

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

**Développement d'une méthode de post-traitement adaptée aux valeurs
extrêmes pour les précipitations simulées par un modèle climatique**

GABRIEL GOBEIL

Département de mathématiques et de génie industriel

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

Août 2021

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Ce mémoire intitulé :

**Développement d'une méthode de post-traitement adaptée aux valeurs
extrêmes pour les précipitations simulées par un modèle climatique**

présenté par **Gabriel GOBEIL**

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

Philippe ROY, membre et codirecteur de recherche

Alejandro DI LUCA, membre

DÉDICACE

*À mes parents, Serge et Marie,
qui m'ont toujours encouragé
et soutenu dans mes études.*

REMERCIEMENTS

Mes remerciements vont à tous ceux et celles qui m'ont aidé de près ou de loin dans la réalisation de ce projet.

Tout d'abord, un gros merci à mes directeurs de recherche Jonathan Jalbert et Philippe Roy pour leur encadrement, leur soutien et leur disponibilité. Votre aide m'a été très précieuse au cours de ces deux années et j'ai beaucoup de plaisir à travailler avec vous.

Je tiens aussi à remercier Émilie Bresson de m'avoir introduit au monde de la recherche scientifique et sans qui je n'aurais probablement pas envisagé de faire cette maîtrise.

De façon plus formelle, je remercie Polytechnique Montréal, Hydro-Québec, Ouranos et Mitacs d'avoir financé ce projet.

Également, merci à Philémon, Marilou, Duy-Anh et Isabelle avec qui j'ai eu beaucoup de plaisir à échanger à chaque rencontre du lundi. Vous avez rendu le travail à distance plus agréable.

Enfin, merci à ma famille et surtout, merci à Lisa, mon amour, pour tout le soutien et les encouragements.

RÉSUMÉ

Les événements de précipitations intenses sont susceptibles d'augmenter en intensité et en fréquence au Canada. Au Québec, cette augmentation aura des impacts considérables sur l'ensemble des secteurs d'activité, tels l'agriculture, l'énergie et l'aménagement du territoire. La quantification de cette augmentation ainsi que l'étude de ces impacts se font généralement à l'aide des précipitations simulées par les modèles climatiques. Les précipitations simulées comportent toutefois des biais par rapport aux observations qu'il est nécessaire de corriger avant de les utiliser. Les méthodes de post-traitement actuellement utilisées pour corriger ces biais modifient la distribution des précipitations simulées en fonction d'une distribution de référence. Les distributions sont généralement estimées de façon non-paramétrique et la précision de l'estimation diminue dans les queues de la distribution. Pour bien modéliser les extrêmes, des méthodes paramétriques sont requises. En général, celles-ci divisent artificiellement les précipitations en deux parties, soit les grandes intensités et le reste, ce qui impose d'utiliser des modèles statistiques différents pour corriger les biais de chaque gamme. Ainsi, l'objectif de ce projet de recherche consiste à développer une méthode de post-traitement paramétrique adaptée pour les valeurs extrêmes et le cœur de la distribution.

Pour atteindre cet objectif, les familles de distributions paramétriques développées par Naveau *et al.* (2016), familles de lois modélisant à la fois le cœur et les queues de la distribution en accord avec la théorie des valeurs extrêmes, ont été adaptées au contexte du post-traitement. Ces familles de modèles ont été ajustées sur les précipitations journalières observées à Montréal et sur les précipitations journalières simulées de l'ensemble ClimEx entre 1980 et 2009. Il a été déterminé qu'une de ces familles était adéquate pour modéliser les observations et les simulations. Une extension non stationnaire a été développée pour ce modèle afin de prendre en compte la non-stationnarité des précipitations simulées sur la période 1955-2009. Puis, la mise en correspondance de ces modèles paramétriques a permis le développement d'une méthode d'ajustement de biais adéquate autant pour le cœur de la distribution que pour les queues.

La méthode a été validée sur les observations par le calcul d'indices adaptés pour le cœur et les extrêmes. Les résultats montrent une amélioration par rapport aux simulations brutes. En somme, la réalisation de ce projet a permis d'ajouter une composante non stationnaire à un modèle de Naveau *et al.* (2016) et son adaptation au contexte du post-traitement des précipitations simulées. La méthode développée pourra servir à post-traiter les précipitations simulées strictement positives.

ABSTRACT

Intense precipitation events are likely to increase in intensity and frequency in Canada. In Quebec, this increase will have considerable impacts on all sectors of activity, such as agriculture, energy and land use. The quantification of this increase and the study of these impacts are generally done using precipitation simulated by climate models. However, simulated precipitation has biases with respect to observations that need to be corrected before it can be used. The post-processing methods currently used to correct these biases modify the distribution of simulated precipitation based on a reference distribution. The distributions are usually estimated non-parametrically and the accuracy of the estimate decreases in the tails of the distribution. To properly model the extremes, parametric methods are required. In general, these methods artificially divide the precipitation into two parts, i.e. the high intensities and the rest, which requires the use of different statistical models to correct the biases of each range. Thus, the objective of this research project is to develop a parametric post-processing method adapted for extreme values and the core of the distribution.

To achieve this goal, the families of parametric distributions developed by Naveau *et al.* (2016), families of laws modeling both the core and tails of the distribution in line with extreme value theory, were adapted to the post-processing context. These families of models were fitted to observed daily precipitation in Montreal and to simulated daily precipitation from the ClimEx ensemble between 1980 and 2009. It was determined that one of these families was adequate to model the observations and simulations. A non-stationary extension was developed for this model to take into account the non-stationarity of the simulated precipitation over the period [1955, 2099]. Then, the mapping of these parametric models allowed the development of an adequate bias adjustment method for both the core and the tails of the distribution.

The method was validated on the observations by computing suitable indices for the core and the tails. The results show an improvement compared to the raw simulations. In sum, the completion of this project added a non-stationary component to a model of Naveau *et al.* (2016) and its adaptation to the context of post-processing simulated precipitation. The developed method can be used to post-process strictly positive simulated precipitation.

TABLE DES MATIÈRES

DÉDICACE	iii
REMERCIEMENTS	iv
RÉSUMÉ	v
ABSTRACT	vi
TABLE DES MATIÈRES	vii
LISTE DES TABLEAUX	x
LISTE DES FIGURES	xi
LISTE DES SIGLES ET ABRÉVIATIONS	xiv
LISTE DES ANNEXES	xv
CHAPITRE 1 INTRODUCTION	1
CHAPITRE 2 REVUE DE LITTÉRATURE	5
2.1 Modèles numériques de climat	5
2.2 Post-traitement statistique	6
2.3 Post-traitement des extrêmes	9
2.4 Validation des méthodes de post-traitement	10
2.5 Modélisation statistique des précipitations extrêmes	12
CHAPITRE 3 CADRE THÉORIQUE	14
3.1 Théorie des valeurs extrêmes	14
3.1.1 Loi limite du maximum	14
3.1.2 Modélisation statistique des excès au-dessus d'un seuil	16
3.1.3 Extension de la loi de Pareto généralisée	17
3.2 Sélection de modèle	21
3.2.1 Critère d'information bayésien	21
3.3 Méthodes de post-traitement statistique	22
3.3.1 Méthode de correction par quantiles	24

3.3.2	Méthode de la CDF- <i>transform</i>	25
3.3.3	Méthode de la CDF- <i>transform</i> pour les extrêmes non stationnaires . .	26
3.4	Validation des méthodes de post-traitement	28
3.4.1	Indices climatiques	28
3.4.2	Mesures de performance	29
CHAPITRE 4	DONNÉES	31
4.1	Précipitations observées	31
4.2	Ensemble de simulations	33
CHAPITRE 5	MÉTHODOLOGIE ET RÉSULTATS	35
5.1	Modélisation des précipitations observées	35
5.1.1	Estimation des paramètres par maximum de vraisemblance	35
5.1.2	Résultats	37
5.1.3	Sélection du modèle	38
5.2	Modélisation des précipitations simulées	40
5.2.1	Estimation des paramètres par maximum de vraisemblance	40
5.2.2	Résultats	41
5.2.3	Sélection du modèle	42
5.2.4	Analyse des 50 membres	42
5.2.5	Étude de la non-stationnarité	43
5.2.6	Choix d'une variable explicative	46
5.2.7	Modèles d'ajustement	47
5.2.8	Sélection du modèle	48
5.2.9	Analyse des 50 membres	49
5.3	Développement de la méthode de post-traitement	50
5.3.1	Ajustement des modèles	50
5.3.2	Mise en correspondance des fonctions de répartition	51
5.3.3	Correction des valeurs	51
5.4	Évaluation de la méthode	52
5.4.1	Résultats pour le membre KDI	52
5.4.2	Résultats pour les 50 membres	54
CHAPITRE 6	DISCUSSION	57
CHAPITRE 7	CONCLUSION ET RECOMMANDATIONS	60
7.1	Synthèse des travaux	60

7.2 Limitations et améliorations futures	61
RÉFÉRENCES	62
ANNEXES	68

LISTE DES TABLEAUX

Tableau 3.1	Interprétation du facteur de Bayes. Adapté de [1].	23
Tableau 5.1	Paramètres estimés par maximum de vraisemblance sur les précipitations observées. Pour le modèle 3, $\kappa_1 = \delta$ et pour le modèle 4, $\kappa_1 = \delta$ et $\kappa_2 = \kappa$. Les intervalles de confiance à 95% (entre parenthèses) sont obtenus par bootstrap non paramétrique sur 500 échantillons.	37
Tableau 5.2	Valeurs du BIC des quatre modèles sur les données observées de 1980 à 2009.	38
Tableau 5.3	Paramètres estimés par maximum de vraisemblance sur les précipitations simulées du membre KDI. Pour le modèle 3, $\kappa_1 = \delta$ et pour le modèle 4, $\kappa_1 = \delta$ et $\kappa_2 = \kappa$. Les intervalles de confiance à 95% (entre parenthèses) sont obtenus par bootstrap non paramétrique sur 500 échantillons.	41
Tableau 5.4	Valeurs du BIC des quatre modèles sur les données simulées du membre KDI de 1980 à 2009.	42
Tableau 5.5	Valeurs du BIC des huit modèles sur les données simulées du membre KDI de 1955 à 2009.	48
Tableau 5.6	Valeurs des indices calculés sur les observations, le membre KDI avant le post-traitement (sim. brute) et après le post-traitement (sim. corrigée) et de l'indice de comparaison (skill score).	53
Tableau 5.7	Différence relative moyenne en % (sur 50 membres) entre les indices calculés sur les observations et la simulation brute (Sim. brute), la méthode développée (EGPQM) et la méthode empirique (EQM) et fréquences d'amélioration du biais (sur 50 membres).	55

LISTE DES FIGURES

Figure 2.1	La distribution gamma d'une variable biaisée (en bleu) est mise en correspondance avec la gamma de référence (en rouge). La fonction de correction (en pointillé noir) dépend du quantile de référence. Figure adaptée de [2].	8
Figure 3.1	Fonction de densité correspondant au modèle 1 pour $\sigma = 1$, $\xi = 0.5$ et $\kappa = 1, 2, 5$ (rouge, bleu, vert, respectivement). Le cas $\kappa = 1$ correspond à une densité GP. La courbe en orange représente une densité $Gamma(1.4, 1.4)$. Figure adaptée de [3].	19
Figure 3.2	Fonction de densité correspondant au modèle 2 pour $p = 0.5$, $\sigma = 1$, $\xi = 0.5$, $\kappa_1 = 2$, et $\kappa_2 = 2, 5, 10$ (rouge, bleu, vert, respectivement). Le cas $\kappa = 1$ correspond à une densité GP. La courbe en orange représente une densité $Gamma(1.4, 1.4)$. Figure adaptée de [3].	20
Figure 3.3	Fonction de densité correspondant au modèle 3 pour $\sigma = 1$, $\xi = 0.5$, et $\delta = 1, 5, \infty$ (rouge, bleu, vert, respectivement). Le cas $\delta = \infty$ correspond à une densité GP. La courbe en orange représente une densité $Gamma(1.4, 1.4)$. Figure adaptée de [3].	21
Figure 3.4	Fonction de densité correspondant au modèle 3 pour $\sigma = 1$, $\xi = 0.5$, $\delta = 1$ et $\kappa = 1, 2, 5$ (rouge, bleu, vert, respectivement). Le cas $\delta = \infty$ correspond à une densité GP. La courbe en orange représente une densité $Gamma(1.4, 1.4)$	22
Figure 3.5	La fonction de répartition d'une variable biaisée (en bleu) est mise en correspondance avec la fonction de répartition de référence (en rouge). La fonction de correction (en pointillé noir) dépend du quantile de référence. Figure adaptée de [4].	24
Figure 3.6	Pour la période de calibration, la fonction de répartition d'une variable biaisée (en bleu) est mise en correspondance avec la fonction de répartition de référence (en rouge). Pour la période de projection, la fonction de répartition d'une variable biaisée (en orange) est mise en correspondance avec la fonction de répartition de référence obtenue par la CDF- <i>transform</i> (en pointillé vert). Les fonctions de correction (en pointillé noir) dépendent du quantile de référence.	26
Figure 4.1	Diagrammes en boîte des précipitations journalières (mm) par saison observées à Montréal entre 1980 et 2009.	32

Figure 4.2	Précipitations cumulatives (mm) par saison observées à Montréal pour la période 1980-2009.	32
Figure 4.3	Précipitations cumulatives (mm) normalisées par le nombre de jours de pluie ($>1\text{mm/jour}$) par saison observées à Montréal pour la période 1980-2009.	33
Figure 5.1	Diagrammes diagnostiques pour l'ajustement du modèle 1 aux précipitations observées à Montréal.	39
Figure 5.2	Diagrammes diagnostiques pour l'ajustement du modèle 1 aux précipitations simulées du membre KDI pour la période 1980-2009.	43
Figure 5.3	Diagrammes diagnostiques pour l'ajustement du modèle 1 stationnaire aux précipitations simulées du membre KDI pour la période 1955-2009.	44
Figure 5.4	Évolution du paramètre d'échelle σ du modèle 1.	45
Figure 5.5	Évolution du paramètre de queue supérieure ξ du modèle 1.	45
Figure 5.6	Évolution du paramètre de forme de la queue inférieure κ du modèle 1.	46
Figure 5.7	Concentration annuelle des GES en équivalent CO_2 (ppm) en fonction des années selon de scénario RCP8.5.	47
Figure 5.8	Diagrammes quantile-quantile des résidus des modèles non stationnaires à l'échelle exponentielle. À gauche : σ dépend de la concentration en CO_2 . À droite : ξ dépend de la concentration en CO_2	49
Figure 5.9	Fréquence de sélection avec le BIC des huit modèles.	50
Figure 5.10	Maxima annuels de la simulation corrigée (bleu) par rapport à la simulation brute (rouge) pour le membre KDI de 1955 à 2013.	52
Figure 5.11	Différence relative (%) entre les indices calculés sur les observations et la simulation KDI brute (bleu) et la simulation KDI corrigée (or).	54
Figure 5.12	Maxima annuels des simulations corrigées (bleu) par rapport aux simulations brutes (rouge). Les lignes pleines représentent la médiane des maxima annuels des 50 membres pour chaque ensemble.	56
Figure 6.1	Maxima annuels des simulations corrigées par EGPQM (bleu) et par EQM (rouge). Les lignes pleines représentent la médiane des maxima annuels des 50 membres corrigés avec chaque méthode.	58
Figure A.1	Différence relative (%) entre les indices calculés sur les observations et la simulation brute (bleu) et la simulation corrigée (or) pour les membres KDA à KDL.	68
Figure A.2	Différence relative (%) entre les indices calculés sur les observations et la simulation brute (bleu) et la simulation corrigée (or) pour les membres KDM à KDX.	69

Figure A.3	Différence relative (%) entre les indices calculés sur les observations et la simulation brute (bleu) et la simulation corrigée (or) pour les membres KDY à KEJ.	70
Figure A.4	Différence relative (%) entre les indices calculés sur les observations et la simulation brute (bleu) et la simulation corrigée (or) pour les membres KEK à KEV.	71
Figure A.5	Différence relative (%) entre les indices calculés sur les observations et la simulation brute (bleu) et la simulation corrigée (or) pour les membres KEW à KEX.	72

LISTE DES SIGLES ET ABRÉVIATIONS

BIC	<i>Bayesian Information Criterion</i>
CC	Changements climatiques
ClimEx	<i>Climate Change and Hydrological Extremes</i>
CO ₂	Dioxyde de carbone
EGPD	<i>Extended Generalized Pareto Distribution</i>
EGPQM	<i>Extended Generalized Pareto Quantile Mapping</i>
EQM	<i>Empirical Quantile Mapping</i>
GIEC	Groupe d'experts intergouvernemental sur le climat
GP	<i>Generalized Pareto</i>
GPQM	<i>Generalized Pareto Quantile Mapping</i>
MGC	Modèle global de climat
MMC	Mélange de modèle conditionnel
MOS	<i>Model Output Statistics</i>
MRC	Modèle régional de climat
ppm	Parties par million
PQM	<i>Parametric Quantile Mapping</i>
QM	<i>Quantile Mapping</i>
RCP	<i>Representative Concentration Pathway</i>

LISTE DES ANNEXES

Annexe A	Résultats pour les 50 membres	68
----------	---	----

CHAPITRE 1 INTRODUCTION

Selon le plus récent rapport du Groupe d'experts intergouvernemental sur le climat (GIEC), les événements extrêmes tels que les précipitations intenses sont susceptibles d'augmenter en intensité et en fréquence au Canada [5]. Au Québec, cette augmentation aura des impacts considérables sur l'ensemble des secteurs d'activité de la province [6], tels que l'agriculture, l'énergie et l'aménagement du territoire.

L'étude des impacts des changements climatiques (CC) basée sur la physique du système climatique se fait à l'aide de simulations issues de modèles climatiques [7]. Un modèle climatique est une représentation numérique du climat terrestre basée sur des équations mathématiques représentant de manière simplifiée les processus physiques et la dynamique du système climatique. Ces équations sont résolues de manière numérique à l'aide de superordinateurs et de méthodes numériques [8]. Puisqu'il serait impossible de résoudre ces équations en tout point, notamment en raison des contraintes en temps de calcul, les équations sont discrétisées sur des grilles qui déterminent la résolution des modèles.

Les modèles climatiques se distinguent généralement en deux types selon leur domaine spatial et leur façon de concentrer le temps de calcul, soit les modèles globaux de climat (MGC) et les modèles régionaux de climat (MRC). Les MGC sont des modèles dont le domaine couvre l'entièreté du globe. La circulation générale de l'atmosphère, les interactions Terre-océans-atmosphère, les cycles biogéochimiques et d'autres interactions y sont représentés [8]. En raison de la puissance du calcul informatique, la résolution horizontale des MGC actuels est conditionnée par la complexité du modèle. Par exemple, la résolution horizontale des modèles de la sixième phase du projet d'intercomparaison des modèles couplés (CMIP6) [9] varie entre environ 80 et 310 km. Par conséquent, les phénomènes physiques de taille inférieure à cette échelle ne peuvent être simulés par les MGC. Ces phénomènes doivent être représentés à l'aide de paramétrages, qui sont une représentation simplifiée par des paramètres des processus physiques importants qui ne peuvent être résolus directement par un modèle numérique.

Ces paramétrages ainsi que plusieurs autres facteurs tels que le choix du schéma numérique, de la grille, du degré de simplification et des schémas de représentation du sol et de la végétation font en sorte qu'il existe de nombreux modèles climatiques différents développés par différents centres de modélisation climatique. Ces différences impliquent que chaque modèle est unique et génère des réalisations du climat futur différentes avec les mêmes données entrantes, sans toutefois être nécessairement indépendant [10].

Pour concentrer les capacités de calcul sur un domaine plus petit afin de pouvoir simuler

les processus à plus petite échelle, les MRC ont été développés [11]. Conceptuellement, un MRC est similaire à un MGC pour lequel le domaine de simulation est limité à une sous-région de la Terre, permettant ainsi d'augmenter la résolution spatiale. Puisque le domaine des MRC ne couvre qu'une partie du globe, les conditions aux frontières du modèle doivent être fournies. Il peut s'agir de données des MRC ou d'autres sources comme les réanalyses. Le domaine réduit des MRC permet de résoudre les équations du modèle climatique pour une résolution spatiale plus fine. Il est à noter qu'il existe également des modèles globaux à résolution variable (MGRV) qui simulent le climat régional pour des régions spécifiques au moyen d'un raffinement régional dans le domaine global [12].

Bien que la résolution des MRC soit plus fine que celle des MGC, certains phénomènes physiques ne peuvent toujours pas être résolus directement par les équations fondamentales des modèles parce qu'ils se produisent à des échelles spatiales encore plus fines que la résolution effective des modèles régionaux. C'est le cas notamment pour les cellules orageuses à haute intensité et de courte durée responsables des précipitations extrêmes. Pour cette raison, certains modèles climatiques sont généralement biaisés dans leur représentation des événements extrêmes [13].

Malgré les améliorations continues dans la représentation des processus physiques, la présence d'erreurs systématiques ou d'autres biais persiste dans les modèles climatiques [14]. Par exemple, si on considère les précipitations observées et simulées comme une variable aléatoire issue d'une distribution multivariée ayant des aspects marginaux, temporels, spatiaux et multivariés représentant le système climatique, les simulations climatiques vont différer des observations sur certains de ces aspects. Autrement dit, les simulations climatiques sont biaisées par rapport aux observations.

Les précipitations simulées par les modèles climatiques servent, entre autres, pour les études d'impacts hydrologiques. Cependant, en raison de leur biais, les simulations climatiques sont rarement directement utilisées car les modèles d'impacts sont souvent des modèles statistiques calibrés empiriquement à partir d'observations. Ainsi, avant de pouvoir utiliser les simulations des modèles climatiques, les biais doivent d'abord être corrigés [14]. Cette étape est accomplie en ayant recours à des méthodes de post-traitement visant à faire correspondre les statistiques des données de simulations climatiques aux statistiques observées pour une période de référence donnée [7].

Plusieurs techniques de correction de biais des modèles climatiques ont été développées au cours des dernières années. Une méthode corrigeant l'ensemble de la distribution de la variable à corriger est l'ajustement par quantiles (Quantile Mapping, QM) [15]. Cette méthode ajuste la fonction de répartition de la précipitation simulée pour correspondre à la fonction de

répartition de la précipitation observée sur une période de calibration. Il existe différentes formulations pour l’ajustement par quantiles. L’approche non-paramétrique considère les quantiles obtenus empiriquement alors que l’approche paramétrique va plutôt considérer les quantiles obtenus par l’ajustement de lois de probabilité.

La qualité du post-traitement des précipitations simulées a notamment une influence sur la qualité des simulations hydrologiques, surtout en ce qui concerne la simulation des événements extrêmes tels que les crues. Des méthodes de correction performantes sont donc utiles à la conception des infrastructures et nécessaires pour assurer une meilleure gestion du risque climatique. L’ajustement par quantiles avec une approche non-paramétrique figure parmi les méthodes d’ajustement de biais les plus communément utilisées dans la littérature et par les centres de services climatiques [16]. Cette méthode fonctionne bien pour le cœur de la distribution où l’estimation non-paramétrique des quantiles est précise, mais elle présente des inconvénients majeurs lors du post-traitement des valeurs extrêmes où l’estimation non-paramétrique devient très imprécise [17]. Cela s’explique par le fait que pour les quantiles élevés, les méthodes non-paramétriques de QM appliquent des corrections imprécises en raison du bruit d’échantillonnage élevé [7].

Une méthode paramétrique est requise pour bien modéliser les extrêmes. De façon générale, les théories statistiques s’appliquant aux queues de la distribution diffèrent de celles s’appliquant au cœur de la distribution. La théorie des valeurs extrêmes (voir par exemple, [18]) est adaptée pour l’étude des queues de distribution tandis que la statistique des tendances centrales s’applique au cœur de la distribution. Ainsi, pour la modélisation paramétrique, la gamme des précipitations est artificiellement divisée en deux parties, soit les grandes intensités et le reste, ce qui impose nécessairement l’utilisation de modèles statistiques différents pour chaque partie. De plus, la sélection du seuil optimal pouvant distinguer correctement les intensités de précipitations faibles et fortes est difficile à faire en pratique.

Pour unifier la modélisation du cœur et des queues de la distribution, plusieurs familles de modèles statistiques ont été développées par [3] dans lesquelles les résultats de la théorie des valeurs extrêmes s’appliquent aux queues de la distribution. Ces distributions étendues pour les valeurs extrêmes permettent une transition en douceur entre les deux queues de la distribution, tout en gardant un faible nombre de paramètres. Les performances et la flexibilité de cette approche ont été vérifiées pour les précipitations enregistrées à Lyon, France.

L’objectif général du mémoire est de développer une méthode de post-traitement paramétrique adaptée pour les valeurs extrêmes et le cœur de la distribution. Cette méthode permettra de corriger les valeurs simulées par un modèle climatique.

L'objectif secondaire de ce projet est le développement d'une librairie performante offrant des outils pour utiliser les extensions de la loi de Pareto généralisée ainsi que la méthode de post-traitement développée. Au regard de cet objectif, le langage de programmation Julia [19] a été choisi. C'est un langage de haut niveau, performant et dynamique pour le calcul scientifique, qui permet notamment de contourner les difficultés logistiques que peut représenter le volume des ensembles de simulations climatiques. La librairie développée est accessible à l'adresse suivante : <https://github.com/houton199/EGPD.jl>.

Dans un premier temps, il faudra vérifier si la distribution des précipitations observées au Québec peut être modélisée par l'une des familles étendues de valeurs extrêmes. Une extension non stationnaire devra ensuite être développée afin d'incorporer les changements projetés des précipitations simulées par les modèles climatiques causés par les CC. Puis, la méthode de correspondance des quantiles paramétriques devra être établie pour compléter le post-traitement. Enfin, la méthode de post-traitement développée sera évaluée.

Ce mémoire est divisé de la manière suivante. Le chapitre 2 présente une revue de littérature. Le chapitre 3 dresse la base théorique nécessaire à la compréhension du projet de recherche. La théorie des valeurs extrêmes ainsi que les méthodes de post-traitement y sont expliquées. Le chapitre 4 présente les données et le chapitre 5, la méthodologie et les résultats. Ces derniers sont présentés et discutés pour un point de grille en particulier. Finalement, le chapitre 6 discute des résultats et le chapitre 7 contient la conclusion générale du projet de recherche et discute des possibles travaux futurs.

CHAPITRE 2 REVUE DE LITTÉRATURE

Ce chapitre présente une revue de littérature des méthodes de post-traitement statistique des modèles climatiques et de la modélisation des précipitations extrêmes en plus d’une brève présentation des modèles numériques de climat. La validation des méthodes de post-traitement est également introduite dans de ce chapitre.

2.1 Modèles numériques de climat

Les modèles globaux de climat (MGC) sont l’outil principal dont dispose la science pour comprendre les changements climatiques (CC) et leurs effets. Toutefois, la résolution spatiale effective des MGC est limitée par la puissance du calcul informatique. Par exemple, la résolution horizontale des MGC actuels est d’environ 200 km [8]. Conséquemment, les processus de petite échelle, soit temporelle ou spatiale, qui génèrent notamment les précipitations extrêmes tels que les orages ne peuvent être simulés avec les résolutions actuelles et sont plutôt approximatés par des paramétrages. Pour résoudre les équations liées à ces phénomènes météorologiques, la résolution des modèles climatiques doit être de moins de 4 km [20]. C’est le cas notamment des modèles permettant la convection (*convection-permitting models*, CPM) qui montrent des avancées prometteuses dans la simulation des précipitations extrêmes [21].

Afin d’accroître la résolution des simulations des MGC, des approches de réduction d’échelle (*downscaling*) ont été développées. On distingue ces approches en deux grandes familles : la réduction d’échelle dynamique basée sur les modèles régionaux de climat (MRC) et la réduction d’échelle statistique qui est plutôt une manière d’intégrer de l’information de plus haute résolution en utilisant des relations statistiques entre les variables simulées à l’échelle globale et le climat à l’échelle régionale ou locale.

Conceptuellement, un MRC est similaire à un MGC pour lequel le domaine de simulation est limité à une sous-région de la Terre, permettant ainsi d’augmenter la résolution spatiale. Toutefois, un MRC doit être piloté par un MGC ou par d’autres sources de données comme les réanalyses pour fournir les conditions aux frontières [11]. Cependant, la réduction d’échelle dynamique peut propager, accentuer ou compenser les biais contenus dans les données pilotes. De plus, les MRC peuvent contenir leurs propres biais. Par conséquent, même après une réduction d’échelle dynamique, les simulations climatiques sont en général biaisées par rapport aux observations et ne peuvent pas être utilisées directement pour les études d’impacts sans qu’une correction de biais soit appliquée [7].

Une revue des méthodes de réduction d'échelle dynamique et statistique pour les précipitations est présentée par Maraun *et al.* (2010) [22]. À titre d'exemple, la formulation d'un MRC, le Modèle Régional Canadien du Climat (MRCC) est présentée dans Laprise (2008) [11].

2.2 Post-traitement statistique

Les techniques de correction de biais font partie de ce qu'on appelle les techniques de post-traitement statistique (*Model Output Statistics*, MOS). Ces techniques utilisent une relation statistique entre une variable observée et une variable prédite par un modèle numérique pour que leurs distributions statistiques concordent. Elles doivent leur origine au domaine de la prévision numérique du temps. Une revue de ces méthodes est présentée dans Glahn et Lowry (1972) [23]. Klein et Glahn (1974) [24] discutent de la dérivation et l'application des MOS de façon opérationnelle pour certains éléments météorologiques tels que la précipitation, la température et les vents. En faisant correspondre les observations météorologiques locales aux sorties de modèles numériques, le biais et l'imprécision du modèle numérique ainsi que la climatologie locale peuvent être intégrés au système de prévision numérique du temps. Les modèles de prévision numérique du temps simulent des variables météorologiques à court terme qui peuvent être comparées directement aux observations une fois passé le moment de la prévision car il existe un synchronisme entre les simulations (prévisions) et les observations.

Avec le besoin grandissant d'analyser les impacts des CC, les MOS ont été adaptées à la modélisation climatique. Toutefois, en raison de la nature chaotique du système climatique, la séquence des événements météorologiques historiques observés ne peut être reproduite, même par un modèle climatique parfait [8]. Autrement dit, les simulations climatiques et les observations ne sont pas synchronisées. Les simulations climatiques sont des réalisations probables du climat dans le sens où elles visent à reproduire les propriétés statistiques des observations. Dans ce contexte, la calibration ne peut se faire qu'en fonction de la distribution.

L'article de Maraun (2016) [7] présente une revue des éléments conceptuels et de l'origine des techniques de post-traitement. Plusieurs techniques de correction de biais des modèles climatiques ont été développées au cours des dernières années. Le livre de Maraun et Widmann (2018) [4] fournit une référence complète des approches largement utilisées et Teuschbein et Seibert (2012) [2] présentent en détail six de ces techniques.

Il existe différentes philosophies du post-traitement. Certaines méthodes vont effectuer le post-traitement en appliquant une correction aux observations pour inclure le signal de CC estimé à partir des simulations alors que d'autres vont modifier les simulations. Par exemple, au lieu d'appliquer une correction de biais aux séries chronologiques simulées par un modèle

climatique, certaines approches modifient plutôt les données d’observation en utilisant le signal de changement climatique du modèle. C’est le cas pour la méthode delta, aussi appelée méthode de perturbation, dont se servent notamment Lenderink *et al.* (2007) [25]. Pour cette raison, certains auteurs (p. ex., Maraun et Widmann, 2018) ne considèrent pas cette méthode comme étant une méthode de correction de biais. Pour bien distinguer la méthode delta des méthodes de correction de biais proprement dites, ces dernières sont appelées méthodes directes parce qu’elles utilisent directement les séries temporelles simulées.

Parmi les méthodes directes, la technique de correction de biais la plus simple pour la précipitation est la mise à l’échelle (*scaling*) formulée par Widmann et Bretherton (2000) [26] où les précipitations simulées sont multipliées par le rapport entre la moyenne des précipitations observées et la moyenne des précipitations simulées (e.g., Lenderink *et al.* (2007) [25]). Cette technique suppose que le biais se retrouve uniquement sur la tendance centrale et non sur les moments supérieurs. Cependant, cela n’est pas toujours le cas car les biais des modèles climatiques affectent souvent non seulement la moyenne, mais aussi les moments d’ordre supérieur, en l’occurrence, la variance et l’aplatissement. Des méthodes de correction de biais plus flexibles sont donc nécessaires pour ajuster l’ensemble de la distribution du modèle pour mieux correspondre à la variance observée [2].

Une méthode corrigeant l’ensemble de la distribution est l’ajustement par quantiles (*Quantile Mapping*, QM) développée par Panofsky et Brier (1968) [15]. Pour une période de calibration, la fonction de répartition de la précipitation simulée est ajustée pour correspondre à la fonction de répartition de la précipitation observée.

Dans la littérature, il existe plusieurs formulations différentes pour l’ajustement par quantiles. Certains auteurs comme Themeßl *et al.* (2011) [27] considèrent les quantiles empiriques interpolés linéairement (*Empirical Quantile mapping*, EQM) alors que d’autres comme Piani *et al.* (2010) [28] font correspondre les quantiles de lois gamma ajustées sur les précipitations observées et simulées (*Parametric Quantile mapping*, PQM).

Par exemple, pour la correction par EQM, les fonctions de répartition empiriques sont construites pour les précipitations observées et simulées pour tous les jours d’un mois donné. Ensuite, la valeur des précipitations simulées pour le jour d du mois m est recherchée sur la fonction de répartition empirique des simulations avec sa probabilité cumulative correspondante. Puis, la valeur de précipitation de la même probabilité cumulative est localisée sur la fonction de répartition empirique des observations. Finalement, cette valeur est utilisée comme valeur corrigée pour la simulation.

Pour la correction des précipitations avec une méthode paramétrique qui utilise des lois gamma (*Gamma Quantile Mapping*, GQM), la procédure pour obtenir la valeur corrigée $\tilde{x}$

de la précipitation journalière x peut être exprimée de la façon suivante :

$$\tilde{x} = F_{Y_m}^{-1}\{F_{X_m}(x_d)\} \quad (2.1)$$

où F_{Y_m} est la fonction de répartition de la loi gamma de paramètres $(\alpha_{Y,m}, \beta_{Y,m})$ et où F_{X_m} est la fonction de répartition de la loi gamma de paramètres $(\alpha_{X,m}, \beta_{X,m})$. Cette procédure est illustrée par la figure 2.1.

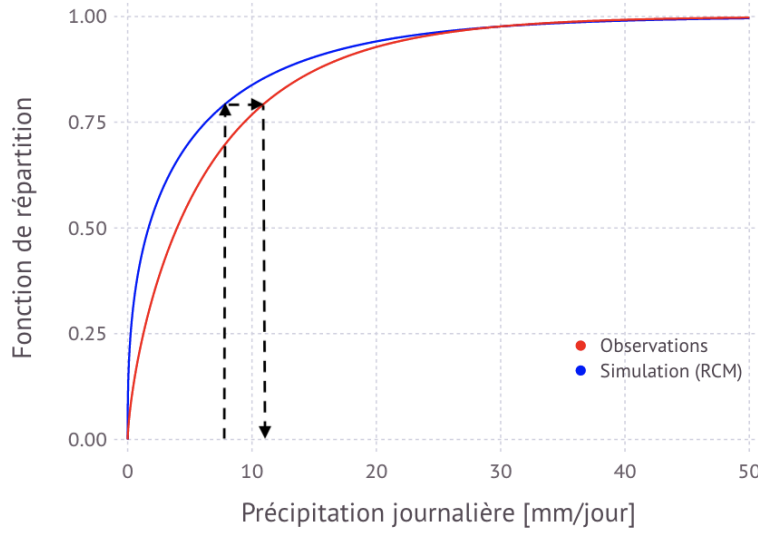


Figure 2.1 La distribution gamma d'une variable biaisée (en bleu) est mise en correspondance avec la gamma de référence (en rouge). La fonction de correction (en pointillé noir) dépend du quantile de référence. Figure adaptée de [2].

Michelangeli *et al.* (2009) [29] présentent une méthode de post-traitement non paramétrique permettant de tenir compte des changements entre la période de calibration et la période de projection dans la fonction de répartition au moyen d'une fonction de transfert $T(\bullet) : [0, 1] \rightarrow [0, 1]$. Leur méthode, *CDF-transform*, permet l'obtention de $F_{Y_{proj}}$, la fonction de répartition des observations pour une période projetée, c'est-à-dire qui n'a pas encore été observée, en fonction des simulations selon l'équation suivante :

$$F_{Y_{proj}}(x) = T\{F_{X_{proj}}(x)\} = F_{Y_{calib}}\left[F_{X_{calib}}^{-1}\{F_{X_{proj}}(x)\}\right] \quad (2.2)$$

où $T(\bullet) = F_{Y_{calib}}\{F_{X_{calib}}^{-1}(\bullet)\}$ est une fonction croissante et $F_{X_{proj}}$, $F_{X_{calib}}$ et $F_{Y_{calib}}$ sont des fonctions de répartition obtenues empiriquement.

2.3 Post-traitement des extrêmes

Étant donné que les extrêmes se situent dans les queues de distribution des variables et qu'il s'agit par définition d'événements rares, les méthodes actuelles de correction de biais ne sont pas adaptées [30]. Au cours des dernières années, quelques techniques de correction de biais conçues spécifiquement pour les extrêmes ont été développées. Cependant, selon Maraun *et al.* (2017) [31], les approches actuelles sont incapables de générer la variabilité quotidienne inexpliquée à une échelle de taille inférieure aux mailles des MRC et peuvent même introduire des artefacts, entre autres dans la représentation des précipitations extrêmes. Plus particulièrement, Maraun (2016) [7] explique que pour les quantiles élevés, les méthodes non-paramétriques de QM appliquent des corrections imprécises en raison du bruit d'échantillonnage élevé.

Parmi les techniques de correction de biais conçues spécifiquement pour les extrêmes, Kallache *et al.* (2011) [32] proposent une approche, *XCDF-t*, où les fonctions de répartition sont issues de la théorie des valeurs extrêmes. L'approche des auteurs se base sur l'article de Michelangeli *et al.* (2009) [29] pour relier la fonction de répartition de modèles de valeurs extrêmes ajustés sur les observations et la fonction de répartition de modèles de valeurs extrêmes ajustés sur les simulations. La même fonction de transfert est utilisée pour tenir compte des changements entre la période de calibration et la période de projection dans la fonction de répartition afin de prendre en compte les changements dans la distribution des extrêmes occasionnés par les CC. Cependant, plutôt que de considérer les fonctions de répartition empiriques, cette approche paramétrique modélise les excès de précipitations observées et simulées au-delà d'un seuil élevé avec des lois de Pareto généralisées (GP) en adéquation avec la théorie des valeurs extrêmes (voir section 3.1).

Les auteurs définissent également un modèle non stationnaire en reliant les paramètres de la loi GP à d'autres variables explicatives. En mode stationnaire, leur approche montre des bons résultats et surpasse l'approche *CDF-transform* non-paramétrique de Michelangeli *et al.* (2009) [29] et EQM pour certaines stations. Pour le cas non stationnaire, les auteurs montrent que le choix des variables explicatives doit être fait avec prudence car cela améliore les résultats dans certains cas seulement.

L'inconvénient majeur d'une telle technique est qu'elle ne fournit aucune information sur la partie centrale de la distribution car elle est conçue pour les valeurs extrêmes seulement. Pour y remédier, Gutjahr *et al.* (2013) [33] proposent une méthode paramétrique (*Generalized Pareto parametric quantile mapping*, GPQM) qui ajuste une loi Gamma à la précipitation en-dessous du quantile d'ordre 0.95 et une loi GP à celle au-dessus. L'article compare cette

technique avec deux autres techniques existantes, une empirique (EQM) et une paramétrique qui utilise des lois gamma (GQM). Selon les auteurs, la méthode empirique surpasse les deux méthodes paramétriques. Ils notent toutefois que la durée limitée de la série temporelle utilisée contient des valeurs aberrantes et que les trois techniques ont des difficultés à corriger les valeurs extrêmes.

Li *et al.* (2018) [34] étudient l'ajout du type d'événement de précipitations (p. ex., convectif ou non-convectif) à la correction de biais des précipitations extrêmes en identifiant des patrons météorologiques d'échelle synoptique (de l'ordre de 1000 km ou plus) à l'aide de cartes autoadaptatives. Cet ajout est évalué en comparant la distribution de la précipitation extrême future corrigée par leur technique avec celle corrigée par la technique conventionnelle d'ajustement par quantiles. Les auteurs concluent que la différence entre les changements obtenus par les deux méthodes de correction n'est pas statistiquement significative.

Une comparaison des méthodes de réduction d'échelle statistique, y compris le post-traitement statistique, par rapport aux événements extrêmes en Europe est présentée par Hertig *et al.* (2019) [30]. Les auteurs concluent que deux aspects affectent fortement la qualité des méthodes de post-traitement. Premièrement, l'inclusion d'une composante saisonnière améliore la performance à représenter les aspects marginaux des extrêmes. Deuxièmement, toutes les méthodes utilisant des distributions paramétriques nécessitent des lois non-classiques pour représenter correctement les aspects marginaux des extrêmes.

2.4 Validation des méthodes de post-traitement

La validation d'une méthode de post-traitement consiste essentiellement à comparer des indices climatiques calculés à partir des sorties du modèles à des indices de référence calculés à partir des observations et à quantifier l'adéquation à l'aide de mesures de performance appropriées [35]. Le terme « indice climatique » est utilisé de manière très générale de façon à inclure non seulement des scalaires mais aussi des vecteurs tels que des séries chronologiques [35]. Les moments statistiques comme la moyenne et l'écart-type peuvent également être vus comme des indices climatiques. Dans le contexte de la correction de biais, les indices sont calculés à partir des sorties de modèles post-traitées. Si on considère les précipitations comme une variable aléatoire issue d'une distribution multivariée représentant le système climatique, certaines mesures comme la moyenne, la variance ou les quantiles vont permettre de quantifier les aspects marginaux de cette distribution. Pour quantifier les aspects temporels, des mesures donnant de l'information sur la dépendance temporelle telles que l'autocorrélation ou la saisonnalité vont plutôt être utilisées [30].

Il existe un grand nombre d'études décrivant des méthodes ou des applications spécifiques de la réduction d'échelle statistique dans différentes régions du monde, utilisant différentes méthodologies de validation [36]. L'*Action* VALUE [35] du programme de coopération européenne dans le domaine de la recherche scientifique et technique (EU Cooperation in Science and Technology, COST) est la première initiative internationale visant à créer un cadre pour l'intercomparaison des méthodes de réduction d'échelle, y compris la correction de biais [36].

L'importance d'un tel cadre réside dans le fait qu'il permet aux utilisateurs de déterminer si une méthode particulière est appropriée pour leur application et de comparer les différentes méthodes existantes. Pour ce faire, VALUE propose un cadre de validation systématique permettant d'évaluer et de comparer les méthodes de réduction d'échelle dynamiques et statistiques [35] et des outils pour différents aspects de la validation [36].

Une validation de tous les aspects serait pratiquement impossible car on ne peut s'attendre à ce qu'un modèle climatique ou une méthode de réduction d'échelle reproduise parfaitement tous les aspects du système climatique [35]. Ainsi, la conception de l'expérience de validation, le choix des variables météorologiques et les aspects à valider doivent être guidés par les besoins des utilisateurs [35]. Une approche de validation axée sur l'utilisateur doit commencer par l'identification des phénomènes et des échelles d'intérêt pour ensuite identifier les principales forces et faiblesses d'une méthode par rapport à ceux-ci [35]. Afin de guider la sélection d'indices pertinents pour les différents problèmes des utilisateurs, VALUE a développé un arbre de décision pour la sélection d'indices et de mesures de performance pertinents [35]. La liste des indices et des mesures de performance correspondantes est disponible sur le site web de VALUE : www.value-cost.eu/indices.

Généralement, une validation cherchera à valider au moins les aspects marginaux et temporels [35]. Un test simple mais très efficace est la comparaison visuelle des séries chronologiques simulées et observées : une différence majeure dans les caractéristiques des séries chronologiques indique un problème de crédibilité ou de représentativité ou une perte de cohérence physique [35]. De même, un diagramme de quantiles entre les quantiles simulés et observés peut révéler une correction de biais non physique : là encore, si les distributions marginales simulées et observées non corrigées sont très différentes, il y a probablement un problème de crédibilité ou de représentativité [35]. Les deux tests doivent être effectués au moins pour certains sites sélectionnés.

Hertig *et al.* (2019) [30] comparent différentes techniques statistiques de réduction d'échelle par rapport aux événements extrêmes en Europe. Pour évaluer les aspects marginaux des précipitations, les indices suivants sont analysés : l'asymétrie, la fréquence relative des jours avec des précipitations supérieures à 10 mm (*R10*), le quantile d'ordre 0,98 des jours humides

en considérant les jours avec des précipitations supérieures à 1 mm ($P98Wet$), la quantité des précipitations des jours supérieures au quantile d'ordre 0,98 ($P98WetAmount$), et la valeur de retour sur 20 ans des fortes précipitations ($RV20max$). Pour ce qui est des propriétés statistiques temporelles des précipitations, les statistiques sur la durée des périodes humides ($WetAnnualMaxSpell$) et sèches ($DryAnnualMaxSpell$) sont analysées. La performance est évaluée en regardant la différence entre la prédiction et les observations pour chacun de ces indices.

Casanueva *et al.* (2020) [37] présentent une étude d'intercomparaison de huit techniques d'ajustement de biais avec et sans préservation des tendances de CC pour la température et la précipitation dans le but d'évaluer leurs effets sur le signal brut (provenant des modèles globaux et régionaux) du CC. La comparaison est effectuée en évaluant un certain nombre d'indices de validation englobant les aspects marginaux, temporels et extrêmes. Les auteurs précisent que, par construction, certains indices (p. ex., la température moyenne, les percentiles) sont directement ajustés par certaines méthodes, ce qui doit être pris en compte dans le cadre d'une comparaison équitable [37]. Pour la précipitation, ils font l'analyse des indices suivants : la fréquence relative des jours avec des précipitations supérieures à 1 mm ($R01$), la précipitation moyenne ($SDII$), la médiane des maxima annuels des périodes sèches ($DryAnnualMaxSpell$), la probabilité de transition sèche-humide ($DWProb$) et le quantile d'ordre 0,98 des jours humides en considérant les jours avec des précipitations supérieures à 1 mm ($P98Wet$).

2.5 Modélisation statistique des précipitations extrêmes

Les méthodes paramétriques pour le post-traitement des précipitations nécessitent une loi de probabilité. Pour modéliser le cœur et les queues de distribution, différentes lois sont généralement utilisées. Ainsi, les mélanges de lois s'avèrent utiles. Vrac et Naveau (2007) [38] proposent un mélange de loi gamma et de loi GP pour modéliser les précipitations faibles, modérées et intenses. Les auteurs soulignent l'importance de modèles de valeurs extrêmes pour améliorer la modélisation et la mise à l'échelle des précipitations extrêmes. Carreau et Bengio (2009) [39] présentent un estimateur de densité non paramétrique qui assemble les avantages de l'estimation non paramétrique pour le cœur avec la théorie des valeurs extrêmes pour la queue à droite. Une loi de Pareto généralisée hybride qui peut être utilisée dans un modèle de mélange pour étendre la loi de Pareto généralisée aux réels positifs est formulée. Carreau et Vrac (2011) [40] présentent une nouvelle classe de modèles stochastiques de mise à l'échelle, les modèles de mélange conditionnel basés sur un modèle de réseau de neurones. Une composante discrète est incluse dans le mélange pour prendre en charge les jours sans pluie.

Les auteurs modélisent la pluie avec un mélange de lois continues, soit des lois normales, log-normales ou une loi de Pareto généralisée hybride. La flexibilité additionnelle des mélanges de lois permet de mieux représenter la distribution de la pluie dans la partie centrale et la queue supérieure. Cependant, certains inconvénients font en sorte que ces mélanges de lois sont mal adaptés pour la précipitation [3]. D’abord, comme la partie basse de la distribution est basée sur une loi normale, elle peut produire des valeurs négatives. Ensuite, pour faire l’union des deux lois, une forte contrainte est imposée aux paramètres de chacune d’entre elles et le modèle final peut s’avérer être peu flexible.

La sélection du seuil optimal pouvant distinguer correctement les intensités de précipitations faibles et fortes est difficile à faire en pratique. Papastathopoulos et Tawn (2013) [41] ont proposé des extensions à la loi de Pareto généralisée qui incorporent un paramètre de forme additionnel sans affecter le comportement de la queue à droite. L’inclusion de ce paramètre offre une structure additionnelle permettant de modéliser les excédents au-dessus d’un seuil moins élevé. Cette approche diminue la sensibilité au choix du seuil de l’estimation du paramètre d’échelle et du paramètre de queue et des niveaux de retours et permet la sélection d’un seuil moins élevé. Naveau *et al.* (2016) [3] ont proposé quatre extensions aux lois de Papastathopoulos et Tawn pour l’ensemble de la distribution, sans choix de seuil. Leurs modèles sont en adéquation avec la théorie des valeurs extrêmes dans les deux extrémités du spectre et permettent une transition en douceur entre les deux queues pour le cœur de la distribution tout en gardant un faible nombre de paramètres. De plus, le choix d’un seuil n’est pas requis pour définir les excès faibles et élevés.

Ces quatre modèles proposés seront utilisés dans ce mémoire et adaptés au contexte du post-traitement. Il sera alors possible de développer une méthode d’ajustement de biais paramétrique adéquate autant pour le cœur de la distribution que pour les queues sans avoir à définir de seuil.

CHAPITRE 3 CADRE THÉORIQUE

Cette section introduit les bases théoriques essentielles à la compréhension du projet. La première section de ce chapitre introduit les bases de la théorie des valeurs extrêmes. L'essentiel y est présenté dans le but de développer la méthode de post-traitement. Pour plus de détails, un lecteur intéressé à en apprendre davantage peut se référer au livre de Coles [18]. La seconde section présente l'indice utilisé pour la sélection de modèle et la troisième, le post-traitement statistique des simulations climatiques. Des méthodes de correction de biais sont présentées, notamment la correction par quantiles, qui constitue la base de la méthode développée dans ce projet de maîtrise. Il existe plusieurs implémentations de la méthode de correction par quantiles (p. ex., *EQM*, *PQM*, *GPQM*). Celles-ci diffèrent notamment quant au modèle utilisé pour estimer les fonctions de répartition. En particulier, pour les queues de distribution, il est possible d'utiliser des modèles de valeurs extrêmes. Pour plus de détails à propos du post-traitement statistique, le livre de Maraun et Widmann [4] est suggéré. Finalement, la dernière section présente les indices climatiques utilisés pour évaluer la méthode de post-traitement développée dans ce mémoire.

3.1 Théorie des valeurs extrêmes

Cette section contient les bases de la théorie des valeurs extrêmes, la discipline de la statistique qui concerne spécifiquement les valeurs dans les queues de distribution. Elle est souvent utilisée en actuariat et en hydrologie pour estimer le risque associé aux événements extrêmes.

3.1.1 Loi limite du maximum

Soit $X_1, \dots, X_n$ une séquence de variables aléatoires indépendantes et identiquement distribuées (i.i.d.) ayant comme fonction de répartition F . Le maximum des n premières variables aléatoires est défini comme suit :

$$M_n = \max\{X_1, \dots, X_n\} \quad (3.1)$$

La distribution de ce maximum peut être dérivée de la façon suivante :

$$P\{M_n \leq z\} = P\{X_1 \leq z, \dots, X_n \leq z\} = P\{X_1 \leq z\} \times \dots \times P\{X_n \leq z\} = F^n(z) \quad (3.2)$$

Cependant, en pratique, la distribution de F est inconnue et il faut trouver une façon de l'estimer. Une façon de faire serait d'utiliser les techniques statistiques usuelles pour estimer F à partir des données observées, puis de substituer cette estimation dans l'équation (2) pour obtenir la loi du maximum. Toutefois, l'incertitude associée à l'estimation de F peut devenir beaucoup trop grande lors de l'estimation de F^n .

L'approche alternative consiste à accepter que la distribution de F soit inconnue et à estimer F^n directement à l'aide d'une distribution limite. Avec cette dernière méthode, une difficulté survient lors de l'analyse du comportement asymptotique de $F^n(z)$. En effet, comme $F^n(z) \rightarrow 0$ lorsque $n \rightarrow \infty$, la loi asymptotique de M_n est dite dégénérée. Une transformation linéaire de M_n doit être effectuée pour contourner cette difficulté. Cette transformation est présentée dans le théorème 3.1.1 (chap. 3, [18]).

Théorème 3.1.1

S'il existe des suites de constantes $\{a_n > 0\}$ et $\{b_n\}$ telles que

$$P\left\{\frac{M_n - b_n}{a_n} \leq z\right\} \rightarrow G(z) \quad (3.3)$$

pour une distribution G non dégénérée, alors G possède la fonction de répartition suivante

$$G(z) = \begin{cases} \exp\left\{-\left[1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} & \text{si } \xi \neq 0 \\ \exp\left\{-\exp\left(\frac{z-\mu}{\sigma}\right)\right\} & \text{si } \xi = 0 \end{cases} \quad (3.4)$$

définie sur $\{z : 1 + \frac{\xi(z-\mu)}{\sigma} > 0\}$, où $-\infty < \mu < \infty, \sigma > 0$ et $-\infty < \xi < \infty$. Cette loi est appelée la loi des valeurs extrêmes généralisées (GEV).

Pour déduire le comportement asymptotique de M_n , on a besoin des constantes $\{a_n > 0\}$ et $\{b_n\}$ qui sont inconnues en pratique. Cependant, tel qu'illustré dans la relation (3.5), ce n'est pas un problème puisque ces constantes peuvent être absorbées dans les paramètres à estimer :

$$\begin{aligned}
P\{M_n \leq z\} &= P\left\{\frac{M_n - b_n}{a_n} \leq \frac{z - b_n}{a_n}\right\} \\
&= G\left(\frac{z - b_n}{a_n}\right) \\
&= G^*(z)
\end{aligned} \tag{3.5}$$

où G^* est aussi un membre de la famille GEV [18]. Estimer la distribution d'une transformation linéaire de M_n revient donc à estimer la distribution de M_n directement :

$$\text{si } M_n \sim GEV(\mu, \sigma, \xi), \text{ alors } \frac{M_n - b_n}{a_n} \sim GEV\left(\frac{\mu - b_n}{a_n}, \frac{\sigma}{a_n}, \xi\right). \tag{3.6}$$

Le théorème est toujours valide dans le cas dépendant (voir [18] ; chap. 5) à condition que cette dépendance dans la suite soit à courte portée. La modélisation statistique consiste donc à estimer la loi GEV à partir des maxima annuels.

3.1.2 Modélisation statistique des excès au-dessus d'un seuil

Dans la théorie des valeurs extrêmes, une méthode souvent utilisée pour la modélisation est la méthode des excès. Elle consiste à utiliser un seuil pour isoler les valeurs considérées extrêmes par rapport aux autres valeurs de l'ensemble des données puis à créer un modèle pour les valeurs extrêmes en modélisant la queue de distribution de toutes les valeurs qui excèdent ce seuil.

Théorème 3.1.2

Soit $X_1, \dots, X_n$ une séquence de variables aléatoires indépendantes et identiquement distribuées (i.i.d.) ayant comme fonction de répartition F . Soit X_i un terme arbitraire dans la séquence et supposons que F satisfasse le Théorème 3.1.1 de sorte que pour $n \rightarrow \infty$,

$$Pr\{M_n \leq z\} \approx G(z), \tag{3.7}$$

où G est la loi GEV de paramètres $\mu, \sigma > 0$ et ξ . Alors, pour un u suffisamment grand, la fonction de répartition de l'excédent $(X - u)$, conditionnelle à $X > u$, est approximativement

$$H(y) \approx 1 - \left(1 + \frac{\xi y}{\tilde{\sigma}}\right)^{-1/\xi} \tag{3.8}$$

définie sur $\{y : y > 0 \text{ et } (1 + \xi y/\tilde{\sigma}) > 0\}$. Cette loi est appelée la loi GPD.

Dans ce résultat, le paramètre de forme ξ est le même que pour celui de la loi GEV du maximum M_n . Cependant, le paramètre d'échelle $\tilde{\sigma}$ est variable et dépend du seuil u selon l'équation

$$\tilde{\sigma} = \sigma + \xi(u - \mu). \quad (3.9)$$

Le seuil u doit être suffisamment élevé pour que l'approximation (3.8) soit raisonnable.

En pratique, le choix du seuil est critique : s'il est trop bas, l'équation (3.9) n'est pas valide, menant à un grand biais. Inversement, s'il est trop haut, il y a peu de données, menant à une grande variance.

3.1.3 Extension de la loi de Pareto généralisée

La loi de Pareto généralisée peut être étendue en utilisant la méthode de la transformée inverse. Soit U une variable aléatoire uniformément distribuée sur $[0, 1]$ et H_ξ^{-1} la fonction de répartition inverse de la GPD de paramètre $(1, \xi)$. Une réalisation de cette loi peut être obtenue avec

$$\sigma H_\xi^{-1}(u). \quad (3.10)$$

où u est une réalisation de la loi uniforme.

[41] ont proposé de remplacer la loi uniforme U dans l'équation précédente par $V = G^{-1}(U)$, où G est une fonction de répartition continue sur $[0, 1]$. En respectant certaines conditions pour G afin de s'assurer d'avoir le même comportement asymptotique, il est possible d'étendre la loi de Pareto généralisée à des seuils plus bas et de couvrir une famille plus riche de variables aléatoires en définissant

$$X = \sigma H_\xi^{-1} G^{-1}(U). \quad (3.11)$$

Différents choix de fonctions G sont proposés par [41] et mènent à des extensions de la distribution de Pareto généralisée permettant d'utiliser un seuil sous-asymptotique. Ces choix de fonctions G améliorent l'estimation du paramètre de queue supérieure, notamment dans les cas où trop peu de valeurs sont observées conduisant à presque aucun excès au-dessus du seuil sélectionné. Cependant, la sélection d'un seuil demeure nécessaire. Cela présente au moins deux inconvénients majeurs. Premièrement, les observations en-dessous du seuil sont complè-

tement ignorées. La gamme des précipitations est artificiellement divisée en deux parties, soit les grandes intensités et le reste, ce qui impose nécessairement l'utilisation de modèles statistiques différents pour chaque partie. Deuxièmement, la sélection du seuil optimal pouvant distinguer correctement les intensités de précipitations faibles et fortes est difficile à faire en pratique.

Pour pallier ces inconvénients, [3] proposent un modèle statistique dans lequel les résultats de la théorie des valeurs extrêmes s'appliquent non seulement aux fortes quantités de précipitations, mais aussi aux faibles, à l'exclusion des zéros. Ce modèle permet une transition en douceur entre les deux queues de distribution, tout en conservant un faible nombre de paramètres. Leur objectif est différent de celui de [41] dans le sens où ils veulent modéliser de manière adéquate toute la gamme des précipitations, et pas seulement améliorer l'inférence pour les quantiles élevés.

Cela exige de trouver une classe de distributions G dans (3.11) telle que le comportement de la queue supérieure avec le paramètre de forme ξ soit préservé et que la fonction de répartition de X pour les valeurs proches de zéro se comporte comme x^κ afin de modéliser correctement les faibles intensités de précipitations. [3] présentent trois conditions pour lesquelles ces contraintes peuvent être remplies. Quatre familles paramétriques possibles pour $G(v)$, $v \in [0, 1]$, remplissant ces trois conditions y sont présentées :

Famille 1

La famille 1 avec la loi de puissance $G(v) = v^\kappa$ est le choix le plus simple. Elle conduit à un modèle avec trois paramètres : κ contrôle la forme de la queue inférieure, σ est un paramètre d'échelle et ξ contrôle le taux de décroissance de la queue supérieure. La fonction de répartition de la loi EGPD correspond à :

$$F(x) = \left\{ H_\xi \left(\frac{x}{\sigma} \right) \right\}^\kappa \quad (3.12)$$

La figure 3.1 présente la fonction de densité correspondant à ce modèle pour différentes valeurs de paramètres.

Famille 2

La famille 2 est un mélange de lois de puissance $G(v) = pv^{\kappa_1} + (1 - p)v^{\kappa_2}$. Elle conduit à un modèle avec cinq paramètres : κ_1 contrôle le comportement de la queue inférieure alors que κ_2 modifie la forme de la densité dans la partie centrale. C'est une façon d'accroître la

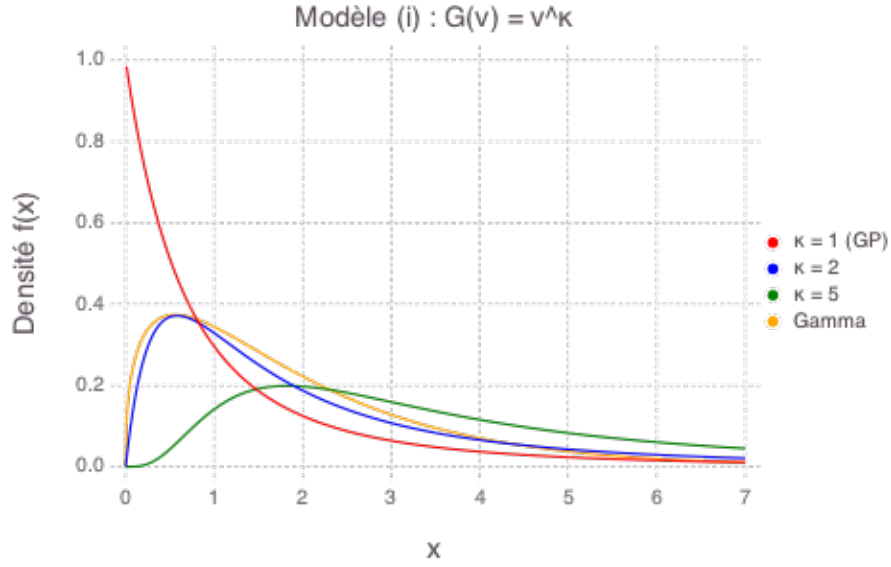


Figure 3.1 Fonction de densité correspondant au modèle 1 pour $\sigma = 1$, $\xi = 0.5$ et $\kappa = 1, 2, 5$ (rouge, bleu, vert, respectivement). Le cas $\kappa = 1$ correspond à une densité GP. La courbe en orange représente une densité $Gamma(1.4, 1.4)$. Figure adaptée de [3].

flexibilité des modèles de la famille 1. La fonction de répartition correspond à :

$$F(x) = p \left\{ H_{\xi} \left(\frac{x}{\sigma} \right) \right\}^{\kappa_1} + (1 - p) \left\{ H_{\xi} \left(\frac{x}{\sigma} \right) \right\}^{\kappa_2} \quad (3.13)$$

La figure 3.2 présente la fonction de densité correspondant à ce modèle pour différentes valeurs de paramètres.

Famille 3

La famille 3 avec $G(v) = 1 - Q_{\delta}\{(1 - v)^{\delta}\}$ où Q_{δ} est la fonction de répartition d'une variable aléatoire beta de paramètres $1/\delta$ et 2,

$$Q_{\delta}(v) = \frac{1 + \delta}{\delta} v^{1/\delta} \left(1 - \frac{v}{1 + \delta} \right), \quad (3.14)$$

conduit à un modèle avec trois paramètres : σ est un paramètre d'échelle, ξ contrôle la queue supérieure et δ pourrait être interprété comme un paramètre de réglage du seuil qui doit être estimé à partir des données disponibles. Ce paramètre augmente la flexibilité de la modélisation pour la partie centrale de la distribution. Pour cette famille, le comportement de la queue inférieure n'est pas estimé à partir des données mais imposé par le choix du

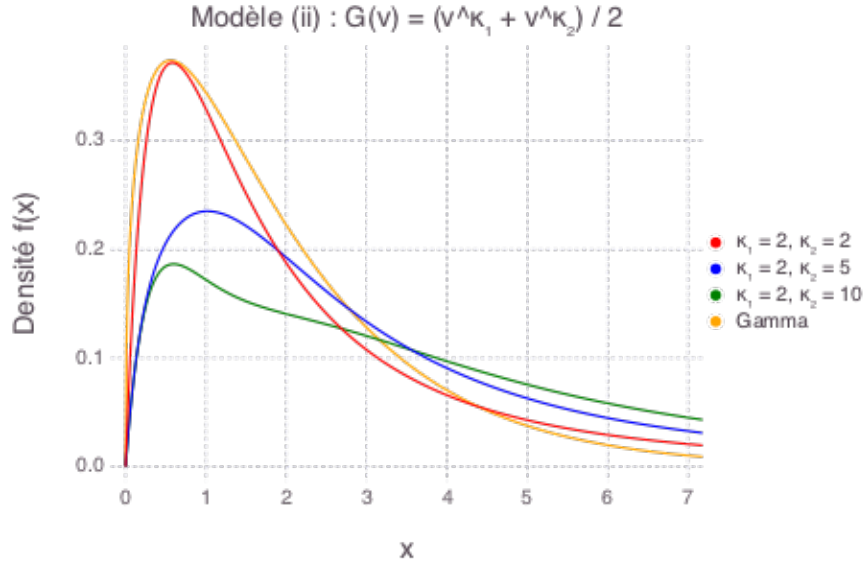


Figure 3.2 Fonction de densité correspondant au modèle 2 pour $p = 0.5$, $\sigma = 1$, $\xi = 0.5$, $\kappa_1 = 2$, et $\kappa_2 = 2, 5, 10$ (rouge, bleu, vert, respectivement). Le cas $\kappa = 1$ correspond à une densité GP. La courbe en orange représente une densité $Gamma(1.4, 1.4)$. Figure adaptée de [3].

modèle. La fonction de répartition est la suivante :

$$F(x) = 1 - Q_\delta \left[\left\{ 1 - H_\xi \left(\frac{x}{\sigma} \right) \right\}^\delta \right] \quad (3.15)$$

La figure 3.3 présente la fonction de densité correspondant à ce modèle pour différentes valeurs de paramètres.

Famille 4

La famille 4, $G(v) = [1 - Q_\delta\{(1 - v)^\delta\}]^{\kappa/2}$, remédie au problème que pose l'imposition du comportement de la queue inférieure par le choix du modèle de la famille 3 en ajoutant un paramètre supplémentaire contrôlant le comportement de la queue inférieure, $\kappa > 0$. Cela conduit à un modèle avec quatre paramètres : κ , δ , ξ décrivent respectivement les parties inférieures, centrales et supérieures de la distribution, et σ est un paramètre d'échelle. En particulier, les queues inférieure et supérieure sont, par construction, des lois de Pareto généralisées avec respectivement les paramètres de forme κ et ξ . La fonction de répartition est la suivante :

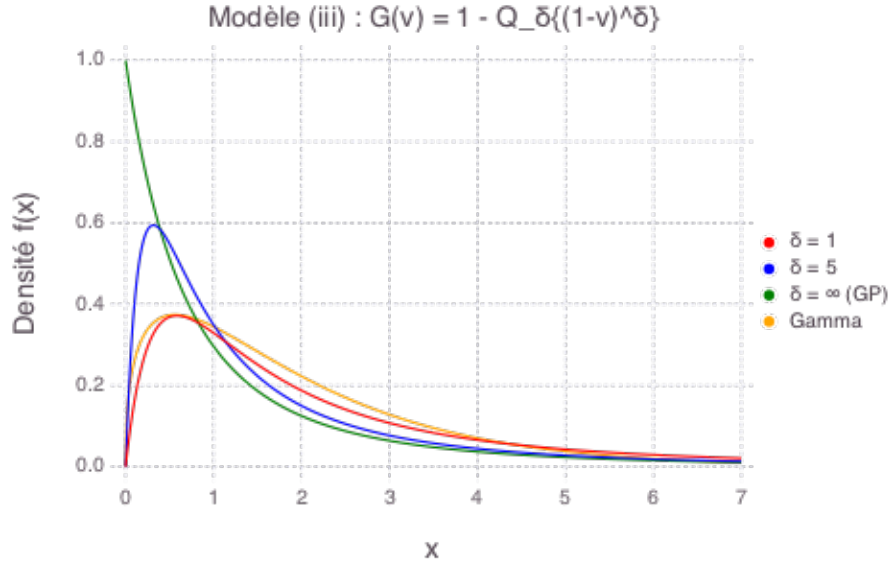


Figure 3.3 Fonction de densité correspondant au modèle 3 pour $\sigma = 1$, $\xi = 0.5$, et $\delta = 1, 5, \infty$ (rouge, bleu, vert, respectivement). Le cas $\delta = \infty$ correspond à une densité GP. La courbe en orange représente une densité $Gamma(1.4, 1.4)$. Figure adaptée de [3].

$$F(x) = \left(1 - Q_{\delta} \left[\left\{ 1 - H_{\xi} \left(\frac{x}{\sigma} \right) \right\}^{\delta} \right] \right)^{\kappa/2} \quad (3.16)$$

La figure 3.4 présente la fonction de densité correspondant à ce modèle pour différentes valeurs de paramètres.

3.2 Sélection de modèle

Afin de faire la sélection du modèle le mieux adapté aux données, les différents modèles peuvent être comparés à l'aide du critère d'information bayésien (BIC). La procédure usuelle de sélection consiste à choisir le modèle qui possède le plus grand BIC.

3.2.1 Critère d'information bayésien

Le BIC est un indice développé par Schwarz en 1978 souvent utilisé dans la littérature pour la sélection de modèles. Il est défini de la façon suivante :

$$BIC = \ln f_{(Y|\hat{\theta}_{MV})}(y) - \frac{k}{2} \ln n \quad (3.17)$$

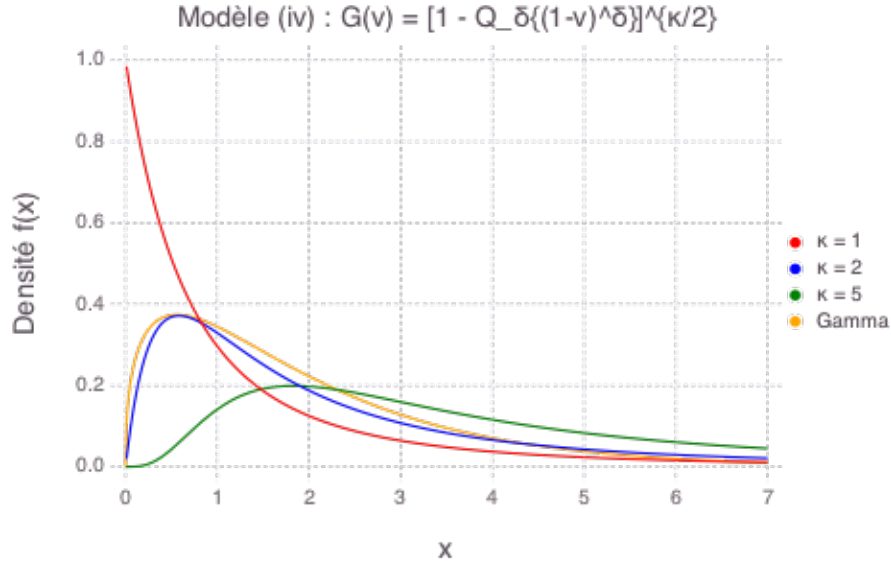


Figure 3.4 Fonction de densité correspondant au modèle 3 pour $\sigma = 1$, $\xi = 0.5$, $\delta = 1$ et $\kappa = 1, 2, 5$ (rouge, bleu, vert, respectivement). Le cas $\delta = \infty$ correspond à une densité GP. La courbe en orange représente une densité $Gamma(1.4, 1.4)$.

où $\hat{\theta}_{MV}$ dénote l'estimateur du maximum de la vraisemblance de θ et k le nombre de paramètres du modèle. Ce critère pénalise les modèles plus complexes en tenant compte du nombre de paramètres.

Pour comparer les modèles entre eux, le logarithme du facteur de Bayes entre les modèles M_2 et M_1 peut être approximé par

$$\ln(B_{21}) \approx \text{BIC}_2 - \text{BIC}_1, \quad (3.18)$$

où BIC_j correspond au BIC du modèle M_j .

Dans un contexte où le choix du modèle entraîne d'importantes conséquences, le tableau 3.1 d'interprétation du facteur de Bayes proposée par Kass et Raftery (1995) [1] s'avère utile.

3.3 Méthodes de post-traitement statistique

Le post-traitement statistique vise à corriger le biais d'une variable aléatoire par rapport à une référence. Pour les simulations climatiques, cela consiste à corriger les séries temporelles simulées pour qu'elles correspondent aux statistiques observées aux stations météorologiques correspondant au même emplacement géographique ou à d'autres produits de référence tels

Tableau 3.1 Interprétation du facteur de Bayes. Adapté de [1].

Facteur de Bayes	Évidence en faveur de M_2
$0 < \ln(B_{21}) \leq 1/2$	faible
$1/2 < \ln(B_{21}) \leq 1$	substantielle
$1 < \ln(B_{21}) \leq 2$	forte
$\ln(B_{21}) > 2$	décisive

que les données sur grille (p. ex., les réanalyses) ou les données de télédétection. Par conséquent, dans le cas où les séries temporelles simulées et observées présentent des différences d'échelle spatio-temporelle, cette correction implique également un ajustement d'échelle.

Il existe plusieurs méthodes de post-traitement statistique. La plus simple est la méthode delta, aussi appelée méthode de perturbation. Cette approche modifie les données d'observation par un signal de changement climatique plutôt que d'appliquer une correction de biais aux données du modèle. Pour cette raison, certains auteurs ne considèrent pas la méthode delta comme une méthode de correction de biais [4].

Pour distinguer la méthode delta des méthodes de correction de biais proprement dites, ces dernières sont souvent appelées méthodes directes parce qu'elles utilisent directement les séries temporelles simulées [7]. On peut donc distinguer les méthodes d'ajustement de biais en deux catégories : la méthode delta et les méthodes directes.

Parmi les méthodes directes, les méthodes de correction de biais les plus simples sont la correction additive et la mise à l'échelle (*scaling*) [4]. La correction additive corrige le biais d'une série future simulée en soustrayant le biais du modèle pour la période actuelle [26] alors que la mise à l'échelle corrige le biais d'une variable en multipliant la série simulée par le rapport entre la moyenne des observations et la moyenne des simulations [25]. Avec ces méthodes, les moments d'ordre supérieur tels que la variance et l'asymétrie sont supposés sans biais. Cependant, en pratique, cela n'est pas toujours le cas puisque les biais des modèles climatiques n'affectent souvent pas seulement la moyenne, mais aussi les moments statistiques d'ordre supérieur. Des méthodes de post-traitement plus flexibles sont donc nécessaires pour ajuster la variance de la distribution du modèle pour mieux correspondre à la variance observée. Une généralisation de ces approches plus flexibles est l'ajustement par quantiles avec une approche non-paramétrique (quantile mapping, QM).

3.3.1 Méthode de correction par quantiles

La méthode de correction par quantiles consiste à faire correspondre la fonction de répartition de la variable aléatoire à celle de la variable aléatoire de référence. Dans une variante largement utilisée, un quantile de la distribution d'une variable explicative biaisée pour une période de calibration est remplacé par le même quantile de la distribution de référence pour la même période de calibration [7] :

$$F_Y(\tilde{x}_i) = F_X(x_i) \quad (3.19)$$

où F_Y est la fonction de répartition de la variable de référence, F_X la fonction de répartition de la variable biaisée et x_i la valeur à corriger.

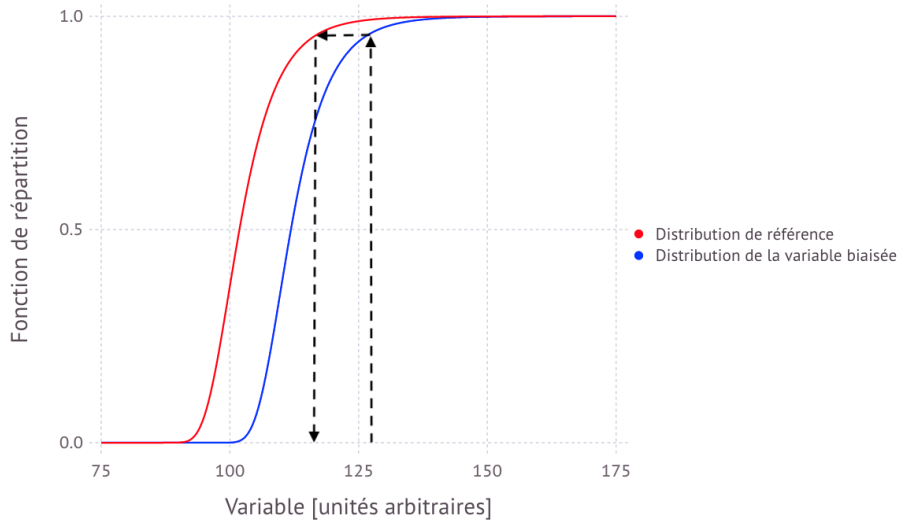


Figure 3.5 La fonction de répartition d'une variable biaisée (en bleu) est mise en correspondance avec la fonction de répartition de référence (en rouge). La fonction de correction (en pointillé noir) dépend du quantile de référence. Figure adaptée de [4].

La figure 3.5 présente une représentation simplifiée de cette méthode. La courbe en rouge représente la distribution d'une variable de référence, la loi GEV avec $\mu = 100$, $\sigma = 5$, $\xi = 0.05$. La courbe en bleu représente la distribution d'une variable biaisée, la loi GEV avec $\mu = 110$, $\sigma = 5$, $\xi = 0.05$, le paramètre μ se trouve à être biaisé d'une valeur de 10 par rapport à la distribution de référence. Il est important de noter que pour cet exemple, le biais se retrouve uniquement sur la valeur du paramètre μ . Cependant, la correction fonctionne aussi lorsque l'échelle σ et la forme ξ sont biaisées.

Pour corriger le biais, on remplace un quantile de la distribution biaisée (en bleu) par le

quantile de la distribution de référence (en rouge) correspondant à la même probabilité. Ce cas de figure assume qu'on connaît les distributions de la variable de référence et de la variable biaisée. En pratique, ces distributions doivent être obtenues de manière empirique ou en ajustant des modèles paramétriques.

La correction de biais appliquée aux simulations climatiques s'effectue de la manière suivante. À partir de toutes les valeurs de la variable explicative simulées par un modèle de climat pour une période de calibration, $\mathbf{x} = (x_i : 1 \leq i \leq n)$, et des valeurs de référence observées à une station météorologique pour la même période de calibration, $\mathbf{y} = (y_i : 1 \leq i \leq n)$, les fonctions de répartition correspondantes F_X et F_Y sont estimées. Les valeurs corrigées de la variable simulée $\tilde{\mathbf{x}} = (\tilde{x}_i : 1 \leq i \leq n)$ sont obtenues en appliquant la transformation

$$\tilde{x}_i = F_Y^{-1} \left\{ F_X(x_i) \right\} \quad (3.20)$$

où $\tilde{x}_i$ est la valeur corrigée, X la simulation et Y la référence.

Cette implémentation de la correction par quantiles suppose une fonction de transfert constante, le biais d'une variable simulée ne dépend que de sa valeur et non de sa position dans le temps. Toutefois, des techniques existent pour tenir compte des tendances. Par exemple, il est possible de soustraire la tendance de manière polynomiale et appliquer la correction de biais sur les résidus.

3.3.2 Méthode de la CDF-*transform*

Pour prendre en compte la non-stationnarité induite par les changements climatiques qui peuvent se manifester par un changement des fonctions de répartition, [29] proposent une extension à la méthode d'ajustement par quantiles.

Soit $F_{Y_{calib}}$ la fonction de répartition empirique des données observées pour la période de calibration et $F_{X_{calib}}$ la fonction de répartition empirique de la variable simulée pour la même période de calibration. Également, posons $F_{X_{proj}}$ la fonction de répartition empirique de la variable simulée pour la période de projection (future). Le but de la méthode de la CDF-*transform* est d'obtenir $F_{Y_{proj}}$ à partir des trois autres fonctions de répartition au moyen d'une fonction de transfert $T(\bullet) : [0, 1] \rightarrow [0, 1]$

$$F_{Y_{proj}}(x) = T \left\{ F_{X_{proj}}(x) \right\} = F_{Y_{calib}} \left[F_{X_{calib}}^{-1} \left\{ F_{X_{proj}}(x) \right\} \right] \quad (3.21)$$

où $T(\bullet) = F_{Y_{calib}} \left\{ F_{X_{calib}}^{-1}(\bullet) \right\}$ est une fonction croissante.

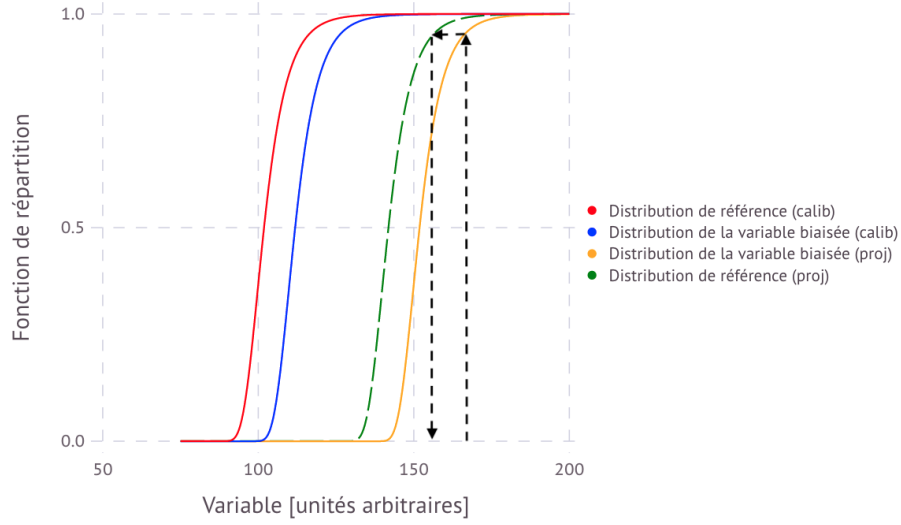


Figure 3.6 Pour la période de calibration, la fonction de répartition d'une variable biaisée (en bleu) est mise en correspondance avec la fonction de répartition de référence (en rouge). Pour la période de projection, la fonction de répartition d'une variable biaisée (en orange) est mise en correspondance avec la fonction de répartition de référence obtenue par la *CDF-transform* (en pointillé vert). Les fonctions de correction (en pointillé noir) dépendent du quantile de référence.

La figure 3.6 présente une représentation simplifiée de cette méthode avec des lois GEV plutôt que des distributions empiriques. La courbe en rouge représente la distribution d'une variable de référence pour la période de calibration, la loi GEV avec $\mu = 100$, $\sigma = 5$ et $\xi = 0.05$. La courbe en bleu représente la distribution d'une variable explicative simulée pour la même période de calibration, la loi GEV avec $\mu = 110$, $\sigma = 5$ et $\xi = 0.05$, le paramètre μ se trouve à être biaisé d'une valeur de 10 par rapport à la distribution de référence. La courbe en orange représente la distribution d'une variable explicative simulée pour la période de projection, la loi GEV($\mu = 150, \sigma = 5, \xi = 0.05$). Dans ce cas-ci, le biais de la distribution de la variable explicative simulée ne peut pas être corrigé en remplaçant un quantile d'intérêt par le quantile de la distribution de référence puisqu'il n'y a pas d'observations pour la période de projection. La méthode de la *CDF-transform* nous permet d'obtenir une fonction de répartition projetée pour la variable de référence. La courbe en pointillé vert représente la fonction de répartition projetée pour la variable de référence obtenue en appliquant la transformation $T(\bullet)$.

3.3.3 Méthode de la *CDF-transform* pour les extrêmes non stationnaires

La méthode de la *CDF-transform* est reprise par [32] en tenant compte des résultats de la théorie des valeurs extrêmes. Plutôt que d'utiliser des fonctions de répartition empiriques

dans (3.21), des lois paramétriques adaptées aux valeurs extrêmes sont utilisées. Plus particulièrement, [32] considèrent les extrêmes comme les excès d'un seuil pouvant être modélisés par une loi de Pareto généralisée (voir section 3.1.2). En procédant ainsi, une méthode de *CDF-transform* adaptée spécifiquement aux extrêmes est développée, *XCDF-t*.

Avec cette méthode, des variables explicatives (p. ex., la température, la pression atmosphérique, la concentration en CO_2) permettant la définition d'un modèle non stationnaire peuvent être ajoutées. Le développement d'un modèle non stationnaire offre la possibilité d'ajuster un seul modèle pour l'ensemble de la période simulée. Par exemple, si on définit un modèle non stationnaire qui dépend de la concentration en CO_2 , on pourra obtenir une fonction de répartition pour chaque année en connaissant la concentration en CO_2 annuelle. L'exemple suivant présente la procédure pour obtenir la fonction répartition projetée pour la variable de référence en 2050.

Exemple

Soit Y une série d'observations allant de 1950 à 2000 et X une série simulée allant de 1950 à 2050. On cherche à obtenir $F_{Y_{2050}}$ la fonction répartition projetée pour la variable de référence en 2050 en appliquant la méthode *XCDF-t*.

La première étape consiste à obtenir $F_{Y_{calib}}$ la fonction de répartition des observations pour la période de calibration en ajustant une loi GP sur les données observées.

Ensuite, pour obtenir la fonction de répartition de la variable simulée F_X , une loi GP est ajustée sur l'ensemble des données simulées en utilisant la concentration en CO_2 annuelle comme variable explicative. Puisqu'il s'agit d'un modèle non stationnaire qui dépend de la concentration en CO_2 de chaque année, il y a une fonction de répartition différente pour chaque année. Dans ce cas-ci, $F_{X_{calib}}$ sera obtenue en évaluant la fonction de répartition avec la concentration en CO_2 de la dernière année observée disponible ($F_{X_{2000}}$). Pour chaque année projetée, $F_{X_{proj}}$ est obtenue en évaluant la fonction de répartition avec la concentration projetée en CO_2 pour cette année.

Finalement, pour obtenir $F_{Y_{proj}}$ associée à chaque année de la période projetée, l'équation (3.21) est appliquée en considérant $F_{Y_{calib}}$, $F_{X_{2000}}$ et $F_{X_{proj}}$ associée à cette même année. Ainsi, pour obtenir $F_{Y_{2050}}$, l'équation (3.21) est appliquée en utilisant $F_{Y_{calib}}$, $F_{X_{2000}}$ et $F_{X_{2050}}$.

3.4 Validation des méthodes de post-traitement

3.4.1 Indices climatiques

Ce projet porte sur le développement d'une méthode de post-traitement adaptée pour les valeurs extrêmes et le cœur de la distribution. Par conséquent, huit indices climatiques visant à valider les extrêmes et le cœur de la distribution ont été sélectionnés : le coefficient d'asymétrie, le *R10*, le *SDII*, le *P90Wet*, le *P98Wet*, le *P98WetAmount*, le *RV20Max* et le *RV100Max*.

Ces indices ont été sélectionnés pour évaluer les caractéristiques marginales des précipitations observées et simulées et post-traitées. Ils sont définis de la façon suivante :

Coefficient d'asymétrie

Le coefficient d'asymétrie ou *Skewness* est une mesure de l'asymétrie de la distribution définie de la façon suivante :

$$\text{Skew} = E\left[\left(\frac{X - \mu}{\sigma}\right)^3\right] \quad (3.22)$$

L'asymétrie d'une distribution est positive si la queue à droite (valeurs élevées) est plus longue ou lourde, et négative si la queue à gauche (valeurs faibles) est plus longue ou lourde. Il est de zéro pour une loi symétrique.

R10

Le R10 est défini comme la fréquence relative des jours avec des précipitations supérieures à 10 mm.

$$R10 = \frac{\sum_{i: x_i > 10mm}^N x_i}{N} \quad (3.23)$$

SDII

Le SDII ou *Simple Precipitation Intensity Index* est un indice d'intensité des précipitations qui est défini comme la quantité de précipitations en jours humides divisée par le nombre de jours humides.

P90Wet et P98Wet

Le P90Wet est défini comme le quantile d'ordre 0,90 des jours humides en considérant les jours avec des précipitations supérieures à 1 mm.

Similairement, le P98Wet est défini comme le quantile d'ordre 0,98 des jours humides en considérant les jours avec des précipitations supérieures à 1 mm.

P98WetAmount

Le P98WetAmount est défini comme l'accumulation en mm des précipitations supérieures au quantile d'ordre 0,98 des jours humides en considérant les jours avec des précipitations supérieures à 1 mm.

RV20Max et RV100Max

Le RV20Max est le niveau de retour associé à une période de retour de 20 ans pour les fortes précipitations. Cette quantité est obtenue en ajustant une loi GEV par la méthode des maxima par blocs sur les maxima annuels.

Similairement, le RV100Max est le niveau de retour associé à une période de retour de 100 ans pour les fortes précipitations.

3.4.2 Mesures de performance

Deux mesures de performance sont utilisées pour quantifier l'adéquation entre les indices calculés sur les observations et les simulations : la différence relative et un indice de comparaison.

Différence relative

Pour quantifier l'adéquation entre deux indices, la mesure de performance utilisée est la différence relative, définie de la façon suivante :

$$\text{Différence relative}_\theta (\%) = \frac{\theta_{sim} - \theta_{obs}}{\theta_{obs}} \times 100 \quad (3.24)$$

où θ représente différents indices climatiques calculés sur les observations (*obs*) et sur la simulation (*sim*). Dans ce cas-ci, il peut s'agir de la simulation brute ou corrigée.

Indice de comparaison

L'indice de comparaison [42] ou *Skill Score* est un indice souvent utilisé en prévision météorologique. Il est établi en comparant la précision d'un ensemble de prévisions avec celle d'un ensemble correspondant obtenu par une méthode déterminée, comme la prévision au hasard, la prévision fondée sur la persistance ou la prévision climatologique. Il a été adapté au contexte du post-traitement en comparant les erreurs entre les indices calculés sur les simulations brute et corrigée et les indices calculés sur les observations. Cet indice est défini de la façon suivante :

$$\text{Indice de comparaison}_\theta = \frac{\text{score de la simulation corrigée} - \text{score de la simulation brute}}{\text{score parfait} - \text{score de la simulation brute}}.$$

Le score utilisé est le carré de l'erreur entre l'indice calculé sur la simulation et l'indice calculé sur les observations. La définition de cet indice est donc la suivante :

$$\text{Indice de comparaison}_\theta = \frac{(\theta_{\text{corrigée}} - \theta_{\text{obs}})^2 - (\theta_{\text{brute}} - \theta_{\text{obs}})^2}{0 - (\theta_{\text{brute}} - \theta_{\text{obs}})^2}. \quad (3.25)$$

Le domaine de l'indice de comparaison s'étend de $-\infty$ à $+1$. Une valeur positive indique que le post-traitement présente une amélioration par rapport à la simulation brute. Une valeur négative indique que le post-traitement n'améliore pas la simulation brute ou, autrement dit, que la simulation brute est plus près des observations que la simulation corrigée.

CHAPITRE 4 DONNÉES

Ce chapitre présente les deux ensembles de données utilisés dans ce mémoire : les données historiques de précipitations observées à la station météorologique de l’aéroport Montréal-Trudeau et les précipitations simulées de l’ensemble de simulations climatiques ClimEx [43].

4.1 Précipitations observées

Le premier ensemble de données est composé des données historiques de précipitations enregistrées à la station météorologique de l’aéroport de Montréal-Trudeau (45.47°N, -73.75°E). Ces données sont constituées d’observations journalières de précipitations en mm entre 1941 et 2013. Parmi cet ensemble, un sous-ensemble de valeurs pour une période de 30 ans s’étendant de 1980 à 2009 inclusivement a été extrait. Pour cette sous-période plus courte, il est raisonnable de supposer le climat stationnaire, ce qui peut ne pas être le cas pour la période 1941-2013. Pour vérifier cette hypothèse, le test de tendance de Mann-Kendall a été effectué sur les données de 1980 à 2009. L’hypothèse nulle voulant qu’il n’y ait pas de tendance pour les données de cette période n’a pas été rejetée. Conséquemment, il est raisonnable de supposer le climat stationnaire. Ce sous-ensemble constitue la référence pour l’analyse des méthodes. Une analyse exploratoire des données observées a été effectuée afin de vérifier la présence d’un signal saisonnier. La figure 4.1 présente les diagrammes en boîte des précipitations journalières pour chaque saison. En ce qui concerne l’intensité, l’été (juin-juillet-août, JJA) est la saison pour laquelle on enregistre les plus fortes intensités, suivie par l’hiver (décembre-janvier-février, DJF). Cependant, les propriétés moyennes sont similaires pour le printemps (mars-avril-mai, MAM) et l’automne (septembre-octobre-novembre, SON).

La figure 4.2 montre les précipitations cumulées pour chaque saison. En termes de précipitations cumulées, de la saisonnalité est apparente pour l’hiver qui est une saison beaucoup plus sèche que les autres. L’été (JJA) et l’automne (SON) sont également plus humides que le printemps (MAM). Toutefois, lorsqu’on normalise les précipitations cumulées par le nombre de jours de pluie, la différence entre les saisons s’amenuise comme le montre la figure 4.3.

Puisqu’on vise à post-traiter des précipitations liquides, seules les données pour le printemps, l’été et l’automne seront utilisées pour éviter de traiter des précipitations solides l’hiver. Ces trois saisons seront traitées comme un seul ensemble de données. De plus, les données des mois de mars, avril et novembre ne seront pas utilisées afin réduire les risques d’avoir de la précipitation solide. Une analogie peut être faite avec les maxima annuels de précipitations

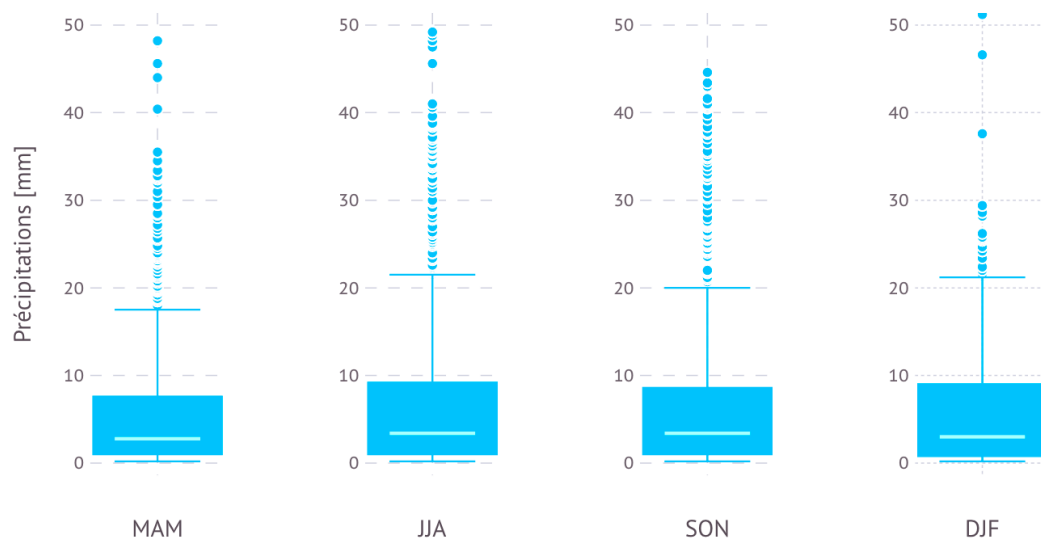


Figure 4.1 Diagrammes en boîte des précipitations journalières (mm) par saison observées à Montréal entre 1980 et 2009.

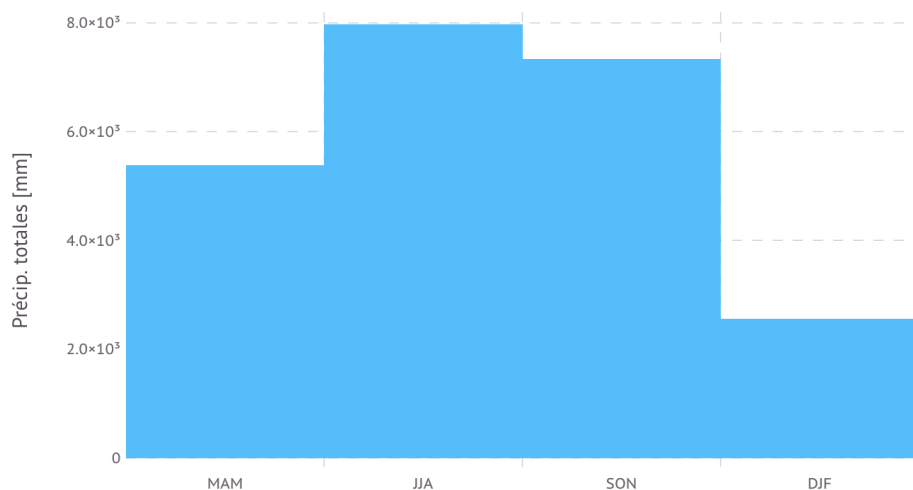


Figure 4.2 Précipitations cumulatives (mm) par saison observées à Montréal pour la période 1980-2009.

de mai à octobre (May to October Annual Maxima, MOAM) qui sont utilisés dans certaines études des précipitations extrêmes au Québec (p. ex., [44]).

En somme, un sous-ensemble des observations de mai à octobre inclusivement pour la période 1980-2009 sera utilisé pour la suite de l'étude.

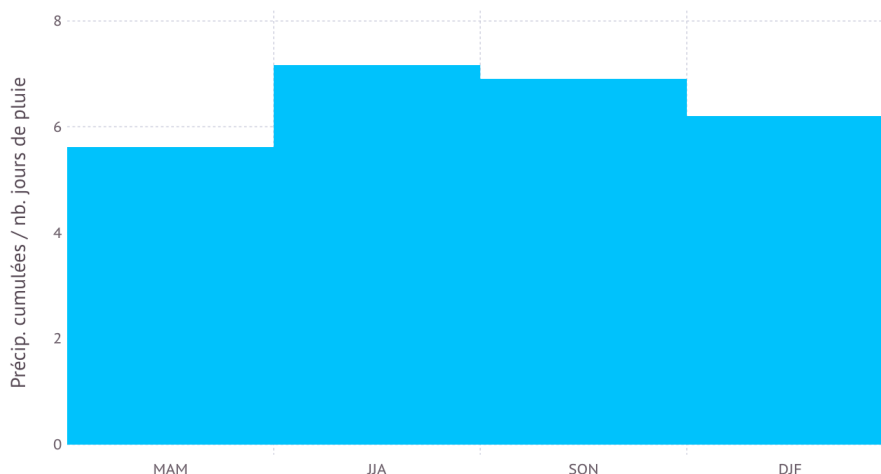


Figure 4.3 Précipitations cumulatives (mm) normalisées par le nombre de jours de pluie (>1mm/jour) par saison observées à Montréal pour la période 1980-2009.

4.2 Ensemble de simulations

L'ensemble de simulations de la précipitation ClimEx est utilisé dans ce projet. ClimEx est composé de 50 membres qui représentent des réalisations équiprobables du climat visant à étudier les extrêmes climatiques et la variabilité naturelle du climat en Bavière et au Québec (voir [43]). Les simulations de cet ensemble ont été réalisées à partir de la cinquième version du Modèle Régional Canadien du Climat (MRCC5) [45,46] développé par le Centre pour l'étude et la simulation du climat à l'échelle régionale (ESCER) et Environnement et Changement climatique Canada (ECCC). Ces simulations à haute résolution horizontale (12 km) ont été pilotées par une simulation du MGC CanESM2 de résolution horizontale plus faible (environ 200 km).

Chaque membre de ClimEx, identifié par une série de trois lettres (p. ex : KDA, KDB, KDC), se distingue par de petites perturbations dans les conditions initiales du MGC qui le pilote. Ces perturbations dans les conditions initiales forcent les simulations à se différencier en raison du chaos du système climatique. Avec plusieurs simulations, il est possible de tenir compte de la variabilité naturelle du climat, c'est-à-dire les variations naturelles issues de la non-linéarité du système climatique. Pour chaque membre, les simulations sont réalisées de 1950 à 2100 selon le scénario de forçage radiatif RCP8.5. Toutefois, les données utilisées pour ce projet vont de 1955 à 2100 afin de permettre à l'hypothèse d'indépendance de se réaliser dans les années suivant l'initialisation [43].

Les données de ClimEx s'étendent sur une grille qui couvre le Nord-Est de l'Amérique du Nord. Pour cette étude, les données d'un point de grille correspondant à la ville de Montréal

sont utilisées. Les données de précipitations de l'ensemble ClimEx étant disponibles dans le format horaire, les accumulations horaires ont été additionnées pour obtenir les accumulations journalières.

Comme pour les données observées, un sous-ensemble des précipitations simulées allant de mai à octobre inclusivement sera utilisé pour la suite de l'étude afin de considérer les précipitations liquides seulement.

CHAPITRE 5 MÉTHODOLOGIE ET RÉSULTATS

Ce chapitre présente les étapes qui ont mené au développement de la méthode de post-traitement paramétrique et les résultats du projet de maîtrise. La méthodologie développée se décompose en quatre étapes. Premièrement, les précipitations journalières observées sont modélisées avec des modèles paramétriques adaptés à l'entièreté de la gamme des précipitations. Deuxièmement, une extension non stationnaire pour modéliser les précipitations simulées par les modèles climatiques est développée. Troisièmement, une méthode de correspondance entre les quantiles des deux modèles statistiques définis aux deux premières étapes est établie pour construire les fonctions de correction. Cette étape permet de développer une méthode de post-traitement adaptée pour les valeurs extrêmes et le cœur de la distribution. Finalement, la méthode de post-traitement développée est validée par le calcul et la comparaison d'indices climatiques appropriés pour évaluer les valeurs extrêmes et le cœur de la distribution.

5.1 Modélisation des précipitations observées

Afin d'obtenir la distribution de référence qui sera utilisée pour le post-traitement paramétrique exprimé à l'équation (3.21), il est nécessaire de modéliser les précipitations observées avec une loi de probabilité paramétrique. Le premier objectif du projet est de modéliser les précipitations observées avec un modèle statistique paramétrique s'appliquant à l'entièreté de la gamme des précipitations. Pour atteindre cet objectif, les quatre familles de lois développées par Naveau et al. (2016) [3] décrites dans la section 3.1.3 sont ajustées sur la distribution des précipitations observées au Québec et comparées afin de déterminer celle présentant le meilleur ajustement. D'abord, les paramètres de chaque famille sont estimés par maximum de vraisemblance sur les observations. Puis, la famille la plus appropriée est sélectionnée en comparant la qualité de l'ajustement.

5.1.1 Estimation des paramètres par maximum de vraisemblance

Soit Y_{ij} la quantité de précipitations observées le jour j de l'année i . Puisque les jours sans précipitations sont ignorés, le nombre de jours de pluie de l'année i , n_i , est différent pour chaque année. Posons le vecteur $\mathbf{y}_i = (y_{ij} : j = 1, \dots, n_i)$ correspondant aux n_i valeurs de précipitations non nulles de l'année i . Posons également $\mathbf{y} = (\mathbf{y}_i : i = 1, \dots, m)$ le vecteur des précipitations non nulles pour toutes les m années.

Pour modéliser l'intensité des précipitations, les quatre modèles de lois de Pareto généralisées étendues décrits dans la section 3.1.3 sont ajustés sur le sous-ensemble des précipitations observées. Les paramètres inconnus sont estimés par maximum de vraisemblance. Cette technique consiste à maximiser la fonction de vraisemblance suivante :

$$L(\boldsymbol{\theta}) = \prod_{i=1}^m \prod_{j=1}^{n_i} f_{(Y_{ij}|\boldsymbol{\theta})}(y_{ij}) \quad (5.1)$$

où l'expression $f_{(Y|\boldsymbol{\theta})}(y)$ dénote la densité de paramètre $\boldsymbol{\theta}$ de la variable aléatoire Y évaluée en y .

L'estimation par maximum de vraisemblance des paramètres d'un modèle absolument continu avec des données de précipitations peut être problématique en raison de la discrétisation numérique causée par la précision des instruments de mesure. Il est en effet impossible de mesurer des précipitations inférieures à 0,2 mm en raison de l'instrumentation utilisée. Cela affecte particulièrement les valeurs faibles et peut biaiser l'estimation si on omet d'en tenir compte [47]. Pour les données de précipitations à Montréal, l'estimation des paramètres par maximum de vraisemblance présente un ajustement inadéquat en raison de l'absence de valeurs entre 0 et 0,2 mm.

Une solution pour contourner ce problème a été de transformer les données de précipitations. D'abord, les jours de pluie ont été définis comme les jours où la précipitation journalière excède 1 mm comme il est courant de le faire dans la littérature (p. ex. : Hennessy et al., 1999 [48] et Zhang et al. 2011 [49]) pour contrer les biais qui peuvent être introduits par la sous-représentation de petites quantités de précipitations [49]. Ensuite, les valeurs ont subi une translation de 1 mm vers la gauche afin de combler l'écart entre 0 et 1 mm.

Pour améliorer l'estimation par maximum de vraisemblance, [3] proposent d'utiliser la vraisemblance censurée pour les observations plus petites que le seuil $c > 0$. La censure est une façon d'améliorer l'ajustement sur les précipitations élevées qui sont moins fréquentes en réduisant le poids des précipitations faibles, beaucoup plus fréquentes en-dessous d'un certain seuil. Cette technique consiste à maximiser la fonction de vraisemblance suivante :

$$L_c(\boldsymbol{\theta}) = \prod_{\{(i,j):y_{ij}<c\}} F_{(Y_{ij}|\boldsymbol{\theta})}(c) \prod_{\{(i,j):y_{ij}\geq c\}} f_{(Y_{ij}|\boldsymbol{\theta})}(y_{ij}), \quad (5.2)$$

où c est le seuil de censure.

Pour l'ajustement des modèles de lois de Pareto généralisées étendues, [3] notent une claire amélioration pour l'estimation par maximum de vraisemblance lorsque les données sont cen-

surées en dessous de 0,5 mm. Il est important de mentionner que dans [3], des données horaires sont utilisées et qu'il est possible que ce seuil soit différent avec les données journalières comme dans le cas de cette présente étude. [47] estiment plutôt que ce seuil ne devrait pas être inférieur à 5 mm pour les précipitations journalières en France. Par conséquent, le seuil de censure a été établi à 5 mm pour notre étude. Cependant, puisque la censure est appliquée aux données transformées, ce seuil est de 6 mm en réalité.

Les quatre modèles de lois de Pareto généralisées étendues ont été ajustés sur une période de 30 ans allant de 1980 à 2009 inclusivement. La taille de l'échantillon est de 1371 valeurs. Cependant, en censurant les valeurs inférieures à 5 mm lors de l'ajustement des modèles par maximum de vraisemblance censurée, la taille effective de l'échantillon contribuant pleinement à la vraisemblance est réduite à 674 valeurs. Par conséquent, près de 14% du total des précipitations se retrouve sous le seuil de censure.

5.1.2 Résultats

Les résultats de l'estimation des paramètres sont présentés dans le tableau 5.1 avec les intervalles de confiance à 95% obtenus par bootstrap non paramétrique [50].

Tableau 5.1 Paramètres estimés par maximum de vraisemblance sur les précipitations observées. Pour le modèle 3, $\kappa_1 = \delta$ et pour le modèle 4, $\kappa_1 = \delta$ et $\kappa_2 = \kappa$. Les intervalles de confiance à 95% (entre parenthèses) sont obtenus par bootstrap non paramétrique sur 500 échantillons.

	$\hat{\sigma}$	$\hat{\xi}$	$\hat{\kappa}_1$	$\hat{\kappa}_2$	$\hat{p}$
Modèle 1	9.08 (6.99, 11.4)	0.0833 (-0.018, 0.186)	0.767 (0.626, 0.967)		
Modèle 2	9.08 (3.81, 11.3)	0.0833 (-0.0239, 0.219)	0.767 (0.605, 1.03)	0.767 (0.656, 26.0)	0.593 (0.00454, 1.0)
Modèle 3	6.68 (6.09, 7.38)	0.2 (0.14, 0.257)	100.0 (93.4, 100.0)		
Modèle 4	6.41 (4.96, 13.3)	0.102 (-0.066, 0.171)	0.000188 (2.92e-5, 7.89)	0.768 (0.649, 1.08)	

On remarque que les estimations des paramètres du modèle 2 correspondent à celles du modèle 1. Les valeurs des paramètres $\hat{\kappa}_1$ et $\hat{\kappa}_2$ sont identiques. Conséquemment, le paramètre $\hat{p}$ n'a pas d'effet puisque les deux lois du mélange sont les mêmes. Pour le modèle 3, la flexibilité restreinte de la queue inférieure semble affecter l'estimation par maximum de vraisemblance.

Le paramètre de queue supérieure ξ est beaucoup plus grand que pour les autres modèles et le paramètre δ tend vers la valeur maximale permise par l'algorithme de maximisation. Rappelons que lorsque $\delta \rightarrow \infty$, ce modèle correspond à la loi GP. Pour le modèle 4, l'estimation du paramètre δ est quasi nulle et le paramètre de queue supérieure ξ est plus élevé que pour le modèle 1.

5.1.3 Sélection du modèle

Le modèle statistique le plus approprié est sélectionné en comparant la qualité de l'ajustement à l'aide du critère d'information bayésien (BIC) défini à la section 3.2.1 et d'analyses graphiques.

Critère d'information bayésien

Le BIC a été calculé pour les quatre modèles puis les résultats sont comparés. Les résultats sont présentés dans le tableau 5.2.

Tableau 5.2 Valeurs du BIC des quatre modèles sur les données observées de 1980 à 2009.

	Modèle 1	Modèle 2	Modèle 3	Modèle 4
BIC	-3167.61	-3174.83	-3170.38	-3171.29

Le modèle 1 est celui qui présente la plus grande valeur du BIC. Le modèle 3 est celui qui s'en rapproche le plus. L'approximation du facteur de Bayes entre ces deux modèles est de 2,8. Conséquemment, selon le tableau 3.1, la certitude en faveur du modèle 1 est décisive. Pour cette raison, le modèle 1 est celui qui a été sélectionné. En somme, la classification suivante est déterminée selon le BIC : Modèle 1, Modèle 3, Modèle 4 puis Modèle 2. On peut remarquer que cette classification favorise les modèles avec moins de paramètres.

Validation graphique

Plusieurs techniques de validation graphique peuvent être utilisées pour vérifier l'ajustement d'un modèle aux valeurs déjà observées. Parmi celles-ci, le diagramme de probabilité et le diagramme quantile-quantile sont couramment utilisés, ainsi que la superposition de la densité à l'histogramme des données et le diagramme de niveau de retour.

La figure 5.1 présente les différents diagrammes diagnostiques pour valider l'ajustement du modèle aux données de précipitations observées.

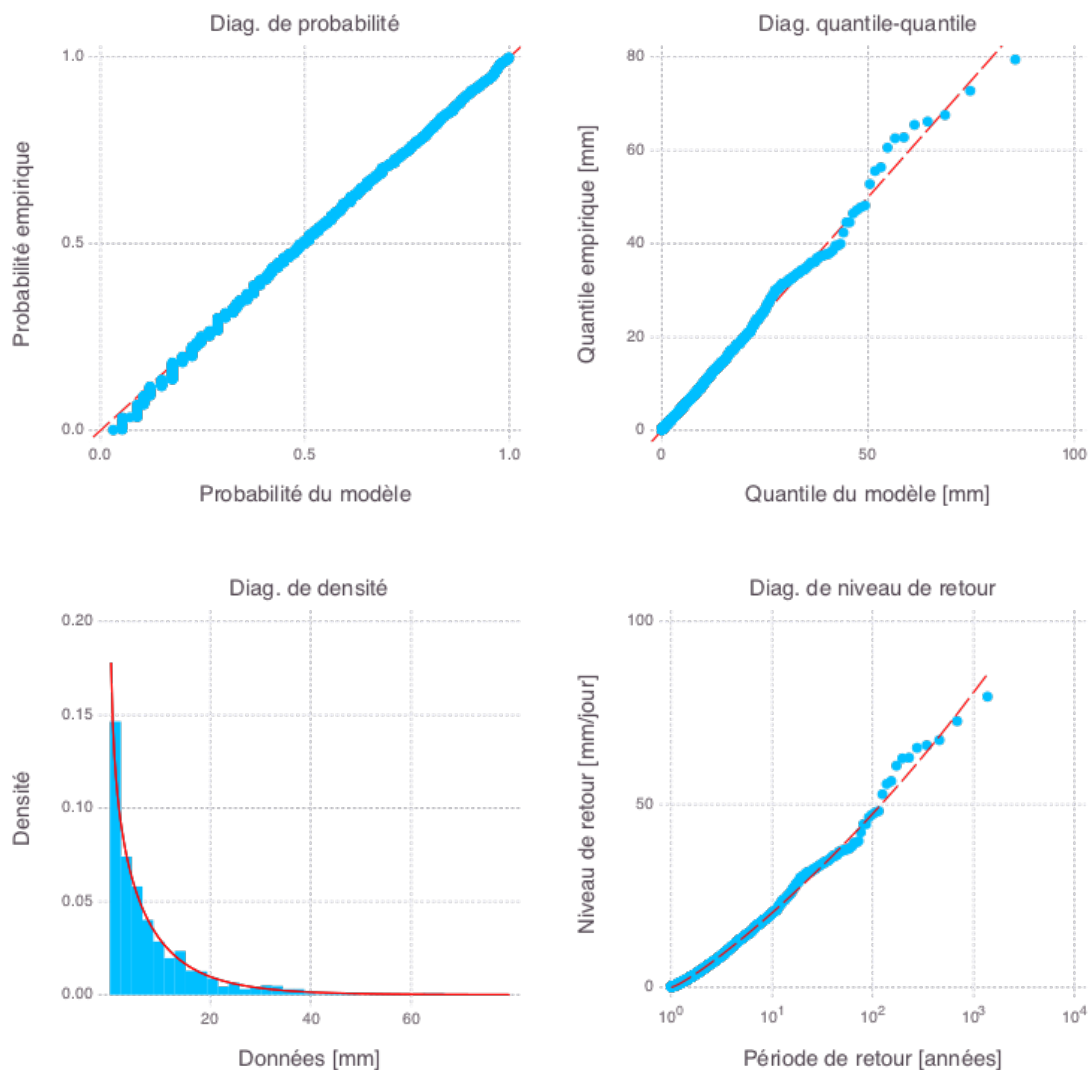


Figure 5.1 Diagrammes diagnostiques pour l'ajustement du modèle 1 aux précipitations observées à Montréal.

D'abord, on remarque que la linéarité des ensembles de points tracés sur les diagrammes de probabilité et quantile-quantile est raisonnable malgré quelques points s'écartant à l'extrémité supérieure du diagramme de quantiles. Bien que les données inférieures à 5 mm aient été censurées lors de l'estimation des paramètres, le diagramme quantile-quantile montre que le modèle parvient bien à modéliser les faibles précipitations en plus de celles modérées et intenses. Toutefois, l'effet de la discrétisation numérique causée par la précision des instruments de mesure est perceptible sur le diagramme de probabilité. La courbe de niveau de retour fournit une représentation satisfaisante des estimations empiriques. La forme convexe implique qu'elle tend asymptotiquement vers des niveaux infinis. Cela s'explique par l'estimation positive du paramètre de forme de la queue supérieure ξ . Enfin, l'estimation de la

densité correspondante semble cohérente avec l’histogramme des données. Conséquemment, les quatre figures diagnostiques soutiennent l’ajustement du modèle 1 aux données de précipitations observées à Montréal.

Pour résumer cette section, après avoir ajusté les quatre modèles de lois de Pareto généralisées étendues par maximum de vraisemblance sur les données de précipitations observées censurées à 5 mm, il a été déterminé par le BIC et par validation graphique que celui qui s’ajuste le mieux est le modèle 1.

5.2 Modélisation des précipitations simulées

Cette section porte sur le second objectif du projet qui est de modéliser les précipitations simulées de l’ensemble ClimEx avec un modèle paramétrique adapté à l’entièreté de la gamme des précipitations. Cet objectif est similaire à celui de la section précédente. Les quatre mêmes familles de modèles sont ajustées et comparées dans le but de déterminer celle présentant la meilleure adéquation aux données de ClimEx. Pour tenir compte du fait que les précipitations sont influencées par les changements climatiques, une analyse de la stationnarité est réalisée et est suivie du développement d’une extension non stationnaire au modèle sélectionné. Pour conclure, le modèle non stationnaire développé est ajusté par maximum de vraisemblance sur la distribution des précipitations simulées d’un membre et la qualité de son ajustement est évaluée.

5.2.1 Estimation des paramètres par maximum de vraisemblance

Les précipitations peuvent évoluer dans le temps en raison des CC (voir p. ex., [51]). Afin de s’assurer que les familles de modèles statistiques peuvent s’appliquer aux précipitations simulées sur toute la période, incluant la non-stationnarité, l’ajustement des modèles est d’abord fait sur un sous-ensemble de données pour une période de 30 ans allant de 1980 à 2009 inclusivement. Il est raisonnable de supposer que la non-stationnarité des précipitations est négligeable sur cette période relativement courte. Par souci de concision, l’analyse est présenté pour un seul membre de ClimEx, KDI. Ce membre a été choisi car il présentait un biais important par rapport aux observations lors de l’analyse exploratoire.

Soit X_{ij} la quantité de précipitations simulées le jour $j \in \{1, \dots, n_i\}$ de l’année $i \in \{1, \dots, m+k\}$. De façon générale, les précipitations simulées sont projetées pour k années supplémentaires depuis la fin de la période d’observation. Posons le vecteur $\mathbf{x}_i = (y_{ij} : j = 1, \dots, n_i)$ correspondant aux n_i valeurs de l’année i . Posons également $\mathbf{x} = (\mathbf{x}_i : i = 1, \dots, m+k)$ le vecteur des données pour toutes les $m+k$ années.

Les données simulées ne sont pas affectées par la discrétisation numérique des instruments de mesure. Cependant, un effet de bruine, ou *drizzle effect*, peut affecter les précipitations simulées [7]. Afin de corriger le problème, les données ont été transformées de la même façon qu'à la section 5.1.1, c'est-à-dire en considérant seulement les valeurs supérieures à 1 mm et en effectuant une translation de 1 mm vers la gauche. La taille de l'échantillon est de 1510 valeurs. Les modèles ont été ajustés par maximum de vraisemblance censurée selon l'équation (5.2). Les valeurs inférieures à 5 mm ont été censurées selon le seuil défini à la section 5.1.1. La taille effective de l'échantillon contribuant pleinement à la vraisemblance est de 670 valeurs.

5.2.2 Résultats

Le tableau 5.3 présente les paramètres estimés pour chacun des modèles par maximum de vraisemblance censurée à 5 mm avec les intervalles de confiance à 95% obtenus par bootstrap non paramétrique.

Tableau 5.3 Paramètres estimés par maximum de vraisemblance sur les précipitations simulées du membre KDI. Pour le modèle 3, $\kappa_1 = \delta$ et pour le modèle 4, $\kappa_1 = \delta$ et $\kappa_2 = \kappa$. Les intervalles de confiance à 95% (entre parenthèses) sont obtenus par bootstrap non paramétrique sur 500 échantillons.

	$\hat{\sigma}$	$\hat{\xi}$	$\hat{\kappa}_1$	$\hat{\kappa}_2$	$\hat{p}$
Modèle 1	9.21 (7.21, 11.7)	0.102 (0.000408, 0.199)	0.656 (0.542, 0.81)		
Modèle 2	9.21 (5.29, 11.4)	0.102 (-0.0205, 0.207)	0.656 (7.67e-7, 0.95)	2.84 (0.572, 41.3)	1.0 (0.00922, 1.0)
Modèle 3	5.65 (5.16, 6.19)	0.282 (0.215, 0.343)	100.0 (100.0, 100.0)		
Modèle 4	6.49 (4.86, 13.8)	0.116 (-0.0614, 0.196)	0.000179 (3.12e-5, 5.83)	0.658 (0.553, 0.862)	

Comme pour les résultats présentés à la section 5.1.2, on remarque que les estimations des paramètres du modèle 2 correspondent à celles du modèle 1. La deuxième loi du mélange n'a aucun effet car son poids est nul parce que le paramètre $\hat{p}$ tend vers 1 (apparaît égal à 1 dans le tableau en raison de l'arrondissement). Cela revient à dire que l'estimation des paramètres ont reproduit le modèle 1. De plus, les mêmes problèmes semblent affecter l'estimation du modèle 3. Le paramètre δ tend vers la valeur maximale permise par l'algorithme de maximisation. Cela semble s'être produit systématiquement pour les 500 échantillons comme le montre

l'intervalle de confiance sur ce paramètre. Pour le modèle 4, l'estimation du paramètre δ est quasi nulle et le paramètre de queue supérieure ξ est plus élevé que pour le modèle 1.

5.2.3 Sélection du modèle

Le BIC a été calculé et comparé pour les quatre modèles. Les résultats sont présentés dans le tableau 5.4.

Tableau 5.4 Valeurs du BIC des quatre modèles sur les données simulées du membre KDI de 1980 à 2009.

	Modèle 1	Modèle 2	Modèle 3	Modèle 4
BIC	-3251.38	-3258.70	-3257.04	-3255.09

Parmi ceux-ci, le modèle 1 est celui qui présente la plus grande valeur du BIC. Le modèle 4 est celui qui s'en rapproche le plus. L'approximation du facteur de Bayes entre ces deux modèles est de 3,7. Conséquemment, selon le tableau 3.1, la certitude en faveur du modèle 1 est décisive. En somme, la classification suivante est déterminée selon le BIC : Modèle 1, Modèle 4, Modèle 3 puis Modèle 2.

La figure 5.2 présente les différentes figures diagnostiques pour valider l'ajustement du modèle 1 aux données de précipitations simulées du membre KDI.

D'abord, on remarque que les ensembles de points tracés sur les diagrammes de probabilité et quantile-quantile sont linéaires. La courbe de niveau de retour fournit une représentation satisfaisante des estimations empiriques. Comme on peut s'y attendre pour les données de précipitations, l'estimation positive du paramètre de la queue supérieure ξ explique la forme concave de la courbe qui tend asymptotiquement vers des niveaux infinis. Enfin, l'estimation de la densité correspondante semble cohérente avec l'histogramme des données. Conséquemment, les quatre figures diagnostiques soutiennent l'ajustement du modèle 1 aux données de précipitations simulées de ClimEx sur la période 1980-2009 pour le membre KDI.

5.2.4 Analyse des 50 membres

Les résultats précédents ont été obtenus pour le membre KDI seulement. Pour s'assurer que le modèle sélectionné est le modèle le mieux adapté pour l'ensemble des 50 membres, le processus de sélection par BIC a été répété pour chaque membre sur trois sous-ensembles de 30 ans de données, 1980-2009, 2040-2069 et 2070-2099.

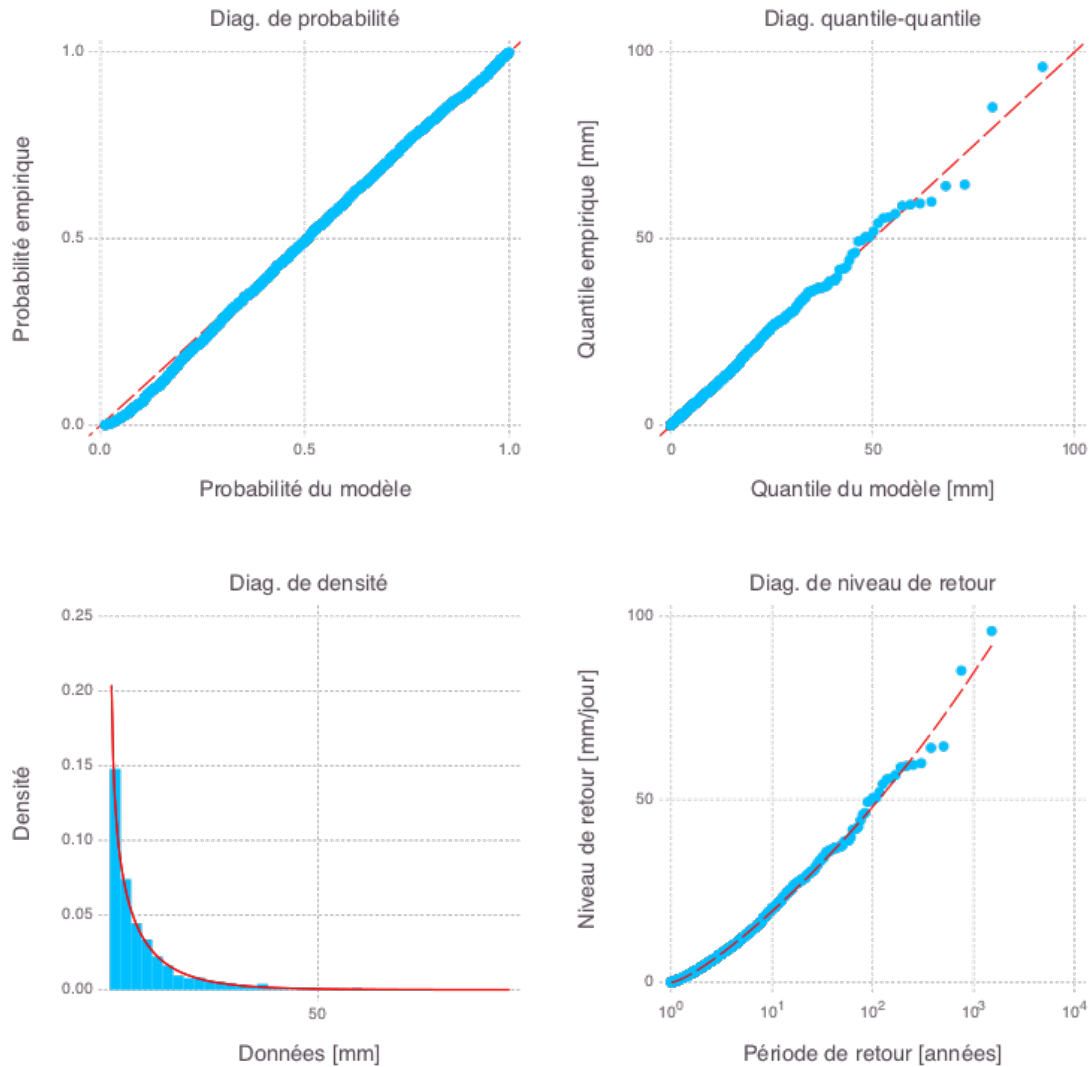


Figure 5.2 Diagrammes diagnostiques pour l'ajustement du modèle 1 aux précipitations simulées du membre KDI pour la période 1980-2009.

Pour chacune de ces trois sous-périodes, le modèle 1 a été favorisé pour chacun des 50 membres. Par conséquent, c'est ce modèle qui sera utilisé pour la suite de l'analyse.

5.2.5 Étude de la non-stationnarité

Le développement d'une extension non stationnaire au modèle 1 pourrait s'avérer nécessaire. Le développement de cette extension permettra d'incorporer la non-stationnarité des précipitations simulées par les modèles climatiques.

Tout d'abord, on doit vérifier si le modèle stationnaire est apte à modéliser les précipitations simulées du membre KDI de 1955 à 2099. La figure 5.3 présente les différentes figures

diagnostics pour valider l'ajustement du modèle aux données de précipitations simulées du membre KDI de 1955 à 2009. La qualité de l'ajustement dans le diagramme de quantiles ne semble pas convaincante, particulièrement dans la queue supérieure. Cela laisse supposer qu'un modèle non stationnaire pourrait s'avérer plus adapté.

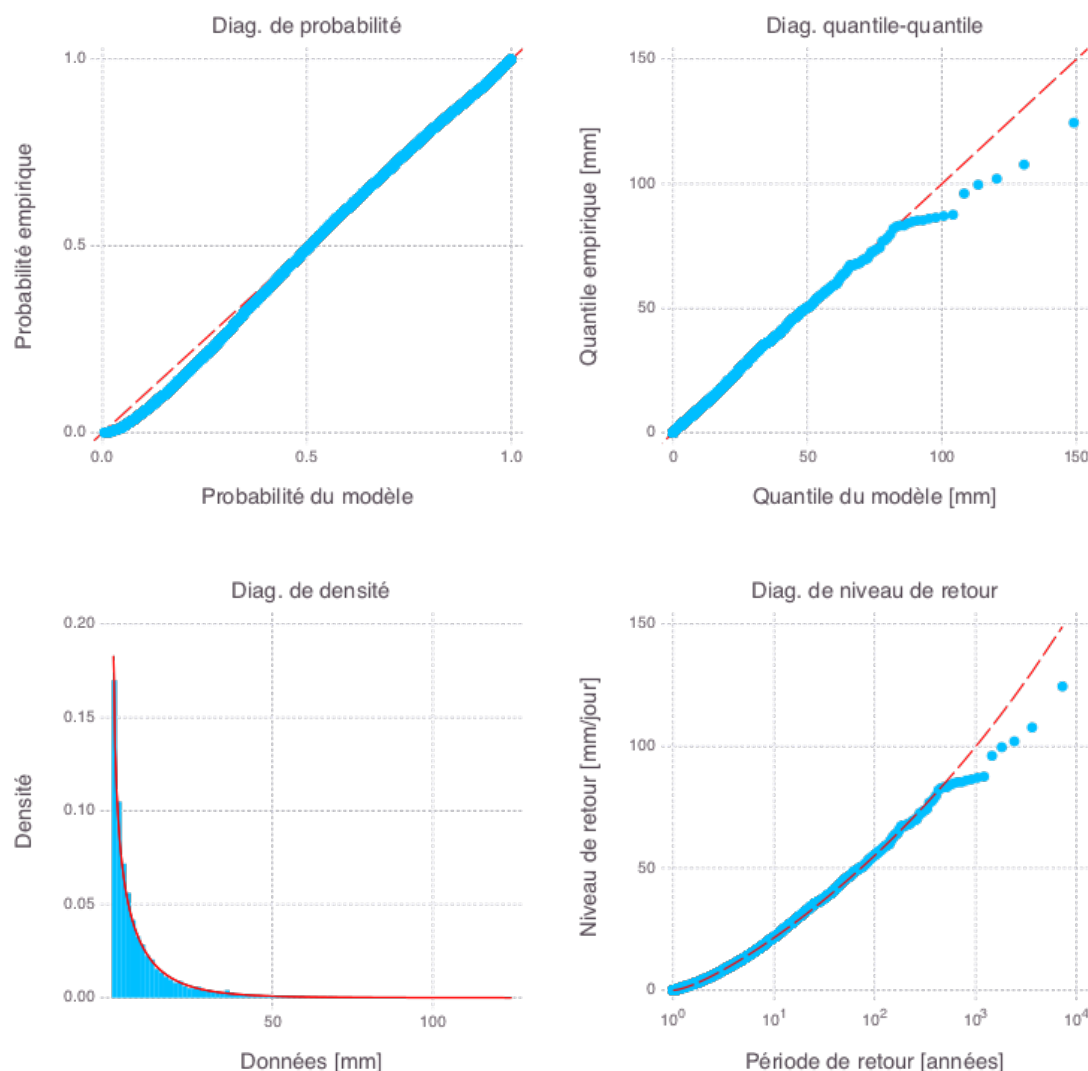


Figure 5.3 Diagrammes diagnostiques pour l'ajustement du modèle 1 stationnaire aux précipitations simulées du membre KDI pour la période 1955-2009.

L'évolution des paramètres du modèle ajustés sur une fenêtre mobile de 30 ans de données pour chaque membre est analysée sur les figures 5.4, 5.5 et 5.6. Plus particulièrement, pour chacun des 50 membres de ClimEx, le modèle 1 a d'abord été ajusté sur les données de 1955 à 1985, puis sur les données de 1956 à 1986 et ainsi de suite jusqu'à couvrir entièrement la période 1955-2009. Ensuite, des diagrammes en boîte ont été construits avec les résultats

de l'estimation pour chaque membre et pour chaque période. Les résultats pour une fenêtre de 30 ans sont indiqués par son année médiane. Par exemple, pour la période 1955-1985, le diagramme en boîte est affiché pour l'année 1970. Cela explique pourquoi l'axe des abscisses s'étend de 1970 à 2084 alors que les données vont de 1955 à 2099. La médiane de chaque boîte est analysée afin d'évaluer l'évolution de chaque paramètre du modèle.

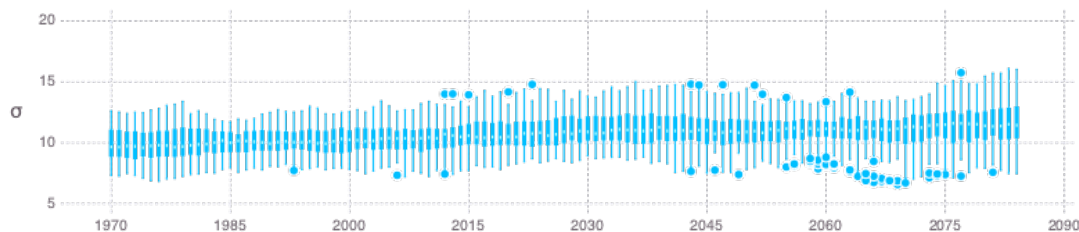


Figure 5.4 Évolution du paramètre d'échelle σ du modèle 1.

La figure 5.4 présente l'évolution du paramètre d'échelle σ du modèle 1. On remarque une faible augmentation de la valeur médiane des boîtes en fonction du temps. Par conséquent, on peut conclure que ce paramètre n'est pas stationnaire.

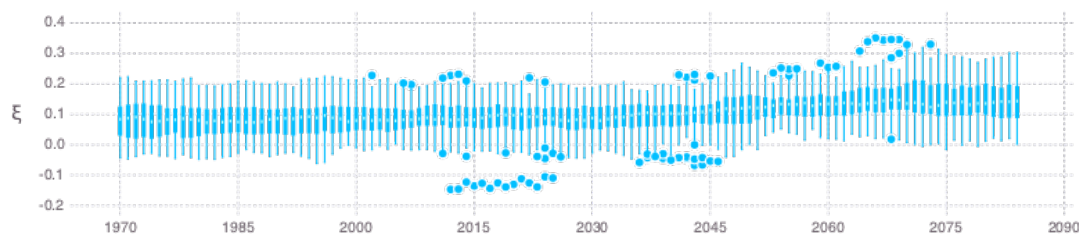


Figure 5.5 Évolution du paramètre de queue supérieure ξ du modèle 1.

La figure 5.5 présente l'évolution du paramètre de queue supérieure ξ du modèle 1. On remarque également une faible augmentation de la valeur médiane entre le début et la fin de la période pour ce paramètre. Il est donc raisonnable de supposer que ce paramètre du modèle ne soit pas stationnaire.

La figure 5.6 présente l'évolution du paramètre de forme de la queue inférieure κ du modèle 1. Une variation négative semble perceptible bien que les moustaches se recoupent. Il est possible que le paramètre κ évolue dans le temps.

L'analyse exploratoire de la stationnarité laisse croire qu'un modèle non stationnaire pourrait être développé en incorporant de la non-stationnarité dans le paramètre d'échelle σ , dans le paramètre de queue supérieure ξ et dans le paramètre de queue inférieure κ .

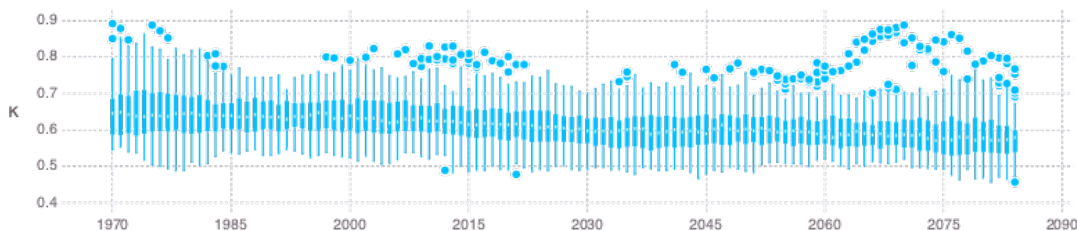


Figure 5.6 Évolution du paramètre de forme de la queue inférieure κ du modèle 1.

5.2.6 Choix d'une variable explicative

Pour modéliser la série non stationnaire des précipitations simulées, il est judicieux d'introduire dans le modèle 1 des variables explicatives pour la non-stationnarité dans les paramètres du modèle. Ceci permet d'utiliser la même technique d'estimation par maximum de vraisemblance tout en permettant aux paramètres de varier en fonction des variables explicatives. Il faut maintenant déterminer quelle variable explicative doit être retenue pour l'ajustement du modèle.

Idéalement, les variables explicatives représentent le processus physique qui cause la non-stationnarité dans les données. Pour la précipitation, les changements projetés résultent de changements dans la température en raison des CC. En effet, le principe de Clausius-Clapeyron suggère que plus la température est élevée, plus le contenu en humidité peut être grand, cela pouvant mener à des précipitations plus fréquentes ou plus intenses. On sait également que le CO_2 joue un rôle majeur dans le changement climatique anthropique et dans l'augmentation des températures. Conséquemment, la concentration équivalente en CO_2 des gaz à effet de serre (GES) donnée par le RCP 8.5, le scénario utilisé dans l'ensemble ClimEx, sera introduite comme variable explicative. Cette variable a d'ailleurs été utilisée avec succès dans d'autres études (p. ex., [52]). Il est à noter que la température aurait également pu être utilisée. Toutefois, une variable simulée additionnelle serait nécessaire. De plus, cela nécessiterait de bien extraire le signal et de ne pas être soumis à la variabilité interannuelle de la température, ce qui rend le traitement plus complexe sans nécessairement ajouter de la valeur à l'ajustement du modèle.

La figure 5.7 présente la concentration en CO_2 selon les années proposées par le scénario RCP8.5, soit le scénario de forçage radiatif qui a servi pour la modélisation de l'ensemble ClimEx. On remarque une augmentation continue sur la période d'étude.

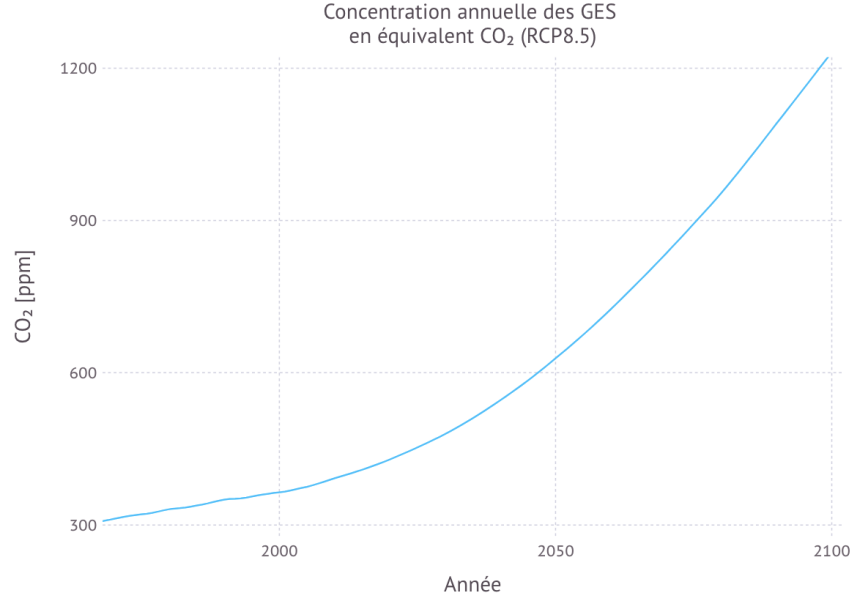


Figure 5.7 Concentration annuelle des GES en équivalent CO₂ (ppm) en fonction des années selon de scénario RCP8.5.

5.2.7 Modèles d'ajustement

Maintenant qu'une possible variable explicative a été déterminée, on peut définir un modèle non stationnaire pour les données de 1956 à 2009. Dans le modèle le plus général, la non-stationnarité est intégrée sur le paramètre d'échelle σ , sur le paramètre de queue supérieure ξ et sur le paramètre de queue inférieure κ via une relation linéaire avec une variable explicative. Ils sont exprimés de la manière suivante, pour l'année t allant de 1956 à 2009 :

$$\begin{aligned}\sigma(t) &= \sigma_0 + \sigma_1 \times \text{CO}_2(t) \\ \xi(t) &= \xi_0 + \xi_1 \times \text{CO}_2(t) \\ \kappa(t) &= \kappa_0 + \kappa_1 \times \text{CO}_2(t)\end{aligned}\tag{5.3}$$

Afin de s'assurer que ce modèle non stationnaire est le plus approprié, il est comparé avec le modèle stationnaire :

$$\begin{aligned}\sigma(t) &= \sigma, \\ \xi(t) &= \xi, \\ \kappa(t) &= \kappa;\end{aligned}\tag{5.4}$$

et avec toutes les combinaisons de sous-modèles. Par exemple, le modèle avec la non-stationnarité sur seulement le paramètre d'échelle σ , celui avec la non-stationnarité sur seulement le para-

mètre de queue supérieure ξ et ainsi de suite. Au total, huit modèles sont comparés.

5.2.8 Sélection du modèle

Le BIC a été calculé et comparé pour les huit modèles. Les résultats sont présentés dans le tableau 5.5.

Tableau 5.5 Valeurs du BIC des huit modèles sur les données simulées du membre KDI de 1955 à 2009.

	BIC
Modèle 1 stationnaire	-16280.55
Modèle 1 avec n.-s. sur σ	-16267.91
Modèle 1 avec n.-s. sur ξ	-16267.45
Modèle 1 avec n.-s. sur κ	-16273.57
Modèle 1 avec n.-s. sur σ et ξ	-16408.45
Modèle 1 avec n.-s. sur σ et κ	-20191.57
Modèle 1 avec n.-s. sur ξ et κ	-16506.79
Modèle 1 avec n.-s. sur σ , ξ et κ	-19979.04

Parmi ceux-ci, le modèle 1 non stationnaire avec seulement ξ qui dépend de la concentration en CO_2 est celui qui présente la plus grande valeur du BIC. Toutefois, la certitude en faveur de ce modèle par rapport à celui dont seul σ dépend de la concentration en CO_2 est faible considérant que l'écart entre les BIC est seulement de 0,46.

Pour s'assurer que le modèle final représente bien les données, une validation graphique est effectuée. Dans le cas non stationnaire, l'absence d'homogénéité dans les hypothèses de distribution pour chaque observation signifie que certaines modifications sont nécessaires par rapport à ce qui a été présenté à la section 5.2.3, *Sélection du modèle*. La validation graphique doit se faire avec une version standardisée des données, conditionnellement aux valeurs des paramètres ajustés.

La figure 5.8 présente le diagramme quantile-quantile résiduel à l'échelle exponentielle des deux modèles non stationnaire. Pour les deux modèles, la linéarité des ensembles de points tracés sur le diagramme quantile-quantile est améliorée par rapport à celle sur le diagramme du modèle non stationnaire, plus particulièrement à l'extrémité supérieure où il n'y a plus de décrochage. Entre ces deux modèles, celui qui semble être en meilleure adéquation avec les données est le modèle 1 non stationnaire dont seul ξ dépend de la concentration en CO_2 .

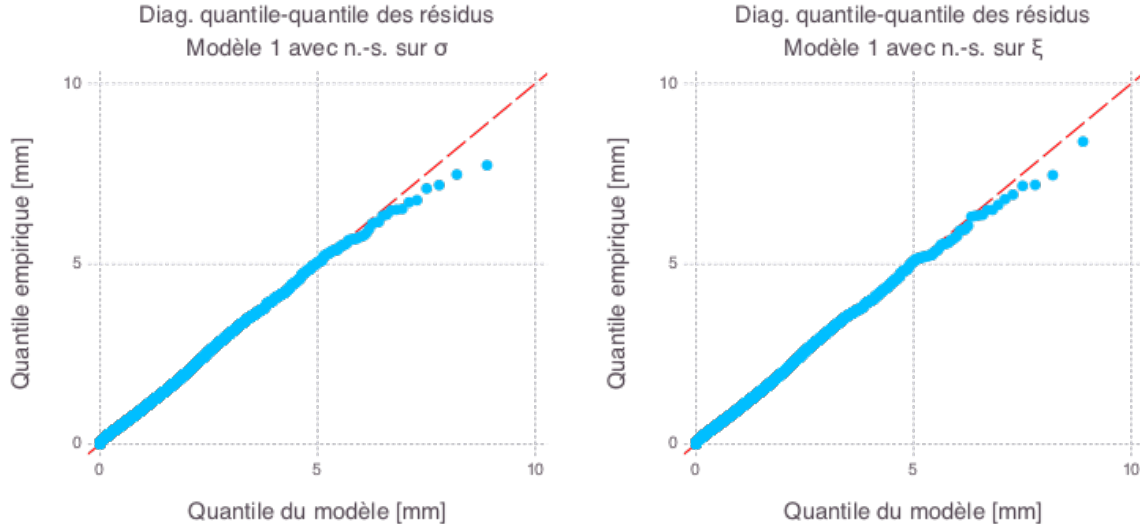


Figure 5.8 Diagrammes quantile-quantile des résidus des modèles non stationnaires à l'échelle exponentielle. À gauche : σ dépend de la concentration en CO_2 . À droite : ξ dépend de la concentration en CO_2 .

5.2.9 Analyse des 50 membres

Les résultats précédents ont été obtenus pour le membre KDI seulement. Pour s'assurer que le modèle non stationnaire sélectionné est le modèle le mieux adapté pour l'ensemble des 50 membres, le processus de sélection avec le BIC parmi les huit possibilités de modèles a été répété pour chaque membre sur la période 1955-2099.

La figure 5.9 présente la fréquence de sélection par BIC des modèles. On remarque que le modèle 1 non stationnaire pour lequel seul ξ dépend de la concentration en CO_2 est favorisé pour la majorité des membres (35 fois sur 50). Le modèle stationnaire, qui est un sous-modèle de $\xi(t)$, est favorisé 10 fois sur 50 alors que le modèle 1 avec la non stationnaire sur σ ne l'est que 5 fois sur 50.

En conclusion, le modèle qui servira à modéliser les précipitations simulées de 1955 à 2099 est le modèle 1 non stationnaire avec les paramètres suivants :

$$\begin{aligned}\sigma(t) &= \sigma \\ \xi(t) &= \xi_0 + \xi_1 \times \text{CO}_2(t) \\ \kappa(t) &= \kappa\end{aligned}\tag{5.5}$$

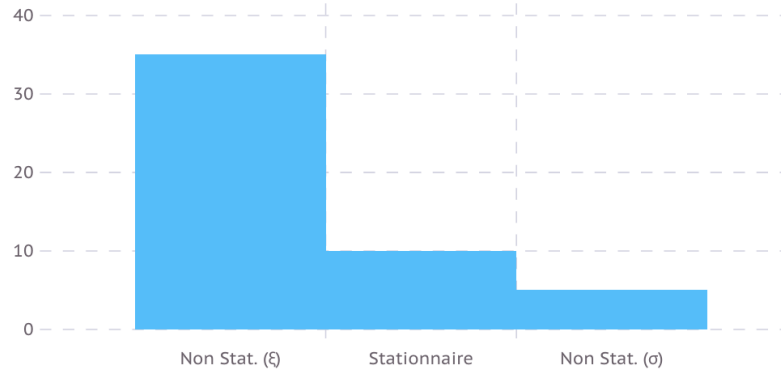


Figure 5.9 Fréquence de sélection avec le BIC des huit modèles.

5.3 Développement de la méthode de post-traitement

Dans les sections précédentes, la modélisation paramétrique des précipitations observées et simulées a été déterminée. Cette section porte sur le troisième objectif du projet qui est de développer une méthode de post-traitement paramétrique des précipitations à partir de ces modèles en mettant en correspondance leurs quantiles. La méthode de post-traitement est constituée de trois étapes. D'abord, les modèles définis aux sections précédentes sont ajustés sur les données observées et simulées. Ensuite, leur fonction de répartition respective est mise en correspondance selon l'équation (3.19) présentée à la section 3.3.1, *Méthode de correction par quantiles*. Finalement, pour chaque valeur à corriger, la valeur de correction est obtenue en calculant le quantile correspondant.

5.3.1 Ajustement des modèles

Soit Y la variable de référence, dans le cas présent, la précipitation observée et X la variable qui doit être corrigée, dans le cas présent, la précipitation simulée par un modèle numérique de climat. La première étape consiste à modéliser ces variables en ajustant les modèles des sections précédentes. Le modèle 1 est ajusté sur les précipitations observées puis le modèle non stationnaire est ajusté sur les précipitations simulées.

La modélisation des précipitations observées permet d'obtenir F_Y , la fonction de répartition des données observées pour une période de calibration. Pour ce qui est des précipitations simulées, celles-ci sont modélisées avec un modèle non stationnaire. Conséquemment, la fonction de répartition F_{X_i} va dépendre de la valeur de la variable explicative. Dans le cas présent, puisqu'il s'agit de la valeur de concentration de CO_2 et que celle-ci dépend de l'année, la fonction de répartition est différente pour chaque année i .

5.3.2 Mise en correspondance des fonctions de répartition

L'étape suivante consiste à mettre en correspondance les fonctions de répartition. Pour les années comprises dans la période de calibration, la fonction de répartition des données observées F_Y peut être obtenue directement. Pour la période projetée, comme on est en dehors de la période d'observation, la fonction de répartition des données observées F_{Y_i} est inconnue car il s'agit d'une période future. Elle est obtenue au moyen de la fonction de transfert proposée par [32] :

$$F_{Y_i}(\tilde{x}_{ij}) = F_{Y_m} \left[F_{X_m}^{-1} \{ F_{X_i}(x_{ij}) \} \right] \quad (5.6)$$

Comme on choisit de post-traiter à partir de la dernière année d'observation m , on a alors deux possibilités pour la fonction de répartition des données observées F_{Y_i} :

$$F_{Y_i}(\tilde{x}_{ij}) = \begin{cases} F_Y(\tilde{x}_{ij}) & \text{si } i \leq m; \\ F_{Y_m} \left[F_{X_m}^{-1} \{ F_{X_i}(x_{ij}) \} \right] & \text{si } i > m. \end{cases} \quad (5.7)$$

Les fonctions de répartition peuvent ainsi être mises en correspondance directement de la façon suivante :

$$F_{Y_i}(\tilde{x}_{ij}) = F_{X_i}(x_{ij}) \quad (5.8)$$

avec F_{X_i} qui va changer selon la concentration en CO_2 de l'année i considérée.

5.3.3 Correction des valeurs

La correction des précipitations simulées peut s'effectuer à partir de la correspondance des fonctions de répartition établie à l'équation (5.8). Pour chaque valeur simulée, la valeur corrigée est obtenue à partir de la fonction de quantile, ou l'inverse de la fonction de répartition.

La valeur post-traitée $\tilde{x}_{ij}$ qui correspond à la valeur simulée x_{ij} pour le jour j de l'année i s'obtient de la façon suivante :

$$\tilde{x}_{ij} = F_{Y_i}^{-1} \{ F_{X_i}(x_{ij}) \} \quad (5.9)$$

où $F_{Y_i}^{-1}$ est obtenue en inversant la fonction exprimée à l'équation (5.7). À la fin de cette étape, la série des valeurs de précipitations corrigées est obtenue. À titre d'exemple, la figure 5.10 présente les maxima annuels de la simulation corrigée en appliquant la méthode développée

par rapport à ceux de la simulation brute pour le membre KDI de 1955 à 2013.

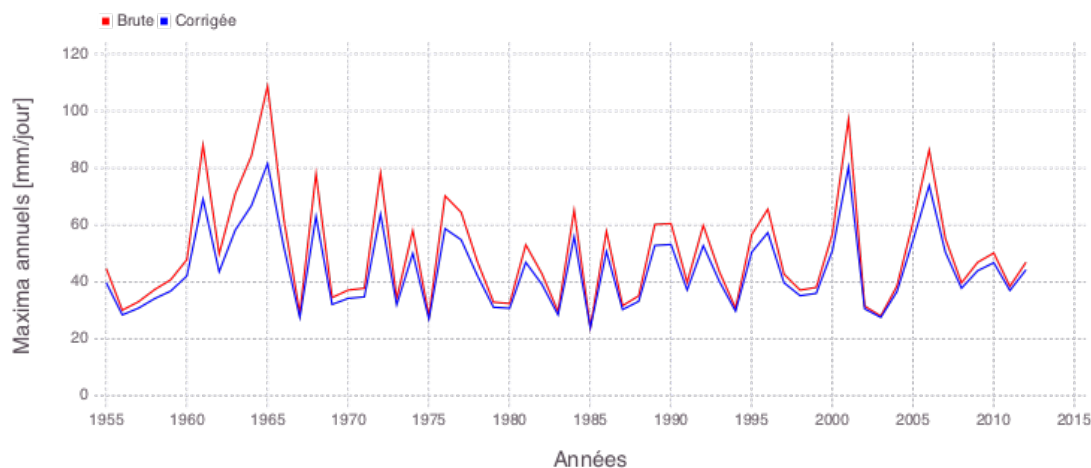


Figure 5.10 Maxima annuels de la simulation corrigée (bleu) par rapport à la simulation brute (rouge) pour le membre KDI de 1955 à 2013.

5.4 Évaluation de la méthode

Cette section présente le dernier objectif du projet qui est d'évaluer la méthode de post-traitement développée dans le but de s'assurer de son bon fonctionnement. L'évaluation de la méthode proposée pour le post-traitement des précipitations simulées s'effectue en deux étapes. D'abord, les indices climatiques présentés à la section 3.4.1 sont calculés sur des observations de précipitations et sur les simulations post-traitées. Ensuite, ces indices sont comparés à l'aide des mesures de performance présentées à la section 3.4.2.

Puisque la méthode est validée par rapport aux données observées, la correction des simulations est effectuée pour la période observée seulement vu l'absence d'observations auxquelles se comparer pour la période future. Les données de précipitations observées et simulées pour une période allant de 1955 à 2013 sont utilisées pour l'évaluation de la méthode.

5.4.1 Résultats pour le membre KDI

Les indices de validation sont calculés sur les données observées et sur le membre KDI avant et après l'application de la méthode de post-traitement développée. Les résultats ainsi que les valeurs de l'indice de comparaison sont présentés dans le tableau 5.6.

Avant la correction, on remarque que la simulation brute présente un certain biais par rapport

Tableau 5.6 Valeurs des indices calculés sur les observations, le membre KDI avant le post-traitement (sim. brute) et après le post-traitement (sim. corrigée) et de l'indice de comparaison (skill score).

Indice	Observations	Sim. brute	Sim. corrigée	Skill Score
Skew	2.3963	3.0722	2.44605	0.99
R10	0.29748	0.271385	0.302918	0.96
SDII	9.02241	8.68228	9.10673	0.94
p90wet	20.52	21.0298	21.1566	-0.56
p98wet	37.284	38.3117	36.0794	-0.37
p98wetamount	47.913	54.9449	48.0994	1.0
RV20max	65.6944	90.5481	72.1404	0.93
RV100max	83.8764	133.151	95.6269	0.94

aux observations pour les huit indices. Le coefficient d'asymétrie indique que la queue à droite est plus lourde pour la simulation brute que pour les observations. Le R10 des observations indique que les jours avec des précipitations supérieures à 10mm sont plus fréquents en réalité que ce qui est calculé sur la simulation brute. Le SDII indique que les précipitations moyennes de la simulation brute sont moins élevées que pour les observations. Finalement, les quantiles empiriques, p90wet et p98wet, le p98wetamount et les niveaux de retour 20 ans et 100 ans sont tous plus élevés pour la simulation brute.

Après l'application de la méthode, on remarque que les indices calculés sur la simulation corrigée sont beaucoup plus près de ceux calculés sur les observations à l'exception du p90wet et p98wet. En effet, l'indice de comparaison montre une amélioration par rapport à la simulation brute pour six des huit indices.

Pour mieux illustrer la performance de la méthode, la différence relative entre les indices calculés sur les observations et la simulation brute et la simulation corrigée est présentée sur la figure 5.11.

Pour les six indices qui montrent une amélioration par rapport à la simulation brute, on remarque une diminution importante de l'écart relatif. De plus, pour certains indices tels que le SDII et p98wetamount, la méthode de post-traitement a été en mesure de réduire presque entièrement l'écart relatif.

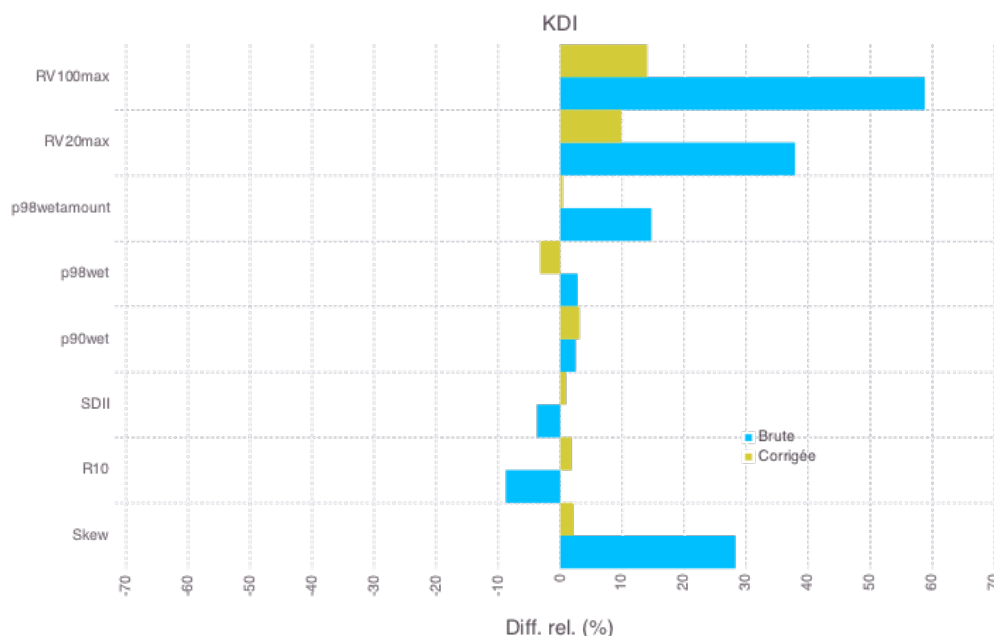


Figure 5.11 Différence relative (%) entre les indices calculés sur les observations et la simulation KDI brute (bleu) et la simulation KDI corrigée (or).

5.4.2 Résultats pour les 50 membres

La même procédure de validation a été répétée pour les 50 membres de ClimEx. Les figures présentant la différence relative entre les indices calculés sur les observations et les simulations brute et corrigée (reproduction de la figure 5.11 pour chacun des membres) sont présentées à l'Annexe A. Dans l'ensemble, la méthode de correction parvient à réduire l'écart entre l'indice calculé sur les observations et l'indice calculé sur la simulation pour la plupart des 50 membres.

Afin de quantifier cette réduction, la moyenne sur les 50 membres de la différence relative entre les indices calculés sur les observations et les simulations brutes et corrigées par la méthode développée a été calculée et est présentée dans le tableau 5.7. On y retrouve également ces mêmes résultats obtenus avec la méthode de correction empirique EQM pour des fins de comparaison. Pour les différencier, la méthode développée dans ce rapport est appelée EGPQM dans le tableau 5.7 pour *Extended Generalized Pareto Quantile Mapping*. Ce tableau présente aussi la fréquence d'amélioration par rapport à la simulation brute calculée sur les 50 membres pour les deux méthodes de correction, c'est-à-dire, le nombre de membres pour lesquels l'indice de comparaison est positif sur les 50 membres, pour chaque indice de validation.

Tableau 5.7 Différence relative moyenne en % (sur 50 membres) entre les indices calculés sur les observations et la simulation brute (Sim. brute), la méthode développée (EGPQM) et la méthode empirique (EQM) et fréquences d'amélioration du biais (sur 50 membres).

Indice	Diff. rel. moyenne (%)			Fréq. d'amélioration	
	Sim. brute	EGPQM	EQM	EGPQM	EQM
Skew	21.1	4.08	12.0	46 / 50	44 / 50
R10	-2.63	1.97	-1.91	32 / 50	32 / 50
SDII	0.42	0.938	-2.74	33 / 50	5 / 50
p90wet	5.37	1.23	-1.23	46 / 50	44 / 50
p98wet	10.6	-0.151	0.707	47 / 50	49 / 50
p98amount	14.7	0.518	3.02	50 / 50	50 / 50
RV20max	27.5	7.18	11.5	50 / 50	50 / 50
RV100max	32.3	6.67	15.8	46 / 50	44 / 50

Par rapport à la différence relative calculée sur les simulations brutes, on constate une réduction significative de l'écart avec la méthode EGPQM pour tous les indices à l'exception du SDII. Toutefois, cet indice semble peu biaisé par rapport aux observations, son écart relatif n'est que de 0,42 % en moyenne.

Quant aux résultats obtenus avec la méthode EQM, la différence relative moyenne est plus petite avec la méthode développée pour tous les indices à l'exception du R10. Finalement, pour ce qui est des indices de comparaison, la méthode développée a une fréquence d'amélioration supérieure ou égale à EQM pour tous les indices à l'exception du p98wet.

Finalement, la figure 5.12 présente l'enveloppe des maxima annuels des simulations corrigées (bleu) par rapport aux simulations brutes (rouge) pour les 50 membres de 1955 à 2013. Les lignes pleines représentent la médiane des maxima annuels pour chaque ensemble. On remarque que l'enveloppe des maxima annuels corrigés est moins large que celle des simulations brutes. Également, les valeurs maximales de l'enveloppe ainsi que la médiane sont moins élevées.

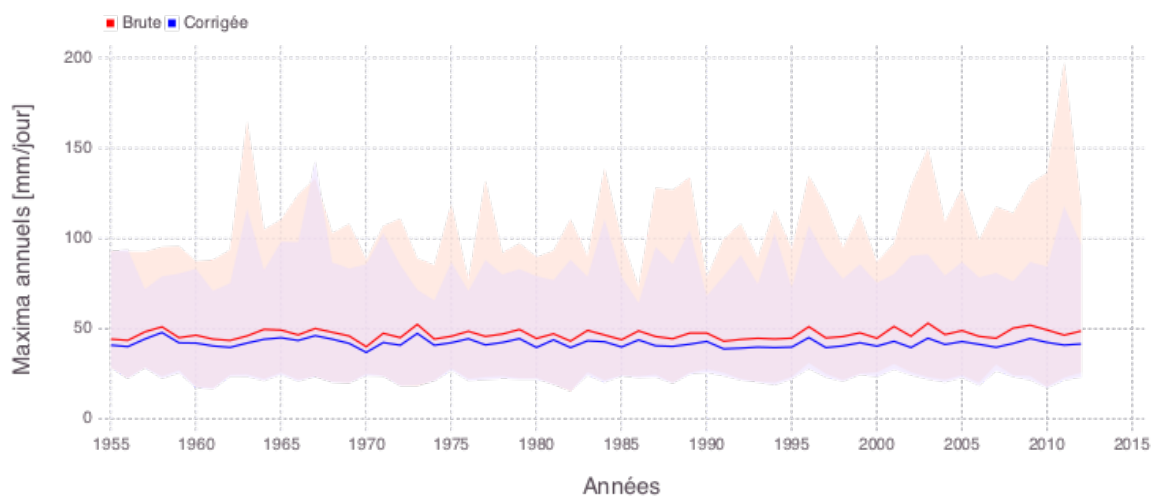


Figure 5.12 Maxima annuels des simulations corrigées (bleu) par rapport aux simulations brutes (rouge). Les lignes pleines représentent la médiane des maxima annuels des 50 membres pour chaque ensemble.

CHAPITRE 6 DISCUSSION

Les résultats de la validation montrent que la méthode développée parvient à reproduire les statistiques observées en ce qui a trait aux indices de validation sélectionnés. Ces indices ont été sélectionnés pour représenter le cœur de la distribution ainsi que les extrêmes. À cet égard, l'objectif de développer une méthode de post-traitement paramétrique adaptée pour les valeurs extrêmes et le cœur de la distribution est atteint. Toutefois, les indices utilisés pour la validation considèrent seulement les aspects marginaux des distributions. Pour valider les aspects temporels tels l'autocorrélation, la saisonnalité, etc., d'autres indices devraient être calculés. Par exemple, le *WetAnnualMaxSpell*, défini comme la médiane des maxima annuels de jours de pluie consécutifs, pourrait servir à évaluer l'habileté de la méthode à corriger des épisodes de précipitations qui s'étendent sur plusieurs jours. Cela met en lumière certains inconvénients de la méthode de correction proposée. Entre autres, le domaine d'application de la méthode proposée est limité à la correction des jours de pluie puisque les modèles paramétriques utilisés pour la correction ne s'appliquent qu'aux valeurs strictement positives. Conséquemment, la méthode de correction n'est pas en mesure de corriger la fréquence des jours secs et humides. L'utilisation de la méthode développée est donc réservée aux nombreux usages qui nécessitent des séries d'intensités de précipitations post-traitées. Cette méthode de post-traitement n'est donc pas recommandée pour l'étude des périodes de sécheresse par exemple.

Les indices de validation ont été calculés pour l'ensemble de la période de validation. Il serait cependant intéressant de considérer les valeurs annuelles des indicateurs pour obtenir des séries temporelles d'indicateurs. Cela permettrait notamment de caractériser l'évolution de certains aspects des précipitations. De plus, le post-traitement des 50 membres de ClimEx a été fait dans le but d'évaluer la méthode pour 50 simulations avec les mêmes observations de référence. En appliquant le post-traitement à chaque simulation séparément, la variabilité de l'ensemble est réduite et diminue son intérêt. D'ailleurs, cet effet est observable par la faible variabilité de la médiane sur la figure 5.12. Il serait préférable de reproduire l'analyse pour d'autres points de grille pour s'assurer que les résultats sont généralisables.

Un facteur à considérer pour l'application de la méthode de post-traitement proposée est la qualité de l'ajustement des modèles paramétriques. Pour cette étude, un ajustement adéquat a été obtenu après avoir établi une censure à 5 mm. Ce seuil de censure a été établi en fonction de la littérature disponible pour ce modèle. Des vérifications ont été effectuées afin de s'assurer qu'il convenait aux données. Pour l'application de la méthode à d'autres données,

ce seuil pourrait être différent et, par conséquent, il s'agit d'un hyperparamètre à déterminer selon l'application. Ce seuil peut s'avérer élevé dans la mesure où la contribution des valeurs inférieures à ce seuil est réduite dans le calcul de la vraisemblance. Conséquemment, une partie importante du total des précipitations est laissée en dehors de l'ajustement, pouvant réduire son adéquation dans la queue inférieure. Puisque le but est d'ajuster un modèle sur l'ensemble de la distribution, ce problème devra être pris en compte en établissant une façon d'estimer le seuil de censure selon les caractéristiques des précipitations de l'endroit à l'étude. Toutefois, il s'agit d'un compromis afin de s'assurer d'avoir un meilleur ajustement dans les extrêmes. Ce compromis est justifié par l'impact plus important des événements de précipitations extrêmes. Il est préférable d'avoir une meilleure adéquation de l'ajustement dans la queue à droite que dans la queue à gauche. Les résultats pour le R10 et le SDII, des indices sélectionnés pour représenter le coeur de la distribution, semblent être davantage affectés par la réduction de l'adéquation de l'ajustement dans la queue inférieure. En particulier, les résultats pour les 50 membres montrent une fréquence d'amélioration limitée de 32/50 et 33/50 de la correction pour du R10 et du SDII respectivement.

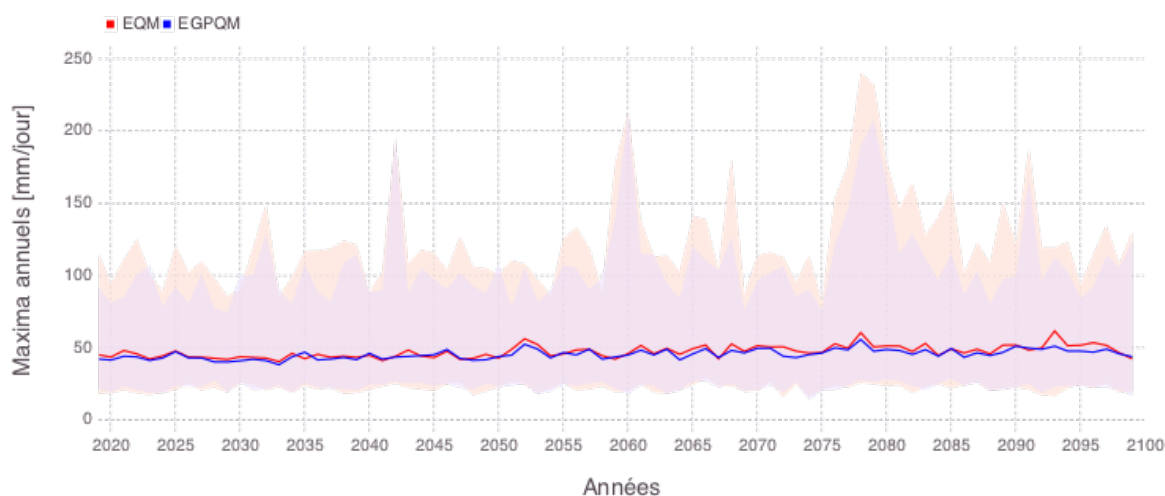


Figure 6.1 Maxima annuels des simulations corrigées par EGPQM (bleu) et par EQM (rouge). Les lignes pleines représentent la médiane des maxima annuels des 50 membres corrigés avec chaque méthode.

Finalement, rappelons que la méthode développée permet également de post-traiter les valeurs futures. Cet aspect n'a pu être validé en raison de l'absence d'observations de précipitations futures. Cependant, il peut être intéressant de regarder les résultats de la méthode pour le post-traitement des projections climatiques. La figure 6.1 présente les enveloppes des maxima annuels des simulations corrigées par la méthode développée (bleu) et par la méthode

EQM (rouge) pour les 50 membres de 2020 à 2099. Les lignes pleines représentent la médiane des maxima annuels des 50 membres corrigés avec chaque méthode. Il semble y avoir peu de variabilité et de différence entre les médianes obtenues pour chacune des méthodes. Toutefois, cette faible variabilité peut être attribuée à l'application du post-traitement à chaque simulation séparément. L'enveloppe des maxima annuels obtenue par la méthode développée est moins large que celle obtenue par EQM. Ces différences pourraient faire l'objet d'une étude plus approfondie puisque le choix de la méthode de post-traitement peut avoir des conséquences, notamment dans un cadre décisionnel.

CHAPITRE 7 CONCLUSION ET RECOMMANDATIONS

7.1 Synthèse des travaux

Dans ce mémoire, une méthode paramétrique de post-traitement des précipitations simulées par les modèles climatiques adaptée pour les valeurs extrêmes et le cœur de la distribution a été développée et validée.

Pour cela, une revue de littérature concernant les méthodes de post-traitement statistique et la modélisation des précipitations a d'abord été présentée. Également, la théorie des valeurs extrêmes ainsi que les techniques de correction de biais ont été introduites. Quatre modèles d'extension à la loi de Pareto généralisée ont été ajustés sur les précipitations observées à la station Montréal-Trudeau entre 1980 et 2009 afin d'établir une méthode de modélisation paramétrique des précipitations observées. Il a été déterminé que la famille 1 proposée par Naveau et al. (2016) est celle présentant la meilleure adéquation pour ces données. Puis, les mêmes quatre modèles ont été ajustés sur les précipitations simulées entre 1980 et 2009 par l'ensemble ClimEx. L'adéquation de cette famille pour ces données a été confirmée. Ensuite, une extension non stationnaire a été développée pour ce modèle afin de modéliser les précipitations simulées entre 1955 et 2009.

Il a été nécessaire de modéliser les précipitations observées et simulées pour obtenir les distributions à utiliser pour le post-traitement. Une fonction de transfert reliant ces distributions a été déterminée afin de développer la méthode paramétrique de post-traitement. Puis, la méthode développée a été validée par le calcul d'indices climatiques spécifiques au cœur et aux queues de distribution. Les résultats de la validation montrent une réduction significative du biais avec la méthode développée.

En somme, la réalisation de ce projet a permis d'ajouter une composante non stationnaire à un modèle de Naveau et al. (2016) et son adaptation au contexte du post-traitement des précipitations simulées afin de proposer une méthode de post-traitement paramétrique adaptée pour l'ensemble des valeurs strictement positives.

Ce projet a également permis le développement d'une librairie performante en Julia offrant des outils pour utiliser les extensions à la loi de Pareto généralisée ainsi que la méthode de post-traitement développée. La librairie développée est accessible à l'adresse suivante : <https://github.com/houton199/EGPD.jl>.

7.2 Limitations et améliorations futures

La méthodologie proposée dans ce mémoire comporte certaines limites. Tout d’abord, les lois EGPD utilisées pour le projet considèrent uniquement les valeurs strictement positives. Par conséquent, les valeurs nulles ne sont pas considérées dans la fonction de répartition. Cela fait en sorte que la fréquence des jours secs/humides n’est pas corrigée par la méthode développée car les jours sans précipitations ne sont pas modélisés. La méthode de post-traitement a donc un domaine d’application limité aux valeurs strictement positives. Autrement dit, la méthode de post-traitement ne corrige que les jours de pluie. De plus, dans notre cas, ce domaine a été réduit aux valeurs supérieures à 1 mm car les jours présentant une accumulation en dessous de 1 mm sont considérés comme des jours secs. L’ajustement de l’aspect temporel serait intéressant à considérer pour des travaux futurs en raison des effets importants que peuvent avoir de longs épisodes de précipitations, particulièrement sur le plan de l’accumulation totale.

Ensuite, la qualité de l’ajustement des modèles paramétriques est à considérer pour l’application de la méthode de post-traitement proposée. Pour cette étude, un ajustement adéquat a été obtenu après avoir établi une censure à 5 mm en fonction de la littérature disponible pour ce modèle. Ce seuil peut s’avérer élevé dans la mesure où la contribution des valeurs inférieures à ce seuil est réduite dans le calcul de la vraisemblance. Pour l’application de la méthode à d’autres données, ce seuil pourrait être différent et, par conséquent, il s’agit d’un hyperparamètre qui devra être déterminé selon l’application. Une façon d’estimer le seuil de censure selon les caractéristiques des précipitations de l’endroit à l’étude devra être déterminée.

Puis, il est important de préciser que les résultats présentés dépendent des membres choisis pour le post-traitement. Dans les cas où certains membres de ClimEx sont peu biaisés par rapport aux observations, l’application d’une méthode de correction ne serait pas nécessaire. Ainsi, il est possible que la méthode de post-traitement n’améliore pas la simulation. La validation, notamment la comparaison avec d’autres méthodes de post-traitement, devra être approfondie et l’analyse devra être répétée pour d’autres points de grille pour s’assurer que les résultats sont généralisables. Également, l’étude de l’effet des périodes utilisées pour la calibration serait intéressante.

Finalement, le modèle non stationnaire a été développé en utilisant comme variable explicative la concentration en CO_2 . D’autres options pourront être explorées dans le futur, par exemple, en utilisant d’autres relations ou d’autres choix de variables explicatives.

RÉFÉRENCES

- [1] R. E. Kass et A. E. Raftery, “Bayes factors,” *Journal of the American Statistical Association*, vol. 90, n°. 430, p. 773–795, 1995. [En ligne]. Disponible : <https://doi.org/10.1080/01621459.1995.10476572>
- [2] C. Teutschbein et J. Seibert, “Bias correction of regional climate model simulations for hydrological climate-change impact studies : Review and evaluation of different methods,” *Journal of Hydrology*, vol. 456-457, p. 12–29, 2012. [En ligne]. Disponible : <https://doi.org/10.1016/j.jhydrol.2012.05.052>
- [3] P. Naveau, R. Huser, P. Ribereau et A. Hannart, “Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection,” *Water Resources Research*, vol. 52, n°. 4, p. 2753–2769, 2016. [En ligne]. Disponible : <https://doi.org/10.1002/2015WR018552>
- [4] D. Maraun et M. Widmann, *Statistical Downscaling and Bias Correction for Climate Research*. Cambridge University Press, 2018.
- [5] IPCC, *Climate Change 2013 : The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, United Kingdom and New York, NY, USA : Cambridge University Press, 2013. [En ligne]. Disponible : <https://doi.org/10.1017/CBO9781107415324>
- [6] Ouranos, *Vers l’adaptation : synthèse des connaissances sur les changements climatiques au Québec*. Montréal, Québec : Ouranos, 2015.
- [7] D. Maraun, “Bias correcting climate change simulations - A critical review,” *Current Climate Change Reports*, vol. 2, n°. 4, p. 211–220, 2016. [En ligne]. Disponible : <https://doi.org/10.1007/s40641-016-0050-x>
- [8] I. Charron, *Guide sur les scénarios climatiques : Utilisation de l’information climatique pour guider la recherche et la prise de décision en matière d’adaptation*. Montréal, Québec : Ouranos, 2016.
- [9] V. Eyring, S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer et K. E. Taylor, “Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization,” *Geoscientific Model Development*, vol. 9, n°. 5, p. 1937–1958, 2016. [En ligne]. Disponible : <https://doi.org/10.5194/gmd-9-1937-2016>
- [10] R. Knutti, D. Masson et A. Gettelman, “Climate model genealogy : Generation CMIP5 and how we got there,” *Geophysical Research Letters*, vol. 40, n°. 6, p. 1194–1199, 2013. [En ligne]. Disponible : <https://doi.org/10.1002/grl.50256>

- [11] R. Laprise, “Regional climate modelling,” *Journal of Computational Physics*, vol. 227, n°. 7, p. 3641–3666, 2008, predicting weather, climate and extreme events. [En ligne]. Disponible : <https://doi.org/10.1016/j.jcp.2006.10.024>
- [12] L. R. Leung, *Regional Climate Models*. New York, NY : Springer New York, 2012, p. 8902–8919. [En ligne]. Disponible : https://doi.org/10.1007/978-1-4419-0851-3_363
- [13] S. Jeon, C. J. Paciorek et M. F. Wehner, “Quantile-based bias correction and uncertainty quantification of extreme event attribution statements,” *Weather and Climate Extremes*, vol. 12, p. 24–32, 2016. [En ligne]. Disponible : <https://doi.org/10.1016/j.wace.2016.02.001>
- [14] A. J. Cannon, “Multivariate quantile mapping bias correction : an N-dimensional probability density function transform for climate model simulations of multiple variables,” *Climate Dynamics*, vol. 50, n°. 1-2, p. 31–49, 2017. [En ligne]. Disponible : <https://doi.org/10.1007/s00382-017-3580-6>
- [15] H. Panofsky et G. Brier, *Some Applications of Statistics to Meteorology*. Earth and Mineral Sciences Continuing Education, College of Earth and Mineral Sciences, The Pennsylvania State University, 1968.
- [16] D. Huard, D. Chaumont, T. Logan, M.-F. Sottile, R. D. Brown, B. G. St-Denis, P. Grenier et M. Braun, “A decade of climate scenarios : The Ouranos consortium modus operandi,” *Bulletin of the American Meteorological Society*, vol. 95, n°. 8, p. 1213 – 1225, 2014. [En ligne]. Disponible : <https://doi.org/10.1175/BAMS-D-12-00163.1>
- [17] T. Q. Murdock, S. R. Sobie et A. J. Cannon, “Bias correction of GCM precipitation by quantile mapping : How well do methods preserve changes in quantiles and extremes?” *Journal of Climate*, vol. 28, n°. 17, p. 6938–6959, 2015. [En ligne]. Disponible : <https://doi.org/10.1175/JCLI-D-14-00754.1>
- [18] S. Coles, *An Introduction to Statistical Modeling of Extreme Values*, ser. Springer Series in Statistics. London : Springer-Verlag, 2001.
- [19] J. Bezanson, A. Edelman, S. Karpinski et V. B. Shah, “Julia : A fresh approach to numerical computing,” *SIAM Review*, vol. 59, n°. 1, p. 65–98, 2017. [En ligne]. Disponible : <https://doi.org/10.1137/141000671>
- [20] A. F. Prein, W. Langhans, G. Fosser, A. Ferrone, N. Ban, K. Goergen, M. Keller, M. Tölle, O. Gutjahr, F. Feser, E. Brisson, S. Kollet, J. Schmidli, N. P. M. van Lipzig et R. Leung, “A review on regional convection-permitting climate modeling : Demonstrations, prospects, and challenges,” *Reviews of Geophysics*, vol. 53, n°. 2, p. 323–361, 2015. [En ligne]. Disponible : <https://doi.org/10.1002/2014RG000475>

- [21] G. Fosser, E. J. Kendon, D. Stephenson et S. Tucker, “Convection-permitting models offer promise of more certain extreme rainfall projections,” *Geophysical Research Letters*, vol. 47, n^o. 13, p. e2020GL088151, 2020, e2020GL088151 2020GL088151. [En ligne]. Disponible : <https://doi.org/10.1029/2020GL088151>
- [22] D. Maraun, F. Wetterhall, A. M. Ireson, R. E. Chandler, E. J. Kendon, M. Widmann, S. Brien, H. W. Rust, T. Sauter, M. Themeßl, V. K. C. Venema, K. P. Chun, C. M. Goodess, R. G. Jones, C. Onof, M. Vrac et I. Thiele-Eich, “Precipitation downscaling under climate change : Recent developments to bridge the gap between dynamical models and the end user,” *Reviews of Geophysics*, vol. 48, n^o. 3, 2010. [En ligne]. Disponible : <https://doi.org/10.1029/2009RG000314>
- [23] H. R. Glahn et D. A. Lowry, “The use of Model Output Statistics (MOS) in objective weather forecasting,” *Journal of Applied Meteorology (1962-1982)*, vol. 11, n^o. 8, p. 1203–1211, 1972. [En ligne]. Disponible : <http://www.jstor.org/stable/26176961>
- [24] W. H. Klein et H. R. Glahn, “Forecasting local weather by means of Model Output Statistics,” *Bulletin of the American Meteorological Society*, vol. 55, n^o. 10, p. 1217–1227, 1974. [En ligne]. Disponible : <http://www.jstor.org/stable/26255464>
- [25] G. Lenderink, A. Buishand et W. van Deursen, “Estimates of future discharges of the river Rhine using two scenario methodologies : direct versus delta approach,” *Hydrology and Earth System Sciences*, vol. 11, n^o. 3, p. 1145–1159, 2007. [En ligne]. Disponible : <https://doi.org/10.5194/hess-11-1145-2007>
- [26] M. Widmann et C. S. Bretherton, “Validation of mesoscale precipitation in the NCEP reanalysis using a new gridcell dataset for the Northwestern United States,” *Journal of Climate*, vol. 13, n^o. 11, p. 1936 – 1950, 2000. [En ligne]. Disponible : [https://doi.org/10.1175/1520-0442\(2000\)013<1936:VOMPIT>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<1936:VOMPIT>2.0.CO;2)
- [27] M. Jakob Themeßl, A. Gobiet et A. Leuprecht, “Empirical-statistical downscaling and error correction of daily precipitation from regional climate models,” *International Journal of Climatology*, vol. 31, n^o. 10, p. 1530–1544, 2011. [En ligne]. Disponible : <https://doi.org/10.1002/joc.2168>
- [28] C. Piani, J. O. Haerter et E. Coppola, “Statistical bias correction for daily precipitation in regional climate models over Europe,” *Theoretical and Applied Climatology*, vol. 99, n^o. 1, p. 187–192, 2010. [En ligne]. Disponible : <https://doi.org/10.1007/s00704-009-0134-9>
- [29] P.-A. Michelangeli, M. Vrac et H. Loukos, “Probabilistic downscaling approaches : Application to wind cumulative distribution functions,” *Geophysical Research Letters*, vol. 36, n^o. 11, 2009. [En ligne]. Disponible : <https://doi.org/10.1029/2009GL038401>

- [30] E. Hertig, D. Maraun, J. Bartholy, R. Pongracz, M. Vrac, I. Mares, J. M. Gutiérrez, J. Wibig, A. Casanueva et P. M. M. Soares, “Comparison of statistical downscaling methods with respect to extreme events over Europe : Validation results from the perfect predictor experiment of the COST Action VALUE,” *International Journal of Climatology*, vol. 39, n°. 9, p. 3846–3867, 2019. [En ligne]. Disponible : <https://doi.org/10.1002/joc.5469>
- [31] D. Maraun, T. G. Shepherd, M. Widmann, G. Zappa, D. Walton, J. M. Gutiérrez, S. Hagemann, I. Richter, P. M. M. Soares, A. Hall et L. O. Mearns, “Towards process-informed bias correction of climate change simulations,” *Nature Climate Change*, vol. 7, n°. 11, p. 764–773, 2017. [En ligne]. Disponible : <https://doi.org/10.1038/nclimate3418>
- [32] M. Kallache, M. Vrac, P. Naveau et P.-A. Michelangeli, “Nonstationary probabilistic downscaling of extreme precipitation,” *Journal of Geophysical Research : Atmospheres*, vol. 116, n°. D5, 2011. [En ligne]. Disponible : <https://doi.org/10.1029/2010JD014892>
- [33] O. Gutjahr et G. Heinemann, “Comparing precipitation bias correction methods for high-resolution regional climate simulations using COSMO-CLM,” *Theoretical and Applied Climatology*, vol. 114, n°. 3, p. 511–529, 2013. [En ligne]. Disponible : <https://doi.org/10.1007/s00704-013-0834-z>
- [34] J. Li, A. Sharma, J. Evans et F. Johnson, “Addressing the mischaracterization of extreme rainfall in regional climate model simulations – A synoptic pattern based bias correction approach,” *Journal of Hydrology*, vol. 556, p. 901–912, 2018. [En ligne]. Disponible : <https://doi.org/10.1016/j.jhydrol.2016.04.070>
- [35] D. Maraun, M. Widmann, J. M. Gutiérrez, S. Kotlarski, R. E. Chandler, E. Hertig, J. Wibig, R. Huth et R. A. Wilcke, “VALUE : A framework to validate downscaling approaches for climate change studies,” *Earth’s Future*, vol. 3, n°. 1, p. 1–14, 2015. [En ligne]. Disponible : <https://doi.org/10.1002/2014EF000259>
- [36] J. M. Gutiérrez, D. Maraun, M. Widmann, R. Huth, E. Hertig, R. Benestad, O. Roessler, J. Wibig, R. Wilcke, S. Kotlarski, D. San Martín, S. Herrera, J. Bedia, A. Casanueva, R. Manzananas, M. Iturbide, M. Vrac, M. Dubrovsky, J. Ribalaygua, J. Pórtolés, O. Räty, J. Räisänen, B. Hingray, D. Raynaud, M. J. Casado, P. Ramos, T. Zerenner, M. Turco, T. Bosshard, P. Štěpánek, J. Bartholy, R. Pongracz, D. E. Keller, A. M. Fischer, R. M. Cardoso, P. M. M. Soares, B. Czernecki et C. Pagé, “An intercomparison of a large ensemble of statistical downscaling methods over Europe : Results from the VALUE perfect predictor cross-validation experiment,” *International Journal of Climatology*, vol. 39, n°. 9, p. 3750–3785, 2019. [En ligne]. Disponible : <https://doi.org/10.1002/joc.5462>

- [37] A. Casanueva, S. Herrera, M. Iturbide, S. Lange, M. Jury, A. Dosio, D. Maraun et J. M. Gutiérrez, “Testing bias adjustment methods for regional climate change applications under observational uncertainty and resolution mismatch,” *Atmospheric Science Letters*, vol. 21, n°. 7, p. e978, 2020. [En ligne]. Disponible : <https://doi.org/10.1002/asl.978>
- [38] M. Vrac et P. Naveau, “Stochastic downscaling of precipitation : From dry events to heavy rainfalls,” *Water Resources Research*, vol. 43, n°. 7, 2007. [En ligne]. Disponible : <https://doi.org/10.1029/2006WR005308>
- [39] J. Carreau et Y. Bengio, “A hybrid Pareto model for asymmetric fat-tailed data : the univariate case,” *Extremes*, vol. 12, n°. 1, p. 53–76, 2009. [En ligne]. Disponible : <https://doi.org/10.1007/s10687-008-0068-0>
- [40] J. Carreau et M. Vrac, “Stochastic downscaling of precipitation with neural network conditional mixture models,” *Water Resources Research*, vol. 47, n°. 10, 2011. [En ligne]. Disponible : <https://doi.org/10.1029/2010WR010128>
- [41] I. Papastathopoulos et J. A. Tawn, “Extended generalised Pareto models for tail estimation,” *Journal of Statistical Planning and Inference*, vol. 143, n°. 1, p. 131–143, 2013. [En ligne]. Disponible : <https://doi.org/10.1016/j.jspi.2012.07.001>
- [42] G.-J. Proulx, *Standard Dictionary of Meteorological Sciences : English-French / French-English*. McGill-Queen’s University Press, 1971. [En ligne]. Disponible : <https://doi.org/10.2307/j.ctt1w6tgct>
- [43] M. Leduc, A. Mailhot, A. Frigon, J.-L. Martel, R. Ludwig, G. B. Brietzke, M. Giguère, F. Brissette, R. Turcotte, M. Braun et J. Scinocca, “The ClimEx Project : A 50-member ensemble of climate change projections at 12-km resolution over Europe and Northeastern North America with the Canadian Regional Climate Model (CRCM5),” *Journal of Applied Meteorology and Climatology*, vol. 58, n°. 4, p. 663 – 693, 01 Apr. 2019. [En ligne]. Disponible : <https://doi.org/10.1175/JAMC-D-18-0021.1>
- [44] A. Mailhot, S. Duchesne, D. Caya et G. Talbot, “Assessment of future change in intensity–duration–frequency (IDF) curves for Southern Quebec using the Canadian Regional Climate Model (CRCM),” *Journal of Hydrology*, vol. 347, n°. 1, p. 197–210, 2007. [En ligne]. Disponible : <https://doi.org/10.1016/j.jhydrol.2007.09.019>
- [45] L. Šeparović, A. Alexandru, R. Laprise, A. Martynov, L. Sushama, K. Winger, K. Tete et M. Valin, “Present climate and climate change over North America as simulated by the fifth-generation Canadian regional climate model,” *Climate Dynamics*, vol. 41, n°. 11, p. 3167–3201, 2013. [En ligne]. Disponible : <https://doi.org/10.1007/s00382-013-1737-5>
- [46] A. Martynov, R. Laprise, L. Sushama, K. Winger, L. Šeparović et B. Dugas, “Reanalysis-driven climate simulation over CORDEX North America domain using

- the Canadian Regional Climate Model, version 5 : model performance evaluation,” *Climate Dynamics*, vol. 41, n°. 11, p. 2973–3005, 2013. [En ligne]. Disponible : <https://doi.org/10.1007/s00382-013-1778-9>
- [47] J. Blanchet, E. Paquet, P. Vaittinada Ayar et D. Penot, “Mapping rainfall hazard based on rain gauge data : an objective cross-validation framework for model selection,” *Hydrology and Earth System Sciences*, vol. 23, n°. 2, p. 829–849, 2019. [En ligne]. Disponible : <https://doi.org/10.5194/hess-23-829-2019>
- [48] K. Hennessy, R. Suppiah et C. Page, “Australian rainfall changes, 1910-1995,” *Australian Meteorological Magazine*, vol. 48, n°. 1, p. 1 – 13, 1999.
- [49] X. Zhang, L. Alexander, G. C. Hegerl, P. Jones, A. K. Tank, T. C. Peterson, B. Trewin et F. W. Zwiers, “Indices for monitoring changes in extremes based on daily temperature and precipitation data,” *WIREs Climate Change*, vol. 2, n°. 6, p. 851–870, 2011. [En ligne]. Disponible : <https://doi.org/10.1002/wcc.147>
- [50] B. Efron, “Bootstrap methods : Another look at the jackknife,” *The Annals of Statistics*, vol. 7, n°. 1, p. 1–26, 1979. [En ligne]. Disponible : <https://doi.org/10.1214/aos/1176344552>
- [51] H. J. Fowler, D. Cooley, S. R. Sain et M. Thurston, “Detecting change in UK extreme precipitation using results from the climateprediction.net BBC climate change experiment,” *Extremes*, vol. 13, n°. 2, p. 241–267, févr. 2010. [En ligne]. Disponible : <https://doi.org/10.1007/s10687-010-0101-y>
- [52] É. Nolet-Gravel, “Changements projetés des précipitations extrêmes au Québec,” Mémoire de maîtrise, Polytechnique Montréal, 2019. [En ligne]. Disponible : <https://publications.polymtl.ca/4096/>

ANNEXE A RÉSULTATS POUR LES 50 MEMBRES

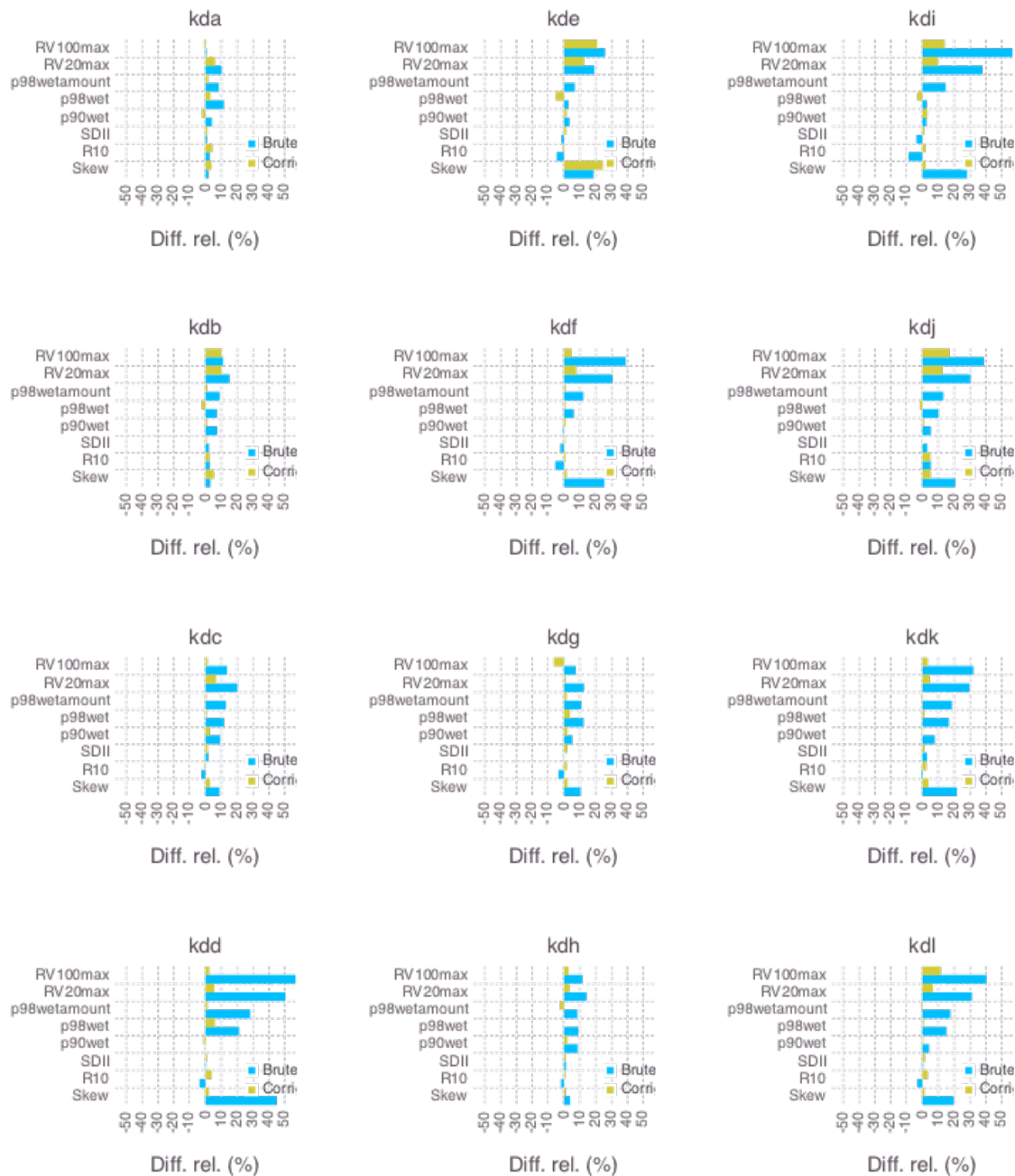


Figure A.1 Différence relative (%) entre les indices calculés sur les observations et la simulation brute (bleu) et la simulation corrigée (or) pour les membres KDA à KDL.

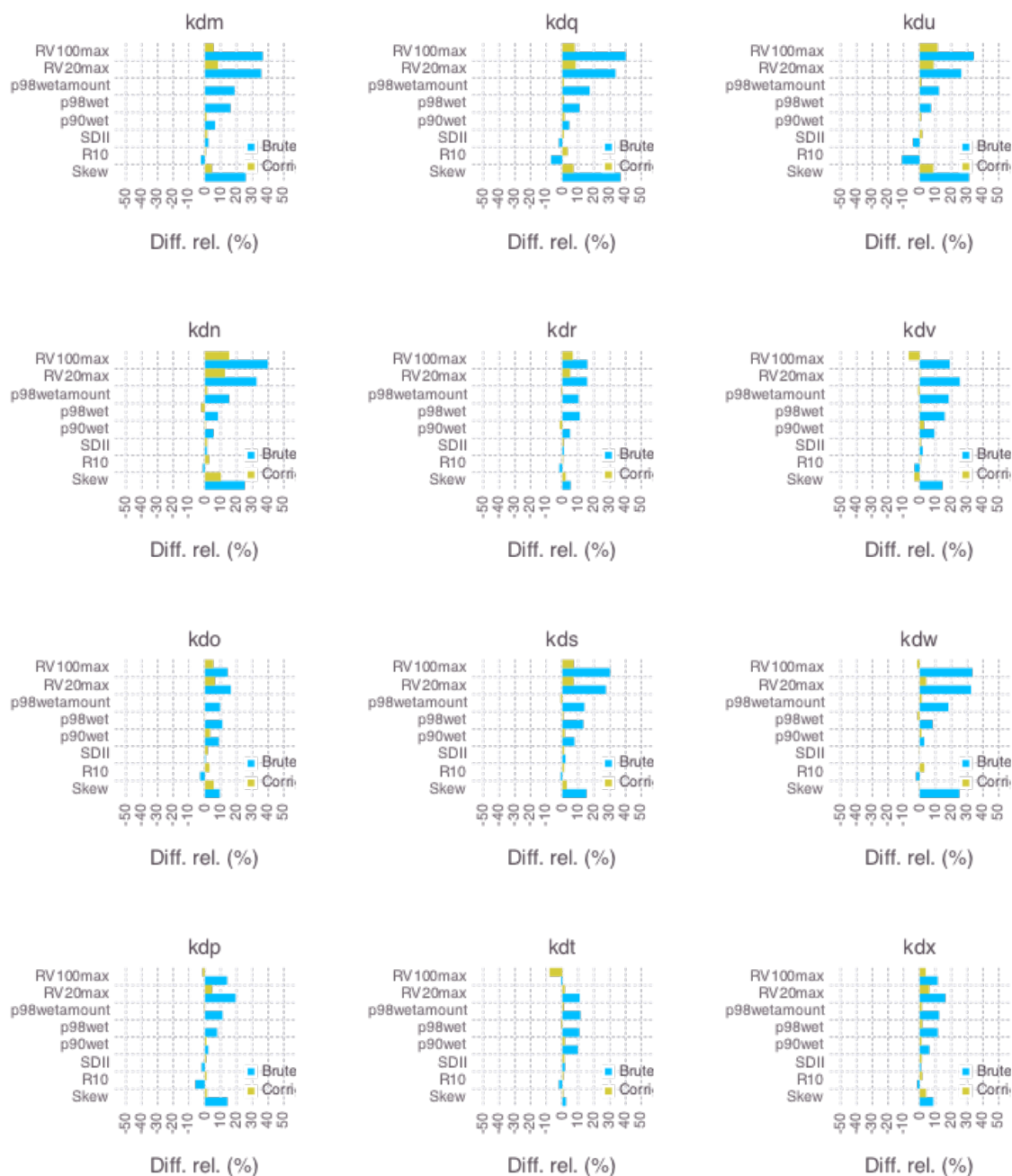


Figure A.2 Différence relative (%) entre les indices calculés sur les observations et la simulation brute (bleu) et la simulation corrigée (or) pour les membres KDM à KDX.

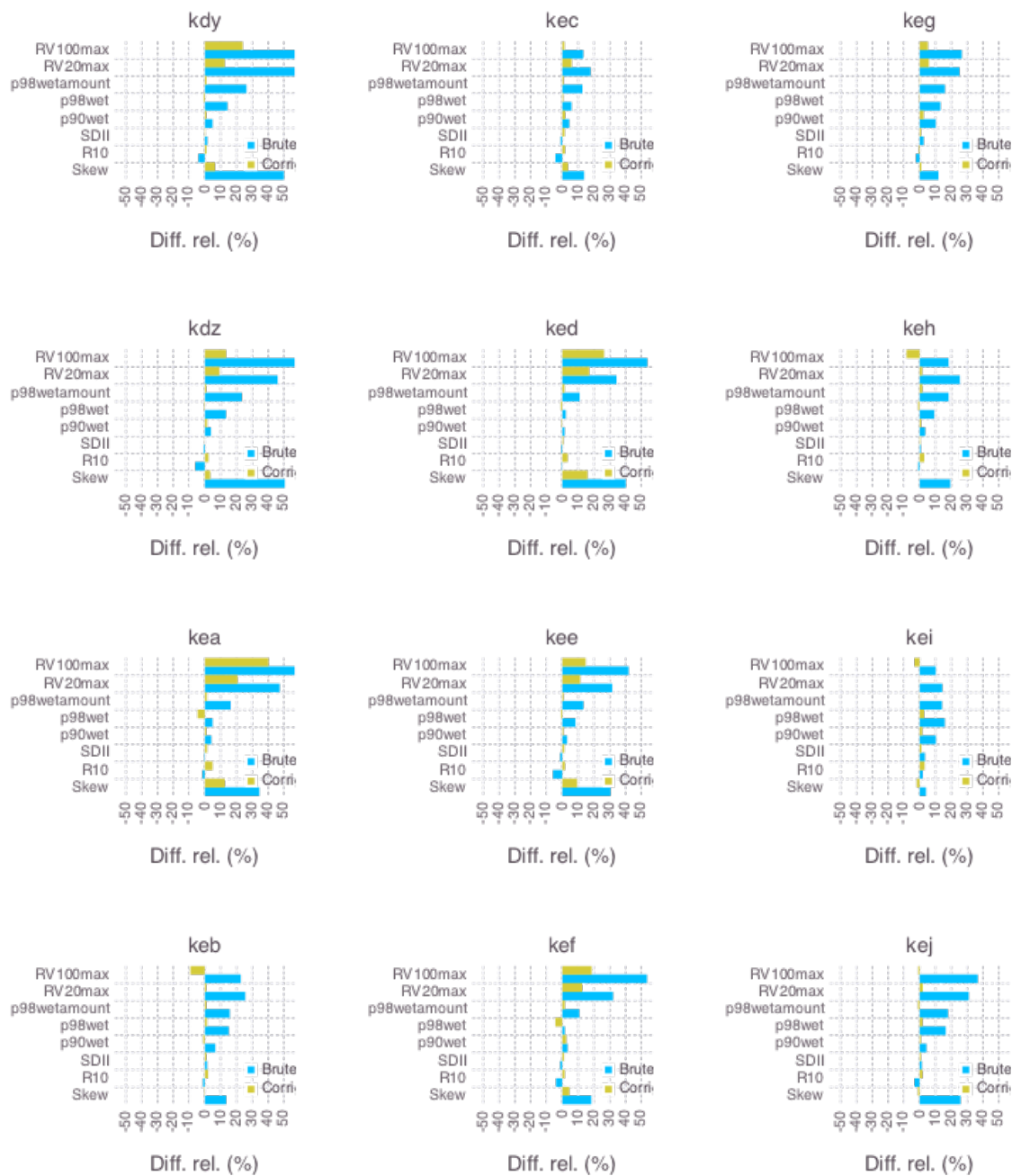


Figure A.3 Différence relative (%) entre les indices calculés sur les observations et la simulation brute (bleu) et la simulation corrigée (or) pour les membres KDY à KEJ.

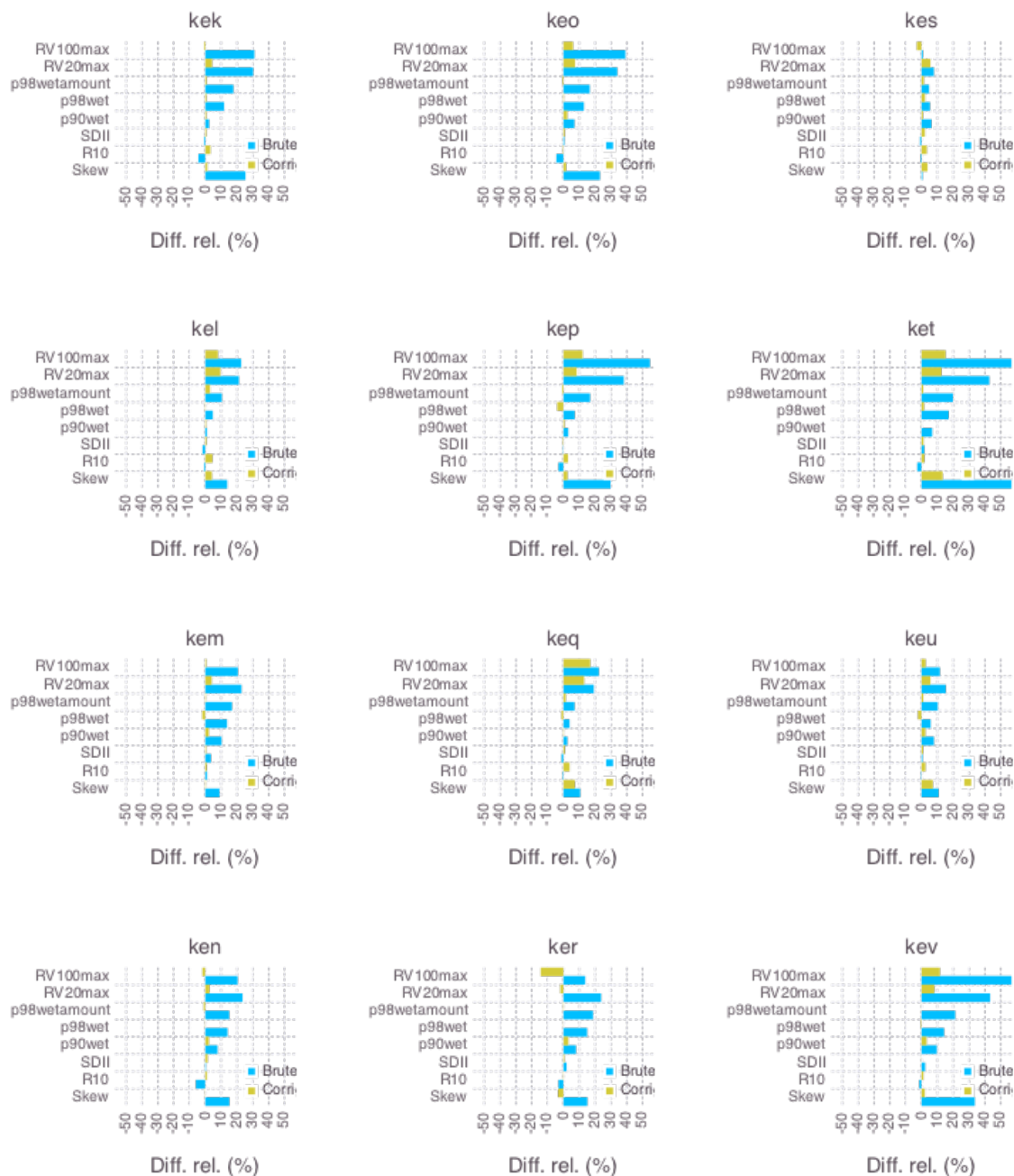


Figure A.4 Différence relative (%) entre les indices calculés sur les observations et la simulation brute (bleu) et la simulation corrigée (or) pour les membres KEK à KEV.

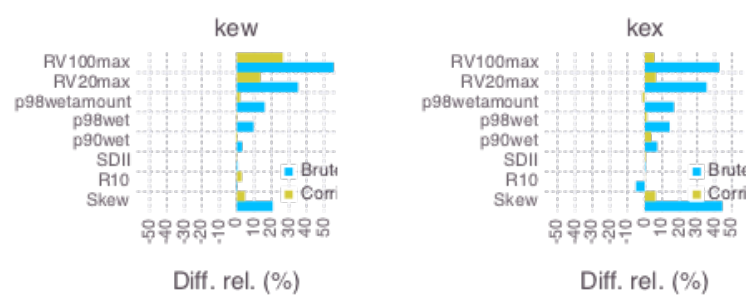


Figure A.5 Différence relative (%) entre les indices calculés sur les observations et la simulation brute (bleu) et la simulation corrigée (or) pour les membres KEW à KEX.