droites rouges correspondent aux quantiles 2.5% et 97.5%.

des méthodes d'estimations dans les tableaux A.1 et A.2 de l'annexe A.

Parce que l'estimation des paramètres est meilleure pour les distributions A et B, on peut

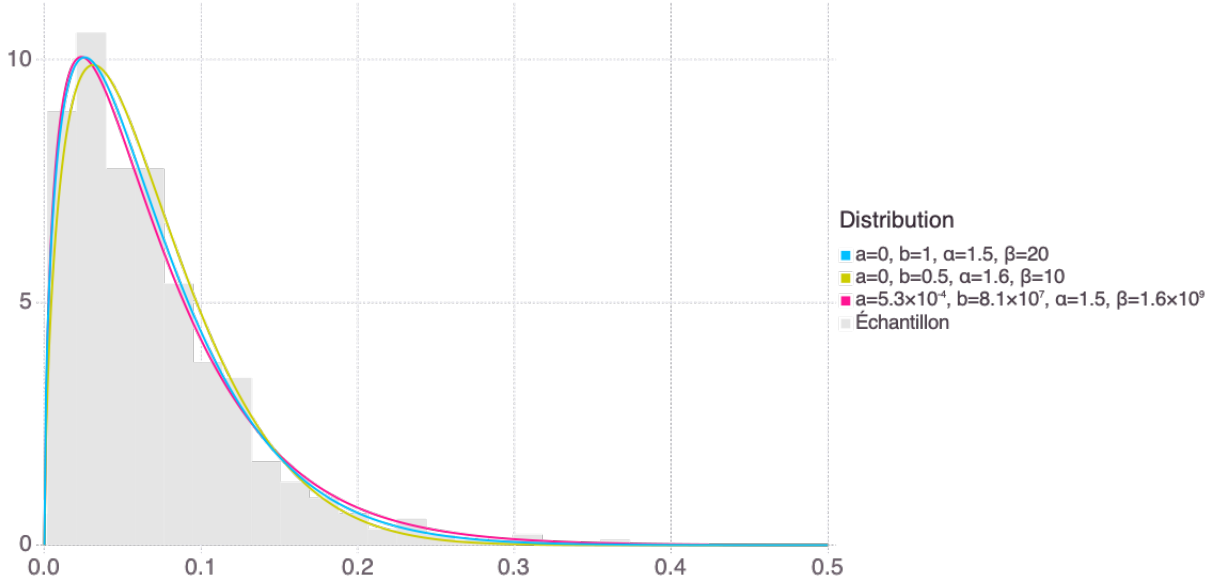


FIGURE 4.4 Illustration de la flexibilité de la distribution de Pearson de type I. L'échantillon, représenté par l'histogramme et composé de 500 données, est tiré d'une distribution de Pearson de type I de paramètres $a = 0$, $b = 1$, $\alpha = 1.5$ et $\beta = 20$, comme la courbe bleue.

remarquer aux figures 4.6 et 4.7 que les nuages de points associés à celles-ci ne varient que peu d'une méthode d'estimation à l'autre. Ce n'est cependant pas le cas pour les distributions C et D, aux figures 4.8 et 4.9. La méthode des moments étant plus stable, l'amplitude des axes est moindre que pour les autres méthodes. Néanmoins, on peut observer que la méthode mixte réduit potentiellement la relation linéaire entre les paramètres b et β .

Ainsi, une mauvaise estimation de l'un des paramètres de forme entraînera une mauvaise estimation du paramètre de borne associé et vice versa. La section 4.3 tentera de résoudre ce problème.

4.2 Pearson de type I à trois paramètres

Afin de réduire le problème d'identifiabilité des paramètres, la borne inférieure a est fixée à 0, ce qui correspond au cas d'absence de précipitation. La densité de probabilité devient

$$f_{(y|b,\alpha,\beta)}(y) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{y^{\alpha-1}(b-y)^{\beta-1}}{b^{\alpha+\beta-1}}, \quad 0 < y < b; \quad (4.1)$$

et il est possible de réitérer la procédure de la section précédente. Les distributions utilisées sont les mêmes que celles affichées sur le graphique de la figure 4.2 et les paramètres théoriques sont encore donnés par le tableau 4.1. Les résultats obtenus sont affichés à la figure 4.10.

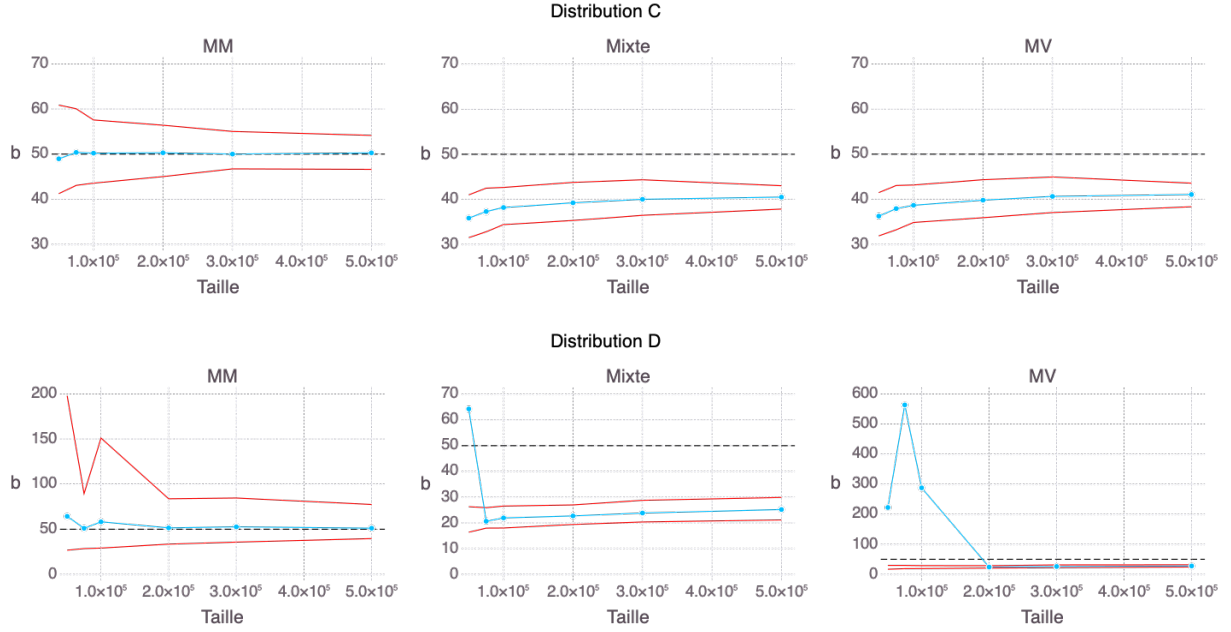


FIGURE 4.5 Ajustement des échantillons de plus grandes tailles tirés des distribution C et D (Pearson de type I à quatre paramètres). De gauche à droite : méthode des moments, estimation mixte, maximisation de la vraisemblance. Les points bleus correspondent à la moyenne des estimations de b pour les 100 échantillons de même taille, tandis que les droites rouges correspondent aux quantiles 2.5% et 97.5%.

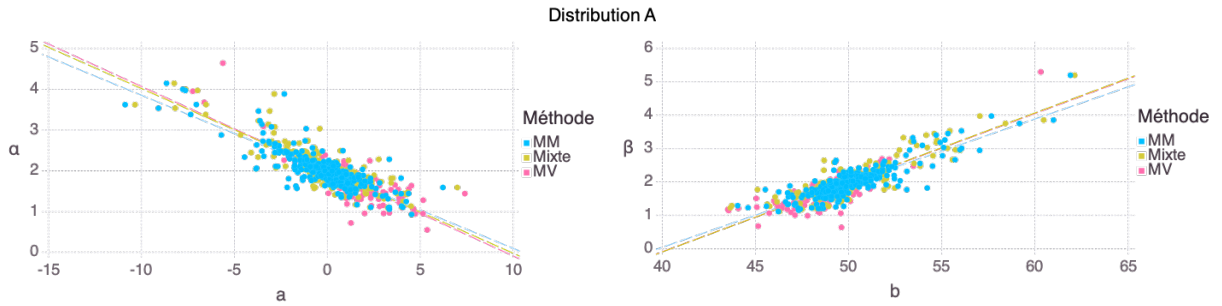


FIGURE 4.6 Corrélation entre les estimations des paramètres (a, α) et (b, β) pour la distribution A. Les différentes couleurs indiquent la méthode utilisée pour l'estimation, soit la méthode des moments, celle mixte ou la maximisation de la vraisemblance.

Encore une fois, les distributions pour lesquelles les valeurs des paramètres de forme α et β sont éloignées ne sont pas concluants, particulièrement dans le cas de la maximisation de la vraisemblance.

Comme dans le cas de la distribution à quatre paramètres, on essaie d'augmenter la taille des échantillons afin d'estimer à partir de quelle grandeur la méthode pourrait être considérée.

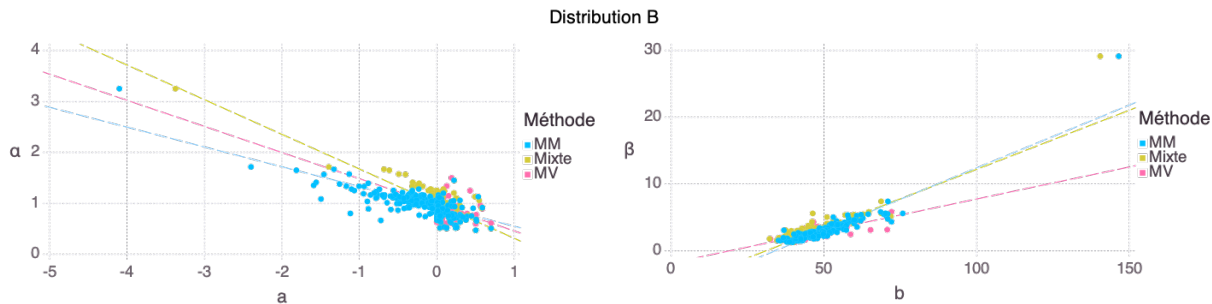


FIGURE 4.7 Corrélation entre les estimations des paramètres (a, α) et (b, β) pour la distribution B. Les différentes couleurs indiquent la méthode utilisée pour l'estimation, soit la méthode des moments, celle mixte ou la maximisation de la vraisemblance.

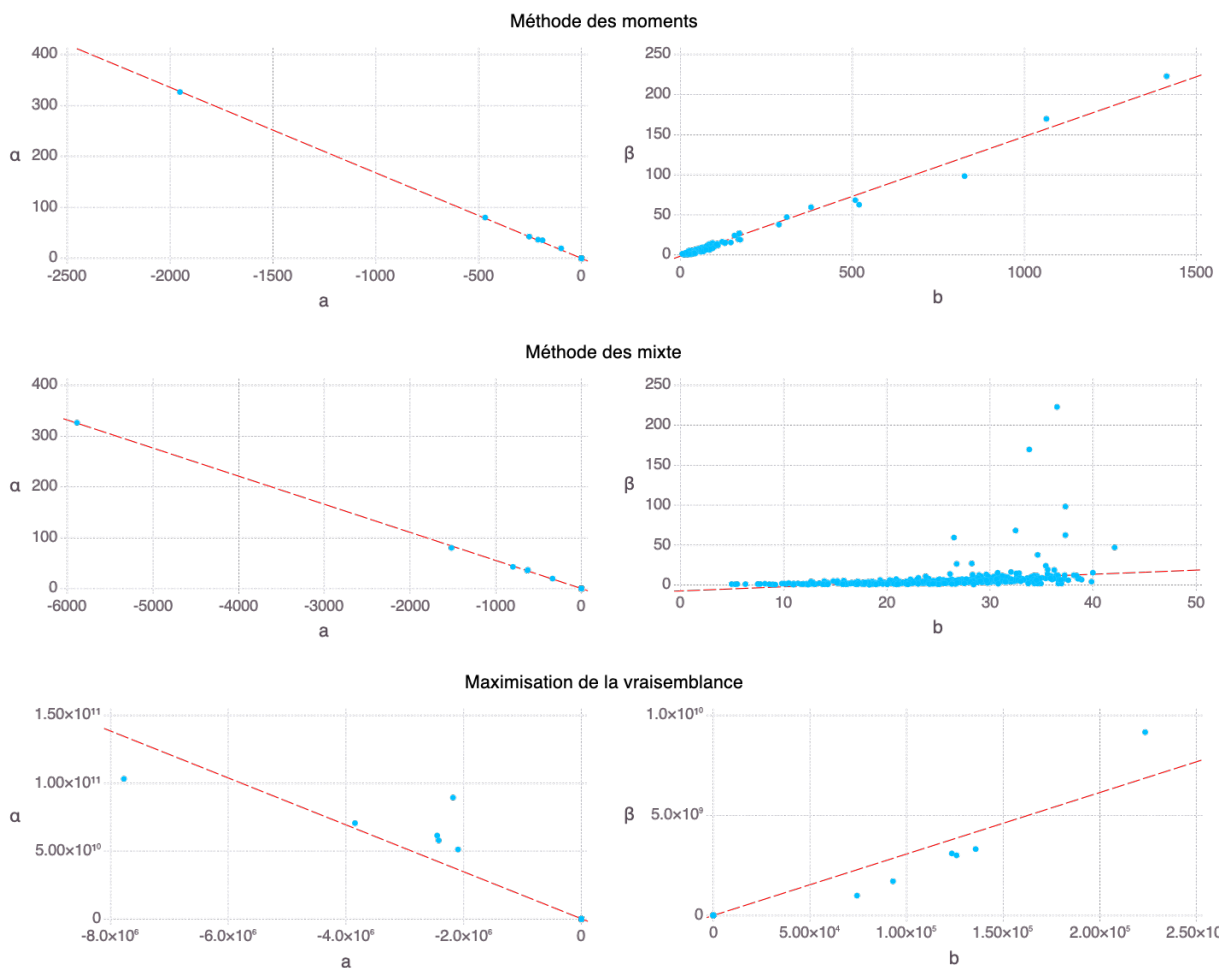


FIGURE 4.8 Corrélation entre les estimations de (a, α) et (b, β) pour la distribution C. Les graphiques associés au maximum de la vraisemblance ont été coupés à des fins de visibilité.

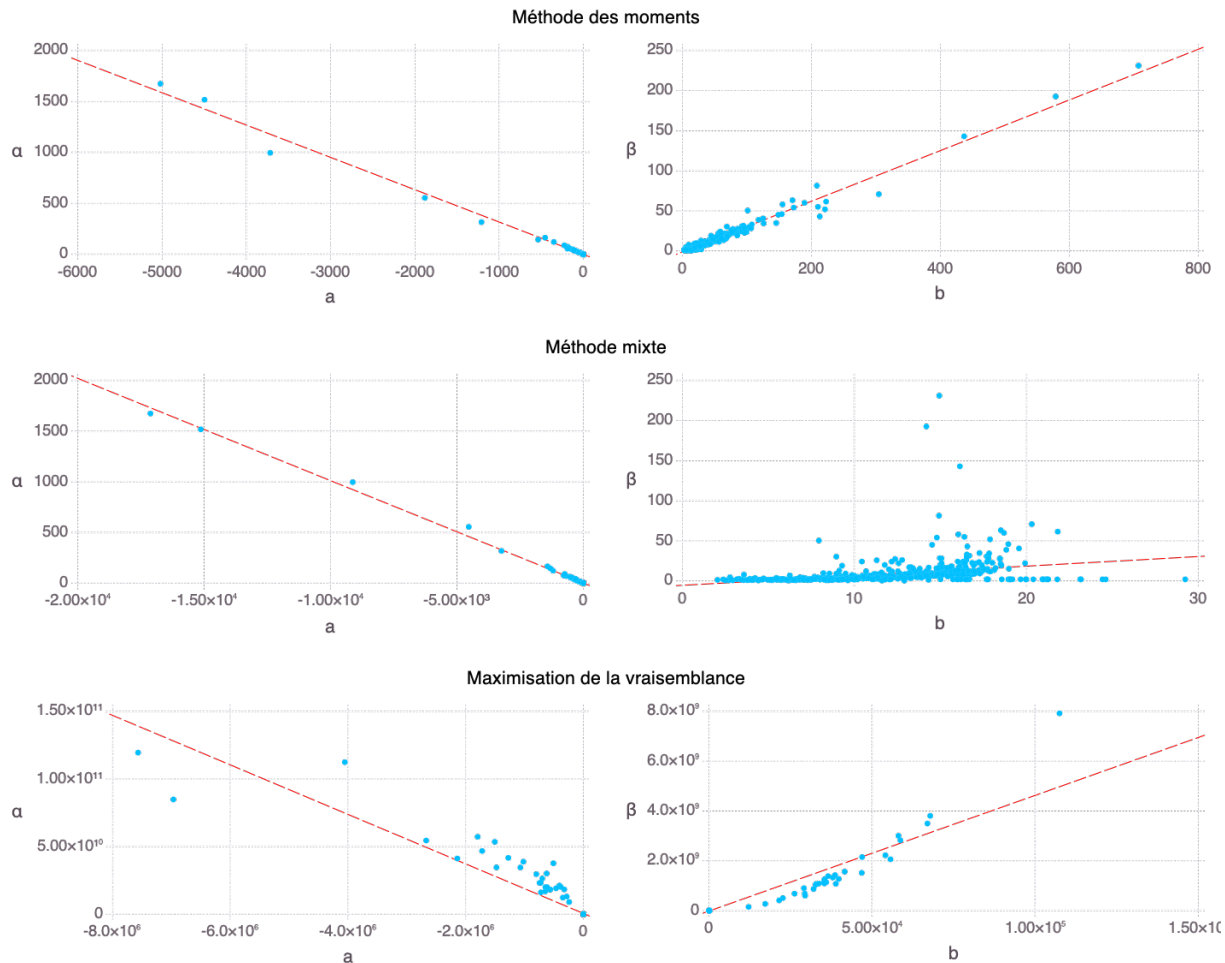


FIGURE 4.9 Corrélation entre les estimations de (a, α) et (b, β) pour la distribution D. Les graphiques associés au maximum de la vraisemblance ont été coupés à des fins de visibilité.

Les résultats se trouvent dans les graphiques de la figure 4.11. Encore une fois, la méthode des moments converge plus rapidement que la maximisation de la vraisemblance vers des estimations raisonnables du paramètre de borne supérieure. La méthode mixte est pour sa part plus performante que son analogue à quatre paramètres et donne des résultats similaires à ceux de la méthode des moments.

Pour la distribution C, on teste des échantillons de taille $\{9\ 000, 15\ 000, 20\ 000, 25\ 000, 45\ 000, 50\ 000\}$ dans les cas d'estimation par méthodes des moments et mixte, puis des échantillons de taille $\{45\ 000, 46\ 000, 47\ 500, 50\ 000, 55\ 000, 60\ 000\}$ pour le maximum de vraisemblance. Les tailles utilisées ainsi que l'écart entre celles-ci diffèrent dû à l'instabilité de la troisième méthode. En effet, on peut voir dans le graphique associé, en haut à gauche de la figure 4.11, que bien que les quantiles 2.5% et 97.5% estimés par maximisation de la

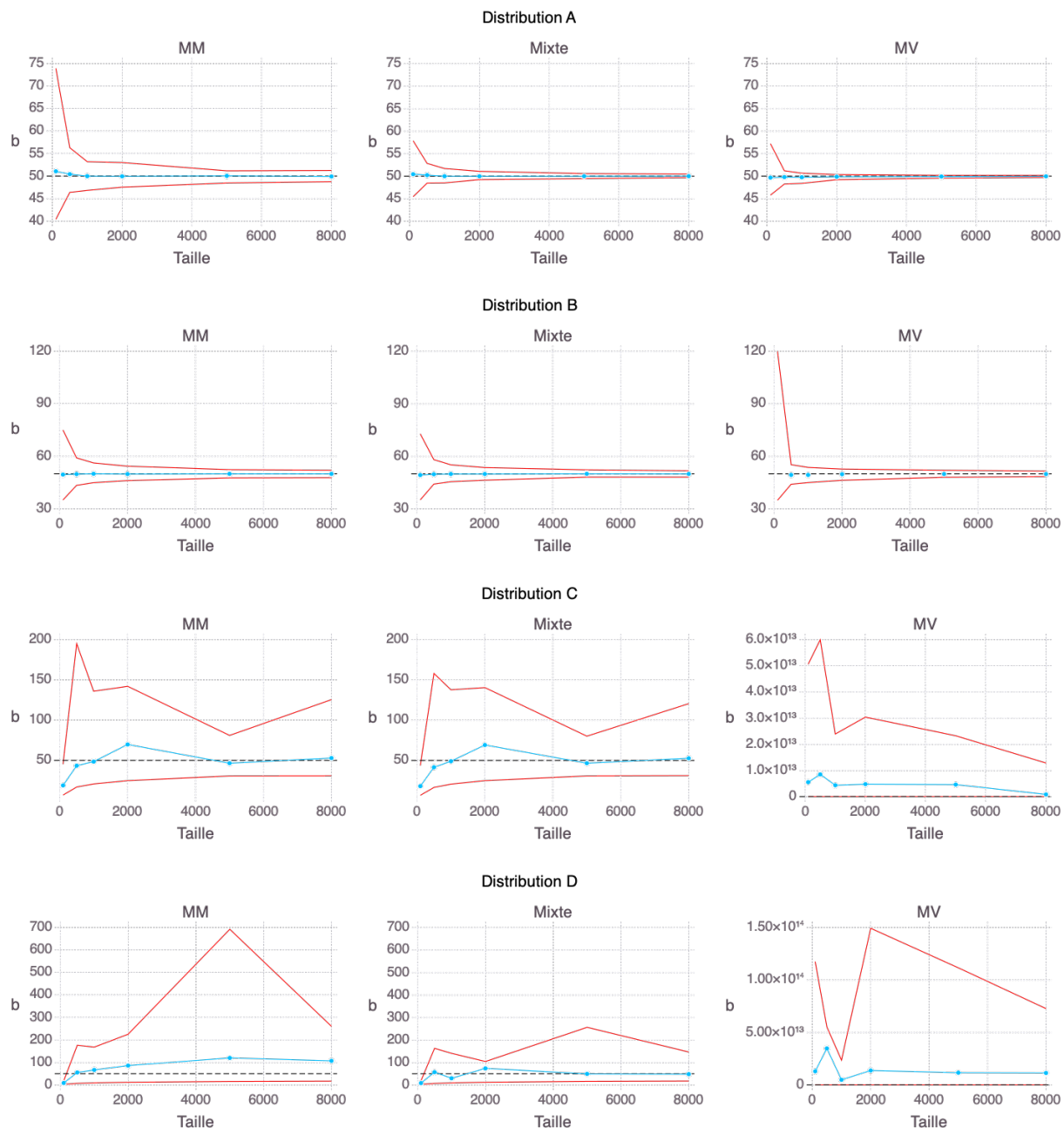


FIGURE 4.10 Estimation de b en fonction de la taille de l'échantillon, pour chaque distribution du modèle bêta à trois paramètres. De gauche à droite : méthode des moments, estimation mixte, maximisation de la vraisemblance. Les points bleus correspondent à la moyenne des estimations de b pour les 100 échantillons de même taille, tandis que les droites rouges correspondent aux quantiles 2.5% et 97.5%.

vraisemblance forment un intervalle plus serré que leurs analogues estimés par les méthodes des moments et mixte pour des tailles similaires, la moyenne des estimations de b ne converge

vers 50 qu'à partir de tailles $\geq 45\,000$. Pour ce qui est de la distribution D, les tailles $\{60\,000, 70\,000, 100\,000, 150\,000, 200\,000, 400\,000\}$ sont utilisées pour les trois méthodes. Dans le cas du maximum de vraisemblance, ni les quantiles 97.5% ni les moyennes ne s'approchent de 50, et ce peu importe la taille de l'échantillon. Le quantile 2.5% varie pour sa part entre 33.9 et 41.4.

On peut conclure que la réduction du nombre de paramètres n'est pas suffisante pour permettre une estimation stable de ceux-ci à partir de données synthétiques, bien que les résultats soient améliorés par rapport à la distribution à quatre paramètres.

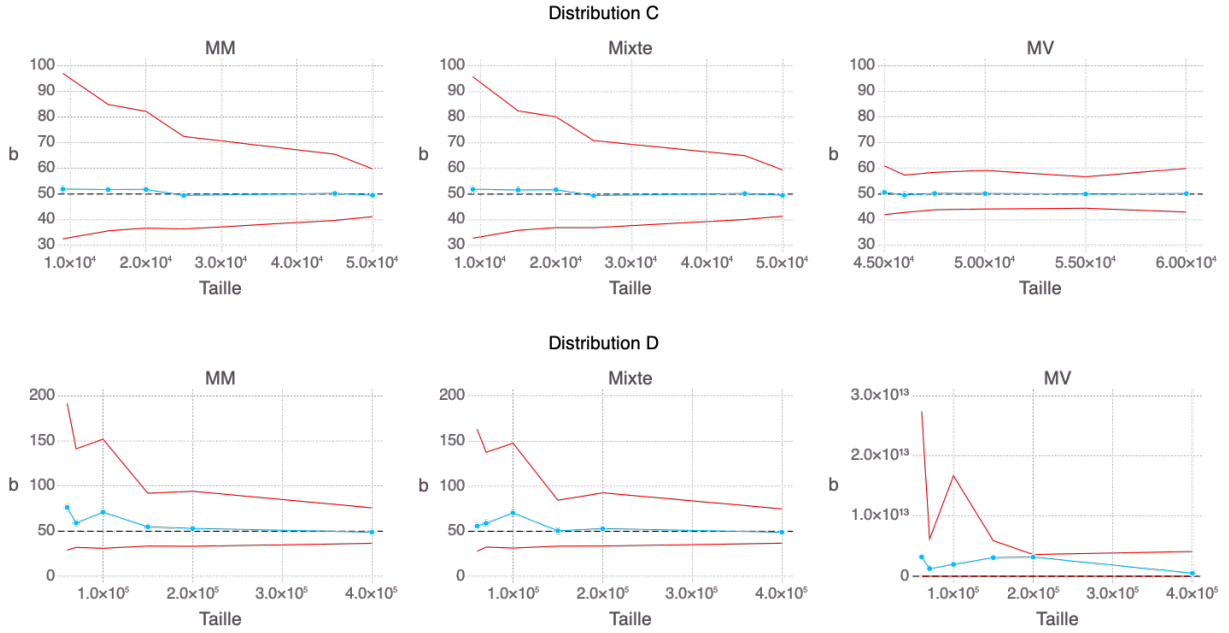


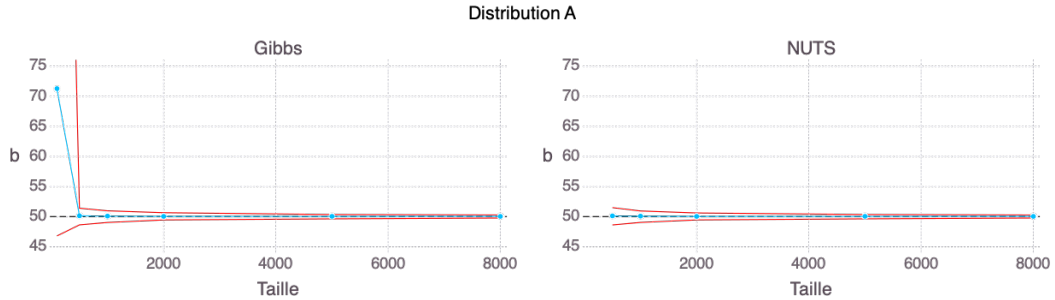
FIGURE 4.11 Ajustement des échantillons de plus grandes tailles tirés des distribution C et D (Pearson de type I à trois paramètres). De gauche à droite : méthode des moments, estimation mixte, maximisation de la vraisemblance. Les points bleus correspondent à la moyenne des estimations de b pour les 100 échantillons de même taille, tandis que les droites rouges correspondent aux quantiles 2.5% et 97.5%.

4.2.1 Estimation bayésienne des paramètres

Comme mentionné en début de chapitre, l'approche bayésienne est moins sensible au manque d'identifiabilité des paramètres que celle fréquentiste. Ce problème, omniprésent dans le cas de la Pearson de type I, est illustré par la figure 4.4. Ainsi, dans le but de réduire davantage les possibilités de paramètres estimés, les algorithmes de Gibbs et NUTS, définis aux sections 3.2.1, 3.2.2 et 3.2.3, sont considérés. Une loi *a priori* impropre et non informative est utilisée. Pour l'échantillonneur de Gibbs, on propose 20 000 candidats desquels les 5 000 premiers sont

utilisés comme période de chauffe, durant laquelle la taille du pas est ajustée afin d'obtenir un taux d'acceptation optimal. De plus, on amincit la chaîne générée en ne gardant qu'une itération par tranche de 10. Du côté de l'algorithme NUTS, on propose 10 000 candidats et seuls les 2 500 premiers sont utilisés comme période de chauffe. La chaîne n'est pas amincie puisque la corrélation entre les candidats consécutifs est moindre. Ainsi, les chaînes finales contiennent 1 500 et 7 500 éléments respectivement.

Encore une fois, on reproduit la même démarche que précédemment, avec les mêmes tailles d'échantillon. Ici, puisque la réponse d'un algorithme MCMC est une chaîne des candidats acceptés pour chacun des paramètres, on approxime ceux-ci par la moyenne de leur chaîne respective. Les résultats des moyennes associées au paramètre de borne supérieure des distributions A et B se trouvent à la figure 4.12. On remarque d'abord que seul l'échantillonneur de Gibbs de la distribution A converge pour tous les échantillons de taille 100. En effet, pour les trois autres cas, les échantillonneurs n'obtiennent des résultats que pour une fraction des échantillons de cette taille. Les résultats ne sont donc pas affichés dans les graphiques. De plus, pour les échantillons de taille 500, on peut voir que la moyenne et le quantile 97.5% estimés par l'algorithme NUTS pour la distribution B ne sont pas non plus tracés sur le graphique associé. Bien qu'il découle des résultats pour les 100 échantillons de cette taille, ceux-ci sont disproportionnés par rapport à la valeur réelle de $b = 50$.



Néanmoins, on note que bien que les algorithmes n'arrivent pas à bien estimer b pour des échantillons de petite taille, ceux-ci sont relativement performants lorsque celle-ci augmente. Les résultats semblent s'apparenter à ceux des méthodes des moments et mixte.

Dans le cas de la distribution C, on peut voir à la figure 4.13 que seul l'échantillonneur de Gibbs converge, et ce uniquement pour les échantillons de taille ≥ 5000 . Aucune des deux approches bayésiennes ne converge pour les 100 échantillons des six tailles testées tirés de la distribution D.

En général, l'algorithme de Gibbs est plus performant que le NUTS. Un problème demeure : si la variance de la distribution instrumentale est trop petite, l'espace des paramètres n'est pas

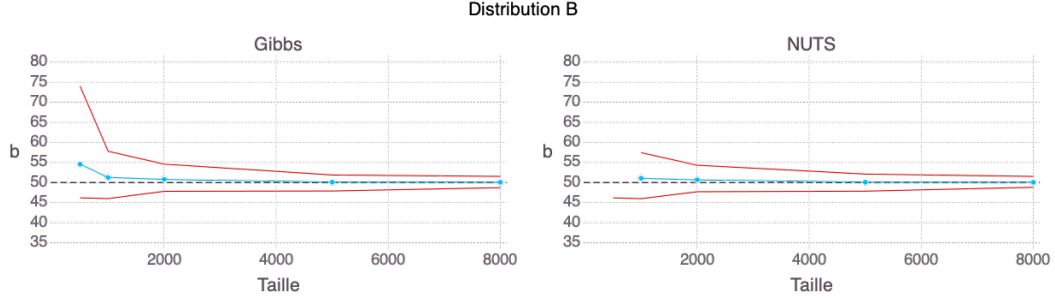


FIGURE 4.12 Estimation bayésienne de b en fonction de la taille de l'échantillon, pour les distribution A et B. Les points bleus correspondent à la moyenne des moyennes des chaînes de Markov associées à b pour les 100 échantillons de même taille, tandis que les droites rouges correspondent aux quantiles 2.5% et 97.5% de ces moyennes. Les graphiques associés à l'échantillonneur de Gibbs de la distribution A et à l'algorithme NUTS de la distribution B ont été coupés à des fins de visibilité. Pour le premier graphique, le quantile 97.5% associé aux échantillons de taille 100 est de 212.36 tandis que pour le second, le quantile supérieur associé aux échantillons de même taille est de 1.89×10^{11} et la moyenne de 4.58×10^{10} .

bien exploré. Si la variance est trop grande (généralement après l'ajustement de la période de chauffe), alors les candidats s'écartent du domaine de définition des paramètres, et ce malgré les accommodements pris à cet égard. Cette complication survient en particulier lorsque la taille de l'échantillon est trop petite. L'utilisation de lois *a priori* sur b ou β pourrait potentiellement réduire l'ampleur du problème.

4.3 Pearson de type I à trois paramètres reparamétrisée

La section 4.1.1 met en lumière la corrélation existante entre les couples (a, α) et (b, β) , peu importe la forme de la distribution et la méthode d'estimation. Réexprimer la distribution de Pearson de type I en fonction de différents paramètres de forme pourrait résoudre le problème de corrélation entre les paramètres. On utilise la paramétrisation donnée par l'équation 3.3 au début du chapitre 3 en posant $a = 0$. Différentes reparamétrisations de la loi existent, mais exigent que la distribution soit concave, ce qui n'est pas le cas lorsque l'un ou les deux paramètres de forme sont ≤ 1 .

Afin de vérifier la corrélation entre le paramètre de borne b et les deux nouveaux paramètres de forme μ et ν , on trace de nouveau les nuages de point des estimations de ceux-ci, de façon semblable aux figures 4.6 à 4.9. On peut facilement voir sur les figures 4.14 et 4.15 que la corrélation est toujours présente, bien que plus nécessairement linéaire. Effectivement, les transformations $\mu = \alpha/(\alpha + \beta)$ et $\nu = \alpha + \beta$ ne sont pas suffisante pour réduire la corrélation. Ainsi, les résultats de l'estimation moyenne de b , affichés aux figures 4.16 et 4.17, ne sont pas

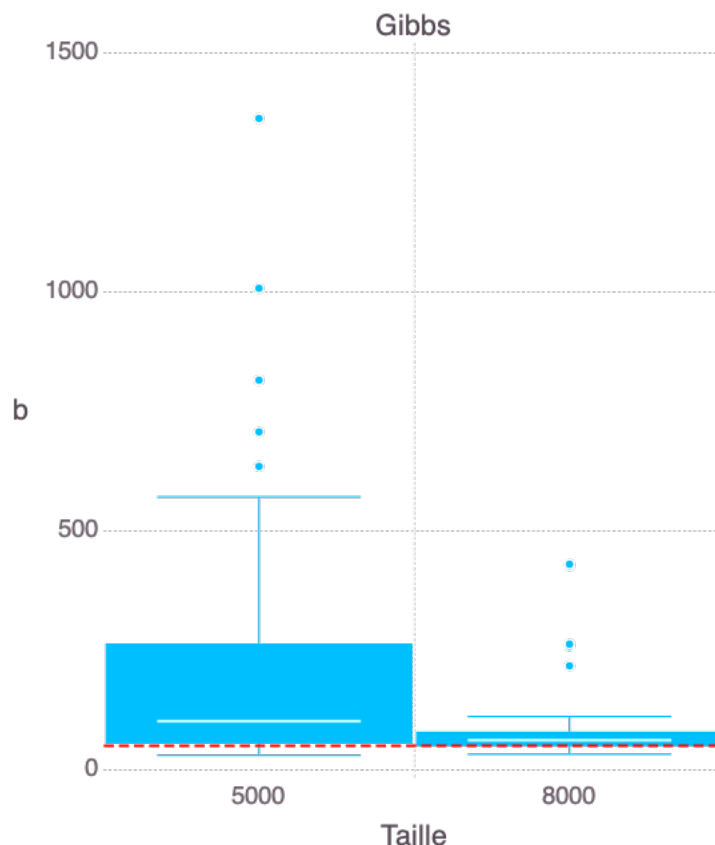


FIGURE 4.13 Boîtes à moustaches des moyennes des chaînes de b proposées par l'échantillonneur de Gibbs à partir d'échantillons de taille 5000 et 8000. La ligne rouge pointillée correspond à la valeur réelle de b , soit 50.

améliorés par rapport à ceux associés à la Pearson de type I de paramétrisation standard, aux figures 4.3 et 4.10. On peut observer que les quantiles extrêmes forment des intervalles similaires à ceux de la figure 4.10, qui trace les estimations de b pour le modèle à trois paramètres de paramétrisation standard.

4.4 Méthodologie choisie

4.4.1 Données

Deux stations sont sélectionnées afin d'en estimer la PMP : l'aéroport international Pierre-Elliott-Trudeau de Montréal et St-Hubert. Ces deux stations sont choisies pour leur proximité ainsi que leur grand nombre de données. Celles-ci sont disponibles sur le site internet des données historiques d'Environnement et Changements Climatiques Canada (ECCC).

Les stations ont les ID climatologiques suivants :

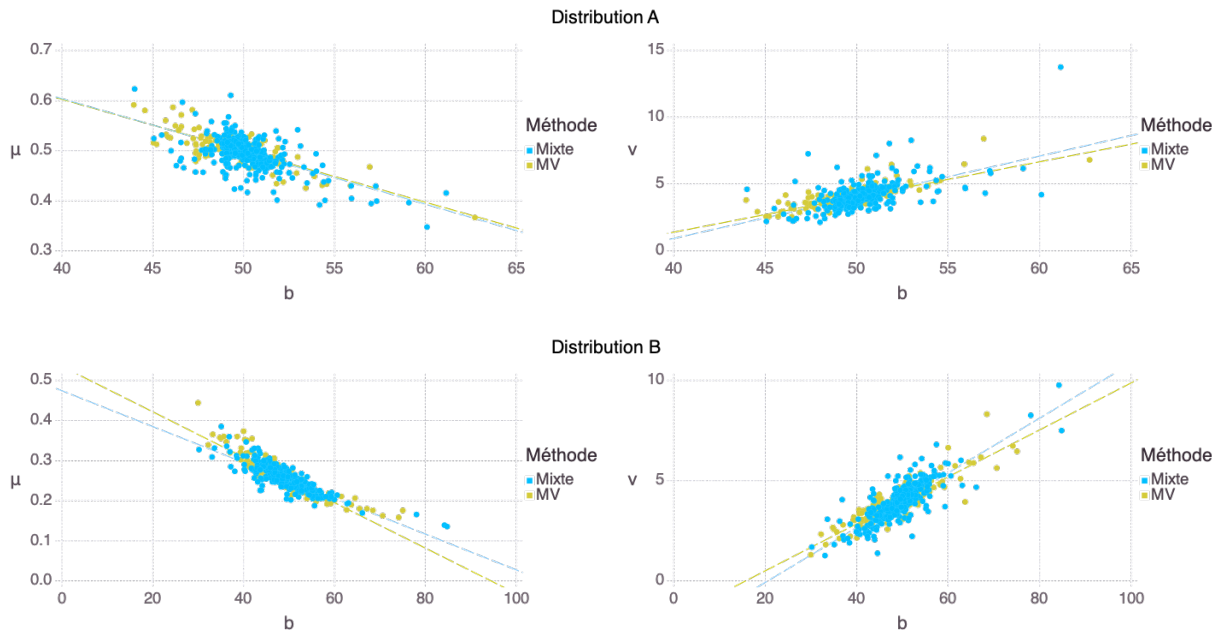


FIGURE 4.14 Corrélation entre les estimations des paramètres (b, μ) et (b, ν) pour les distributions A et B. Les différentes couleurs indiquent la méthode utilisée pour l'estimation, soit la méthode mixte ou la maximisation de la vraisemblance.

- Montréal : 7025250 pour la période 1953-2012 et 7025251 pour 2013-2024
- St-Hubert : 7027320 pour 1949-2004, 7027322 pour 2005-2009 et finalement 7027329 pour 2010-2024.

Les données de précipitations journalières et de points de rosée horaires sont extraites de ces stations. Il est à noter qu'à Montréal comme à St-Hubert, les points de rosée ne sont disponibles qu'à partir de 1953.

Seules les précipitations observées entre le 1er mai et le 31 octobre inclusivement sont utilisées puisque pour la PMP, l'intérêt est porté sur les précipitations liquides. L'ensemble de données de précipitations journalières de Montréal (figure 4.1) contient 5 321 données, avec un maximum de précipitation observé à 81.9mm et celle de St-Hubert comporte 5 303 données dont la valeur maximale est 106.5mm (figure 4.18).

4.4.2 Méthode d'estimation de la PMP bêta

À la lumière des résultats des précédentes sections, on peut conclure que la méthode développée requiert un nombre de données conséquent afin d'améliorer l'identifiabilité des paramètres. Le modèle à trois paramètres est préféré à celui de quatre paramètres, et les méthodes d'estimation des moments et mixte devraient être privilégiées au maximum de vraisemblance.

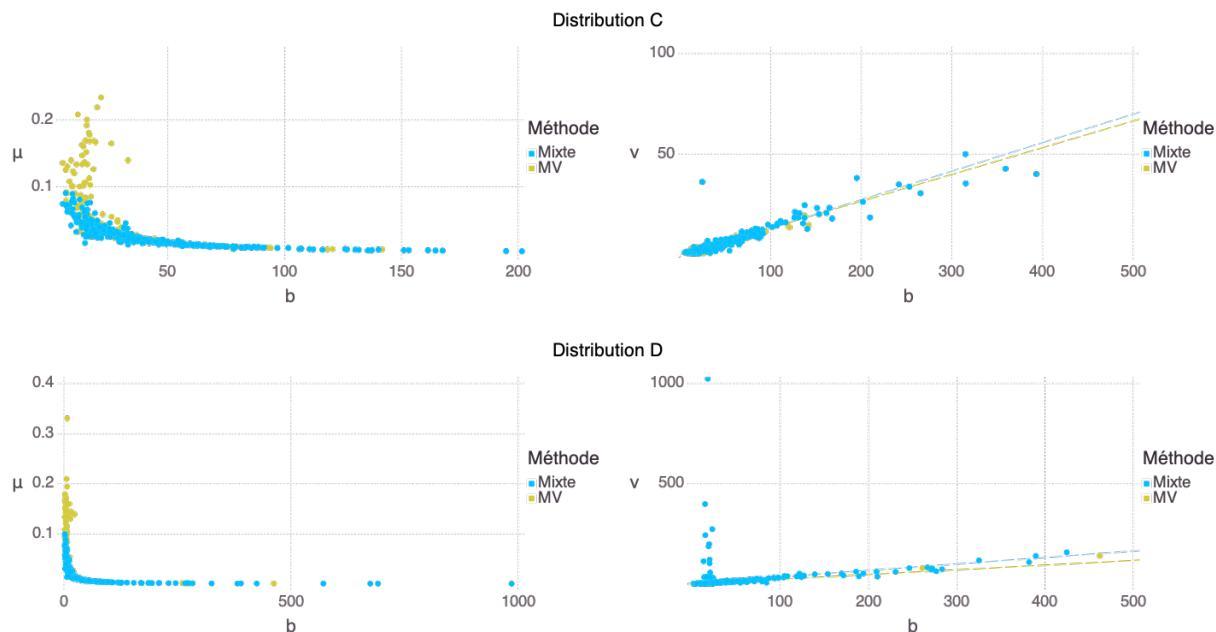


FIGURE 4.15 Corrélation entre les estimations des paramètres (b, μ) et (b, ν) pour les distributions C et D. Les différentes couleurs indiquent la méthode utilisée pour l'estimation, soit la méthode mixte ou la maximisation de la vraisemblance. Les graphiques ont été coupés à des fins de visibilité.

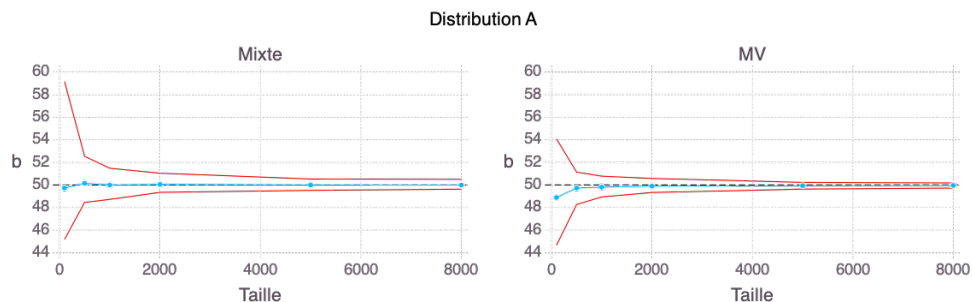


FIGURE 4.16 Estimation de b en fonction de la taille de l'échantillon pour la distribution reparamétrisée A, à partir de l'estimation mixte (à gauche) et du maximum de vraisemblance (à droite). Les points bleus correspondent à la moyenne des estimations de b pour les 100 échantillons de même taille, tandis que les droites rouges correspondent aux quantiles 2.5% et 97.5%.

Dans le cas où la taille de l'échantillon est suffisante, l'échantillonneur de Gibbs permet également de bien caractériser la borne supérieure de la distribution.

Ainsi, les méthodes des moments, mixtes et l'échantillonneur de Gibbs sont utilisés ici afin d'estimer la PMP. Deux modèles de Pearson de type I sont étudiés, soit celui dont la borne

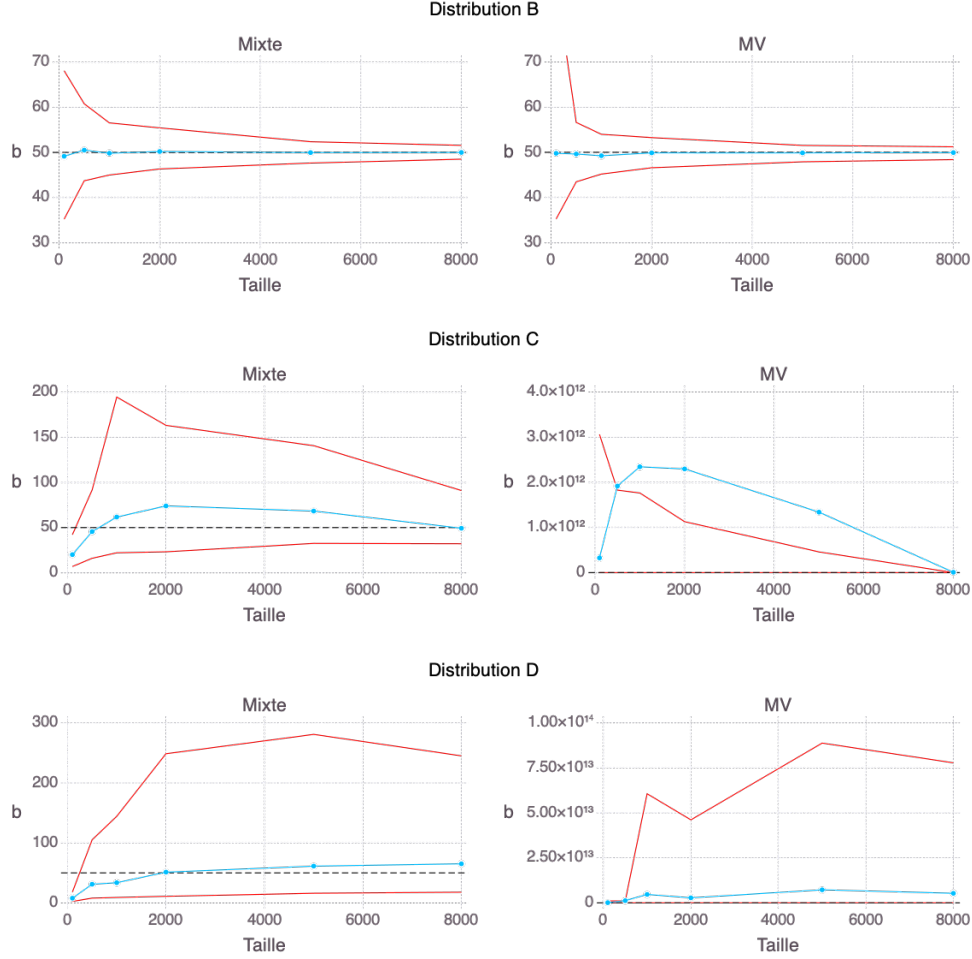


FIGURE 4.17 Estimation de b en fonction de la taille de l'échantillon pour les distributions reparamétrisée B, C et D, à partir de l'estimation mixte (à gauche) et du maximum de la vraisemblance (à droite). Les points bleus correspondent à la moyenne des estimations de b pour les 100 échantillons de même taille, tandis que les droites rouges correspondent aux quantiles 2.5% et 97.5%.

inférieure a est fixée à 0, de densité donnée par l'équation 4.1, ainsi qu'un second dont la borne inférieure est fixée à 0.2mm de densité

$$f_{(Y|b,\alpha,\beta)}(y) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{(y - 0.2)^{\alpha-1}(b - y)^{\beta-1}}{(b - 0.2)^{\alpha+\beta-1}}, \quad 0.2 < y < b. \quad (4.2)$$

Cette valeur de 0.2mm correspond au minorant observable des précipitations journalières par les stations étudiées.

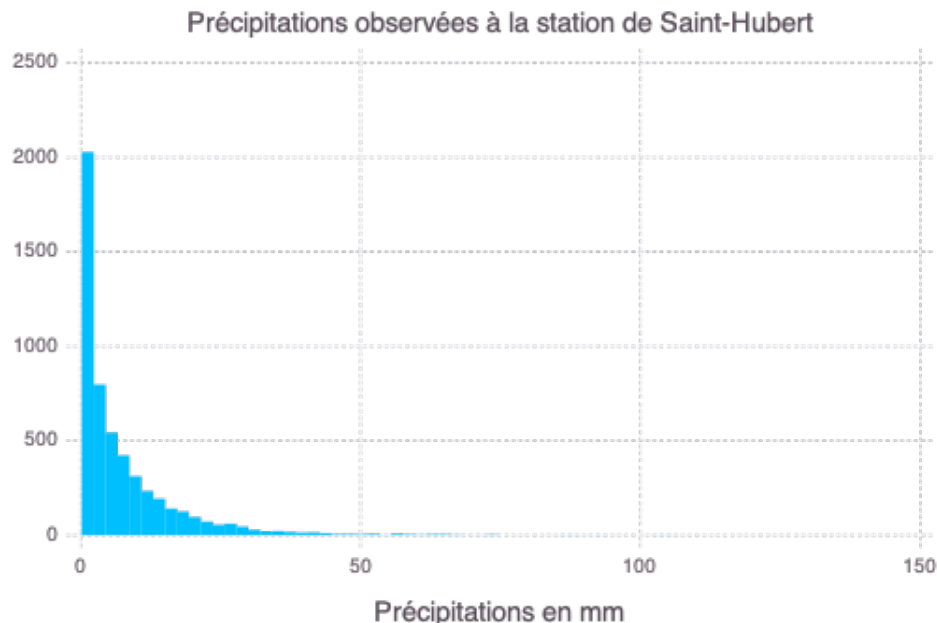


FIGURE 4.18 Précipitations journalières observées à la station de Saint-Hubert du 4 juillet 1949 au 29 mai 2024

4.4.3 Méthodes usuelles d'estimation de la PMP

Afin de comparer les résultats de l'approche développée dans ce travail, il semble nécessaire d'estimer également la PMP à l'aide de méthodes usuelles telles que celles hydrométéorologiques et l'approche empirique d'Hershfield. Pour celle-ci, on utilise un facteur de fréquence $K = 15$, qui est celui suggéré par Hershfield (1961a). L'ajustement en fonction du nombre de données suggéré par l'OMM (2009) n'est pas pris en compte.

Dans le cas des méthodes hydrométéorologiques, seules les données de la station d'intérêt sont utilisées pour l'estimation de la PMP, c'est-à-dire qu'aucune méthode de transposition des tempêtes n'est exploitée. Une sélection simple des tempêtes est cependant appliquée : les quantiles 90% des précipitations journalières de chaque année sont choisies. Puisque les données d'humidité spécifique ne sont pas disponibles, l'eau précipitable est approximée à partir du point de rosée persistant maximum sur douze heures tel que décrit par l'OMM (2009). On note que pour la station de St-Hubert, puisque le début de l'observation du point de rosée horaire est ultérieur à celui de l'observation des précipitations journalières, les tempêtes des années 1949 à 1952 inclusivement ne peuvent être utilisées pour l'estimation de la PMP. Deux PMP sont estimées à partir de la maximisation de l'humidité : la première au moyen de l' EP_{max} déterminée comme le maximum des EP calculées (Ben Alaya *et al.*,

2018) et la seconde à l'aide de l' EP_{100} définie comme la valeur de retour 100 ans de l'eau précipitable (Rousseau *et al.*, 2014).

Une librairie julia et sa documentation permettant d'estimer rapidement les PMP à partir de l'approche bêta, la méthode empirique d'Hershfield ainsi que la maximisation de l'humidité sont disponibles sur le GitHub de JuliaExtremes.

CHAPITRE 5 RÉSULTATS

5.1 Station de l'aéroport international Montréal-Trudeau

On commence par ajuster le modèle à trois paramètres ($a = 0$) aux données de pluie de la station de Montréal (5 321 données). On peut étudier les diagrammes quantile-quantile et probabilité-probabilité des modèles estimés à la figure 5.1. On s'intéresse particulièrement aux diagrammes quantile-quantile puisqu'on recherche un modèle qui s'ajuste bien aux quantiles les plus élevés, la PMP étant la borne supérieure de la distribution.

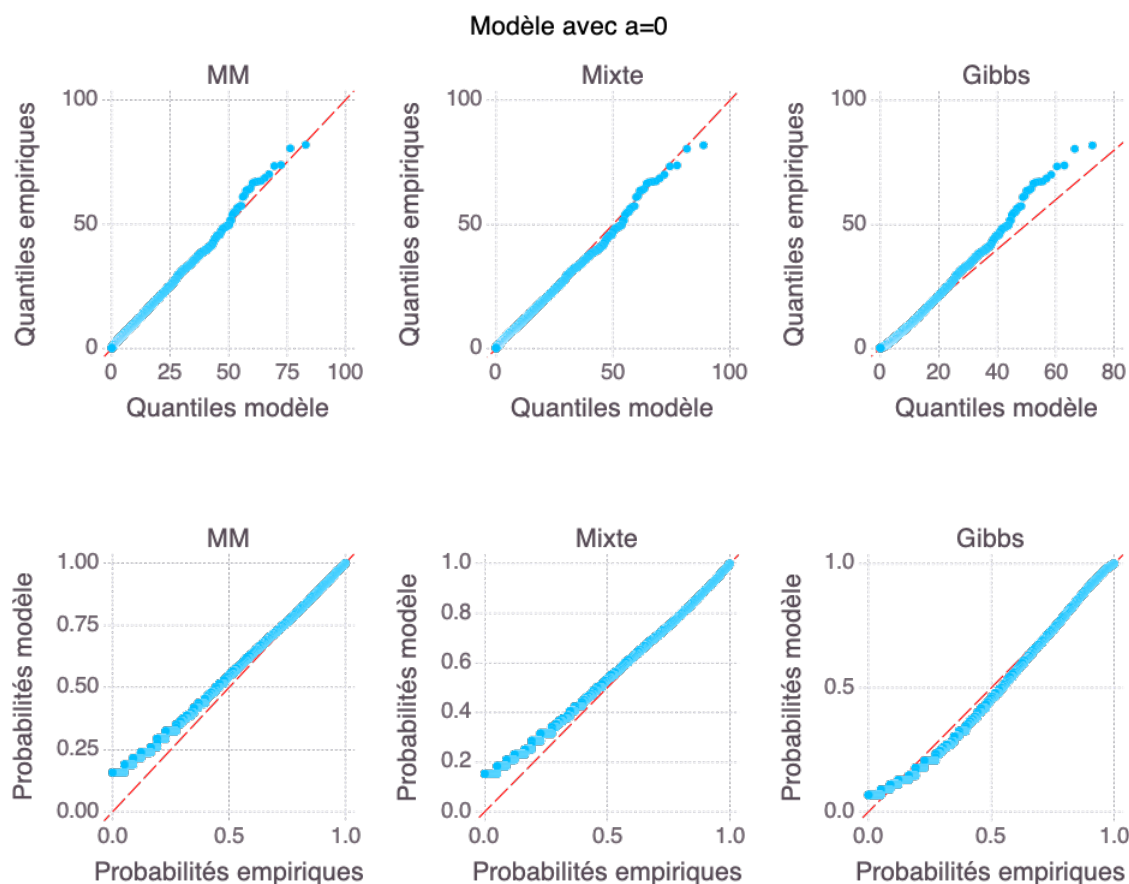


FIGURE 5.1 Diagrammes quantile-quantile et probabilité-probabilité pour le modèle bêta avec a fixé à 0mm et ajustée sur les données de Montréal à l'aide de trois méthodes différentes (moments, mixte et Gibbs)

Pour les trois modèles, les quantiles empiriques et du modèle s'accordent bien. On remarque une légère sous-estimation du modèle ajusté par Gibbs pour les quantiles les plus élevés. Cependant, on peut voir dans les graphiques du bas de la figure 5.1 que ce même modèle est mieux centré pour les petites probabilités.

Pour le modèle ajusté à l'aide de l'échantillonneur de Gibbs, il est possible d'afficher le graphique de la trace et l'histogramme des états de la chaîne. Ces graphiques sont présentés à la figure 5.2. On remarque d'abord que la trace associée au paramètre de forme α est régulière, constante et explore bien l'espace. L'histogramme du même paramètre n'a qu'un maximum et est centré en sa moyenne, indiquée par la droite pointillée.

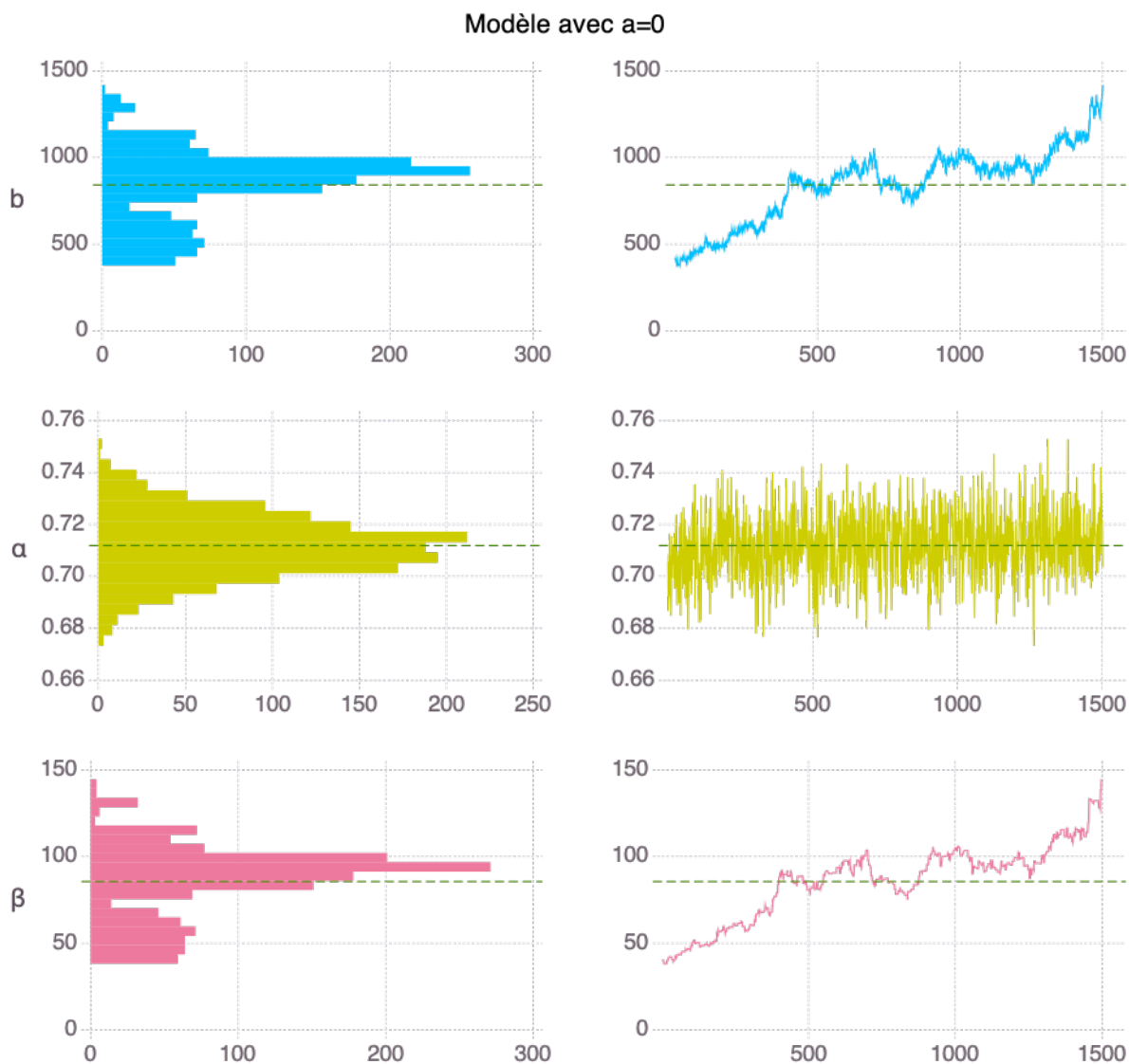


FIGURE 5.2 Traces et histogrammes de la chaîne de Markov générée par Gibbs sur les données de Montréal pour le modèle bêta avec $a = 0$. Les graphiques du haut sont associés à b , ceux du centre à α et ceux du bas à β . Les droites pointillées correspondent aux moyennes des états de chacun des paramètres.

Les graphiques associés aux paramètres b et β sont pour leur part plus problématiques. On peut noter que les traces et histogrammes sont presque identiques : la corrélation entre ces deux paramètres est toujours grande. Les histogrammes présentent plusieurs maxima locaux, et les traces divergent dû au manque d'identifiabilité des paramètres. Par exemple, si l'algorithme propose un état qui devrait être considéré comme trop grand pour b , le paramètre β compensera en acceptant un candidat trop élevé également. Ce phénomène est expliqué par relation linéaire positive entre les deux paramètres, relation illustrée par les graphiques de droite des figures 4.6 à 4.9. Cette divergence de la chaîne implique que l'algorithme pourrait retourner n'importe quelle valeur de moyenne pour b et β , selon le nombre d'itérations total et la taille de la période de chauffe. On peut également remarquer la présence de plateaux sur la trace du paramètre β , ce qui indique que le paramètre de dispersion de la distribution instrumentale de β pourrait être trop grand. Dans ce cas, à chaque itération, les candidats proposés par cette distribution sont potentiellement trop éloignés de l'état actuel de β , généralement dans une région de faible densité et sont donc rejetés. Ainsi, l'algorithme n'accepte des candidats éloignés de β que lorsqu'un état éloigné de b est accepté précédemment.

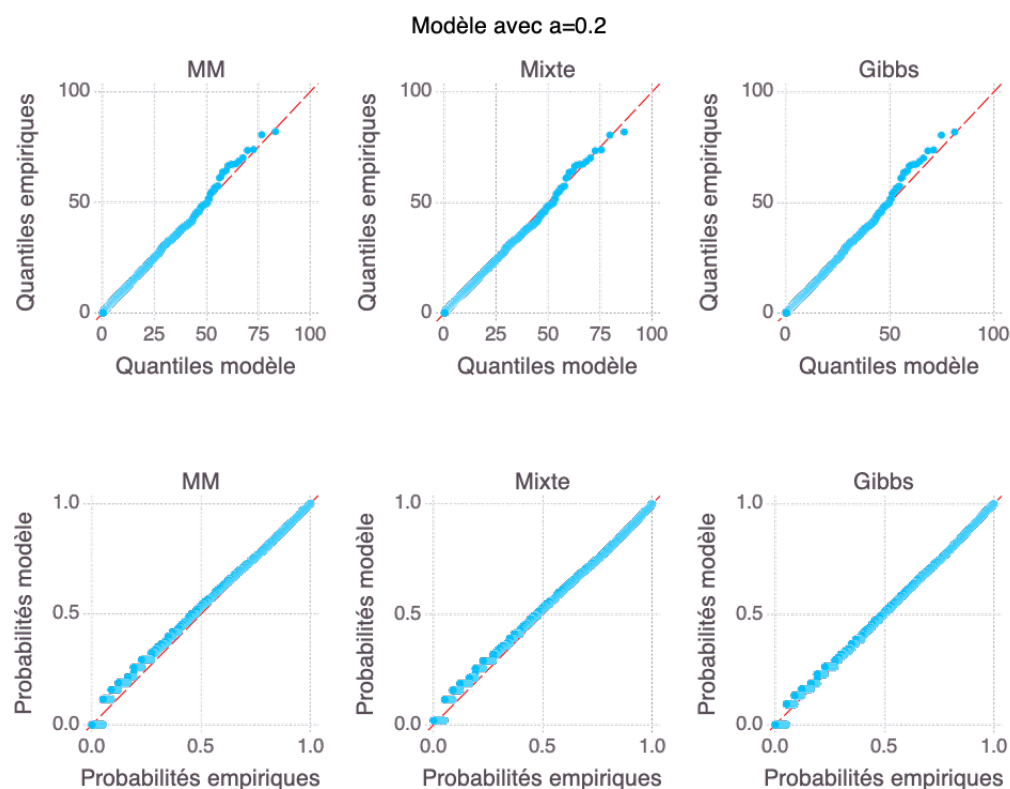


FIGURE 5.3 Diagrammes quantile-quantile et probabilité-probabilité pour le modèle bêta à quatre paramètres avec $a = 0.2$ fixé, ajusté sur les données de Montréal à l'aide des méthodes choisies (moments, mixte et Gibbs)

On ajuste par la suite le modèle pour lequel a est fixée à 0.2mm. Les diagrammes quantile-quantile et probabilité-probabilité associés sont affichés à la figure 5.3. Peu importe la méthode d'estimation des paramètres utilisée, les modèles sont très bien ajustés aux données observées. On peut encore une fois vérifier les traces et histogrammes des chaînes de Markov générées par l'échantillonneur de Gibbs, à la figure 5.4

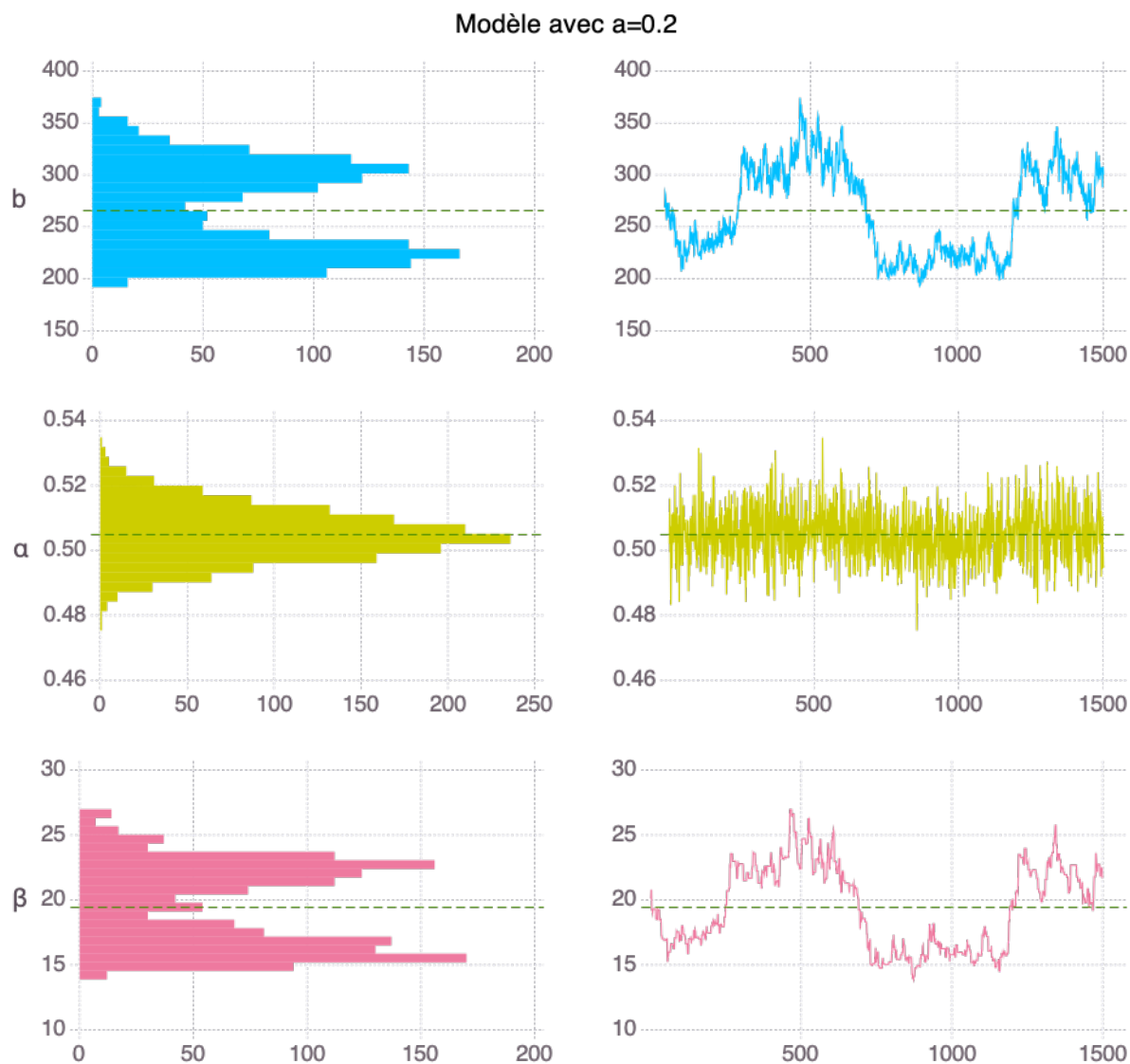


FIGURE 5.4 Traces et histogrammes de la chaîne de Markov générée par l'échantillonneur de Gibbs sur les données de Montréal pour le modèle bêta avec a fixé à 0.2mm. Les graphiques du haut sont associés à b , ceux du centre à α et ceux du bas à β . Les droites pointillées correspondent aux moyennes des états de chacun des paramètres.

Comme pour le modèle avec a fixé à 0, la trace et l'histogramme du paramètre α ne posent aucun problème, on obtient une distribution claire du paramètre. Les graphiques associés à b et β sont toujours suspects, cependant peut-être moins problématiques que ceux de la figure 5.2. En effet, on peut voir que les traces ne divergent plus et que la moyenne des états ne dépend ainsi plus du nombre d'itérations de l'algorithme. Cependant, la dépendance entre les deux paramètres est toujours très grande, et on observe l'apparition de deux points d'équilibre autour de $b = 225$ ($\beta = 15$) et $b = 300$ ($\beta = 22$), ce qui se traduit par la présence de deux maxima clairs sur les histogrammes. De plus, la trace de β affiche encore des plateaux qui indiquent que trop peu de candidats sont acceptés.

Le tableau 5.1 résume toutes les distributions ajustées aux données de la station de Montréal. À l'exception de celui ajusté par l'échantillonneur de Gibbs lorsque $a = 0$ (troisième ligne), les estimations des paramètres sont très stables d'une méthode à l'autre. On constate que pour cette distribution, la borne supérieure dépasse de beaucoup celles des autres distributions. Cependant, le diagramme quantile-quantile associé à ce même modèle, à la figure 5.1, indique une sous-estimation des quantiles élevés. Ceci peut être expliqué par le fait que ce modèle a une queue légère. Ainsi, il sous-estime en effet les valeurs maximales prélevées à la station, mais surestime probablement la borne supérieure.

TABLEAU 5.1 Paramètres estimés pour différents modèles et méthodes d'estimations, à partir des données observées de précipitation journalières entre 1953 et 2024 à la station de Montréal

Modèle	Méthode	a	b	α	β
$a = 0$	Moments	0	269.98	0.46	18.81
	Mixte	0	288.70	0.46	18.81
	Gibbs	0	841.72	0.71	85.72
$a = 0.2$	Moments	0.2	270.18	0.46	18.81
	Mixte	0.2	281.38	0.46	18.81
	Gibbs	0.2	265.49	0.50	19.44

On peut comparer la colonne du paramètre b avec les PMP estimées à partir des méthodes usuelles, qui sont affichées dans le tableau 5.2. On note que ces estimations et les paramètres de borne supérieure estimées sont très rapprochés.

TABLEAU 5.2 PMP estimées à l'aide de méthodes usuelles pour la station de Montréal

Méthode	PMP (mm)
Maximisation de l'humidité avec EP_{max}	282.19
Maximisation de l'humidité avec EP_{100}	284.44
Hershfield avec $K = 15$	260.76

5.2 Station de St-Hubert

Les mêmes étapes sont réalisées sur les données de la station de St-Hubert (5 303 données). Pour le modèle tel que $a = 0$, on peut voir directement dans les diagrammes quantile-quantile et probabilité-probabilité de la figure 5.5 que les deux distributions estimées à partir de méthodes fréquentistes ne sont pas bien ajustées aux données. Celle ajustée par méthode des moments s'accorde bien aux quatre quantiles les plus élevés, mais surestime considérablement tous les autres niveaux. La distribution paramétrée par la méthode mixte sous-estime les quantiles supérieurs, mais sous-estime les plus petites valeurs.

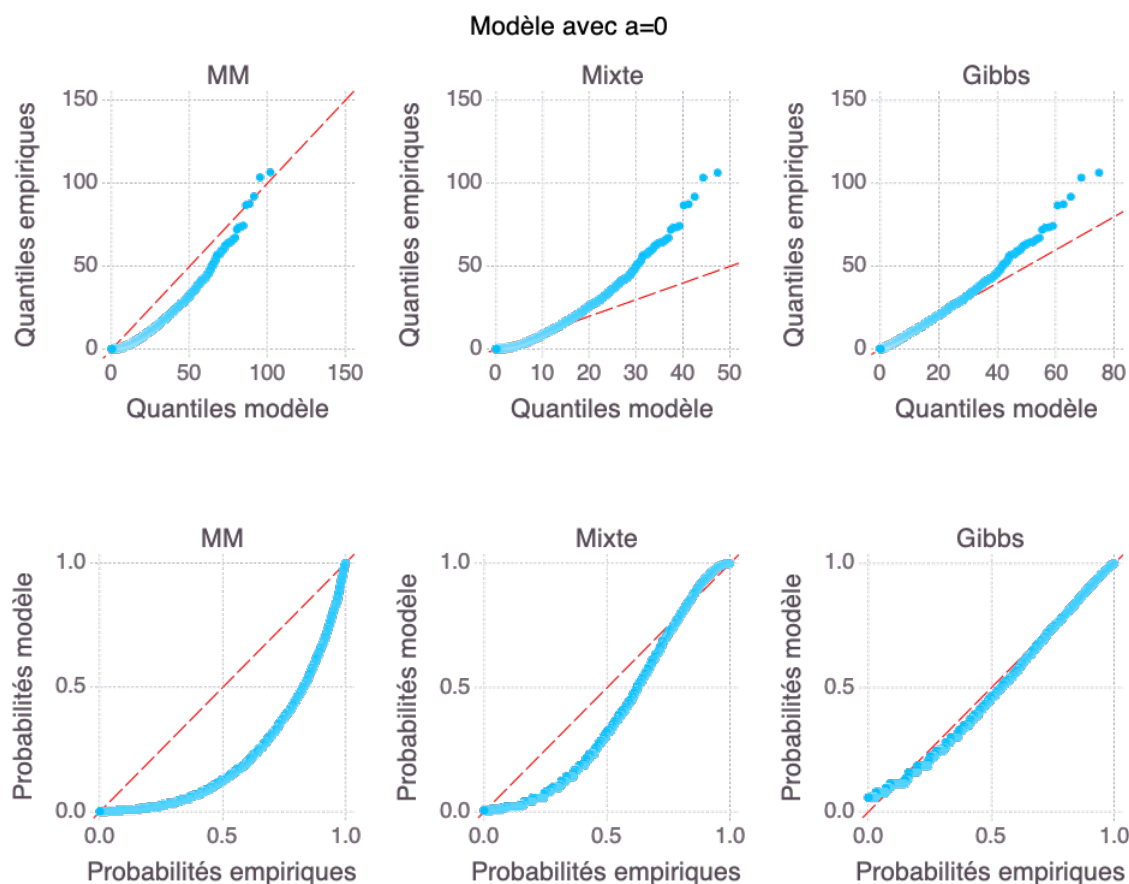


FIGURE 5.5 Diagrammes quantile-quantile et probabilité-probabilité pour la distribution Pearson de type I à trois paramètres ajustée sur les données de St-Hubert à l'aide de trois méthodes différentes (moments, mixte et Gibbs)

La distribution ajustée selon l'algorithme de Gibbs est meilleure que les deux autres, mais sous-estime encore les quantiles les plus élevés. Néanmoins, si on étudie les graphiques de la figure 5.6, on peut conclure que le modèle présente les mêmes problèmes que son analogue

de la station de Montréal et est plutôt médiocre. En effet, les traces de b et β divergent, ce qui indique que les valeurs moyennes de ces paramètres dépendent directement du nombre d'itérations de la période de chauffe et total. La présence de plateaux visibles sur la trace de β signale encore un problème dans l'amplitude du pas généré par la distribution instrumentale.

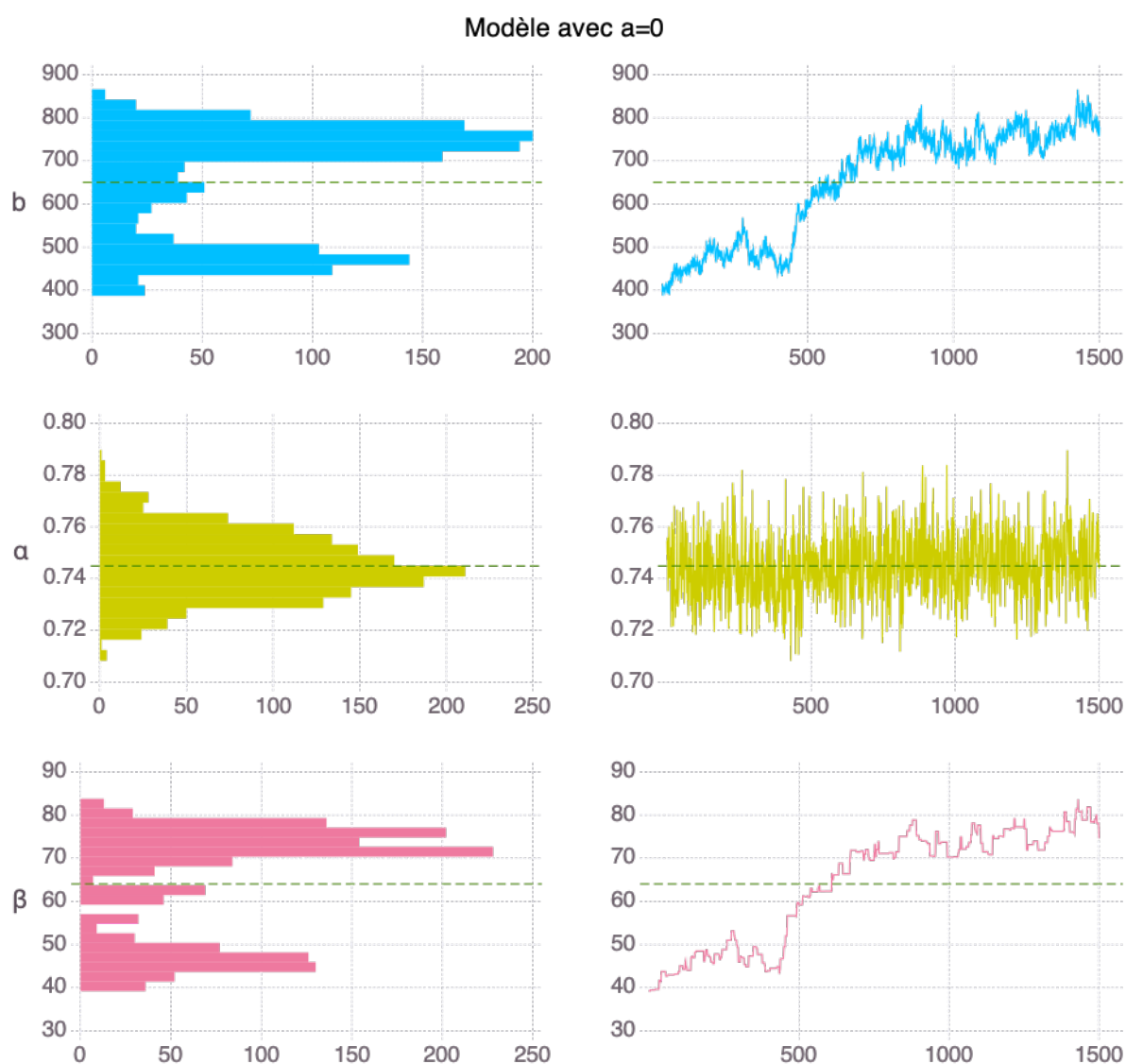


FIGURE 5.6 Traces et histogrammes de la chaîne de Markov générée par l'échantillonneur de Gibbs sur les données de St-Hubert pour le modèle bêta avec a fixé à 0mm. Les graphiques du haut sont associés à b , ceux du centre à α et ceux du bas à β . Les droites pointillées correspondent aux moyennes des états de chacun des paramètres.

Les diagrammes quantile-quantile et probabilité-probabilité de la figure 5.7, associés au modèle avec $a = 0.2\text{mm}$, sont semblables à ceux du modèle précédent, avec un très mauvais ajustement des distributions estimées par méthodes fréquentistes aux observations. Malgré cela, on constate une amélioration de l'ajustement aux données de la distribution paramétrée par l'échantillonneur de Gibbs, particulièrement pour les petites valeurs, mais également vis-à-vis les quantiles plus élevés.

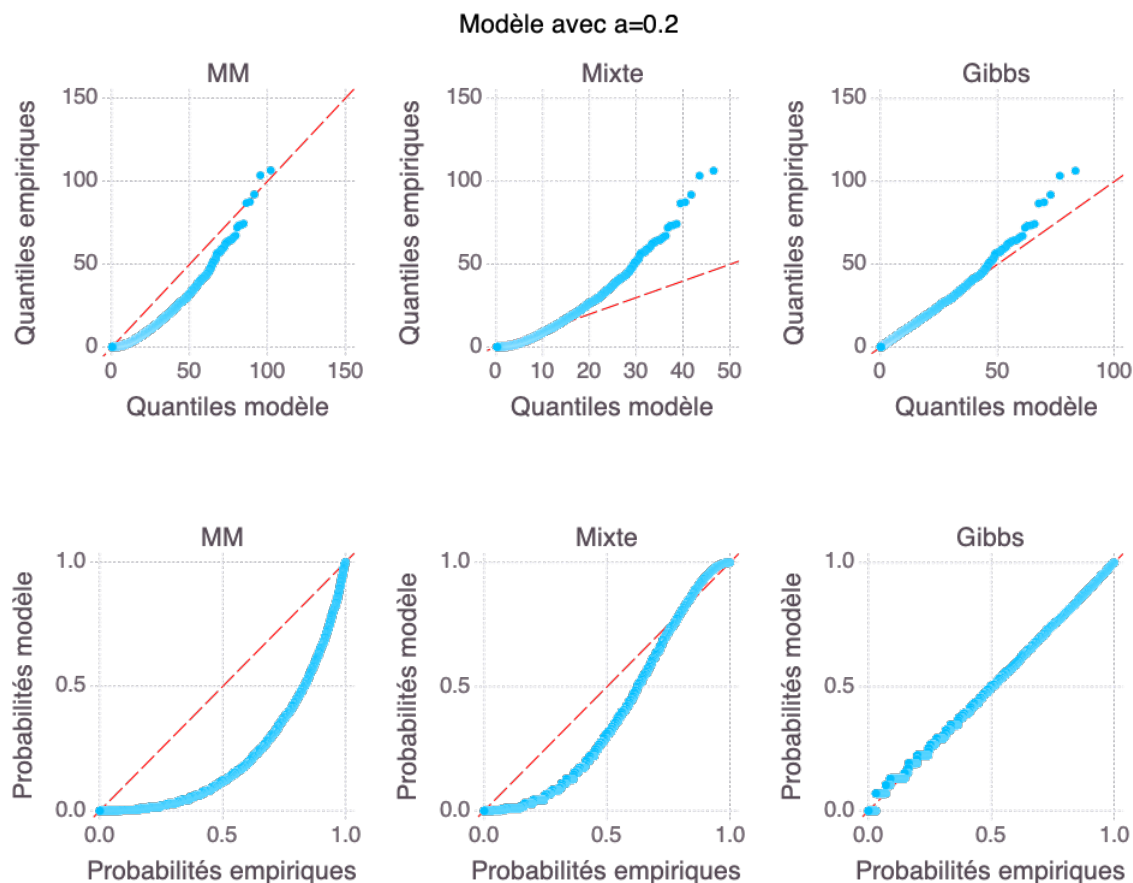


FIGURE 5.7 Diagrammes quantile-quantile et probabilité-probabilité pour la distribution Pearson de type I à quatre paramètres avec $a = 0.2$ fixé, ajustée ajustée sur les données de St-Hubert à l'aide de trois méthodes différentes (moments, mixte et Gibbs)

On confirme ce résultat à l'aide des graphiques de la trace et de l'histogramme de chacun des paramètres de la chaîne de Markov, à la figure 5.8. Ici, les traces des paramètres b et β ne divergent plus, mais oscillent autour de la valeur moyenne (illustrée par la droite pointillée). Les histogrammes ne contiennent chacun qu'un seul maximum, contrairement aux histogrammes des figures précédentes.

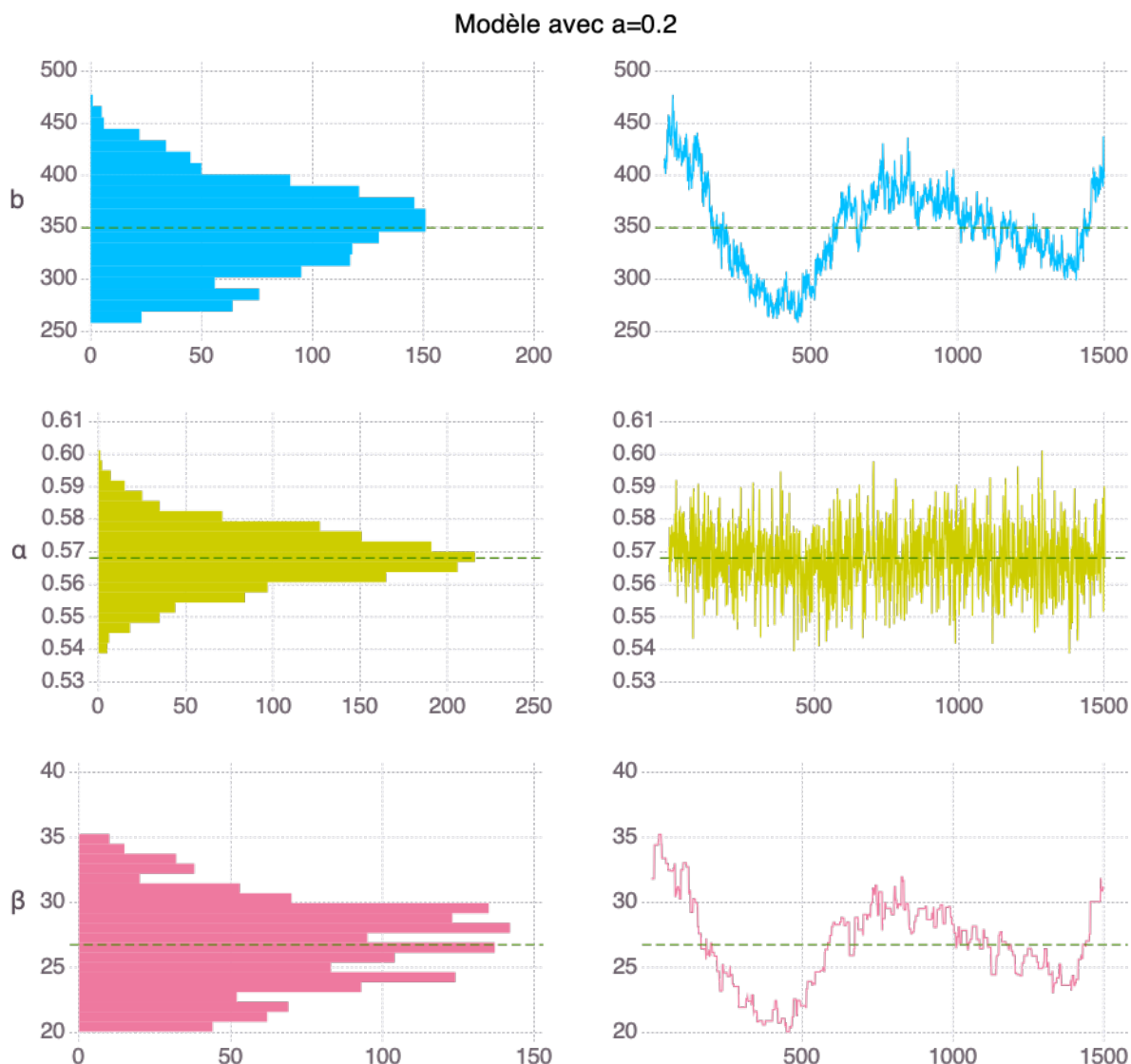


FIGURE 5.8 Traces et histogrammes de la chaîne de Markov générée par l'échantillonneur de Gibbs sur les données de St-Hubert pour le modèle bêta avec a fixé à 0.2mm. Les graphiques du haut sont associés à b , ceux du centre à α et ceux du bas à β . Les droites pointillées correspondent aux moyennes des états de chacun des paramètres.

L'ensemble des paramètres estimés selon les trois méthodes sont affichés dans le tableau 5.3. La colonne associée au paramètre α , qui affiche plusieurs estimations au-dessus de 1, permet de comprendre la faible performance des diagrammes associés aux méthodes fréquentistes des figures 5.5 et 5.7. En effet, on sait en observant la distribution des données de précipitations de la station (figure 4.18) que celle-ci devrait être convexe et donc $\alpha < 1$. Les seules valeurs

estimées de α plus petite que 1 sont associées à l'échantillonneur de Gibbs, méthode ajustant le mieux la distribution aux données selon les graphiques diagnostiques des figures 5.5 et 5.7.

TABLEAU 5.3 Paramètres estimés pour différents modèles et méthodes d'estimations, à partir des données observées de précipitation journalières entre 1949 et 2024 à la station de Saint-Hubert

Modèle	Méthode	a	b	α	β
$a = 0$	Moments	0	416.50	1.46	34.75
	Mixte	0	193.88	1.46	34.75
	Gibbs	0	650.29	0.74	64.05
$a = 0.2$	Moments	0.2	416.70	1.46	34.75
	Mixte	0.2	189.74	1.46	34.75
	Gibbs	0.2	349.66	0.57	26.75

Le tableau 5.4 affiche les PMP estimées à partir des méthodes hydrométéorologiques et d'Hershfield pour la station. Bien que les valeurs s'approchent de celles estimées par la méthode des moments, on sait que ces dernières proviennent de distributions qui ne sont pas adaptées, leur paramètre α étant supérieur à 1. Cependant, on constate que les PMP usuelles s'apparentent également à l'estimation de la borne supérieure obtenue par Gibbs lorsque $a = 0.2$, qui elle pourrait potentiellement être plus pertinente.

TABLEAU 5.4 PMP estimées à l'aide de méthodes usuelles pour St-Hubert

Méthode	PMP (mm)
Maximisation de l'humidité avec EP_{max}	435.54
Maximisation de l'humidité avec EP_{100}	427.27
Hershfield avec $K = 15$	321.75

5.3 Discussion

Un problème récurrent quant à l'inférence bayésienne des précédentes sections est celui de la taille du paramètre de dispersion de la distribution instrumentale de β . Celui-ci est perceptible par l'émergence de plateaux sur les différents graphiques de trace de β . On peut également constater que les traces de b sont problématiques : en effet, bien que celles-ci n'affichent pas de plateaux, elles ne sont pas régulières comme celles du paramètre α . Leur allure, à l'opposé des traces de β , indiquent que la variance instrumentale n'est pas assez grande. Ici, l'ajustement de ce paramètre de dispersion se fait lors de la période de chauffe de l'algorithme. Or, la dépendance entre b et β pourrait rendre cet ajustement arbitraire. Puisque l'acceptation d'un candidat de l'un des paramètre est potentiellement biaisé par l'état actuel de l'autre, le

taux d'acceptation, qui détermine si la variance instrumentale devrait être modifiée, pourrait être erroné. Ainsi, il serait peut-être préférable de produire un grand nombre de chaîne en mettant à jour les distributions instrumentales séparément et *a priori* plutôt qu'au fil des itérations.

Cependant, il ne faut pas oublier le problème soulevé à la fin de la section 4.2.1 : si la variance instrumentale est trop grande, l'algorithme propose des candidats à l'extérieur du domaine de définition des paramètres. Ainsi, un paramètre de dispersion instrumental peut être considéré comme «trop petit» selon les attentes de qualité de la trace, mais «trop grand» pour le support du paramètre d'intérêt. L'équilibre mariant une dispersion instrumentale trop étroite pour β et trop étendue pour b est le compromis accepté pour le moment.

Autant du côté de l'estimation pour la station de Montréal que pour celle de St-Hubert, on note que les modèles pour lesquels a est fixée à 0.2mm sont meilleurs que ceux pour lesquels a est fixée à 0mm puisque la borne est plus représentative de la limite réelle des observations. On peut même observer dans les tableaux 5.1 et 5.3 que l'échantillonneur de Gibbs sur-estime le paramètre de forme α lorsque la borne inférieure choisie n'est pas exacte. Ainsi, le modèle avec $a = 0.2$ est préféré à celui pour lequel $a = 0$.

Aussi bien à partir des résultats affichés dans les tableaux susmentionnés que ceux illustrés par les diagrammes quantile-quantile et probabilité-probabilité (figures 5.1, 5.5, 5.3 et 5.7), il est possible de déterminer que la méthode bayésienne est plus performante que les approches fréquentistes. Or, il est recommandé d'utiliser les trois méthodes (méthode des moments, mixte et Gibbs) afin de pouvoir comparer les résultats et la stabilité de ceux-ci.

En général, les distributions estimées sont mieux ajustées aux données de la station de Montréal qu'à celles de St-Hubert, pour laquelle les paramètres obtenus sont plus instables. Malgré leur proximité, les stations ne semblent pas recevoir la même quantité de pluie et avoir le même nombre de jours pluvieux par année. En effet, il pleut en moyenne 73.9 jours/an à Montréal pour 69.8 jours/an à St-Hubert, alors que le maximum annuel de la précipitation journalière est en moyenne de 44.55mm à Montréal et de 49.54mm à St-Hubert. Au total, il pleut en moyenne 9mm de plus à St-Hubert (519.22mm) qu'à Montréal (510.53mm) par année. Ces différences dans les estimations des PMP obtenues à partir de l'approche bêta entre les stations, couplées à la similarité intra-station entre elles et celles estimées à l'aide de méthodes usuelles, indique que la nouvelle approche est potentiellement aussi dépendante des données que les autres méthodes et ce, malgré l'utilisation de l'ensemble des observations plutôt qu'une sélection des tempêtes plus extrêmes. Néanmoins, bien qu'elle soit dépendante des données, la méthode bêta n'est pas exposée au jugement de la personne qui l'estime.

De plus, l'approche statistique développée permet une certaine évaluation de la justesse des

résultats. Alors que les méthodes usuelles donnent une valeur unique, parfois améliorée d'un intervalle de confiance, il est impossible de savoir si ces résultats sont raisonnables. Dans le cas de la méthode bêta, quelques tests de vérification tels que les diagrammes quantile-quantile et probabilité-probabilité peuvent justifier ou non la légitimité de la borne supérieure produite. L'analyse des estimations de l'ensemble des paramètres et de leur variabilité selon la méthode d'estimation, ainsi que de leurs histogrammes et traces lors de l'inférence bayésienne est également une démarche qui permet de conclure de la qualité de la PMP estimée.

CHAPITRE 6 DISCUSSION ET CONCLUSION

6.1 Synthèse des travaux

À la suite de l'introduction de la PMP et de son état de l'art, on note que ses différentes méthodes d'estimations soulèvent un nombre de critiques. Au cours de ce travail, une méthodologie statistique d'estimation de la PMP est développée en suivant la définition actuelle de l'OMM (2009) et permettant une meilleure estimation de l'incertitude et une adaptation aux CC améliorée par rapport aux méthodes actuelles.

La loi de Pearson de type I, ou loi bêta à quatre paramètres, est choisie pour son domaine de définition borné et sa grande flexibilité afin d'être ajustée aux données de précipitation et ainsi d'en estimer la borne supérieure. Un ensemble d'approches d'estimation fréquentistes telles que la maximisation de la vraisemblance, la méthode des moments et une méthode mixte, ainsi que des méthodes bayésiennes sont introduites afin d'approximer les paramètres de la distribution.

Ces approches sont par la suite affinées à l'aide d'échantillons synthétiques tirés de distributions bêta dont les quatre paramètres sont connus. Le manque d'identifiabilité des paramètres de la distribution entraîne la nécessité de réduire le nombre de paramètres en fixant la borne inférieure a ainsi que l'augmentation de la taille des échantillons testés. Une reparamétrisation de la loi est étudiée afin d'adoucir la corrélation entre les paramètres de borne et de forme de la distribution, sans résultats convaincants. Finalement, les méthodes des moments et mixte de même que l'algorithme de Gibbs sont sélectionnés dans le but d'ajuster la loi bêta aux données de précipitations journalières des stations de l'aéroport international Montréal Trudeau et de St-Hubert.

De cette étude de cas se dégage le potentiel de la méthode malgré ses imperfections. En effet, la méthode bêta propose certaines améliorations aux méthodes usuelles d'estimation de la PMP, entre autres sur le plan de l'évaluation de la qualité de la borne supérieure estimée ainsi que par sa simplicité et son objectivité.

6.2 Limitations de la solution proposée et améliorations futures

La grande flexibilité de la loi de Pearson de type I est un couteau à double tranchant. Bien que celle-ci permette un éventail de possibilités de formes (figure 3.1), le problème du manque d'identifiabilité des paramètres caractérisant la distribution lorsque α et β sont éloignés dégrade la qualité de l'ajustement aux données du modèle pour l'ensemble des méthodes d'estimation des paramètres. La forme des distributions de pluies (figures 4.1 et 4.18) rendent

ainsi, par définition, l'estimation de la borne supérieure b difficile. La rareté des données extrêmes entraîne une grande variabilité dans l'estimation de b . L'inférence bayésienne pourrait être améliorée par l'utilisation de lois *a priori* informatives mais cette approche annulerait l'objectivité recherchée par la méthode statistique développée. La PMP théorique n'ayant jamais été une valeur calculée scientifiquement, les connaissances antérieures de celle-ci sont minces et le choix de distributions *a priori* engendre une subjectivité dans les calculs.

Le nombre de données disponibles aux stations limite également la qualité des estimations. En effet, on a pu voir au fil de ce mémoire que la justesse de l'estimation des paramètres dépend directement de la taille de l'échantillon ajusté. Par exemple, une distribution dont la valeur réelle de α est 0.1 et celle de β est 6 nécessite un échantillon de taille $>10\,000$ afin d'obtenir des estimations raisonnables à partir de la méthode des moments. Selon le nombre de jours de pluie moyen par an aux stations utilisées dans ce travail, il faudrait plus de 135 (Montréal) et 143 (St-Hubert) années d'enregistrement afin de calculer des estimés acceptables.

Dans le chapitre méthodologique, l'estimation des paramètres par maximum de vraisemblance pose régulièrement des problèmes lorsque les valeurs des paramètres de forme α et β s'éloignent. Carnahan (1989) et Wang (2005) mentionnent la difficulté d'utiliser cette approche dans le cas d'une distribution bêta à quatre paramètres convexe dont les bornes sont inconnues. Dans son article, Wang (2005) développe une méthode alliant maximum de vraisemblance et inférence bayésienne et teste celle-ci sur des échantillons de très petite taille (entre 10 et 100) afin d'améliorer l'estimation. Cette approche pourrait être étudiée dans le cas de l'estimation par la méthode bêta de la PMP.

En plus de développer une méthodologie claire pour la caractérisation des incertitudes sur les paramètres estimés, en particulier dans le cas fréquentiste, l'ajout de variables explicatives telles que l'année ou la température dans le modèle statistique sont des avenues d'étude d'amélioration de la démarche d'estimation des PMP futures nécessaires. L'adaptation de la méthode à des données de simulations climatiques est également indispensable, ainsi qu'une démarche permettant la spatialisation du modèle afin de prendre en compte les données sur l'ensemble d'un bassin versant. De plus, l'emploi de modèles climatiques permettrait l'accès à un plus grand nombre de données, ce qui implique des estimations de meilleure qualité.

Il est cependant pertinent de se demander si la définition actuelle de la PMP est toujours adéquate. Comme mentionné au début de la revue de littérature, la PMP n'est pas un concept scientifique. Puisque les études statistiques indiquent que les données de précipitations extrêmes suivent plutôt des distributions à queues lourdes (Papalexiou et Koutsoyiannis, 2013; Papalexiou *et al.*, 2013; Serinaldi et Kilsby, 2014), l'existence d'une borne supérieure pour les

pluies porte à débat. En effet, l'imposition d'une telle borne laisserait à toute approximation de la PMP la possibilité d'être un jour dépassée. Les méthodes usuelles sont également largement critiquées par la communauté scientifique (Hwang et Lall, 2024; National Academies of Sciences, 2024; Papalexiou et Koutsoyiannis, 2006; Rouhani et Leconte, 2020). Ainsi, une nouvelle définition de la PMP est proposée par les National Academies of Sciences (2024) comme étant

« The depth of precipitation for a particular duration, location, and areal extent, such as a drainage basin, with an extremely low annual probability of being exceeded, for a specified climate period. »

En addition à cette définition, la National Academies of Sciences (2024) encourage l'utilisation de méthodes statistiques, en particulier la théorie des valeurs extrêmes, ainsi que de données de modèles climatiques. Ces recommandations ouvrent la porte à l'étude des méthodes d'analyse fréquentielles introduites lors de la revue de littérature. L'approche développée dans ce travail ainsi que l'ensemble des améliorations mentionnées précédemment, loin de devoir être rejetés, peuvent inspirer les travaux futurs sur l'utilisation d'une PMP statistique basée sur l'ajustement d'une loi de probabilité.

RÉFÉRENCES

- BEAUCHAMP, J. (2010). Estimation d'une PMP et d'une CMP d'un bassin versant septentrional, en contexte de changements climatiques. Mémoire de D.E.A., École de technologie supérieure, Montréal.
- BEAUCHAMP, J., LECONTE, R., TRUDEL, M. et BRISSETTE, F. (2013). Estimation of the summer-fall PMP and PMF of a northern watershed under a changed climate. *Water Resources Research*, 49.
- BEN ALAYA, M. A., ZWIERS, F. et ZHANG, X. (2018). Probable maximum precipitation : its estimation and uncertainty quantification using bivariate extreme value analysis. *Journal of Hydrometeorology*, 19(4):679–694.
- BEN ALAYA, M. A., ZWIERS, F. W. et ZHANG, X. (2020a). A bivariate approach to estimating the probability of very extreme precipitation events. *Weather and Climate Extremes*, 30:100290.
- BEN ALAYA, M. A., ZWIERS, F. W. et ZHANG, X. (2020b). Probable maximum precipitation in a warming climate over North America in CanRCM4 and CRCM5. *Climatic Change*, 158(3-4):611–629.
- BETANCOURT, M. (2018). A conceptual introduction to Hamiltonian Monte Carlo. Posted Date : 16 Jul 2018. arXiv :1701.02434v2 [stat.ME].
- CARNAHAN, J. V. (1989). Maximum likelihood estimation for the 4-parameter beta distribution. *Communications in Statistics - Simulation and Computation*, 18(2):513–536.
- CEHQ (2023). Répertoire des barrages.
- CEHQ et SNC-LAVALIN (2004). Estimation de conditions hydrométéorologiques conduisant aux crues maximales probables (CMP) au Québec : Rapport final. Rapport technique 014713-1000-40RT-001-00, Centre d'expertise hydrique du Québec.
- CHEN, L.-C. et BRADLEY, A. A. (2006). Adequacy of using surface humidity to estimate atmospheric moisture availability for probable maximum precipitation. *Water Resources Research*, 42(9).
- CHOW, V. T. (1951). A general formula for hydrologic frequency analysis. *Eos, Transactions American Geophysical Union*, 32(2):231–237.

- CLAVET-GAUMONT, J., HUARD, D., FRIGON, A., KOENIG, K., SLOTA, P., ROUSSEAU, A., KLEIN, I., THIÉMONGE, N., HOUDRÉ, F., PERDIKARIS, J., TURCOTTE, R., LAFLEUR, J. et LAROCHE, B. (2017). Probable maximum flood in a changing climate : An overview for Canadian basins. *Journal of Hydrology : Regional Studies*, pages 11–25.
- COLES, S. (2001). *An introduction to statistical modeling of extreme values*. Springer series in statistics. Springer, London Berlin Heidelberg, 4. printing édition.
- DTN et MGS ENGINEERING (2020). Probable maximum precipitation guidelines for British Columbia. technical report Bulletin 2020-3-PMP, British Columbia Ministry of Forests, Lands, Natural Resources Operations, and Rural Development.
- ELÍASSON, J. (1997). A statistical model for extreme precipitation. *Water Resources Research*, 33(3):449–455.
- EWGR (2020). Assessment of potential impact of climate change on probable maximum precipitation applicable to nuclear facilities in Canada. Final Report R723.1, Canadian Nuclear Safety Commission (CNSC), British Columbia.
- HANSEN, E. M. (1988). Probable maximum precipitation estimates-United States between the Continental Divide and the 103rd meridian. Rapport technique 55A, United States Office of Hydrology., Hydrometeorological Branch ; United States Army., Corps of Engineers, Washington D.C.
- HERSHFIELD, D. M. (1961a). Estimating the probable maximum precipitation. *Journal of the Hydraulics division Proceedings of the American Society of Civil Engineers*.
- HERSHFIELD, D. M. (1961b). Rainfall frequency atlas of the United States. *Technical paper*, 40:1–61.
- HERSHFIELD, D. M. (1965). Method for estimating probable maximum rainfall. *Journal AWWA*, 57(8):965–972.
- HIRAGA, Y., ISERI, Y., WARNER, M. D., FRANS, C. D., DUREN, A. M., ENGLAND, J. F. et KAVVAS, M. L. (2021). Estimation of long-duration maximum precipitation during a winter season for large basins dominated by atmospheric rivers using a numerical weather model. *Journal of Hydrology*, 598:126224.
- HWANG, J. et LALL, U. (2024). Increasing risk of dam failures in the United States due to compound risk of rainfall clusters as climate changes. Preprint. Posted Date : February 12th, 2024. DOI :10.21203/rs.3.rs-3941703/v1, ISSN : 2693-5015.

- JOHNSON, N. L., KOTZ, S. et BALAKRISHNAN, N. (1995). *Continuous Univariate Distributions*, volume 2. Wiley, 2 édition.
- KOUTSOYIANNIS, D. (1999). A probabilistic view of hershfield's method for estimating probable maximum precipitation. *Water Resources Research*, 35(4):1313–1322.
- KUNKEL, K. E., KARL, T. R., EASTERLING, D. R., REDMOND, K., YOUNG, J., YIN, X. et HENNON, P. (2013). Probable maximum precipitation and climate change. *Geophysical Research Letters*, 40(7):1402–1408.
- LAROCQUE, A. (2013). Probable Maximum Flood (PMF). In BOBROWSKY, P. T., éditeur : *Encyclopedia of Natural Hazards*, Encyclopedia of Earth Sciences Series, pages 777–778. Springer Netherlands, Dordrecht.
- MAHONEY, K., LUKAS, J. et MUELLER, M. (2018). Considering climate change in the estimation of extreme precipitations for dam safety. Summary VI, Colorado Division of Water Resources, Department of Natural Resources, Colorado, New Mexico.
- MARTINS, E. et STEDINGER, J. (2000). Generalized maximum likelihood GEV quantile estimators for hydrologic data. *Water Resources Research*, 36:737–744.
- MICOVIC, Z., SCHAEFER, M. G. et TAYLOR, G. H. (2015). Uncertainty analysis for probable maximum precipitation estimates. *Journal of Hydrology*, 521:360–373.
- MILLER, J. F. (1963). Probable maximum precipitation and rainfall-frequency data for Alaska for areas to 400 square miles, durations to 24 hours and return periods from 1 to 100 years. Technical paper 47, Weather Bureau, United States Department of Commerce Washington, D.C.
- National Academies of SCIENCES, Engineering, a. M. (2024). *Modernizing Probable Maximum Precipitation Estimation*. National Academies Press, Washington D.C., the national academies press édition.
- OMM (2009). *Manual on estimation of Probable Maximum Precipitation (PMP)*. World Meteorological Organization (WMO), Geneva.
- PAPALEXIOU, S. M. et KOUTSOYIANNIS, D. (2006). A probabilistic approach to the concept of Probable Maximum Precipitation. *Advances in Geosciences*, 7:51–54.
- PAPALEXIOU, S. M. et KOUTSOYIANNIS, D. (2013). Battle of extreme value distributions : A global survey on extreme daily rainfall. *Water Resources Research*, 49(1):187–201.

- PAPALEXIOU, S. M., KOUTSOYIANNIS, D. et MAKROPOULOS, C. (2013). How extreme is extreme? An assessment of daily rainfall distribution tails. *Hydrology and Earth System Sciences*, 17(2):851–862.
- QUÉBEC (2023). Règlement sur la sécurité des barrages.
- ROUHANI, H. et LECONTE, R. (2016). A novel method to estimate the maximization ratio of the Probable Maximum Precipitation (PMP) using regional climate model output. *Water Resources Research*, 52(9):7347–7365.
- ROUHANI, H. et LECONTE, R. (2018). A methodological framework to assess PMP and PMF in snow-dominated watersheds under changing climate conditions – A case study of three watersheds in Québec (Canada). *Journal of Hydrology*, 561:796–809.
- ROUHANI, H. et LECONTE, R. (2020). Uncertainties of precipitable water calculations for PMP estimates in current and future climates. *Journal of Hydrologic Engineering*, 25(3): 04019066.
- ROUSSEAU, A. N., KLEIN, I. M., FREUDIGER, D., GAGNON, P., FRIGON, A. et RATTÉ-FORTIN, C. (2014). Development of a methodology to evaluate probable maximum precipitation (PMP) under changing climate conditions : Application to southern Quebec, Canada. *Journal of Hydrology*, 519:3094–3109.
- SALAS, J. D., GAVILÁN, G., SALAS, F. R., JULIEN, P. Y. et ABDULLAH, J. (2014). Uncertainty of the PMP and PMF. *Handbook of Engineering Hydrology*.
- SARKAR, S. et MAITY, R. (2020). Increase in probable maximum precipitation in a changing climate over India. *Journal of Hydrology*, 585:124806.
- SCHREINER, L.-C. et RIEDEL, J. T. (1978). Probable maximum precipitation estimates, United States east of the 105th meridian. Rapport technique 51, United States, Office of Hydrology., Hydrometeorological Branch, ;United States, Army., Corps of Engineers, ;, Washington D.C.
- SERINALDI, F. et KILSBY, C. G. (2014). Rainfall extremes : Toward reconciliation after the battle of distributions. *Water Resources Research*, 50(1):336–352.
- SINGH, A., SINGH, V. P. et AR, B. (2018). Computation of probable maximum precipitation and its uncertainty. *International Journal of Hydrology*, 2(4).

- UNITED STATES WEATHER BUREAU (1960). Generalized estimates of probable maximum precipitation for the United States west of the 105th meridian for areas to 400 squares miles and durations to 24 hours. Technical paper 38, Weather Bureau, United States Department of Commerce Washington, D.C.
- VISSER, J. B., KIM, S., WASKO, C., NATHAN, R. et SHARMA, A. (2022). The impact of climate change on operational probable maximum precipitation estimates. *Water Resources Research*, 58(11):e2022WR032247.
- VISWANADHAM, Y. (1981). The relationship between total precipitable water and surface dew point. *Journal of Applied Meteorology and Climatology*, 20(1):3–8.
- WANG, J. Z. (2005). A note on estimation in the four-parameter beta distribution. *Communications in Statistics - Simulation and Computation*, 34(3):495–501.
- ZHOU, H. et HUANG, X. (2022). Bayesian beta regression for bounded responses with unknown supports. *Computational Statistics & Data Analysis*, 167:107345.

ANNEXE A CORRÉLATION DES PARAMÈTRES

TABLEAU A.1 Coefficients de régression entre $\hat{a}$ et $\hat{\alpha}$ pour chacune des distributions et des méthodes d'estimations, avec m_0 l'ordonnée à l'origine et m_1 la pente.

Distribution	Methode		Coefficient	t	P(> t)
A	MM	m_0	1.976(0.008)	257.70	<1e-99
		m_1	-0.188(0.005)	-37.30	<1e-99
	Mixte	m_0	1.998(0.008)	262.99	<1e-99
		m_1	-0.202(0.005)	-37.72	<1e-99
	MV	m_0	1.991(0.007)	297.80	<1e-99
		m_1	-0.208(0.005)	-38.16	<1e-99
B	MM	m_0	0.941(0.005)	201.94	<1e-99
		m_1	-0.39(0.01)	-29.28	<1e-99
	Mixte	m_0	0.995(0.005)	210.46	<1e-99
		m_1	-0.68(0.03)	-26.76	<1e-99
	MV	m_0	0.973(0.004)	251.38	<1e-99
		m_1	-0.51(0.04)	-13.23	<1e-34
C	MM	m_0	0.103(0.009)	11.44	<1e-26
		m_1	-0.1677(1×10^{-4})	-1562.11	<1e-99
	Mixte	m_0	0.087(0.008)	10.93	<1e-24
		m_1	-0.05529(3×10^{-5})	-1752.67	<1e-99
	MV	m_0	1×10^8 (1×10^8)	1.11	0.2694
		m_1	-17300(300)	-61.65	<1e-99
D	MM	m_0	0.1(0.4)	0.30	0.7680
		m_1	-0.317(0.001)	-260.91	<1e-99
	Mixte	m_0	0.4(0.2)	1.56	0.1190
		m_1	-0.1011(0.0002)	-448.77	<1e-99
	MV	m_0	5×10^8 (2×10^8)	3.30	0.001
		m_1	-18300(300)	-58.70	<1e-99

TABLEAU A.2 Coefficients de régression entre $\hat{b}$ et $\hat{\beta}$ pour chacune des distributions et des méthodes d'estimations, avec m_0 l'ordonnée à l'origine et m_1 la pente.

Distribution	Méthode		Coefficient	t	P(> t)
A	MM	m_0	-7.6(0.2)	-33.60	<1e-99
		m_1	0.192(0.005)	42.35	<1e-99
	Mixte	m_0	-8.4(0.2)	-34.73	<1e-99
		m_1	0.208(0.005)	42.94	<1e-99
	MV	m_0	-8.4(0.3)	-28.61	<1e-99
		m_1	0.207(0.006)	35.09	<1e-99
B	MM	m_0	-6.4(0.2)	-36.44	<1e-99
		m_1	0.188(0.003)	53.95	<1e-99
	Mixte	m_0	-5.8(0.2)	-27.68	<1e-99
		m_1	0.177(0.004)	43.21	<1e-99
	MV	m_0	-1.8(0.1)	-14.67	<1e-41
		m_1	0.096(0.003)	37.08	<1e-99
C	MM	m_0	-1.58(0.08)	-19.41	<1e-64
		m_1	$0.1493(8 \times 10^{-4})$	183.75	<1e-99
	Mixte	m_0	-7(2)	-3.76	0.0002
		m_1	0.52(0.07)	7.03	<1e-11
	MV	m_0	$-5 \times 10^6(5 \times 10^6)$	-0.95	0.3431
		m_1	30800(400)	78.89	<1e-99
D	MM	m_0	-1.3(0.1)	-10.31	<1e-22
		m_1	0.316(0.002)	146.07	<1e-99
	Mixte	m_0	-6(2)	-2.91	0.0038
		m_1	1.2(0.1)	8.01	<1e-14
	MV	m_0	$-1.4 \times 10^7(7 \times 10^6)$	-2.12	0.0348
		m_1	46500(700)	68.52	<1e-99