

Titre: Changements projetés des précipitations extrêmes au Québec
Title:

Auteur: Éloïse Nolet-Gravel
Author:

Date: 2019

Type: Mémoire ou thèse / Dissertation or Thesis

Référence: Nolet-Gravel, É. (2019). Changements projetés des précipitations extrêmes au Québec [Mémoire de maîtrise, Polytechnique Montréal]. PolyPublie.
Citation: <https://publications.polymtl.ca/4096/>

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

URL de PolyPublie: <https://publications.polymtl.ca/4096/>
PolyPublie URL:

**Directeurs de
recherche:** Jonathan Jalbert
Advisors:

Programme: Maîtrise recherche en mathématiques appliquées
Program:

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Changements projetés des précipitations extrêmes au Québec

ÉLOÏSE NOLET-GRAVEL

Département de mathématiques et de génie industriel

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

Mathématiques appliquées

Décembre 2019

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Ce mémoire intitulé :

Changements projetés des précipitations extrêmes au Québec

présenté par **Éloïse NOLET-GRAVEL**

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

a été dûment accepté par le jury d'examen constitué de :

Luc ADJENGUE, président

Jonathan JALBERT, membre et directeur de recherche

Richard LABIB, membre

DÉDICACE

*À Huguette,
une femme forte
et inspirante . . .*

REMERCIEMENTS

Ce mémoire est l’aboutissement d’un travail de longue haleine qui n’aurait pas été possible sans le soutien de mon entourage. C’est un grand honneur d’évoluer aux côtés de gens si passionnés et inspirants.

Tout d’abord, un immense merci à mon directeur de maîtrise Jonathan Jalbert pour son encadrement et sa disponibilité tout au long de mon projet. Merci de m’avoir permis de présenter ce projet à travers le monde et de cheminer tout au long de notre collaboration.

Également, je tiens à remercier Philémon Gamet pour son aide durant ma maîtrise et pour ses commentaires pertinents lors de la révision ce mémoire.

Je souhaite ensuite remercier l’équipe d’Ouranos. Ce consortium m’a non seulement fourni les données de ClimEx, mais m’a aussi permis de côtoyer des experts de plusieurs domaines. Un merci spécial à Martin Leduc pour sa disponibilité par rapport à ClimEx ainsi qu’à Philippe Roy pour son outil *Climate Tools* dans Julia et pour tous ses conseils de programmation. Je remercie également Luc Perreault de l’IREQ pour son soutien extérieur. Le travail effectué avec lui et Jonathan Jalbert m’a grandement aidé pour compléter ma maîtrise.

De façon plus formelle, je remercie le CRSNG, Hydro-Québec, Mitacs et Polytechnique Montréal d’avoir financé mon projet de maîtrise. Sans cet appui précieux, ce projet n’aurait pas vu le jour et n’aurait certainement pas été présenté à Zagreb.

Je souhaite dire merci à ma famille et mes amis pour tous les encouragements. Je ne peux exprimer toute la gratitude que j’ai à l’égard de mes parents, qui me supportent depuis le tout début et qui m’inspirent par leurs actions.

Le dernier remerciement s’adresse à une personne très spéciale, mon copain Marc-André. Merci de toujours savoir comment me redonner le sourire dans mes moments de découragement (et il y a en eu plusieurs durant cette maîtrise!) et de faire de moi une personne meilleure au quotidien.

RÉSUMÉ

Le Québec fait partie des régions dans le monde dont les événements de précipitations intenses sont susceptibles d'augmenter considérablement si le réchauffement climatique passe de 1.5°C à 2°C. Les impacts d'événements de pluies intenses sont nombreux et peuvent affecter gravement la santé et l'économie d'une population. Il est donc très important d'estimer adéquatement les précipitations extrêmes dans le futur. Les courbes Intensité-Durée-Fréquence (IDF) permettent d'analyser le risque associé aux événements de pluies intenses. Elles servent notamment pour le dimensionnement des infrastructures exposées aux aléas du climat et à la définition des zones inondables. Pour l'instant, Environnement et Changement climatique Canada (ECCC) fournit les courbes IDF les plus utilisées au pays et suppose que les événements de pluies intenses sont stationnaires dans le temps. Cela signifie qu'ECCC considère que le comportement d'une série de précipitations extrêmes demeure inchangé à travers le temps. Par conséquent, l'objectif du projet est d'introduire les effets des changements climatiques dans les courbes IDF au Québec en incorporant la non-stationnarité due au climat dans la modélisation des séries de précipitations extrêmes. Pour modéliser les extrêmes climatiques de manière non stationnaire, il faut se tourner vers les modèles numériques de climat afin d'estimer les effets des changements climatiques. Ces modèles permettent de faire des projections quantitatives du climat futur. En particulier, les modèles climatiques régionaux permettent de simuler le climat à petite résolution spatiale, puisqu'ils sont limités à une région du globe seulement. Pour ce mémoire, l'ensemble de simulations ClimEx a été utilisé. Ce dernier exploite la cinquième version du Modèle Régional Canadien du Climat ainsi que le scénario RCP8.5 pour les émissions de gaz à effet de serre et d'aérosols. Pour atteindre notre objectif, un modèle hiérarchique bayésien a été développé afin de synthétiser la non-stationnarité contenue dans les séries de précipitations extrêmes de ClimEx. La concentration annuelle des gaz à effet de serre pondérés selon leur pouvoir de réchauffement planétaire a servi de variable explicative dans le modèle pour incorporer les effets des changements climatiques. Un facteur de correction, à appliquer ultérieurement sur les courbes IDF, a été calculé pour une cellule en particulier et permet de voir l'augmentation des niveaux de retour.

ABSTRACT

Extreme precipitations events are expected to increase substantially in the province of Quebec if the global warming goes from 1.5°C to 2°C . Intense rainfall can have big consequences on the health and the economy of a population. Therefore, it is important to obtain accurate estimates of the extreme precipitations to expect in the future. Intensity-Duration-Frequency (IDF) curves represent the main tool to analyse the risk associated to intense rainfall events. They are used for the design of infrastructures exposed to climate hazards and to define flood plains. In Canada, the most used IDF curves are the one provided by Environment and Climate Change Canada (ECCC), which considers extreme precipitations events as stationary processes, meaning their behavior remains unchanged over the years. This project has the goal of introducing the effects of climate change into the estimation of IDF curves in Quebec by adding non-stationarity into extreme precipitations series. Numerical climate models will be very useful to model climate extremes by non-stationary processes. Climate models make quantitative projections of the climate to come. More precisely, regional climate models are restricted to a region of the world and can therefore simulate the climate for a small domain. For this research, the set of simulations ClimEx was utilized. It uses the 5th version of the Canadian Regional Climate Model and the scenario RCP8.5 for the greenhouse gases and aerosols emissions. A bayesian hierarchical model was developed to synthesize the non-stationarity contained in the ClimEx dataset. To incorporate the effects of climate change, the greenhouse gases concentration weighted by their global warming potential was used as a covariate in the model. A correction factor was calculated for a cell to update IDF curves and it shows that the return levels are increasing over the period.

TABLE DES MATIÈRES

DÉDICACE	iii
REMERCIEMENTS	iv
RÉSUMÉ	v
ABSTRACT	vi
TABLE DES MATIÈRES	vii
LISTE DES TABLEAUX	ix
LISTE DES FIGURES	x
LISTE DES SIGLES ET ABRÉVIATIONS	xii
LISTE DES ANNEXES	xiii
CHAPITRE 1 INTRODUCTION	1
CHAPITRE 2 CADRE THÉORIQUE	6
2.1 Théorie des valeurs extrêmes	6
2.1.1 Résultat fondamental	6
2.1.2 Modélisation statistique des maxima par blocs	8
2.1.3 Les cas stationnaire et non stationnaire	9
2.2 Théorie de la décision bayésienne	12
CHAPITRE 3 REVUE DE LITTÉRATURE	14
3.1 Précipitations extrêmes	14
3.2 Ensemble de simulations	17
CHAPITRE 4 MÉTHODOLOGIE ET RÉSULTATS	20
4.1 Extraction des données	20
4.2 Analyse exploratoire	22
4.2.1 Recherche de non-stationnarité	23
4.2.2 Choix du modèle d'ajustement	31
4.2.3 Membre KDA	32

4.2.4	Moyenne des 50 membres	33
4.3	Synthèse de la non-stationnarité	38
4.3.1	Modèle pour maxima non stationnaires	38
4.3.2	Possibilités de facteur de correction	42
4.3.3	Résultats pour une cellule	45
CHAPITRE 5	CONCLUSION ET RECOMMANDATIONS	50
RÉFÉRENCES		52
ANNEXES		56

LISTE DES TABLEAUX

Tableau 4.1	Interprétation de l'écart entre deux indices BIC	27
Tableau 4.2	Estimations des niveaux de retour de la loi GEV selon les deux méthodes	46

LISTE DES FIGURES

Figure 1.1	Exemple de courbe IDF produite par Environnement et Changement climatique Canada (ECCC)	2
Figure 2.1	Échantillons de la loi <i>Generalized Extreme Value</i> (GEV) avec $\mu_0 = 40$, $\sigma = 6$ et $\xi = 0.2$	9
Figure 4.1	Fenêtre glissante permettant d'obtenir la quantité de pluie tombée sur deux heures	22
Figure 4.2	Cellules choisies	23
Figure 4.3	Séries temporelles des maxima pour cinq cellules du membre KDA . .	24
Figure 4.4	Variables explicatives possibles en fonction du temps	25
Figure 4.5	Ajustement selon les deux variables explicatives possibles pour la ville de Salluit	26
Figure 4.6	Maxima annuels de précipitations pour la durée 24h pour la ville de Fermont	28
Figure 4.7	Maxima annuels de précipitations pour la durée 24h pour la ville de Gaspé selon les variables explicatives	29
Figure 4.8	Maxima annuels de précipitations pour la durée 24h pour la ville de Montréal selon les variables explicatives	29
Figure 4.9	Maxima annuels de précipitations pour la durée 24h pour la ville de Radisson selon les variables explicatives	30
Figure 4.10	Estimations des paramètres de la loi GEV pour le membre KDA et la durée 24h	33
Figure 4.11	Moyenne des estimations des paramètres de la loi GEV sur les 50 membres pour la durée 24h	35
Figure 4.12	Moyenne de l'estimation du paramètre de position de la loi GEV sur les 50 membres pour la durée 24h	36
Figure 4.13	Moyenne des estimations des quantiles de la loi GEV pour une période de retour de 20 ans et une durée de 24h	37
Figure 4.14	Structure du modèle hiérarchique bayésien pour le paramètre μ_0 . . .	39
Figure 4.15	Estimation de l'influence du facteur de correction sur le territoire . .	44
Figure 4.16	Distributions marginales des paramètres ν_0 , ν_1 , ν_2 et ν_3	46
Figure 4.17	Estimation du facteur de correction à partir de 1990 pour une période de retour de 20 ans et une durée de 24h avec un intervalle de crédibilité à 95%	48

Figure A.1	Estimation de la moyenne et de l'écart-type pour le paramètre μ_0 pour une durée de 24h	56
Figure B.1	Estimation du facteur de correction à partir de 1990 pour une période de retour de 20 ans et une durée de 24h avec un intervalle de crédibilité à 95% pour la ville de Salluit	57

LISTE DES SIGLES ET ABRÉVIATIONS

ARF	<i>Areal Reduction Factor</i>
BIC	<i>Bayesian Information Criterion</i>
CFC	chlorofluorocarbures
CH ₄	méthane
ClimEx	<i>Climate Change and Hydrological Extremes</i>
CO ₂	dioxyde de carbone
ECCC	Environnement et Changement climatique Canada
GES	gaz à effet de serre
GEV	<i>Generalized Extreme Value</i>
GIEC	Groupe d'experts intergouvernemental sur l'évolution du climat
Go	gigaoctet
GP	<i>Generalized Pareto</i>
IDF	Intensité-Durée-Fréquence
MCMC	Monte Carlo par chaînes de Markov
MOAM	<i>May to October Annual Maxima</i>
MRCC	Modèle Régional Canadien du Climat
NetCDF	<i>Network Common Data Form</i>
N ₂ O	protoxyde d'azote
ppm	parties par million
PRP	potentiel de réchauffement planétaire
RFA	<i>Regional Frequency Analysis</i>
SOI	<i>Southern Oscillation Index</i>
To	téraoctet
TCL	Théorème central limite

LISTE DES ANNEXES

Annexe A	CHAÎNES DE MARKOV	56
Annexe B	FACTEUR DE CORRECTION	57

CHAPITRE 1 INTRODUCTION

Les événements de précipitations extrêmes sont un réel enjeu de société. Ils peuvent avoir des conséquences majeures sur la santé de la population, les infrastructures, l'agriculture ainsi que le développement urbain. L'occurrence de pluies extrêmes peut effectivement influencer le développement du transport, le développement économique et le développement des espaces verts d'une ville. Une pluie intense cause habituellement plus d'accidents routiers et peut mener au débordement des eaux contaminées [1]. Ces conséquences entraînent dans la plupart des cas d'importants coûts économiques et sociaux pour la population. Il est donc nécessaire d'estimer adéquatement le comportement probabiliste des précipitations extrêmes afin d'obtenir des estimations de la quantité de pluie intense à prévoir pour le futur.

Les courbes Intensité-Durée-Fréquence (IDF) constituent le principal outil permettant d'analyser le risque associé aux événements de précipitations extrêmes. Les courbes IDF sont essentielles dans le domaine du génie puisque leur utilisation est obligatoire pour le dimensionnement d'infrastructures. En secteurs urbains comme en secteurs agricoles, les courbes IDF représentent un outil important pour concevoir des ouvrages hydrauliques, puisqu'elles permettent de déterminer la dimension de ces ouvrages selon le niveau de risque déterminé au préalable [2]. Elles servent notamment à la définition des zones inondables et au dimensionnement des infrastructures exposées aux aléas du climat, telles que les digues, les ponceaux et les réseaux d'aqueduc.

En effet, le niveau de risque est déterminé de façon à ce que l'ouvrage à concevoir soit résistant à un niveau attendu en moyenne une fois aux T ans. On appelle T la période de retour et sa valeur varie selon le niveau de risque fixé par les décideurs. On peut donc dire que les courbes IDF servent essentiellement à contrôler le risque selon la période de retour choisie. Par exemple, les postes de transformation de la société d'état Hydro-Québec, qui produit la majorité de l'électricité au Québec, sont équipés de séparateurs eau-huile qui font la gestion des liquides en cas de fuite d'huile. Ces séparateurs eau-huile sont conçus de façon à être capables de résister à la quantité estimée de pluie qui est attendue en moyenne une fois aux 25 ans [3]. En environnement, ces séparateurs sont alors dits dimensionnés selon une période de retour de 25 ans.

Les estimateurs IDF établissent un lien statistique entre les intensités de précipitations extrêmes et la récurrence de ces événements. Ces estimateurs sont habituellement dérivés sous forme de courbes à partir de statistiques donnant la probabilité d'une intensité de pluie pour diverses durées. Les durées utilisées dans le projet sont les suivantes : 1, 2, 6, 12 et 24 heures.

Chaque courbe IDF représente une période de retour (en années) en particulier.

La méthode pour produire des courbes IDF est présentée en 5 étapes dans le rapport de Perreault et Jalbert pour la société d'État Hydro-Québec [2]. Pour une durée D donnée, la première étape consiste à extraire les maxima de l'ensemble des événements de pluie de durée D . Ce sont les maxima qui sont utilisés plutôt que la quantité de pluie quotidienne, étant donné que les courbes IDF portent sur les événements de pluie extrême. Puis, il faut ajuster une loi de probabilité à la série de maxima construite précédemment. De cet ajustement, les paramètres de la distribution sont estimés et les quantiles estimés sont déduits pour chaque période de retour T considérée. Troisièmement, les quantiles sont divisés par la durée D afin d'obtenir l'intensité I_T pour chaque période de retour. Cette intensité I_T calculée pour la période de retour T désigne la quantité de pluie à tomber durant la durée D qui est attendue en moyenne une fois aux T ans. Les périodes de retour utilisées sont généralement 2, 5, 10, 25, 50 et 100 ans. Il est ensuite nécessaire de relier les valeurs d'une même période T , habituellement par ajustement linéaire. La dernière étape consiste à tracer le tout sur des axes log-log.

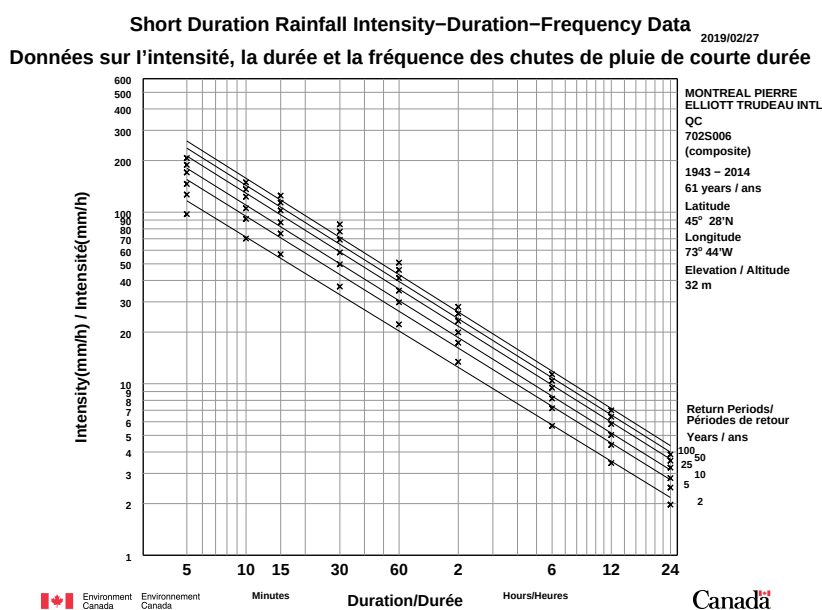


Figure 1.1 Exemple de courbe IDF produite par ECCC

Les courbes IDF les plus utilisées au pays sont celles fournies par Environnement et Changement climatique Canada (ECCC). La figure 1.1 présente un exemple de courbe IDF produite

par ECCC à la station Pierre Elliott Trudeau. Elles permettent donc de relier des quantiles élevés de pluie extrême à des durées fixées pour une période de retour donnée. Les courbes IDF produites par ECCC sont disponibles à l'adresse suivante :

http://climate.weather.gc.ca/prods_servs/engineering_e.html.

Actuellement, les courbes IDF fournies par ECCC supposent que les précipitations intenses sont stationnaires dans le temps, c'est-à-dire qu'elles agissent comme des processus aléatoires dont le comportement demeure inchangé à travers le temps. Or, cette hypothèse n'est pas adaptée en raison des changements climatiques [4]. Selon le plus récent rapport du Groupe d'experts intergouvernemental sur l'évolution du climat (GIEC), le Québec fait partie des régions dans le monde dont les événements de précipitations intenses sont susceptibles d'augmenter dans le futur [5]. En effet, il est montré que le changement climatique a eu pour effet d'augmenter la variabilité naturelle des précipitations intenses et de modifier la distribution des extrêmes climatiques à travers le temps [6]. Dans la plupart des cas, il est donc nécessaire de modéliser les extrêmes climatiques par une série non stationnaire.

La grande majorité de la communauté scientifique s'entend pour dire que les changements climatiques sont causés par l'activité humaine, plus particulièrement par les émissions anthropiques de gaz à effet de serre (GES) [7]. Les GES ont été nommés ainsi puisqu'ils agissent littéralement comme une serre pour la planète. Ils gardent la chaleur rayonnée par la Terre dans l'atmosphère, ce qui a pour effet de réchauffer la planète [8]. Ils jouent donc le rôle central dans le réchauffement climatique observé ces dernières années, spécialement la concentration de dioxyde de carbone (CO_2). Les scientifiques projettent que le réchauffement planétaire causera, entre autres, une météo plus variable ainsi que plus d'événements météorologiques extrêmes [8]. L'intensification du changement climatique est donc intimement liée aux émissions de GES.

Dans cet ordre d'idées, les courbes IDF sont sujettes à évoluer en fonction des changements climatiques, ce qui n'est actuellement pas pris en compte dans les courbes IDF utilisées au Canada [2]. Il devient alors primordial d'inclure leurs effets dans l'estimation des courbes IDF, en incorporant la non-stationnarité induite par les changements climatiques dans les séries de précipitations extrêmes. Comme il est coutume de faire dans diverses études récentes (voir par exemple [9–11]), les événements de précipitations extrêmes étudiés dans ce projet de recherche sont représentés par une série non stationnaire en incorporant des variables expliquant la non-stationnarité.

Pour quantifier les effets des changements climatiques, il faut se tourner vers l'utilisation de modèles numériques de climat. Les simulations climatiques représentent l'unique outil dispo-

nible à ce jour pour faire des projections quantitatives du climat futur d'après le GIEC. Ces modèles « parviennent à simuler de façon assez réaliste une multitude de caractéristiques du climat actuel ainsi que l'évolution observée de la température moyenne de la Terre depuis plus de 100 ans. » [12]. Les modèles climatiques globaux font cela en représentant mathématiquement de manière simplifiée la physique du climat à l'aide d'équations qui proviennent des processus physiques de conservation de masse et d'énergie [7]. Un modèle climatique global résulte donc de la résolution de ces équations pour le système climatique terrestre complet. Les simulations d'un modèle climatique dépendent grandement du scénario RCP choisi. Ces scénarios ont été conçus par le GIEC et représentent la trajectoire de divers scénarios d'émissions futures de GES et d'aérosol [13]. Ils sont utilisés dans les modèles climatiques afin de simuler les changements climatiques associés. Il est important de mentionner qu'un modèle climatique global a une résolution spatiale limitée par la puissance du calcul informatique. Cette résolution spatiale ne permet pas de simuler des événements à petite échelle tels que des cellules orageuses à haute intensité et de courte durée, car ce genre d'événements se produit sur un domaine trop petit.

Les modèles climatiques régionaux ont été développés afin de simuler le climat sur un plus petit domaine. En effet, un modèle climatique régional ressemble à un modèle climatique global, mais se distingue par le fait que le domaine utilisé pour la simulation est limité à une région de la planète [14]. Réduire la taille du domaine de simulation permet d'augmenter la résolution spatiale du modèle climatique. En l'occurrence, les estimations de précipitations extrêmes à petite échelle peuvent être déduites à partir d'un modèle régional. Un modèle climatique régional s'appuie tout de même sur un modèle climatique global pour les conditions aux limites.

La cinquième version du Modèle Régional Canadien du Climat (MRCC) sera utilisée dans ce mémoire pour estimer les précipitations extrêmes, dont le domaine de simulation est limité au Canada. L'efficacité et la performance de ce modèle régional ont été testées à maintes reprises [15, 16] et s'améliorent de version en version.

L'objectif général du mémoire est d'introduire les effets du changement climatique dans les courbes IDF de précipitations au Québec. Cela se traduit par un objectif spécifique bien précis, qui est de développer un modèle statistique non stationnaire de valeurs extrêmes pour un ensemble de simulations climatiques. Ce modèle permettra d'extraire le maximum d'information sur la non-stationnarité des précipitations extrêmes et de développer un facteur de correction à appliquer aux courbes IDF actuelles afin de les mettre à jour. Un ensemble de projections climatiques sera utilisé pour quantifier l'évolution des précipitations intenses dans ce projet. Le sujet de cette recherche a été motivé en grande partie par l'absence de

courbes IDF au Canada prenant en compte les effets des changements climatiques.

Ce mémoire est divisé de la manière suivante. Le chapitre 2 dresse la base théorique nécessaire à la compréhension du projet de recherche. La théorie des valeurs extrêmes ainsi que la théorie de la décision bayésienne y sont expliquées. Le chapitre 3 présente quant à lui une revue de littérature sur les différents travaux de recherche effectués dans le thème des valeurs extrêmes non stationnaires. Également, l'ensemble de simulations climatiques *Climate Change and Hydrological Extremes* (ClimEx) y est présenté. Permettant d'entrer en plein coeur du projet, le chapitre 4 présente la méthodologie et les résultats. La méthodologie comprend entre autres l'extraction des maxima de précipitations et l'analyse exploratoire des données, étape très importante ayant guidé le restant de la recherche. C'est également dans ce chapitre que le modèle hiérarchique bayésien est développé et que les différentes possibilités de facteur de correction sont énoncées. Les résultats sont présentés pour une cellule en particulier. Finalement, le chapitre 5 contient la conclusion générale du projet de recherche ainsi qu'une discussion portant sur des travaux futurs.

CHAPITRE 2 CADRE THÉORIQUE

L'objectif de ce chapitre est d'introduire les bases théoriques essentielles à la compréhension du projet de maîtrise. Pour débiter, les fondements de la théorie des valeurs extrêmes sont présentés. Le lecteur intéressé à en apprendre davantage peut consulter le livre de Coles [17]. En statistique, il existe deux approches pour faire de la modélisation qui sont fondamentalement différentes. Il y a l'approche classique, qui considère les paramètres à estimer comme des quantités fixes et inconnues à estimer. Puis, il y a l'approche bayésienne, qui considère les paramètres à estimer comme des variables aléatoires. Dans ce dernier cas, on cherche alors à estimer la distribution de ces paramètres [18]. Cette approche sera grandement utile pour le développement d'un modèle hiérarchique et c'est pour cette raison qu'elle est privilégiée dans ce projet de recherche. La théorie de la décision bayésienne est donc introduite dans la deuxième section de ce chapitre. Les concepts essentiels à cette approche statistique sont définis, comme les distributions *a priori* et *a posteriori*.

2.1 Théorie des valeurs extrêmes

La théorie des valeurs extrêmes se concentre sur l'étude des événements rares. L'intérêt de prédire l'occurrence d'événements extrêmes remonte au 20e siècle. Ce n'est qu'aux alentours des années 1950 qu'une méthodologie rigoureuse fut développée. Les premières applications de cette théorie sont attribuables au domaine du génie civil. Les ingénieurs ont toujours eu comme mandat de faire des infrastructures capables de résister à des événements très rares dont l'intensité dépasse souvent les plus grandes quantités observées historiquement [17].

2.1.1 Résultat fondamental

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 :

$$M_n = \max\{X_1, \dots, X_n\}.$$

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

$$P\{M_n \leq z\} = P\{X_1 \leq z, \dots, X_n \leq z\} = P\{X_1 \leq z\} \times \dots \times P\{X_n \leq z\} = F^n(z). \quad (2.1)$$

La distribution de F étant habituellement inconnue en pratique, il faut trouver une façon efficace de l'estimer. Une option intuitive serait d'estimer F à partir des données observées et de se servir de cette estimation dans l'équation (2.1) pour déterminer la loi du maximum. Le problème avec cette approche est que l'incertitude associée à l'estimation de F deviendrait beaucoup trop grande lors de l'estimation de F^n . Cette méthode nous empêcherait également de faire de l'inférence au-delà du maximum observé. De plus, comme on s'intéresse aux événements extrêmes, il est fort probable que le nombre d'observations soit faible. Il faut donc estimer F^n directement à l'aide d'une distribution limite et accepter que la distribution de F est inconnue. 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. Afin de contourner cette difficulté, il faut effectuer une transformation linéaire M_n , telle que présentée dans le théorème 2.1.1.

Théorème 2.1.1 (chap. 3 [17]). *S'il existe des suites de constantes $\{a_n > 0\}$ et $\{b_n\}$ telles que*

$$P \left\{ \frac{M_n - b_n}{a_n} \leq z \right\} \rightarrow G(z) \text{ pour } n \rightarrow \infty \quad (2.2)$$

pour une distribution G non dégénérée, alors G est une loi Generalized Extreme Value (GEV)

$$G(z) = \begin{cases} \exp \left[- \left\{ 1 + \xi \left(\frac{z - \mu}{\sigma} \right) \right\}^{-1/\xi} \right] & \text{si } \xi \neq 0, \\ \exp \left[- \exp \left(- \frac{z - \mu}{\sigma} \right) \right] & \text{si } \xi = 0. \end{cases} \quad (2.3)$$

définie sur $\{z : 1 + \frac{\xi(z - \mu)}{\sigma} > 0\}$, où $-\infty < \mu < \infty$, $\sigma > 0$ et $-\infty < \xi < \infty$.

Ce théorème est central dans la théorie des valeurs extrêmes, puisqu'il implique qu'une série de maxima pouvant être stabilisée par des constantes de normalisation $\{a_n\}$ et $\{b_n\}$ converge en loi vers la famille GEV. La loi GEV est composée du paramètre de position $\mu \in \mathbb{R}$, du paramètre d'échelle $\sigma \in \mathbb{R}_+^*$ et du paramètre de forme $\xi \in \mathbb{R}$. Pour bien illustrer l'importance de ce théorème, une analogie peut être faite entre ce théorème et le Théorème central limite (TCL). En effet, le TCL affirme que la moyenne d'un échantillon de taille n suit une loi normale pour un n assez grand, peu importe la loi des observations de l'échantillon. Il en est de même pour le théorème 2.1.1, qui implique qu'une série de maxima de taille suffisante peut être modélisée par la loi GEV et ce, peu importe la distribution de la population d'origine.

Les constantes $\{a_n > 0\}$ et $\{b_n\}$ de l'équation (2.2) sont inconnues, mais ce n'est pas un problème puisqu'elles peuvent être absorbées dans les paramètres à estimer. Ce concept est

illustré dans la relation suivante :

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

où G^* est aussi un membre de la famille GEV [17]. Estimer la distribution d'une transformation linéaire de M_n revient donc à estimer la distribution de M_n directement :

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

Le théorème a été présenté dans le cas indépendant et identiquement distribué, mais il peut également être généralisé à des suites non indépendantes. En pratique, ce sont généralement ces suites qui sont d'intérêt, puisque la propriété d'indépendance est rarement respectée. Les résultats du théorème pour des maxima dépendants restent valables sous certaines conditions spécifiques. Ces dernières sont plutôt techniques et difficiles à vérifier, mais elles sont énoncées dans le Chapitre 5 du livre de Coles [17]. Il s'agit plus particulièrement de la condition $D(u_n)$, qui stipule que la dépendance des valeurs extrêmes doit être à courte portée pour que le théorème tienne. Dans le contexte de notre projet, cela se traduit par le fait que, par exemple, si on sait qu'il y a eu une forte pluie aujourd'hui, cet événement peut influencer la probabilité de pluie intense du lendemain et du surlendemain. Par contre, cet événement n'aura pas d'influence sur un jour spécifique dans, disons, quatre semaines. En somme, la condition $D(u_n)$ permet de modéliser les séries de précipitations extrêmes par la loi GEV.

2.1.2 Modélisation statistique des maxima par blocs

En théorie des valeurs extrêmes, une méthode grandement utilisée est la méthode des maxima par blocs. Elle consiste à partitionner les observations dans des blocs de même taille et d'en générer une série de maxima, sur laquelle la loi GEV peut être ajustée. Si la taille de bloc est suffisamment grande, alors la loi GEV est un bon modèle pour l'ajustement.

La taille des blocs est un choix crucial dans la modélisation statistique d'un jeu de données, comme mentionné dans le livre de Coles [17]. En effet, si la taille de bloc est trop petite, l'approximation par la loi GEV risque d'être de piètre qualité, ce qui introduirait un biais dans l'estimation. D'un autre côté, si la taille de bloc est trop grande, le nombre de maxima est réduit, menant à une grande variance d'estimation. En sciences environnementales, la

taille de bloc habituellement utilisée est une année, ce qui donne des maxima annuels (voir par exemple [11, 14]). C'est d'ailleurs la taille utilisée pour ce projet de recherche.

Dans le contexte des précipitations extrêmes, la saisonnalité des événements de pluie est un aspect à considérer. Si les blocs sont de taille inférieure à une année, la saisonnalité peut être un problème, étant donné que le maximum peut varier selon les mois considérés. Cela pourrait fausser la modélisation statistique, puisque la non-homogénéité des précipitations ne serait pas prise en compte [17]. C'est ce qui a motivé notre choix de prendre des blocs de taille d'une année, puisque cela permet de faire l'hypothèse que les maxima de blocs, qui correspondent à des maxima annuels, ont une distribution commune.

2.1.3 Les cas stationnaire et non stationnaire

Il faut faire la distinction entre une série dite stationnaire et une série dite non stationnaire. Des échantillons de loi GEV ont été tirés pour illustrer les deux cas de séries et sont tracés à la figure 2.1. Le cas stationnaire signifie que tous les paramètres sont constants à travers le temps. Quant à lui, le cas non stationnaire incorpore de la non-stationnarité via une variable explicative fictive, prenant ses valeurs entre 37.5 et 40, à relation linéaire avec le paramètre de position μ . Les deux échantillons sont de taille $n = 500$.

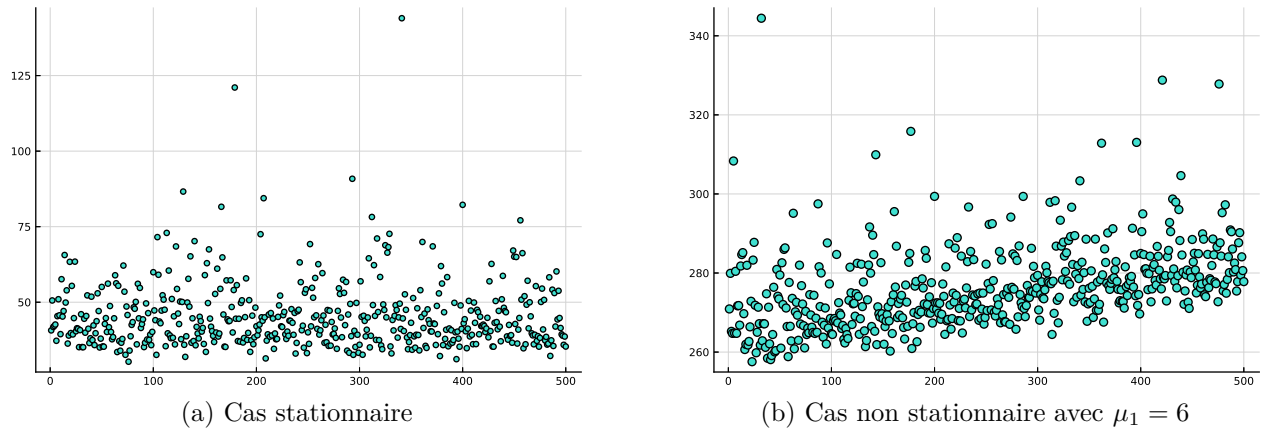


Figure 2.1 Échantillons de la loi GEV avec $\mu_0 = 40$, $\sigma = 6$ et $\xi = 0.2$.

Cas stationnaire

Commençons par introduire une série de maxima dans le cas stationnaire. Cela signifie que le comportement de la série demeure inchangé à travers le temps, comme illustré à gauche de la figure 2.1. La définition formelle d'une série stationnaire est énoncée.

Définition 2.1.2 (chap. 2 [17]) *Un processus aléatoire $X_1, X_2, \dots$ est dit **stationnaire** si, étant donné un ensemble d'entiers $\{i_1, \dots, i_k\}$ et un entier m , la distribution conjointe de $(X_{i_1}, \dots, X_{i_k})$ est identique à celle de $(X_{i_1+m}, \dots, X_{i_k+m})$.*

Soit $X_1, X_2, \dots$ la série de variables aléatoires stationnaire utilisée pour illustrer la modélisation des maxima. Les données sont partitionnées en k blocs de même taille n . Le maximum de chaque bloc est calculé, menant à une série de maxima $M_{n,1}, \dots, M_{n,k}$. Selon le théorème 2.1.1, ces maxima sont approximativement distribués selon la loi GEV. Pour estimer les paramètres de cette loi, la méthode du maximum de vraisemblance est souvent utilisée en pratique. Elle consiste à maximiser la fonction de vraisemblance suivante :

$$\mathcal{L}(\mu, \sigma, \xi) = \prod_{t=1}^k g(m_{n,t} | \mu, \sigma, \xi), \quad (2.4)$$

où l'expression $g(z | \mu, \sigma, \xi)$ dénote la densité de la loi GEV de paramètres (μ, σ, ξ) évaluée en z . De nos jours, la maximisation de la vraisemblance peut se faire aisément à l'aide de méthodes numériques.

Avec des maxima annuels, les estimations de quantiles extrêmes de leur distribution annuelle sont obtenues en inversant l'équation (2.3) où $G(z_p) = 1 - p$ [17]. Voici le résultat :

$$z_p = \begin{cases} \mu - \frac{\sigma}{\xi} \times [1 - \{-\log(1-p)\}^{-\xi}] & \text{si } \xi \neq 0, \\ \mu - \sigma \times \log \{-\log(1-p)\} & \text{si } \xi = 0. \end{cases} \quad (2.5)$$

Le quantile z_p est appelé le niveau de retour correspondant à la période de retour $T = 1/p$. Lorsque M correspond à un maximum annuel, comme c'est le cas dans notre recherche, T correspond à une période de retour en années. Cela signifie que le niveau z_p est approximativement dépassé par le maximum d'une année quelconque avec une probabilité p . Plus généralement, il est attendu que le niveau z_p soit dépassé en moyenne une fois aux T ans.

Cas non stationnaire

Le prochain cas concerne les processus aléatoires dont certaines de leurs caractéristiques changent à travers le temps, appelés alors des processus non stationnaires. Ce concept est illustré à la figure 2.1, où l'on peut voir une augmentation avec le temps de la valeur des données.

Contrairement au cas stationnaire où la théorie des valeurs extrêmes s'appliquait telle quelle, aucune théorie générale n'existe pour les séries non stationnaires. Pour une telle série de maxima, il y a principalement deux manières de la modéliser. Une première option est d'ef-

fectuer une étape de pré-traitement visant à enlever toute trace de non-stationnarité dans le jeu de données, comme dans l'article de Jalbert *et al* [3]. Il y a également l'option d'introduire des variables explicatives dans les paramètres de la loi GEV (voir par exemple [9–11]). C'est cette dernière option qui a été choisie afin de modéliser la non-stationnarité des précipitations extrêmes. Cette technique permet aux paramètres de varier avec le temps et/ou des variables explicatives tout en utilisant la technique des maxima de blocs. Les variables explicatives représentent habituellement le processus physique qui cause la non-stationnarité dans les données [17].

Prenons M_t , le maximum au temps t d'un bloc d'observations $X_1, X_2, \dots$, pour illustrer le concept. Sa distribution est notée comme suit :

$$M_t \sim G(\mu(t), \sigma(t), \xi(t)),$$

où les paramètres peuvent dépendre du temps t , de variables explicatives ou simplement demeurer constants dans le temps. Les variables explicatives et le temps t sont habituellement incorporés dans les paramètres selon une relation linéaire, quadratique ou avec un modèle à point de changement.

L'équation suivante illustre la façon de modéliser le paramètre de position avec une tendance linéaire selon le temps :

$$\mu(t) = \mu_0 + \mu_1 t,$$

pour des paramètres μ_0 et μ_1 . Dans le cas où seul le paramètre de position varie dans le temps (i.e. $Z_t \sim G\{\mu(t), \sigma, \xi\}$), il y aurait maintenant un total de quatre paramètres à estimer (μ_0 , μ_1 , σ et ξ).

C'est le même principe pour une relation quadratique :

$$\mu(t) = \mu_0 + \mu_1 t + \mu_2 t^2,$$

et pour un modèle à point de changement :

$$\mu(t) = \begin{cases} \mu_1 & \text{si } t \leq t_0 \\ \mu_2 & \text{si } t > t_0, \end{cases}$$

pour t_0 un instant dans le temps qui expliquerait un changement dans les données.

On peut considérer l'exemple concernant les maxima annuels du niveau de la mer présenté dans le chapitre 6 de l'ouvrage de Coles [17] pour illustrer cette fois-ci l'utilisation d'un indice climatique comme variable explicative. Étant donné qu'il semble y avoir une relation positive

entre l'Indice de l'Oscillation astrale (*Southern Oscillation Index*, SOI) et les maxima du niveau de la mer, cet indice est utilisé comme variable explicative potentielle afin de modéliser la série de maxima. Une relation linéaire avec le SOI pour le paramètre de position serait représentée de la façon suivante :

$$\mu(t) = \mu_0 + \mu_1 \text{SOI}(t). \quad (2.6)$$

La notation générale pour écrire les paramètres de la loi GEV est la suivante [17, p. 107] :

$$\theta(t) = h(\mathbf{x}^T \boldsymbol{\beta}), \quad (2.7)$$

où θ peut représenter μ , σ ou ξ , h une fonction spécifiée, $\boldsymbol{\beta}$ le vecteur des paramètres et $\mathbf{x}$ le vecteur des variables explicatives.

La fonction h peut être la fonction identité I comme vu pour le paramètre de position dans l'équation (2.6). Pour le paramètre d'échelle, la fonction exponentielle est habituellement utilisée comme dans l'expression suivante :

$$\sigma(t) = \exp \left\{ [1, t] \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \right\}.$$

En effet, l'utilisation de la fonction exponentielle est nécessaire pour s'assurer que l'estimation du paramètre d'échelle demeure positive, de façon à respecter les conditions du théorème 2.1.1. Il est peu recommandé d'inclure une relation avec le temps t ou avec des variables explicatives dans le paramètre de forme à cause de sa grande variance [3]. Pour les paramètres restants, il est commun d'inclure des variables explicatives dans les paramètres de position et d'échelle (voir par exemple [9]) ou d'en inclure dans le paramètre de position seulement (voir [10]).

2.2 Théorie de la décision bayésienne

Cette section contient les bases de l'approche paramétrique de la théorie de la décision bayésienne. Supposons que les variables aléatoires $Y_1, \dots, Y_n$ ont une densité $f_{(\mathbf{Y}|\boldsymbol{\theta})}(\mathbf{y})$, où le vecteur $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)$ est inconnu et la fonction f est connue. L'important théorème concernant l'inversion des probabilités a été fourni par Bayes, donnant l'énoncé suivant :

Théorème 2.2.1 (Théorème de Bayes - version continue, chap. 1 [18]). *Soit X et Y deux variables aléatoires de distributions conditionnelles $f_{(Y|X=x)}(y)$ et marginale $f_X(x)$, alors la distribution conditionnelle de X sachant $(Y=y)$ est*

$$f_{(X|Y=y)}(x) = \frac{f_{(Y|X=x)}(y)f_X(x)}{\int f_{(Y|X=x)}(y)f_X(x)dx}.$$

Ce qui distingue la théorie bayésienne de la statistique classique est le fait que l'incertitude sur le paramètre $\boldsymbol{\theta}$ peut être décrite par une distribution de probabilité $f_{\boldsymbol{\theta}}(\boldsymbol{\theta})$ nommée distribution *a priori*. La distribution de $\boldsymbol{\theta}$ conditionnelle à $(\mathbf{Y} = \mathbf{y})$, pour $\mathbf{y}$ le vecteur des observations, est quant à elle appelée distribution *a posteriori* et est définie comme suit :

$$f_{(\boldsymbol{\theta}|\mathbf{Y}=\mathbf{y})}(\boldsymbol{\theta}) = \frac{f_{(\mathbf{Y}|\boldsymbol{\theta}=\boldsymbol{\theta})}(\mathbf{y})f_{\boldsymbol{\theta}}(\boldsymbol{\theta})}{\int f_{(\mathbf{Y}|\boldsymbol{\theta}=\boldsymbol{\theta})}(\mathbf{y})f_{\boldsymbol{\theta}}(\boldsymbol{\theta})d\boldsymbol{\theta}}. \quad (2.8)$$

L'inférence en statistique bayésienne est basée sur la loi *a posteriori* des paramètres $f_{(\boldsymbol{\theta}|\mathbf{Y}=\mathbf{y})}(\boldsymbol{\theta})$. La contribution principale de cette théorie est le fait de considérer les paramètres à estimer comme des variables aléatoires, et non telles que des valeurs fixes inconnues comme en statistique classique. Le choix de la loi *a priori* d'un paramètre est une étape importante de l'inférence bayésienne. Afin d'en apprendre davantage sur les techniques existantes pour déterminer une loi *a priori*, voir le chapitre 3 du livre de Christian P. Robert [18].

C'est ce qui nous amène à introduire une méthode classique d'approximation : les méthodes de Monte Carlo par chaînes de Markov (MCMC). Elles représentent des méthodes de simulations récentes particulièrement utiles en statistique bayésienne qui permettent de générer un échantillon aléatoire de la loi *a posteriori*, même lorsque la constante de normalisation est inconnue [18]. Le concept derrière les méthodes MCMC est qu'il faut générer une chaîne de Markov $(\boldsymbol{\theta}^{(m)})_m$ ayant comme loi stationnaire la loi *a posteriori*. L'échantillon *a posteriori* permet ensuite de calculer facilement diverses quantités, comme un intervalle de crédibilité.

Les méthodes MCMC seront utilisées dans le projet, étant donné que les distributions *a posteriori* des paramètres de la loi GEV sont inconnues. L'idée générale étant d'utiliser le comportement limite d'une chaîne de Markov, il y a tout de même différentes approches afin de mettre en pratique les méthodes MCMC. Au chapitre 6 du livre de Robert [18], les différentes techniques y sont présentées, comme l'algorithme de Metropolis-Hastings, l'échantillonnage de Gibbs et l'échantillonnage par tranche. Les détails concernant les méthodes MCMC utilisées dans ce projet de maîtrise sont fournis au chapitre 4 lors de la présentation du modèle hiérarchique bayésien.

CHAPITRE 3 REVUE DE LITTÉRATURE

Depuis plusieurs années, l'effet des changements climatiques est une préoccupation centrale dans l'estimation des précipitations extrêmes. De nombreuses études se penchent sur cet enjeu. Dans ce chapitre, une revue de littérature des méthodes statistiques pour modéliser des précipitations extrêmes sous les effets des changements climatiques est présentée et l'ensemble de simulations ClimEx est introduit.

3.1 Précipitations extrêmes

Pour commencer, la méthode d'analyse fréquentielle régionale (*Regional Frequency Analysis*, RFA) est présentée. L'analyse de fréquence désigne le fait d'estimer le nombre d'occurrence d'un certain événement. De ce fait, la méthode RFA fait l'analyse de la fréquence d'événements extrêmes tout en reliant les différents échantillons entre eux de façon à en tirer le maximum d'information. Formellement, la méthode RFA développée par Hosking et Wallis [19] est une approche pour modéliser de façon spatiale un jeu de données. Elle ne modélise pas la dépendance spatiale comme un processus stochastique dans l'espace. Cette méthode est plutôt basée sur l'hypothèse que les points contenus dans une même région homogène ont certaines propriétés en commun, dont une constante de mise à l'échelle. Les régions dites homogènes sont des sous-ensembles du domaine complet à étudier. Leur définition est difficile à mettre en pratique. De plus, l'ajout de variables explicatives est impossible avec la méthode RFA, ce qui empêche de modéliser la non-stationnarité d'un jeu de données en permettant aux paramètres de varier selon le temps et/ou les variables explicatives [20]. Pour l'ensemble de ces raisons, cette méthode n'a pas été retenue pour la recherche présentée. Elle est par contre utilisée dans un article important de Mailhot *et al.* qui est maintenant présenté.

Les simulations du MRCC ont été utilisées dans un article de Mailhot *et al.* il y a de cela plus de 10 ans [14]. Les maxima annuels de précipitations de mai à octobre (*May to October Annual Maxima*, MOAM) étaient analysés avec la méthode RFA. Le modèle régional d'une résolution de 45km × 45km était appliqué à deux périodes de 30 ans chacune. La conclusion était qu'il n'y avait pas d'indication claire démontrant une présence de non-stationnarité dans les séries de MOAM provenant du sud du Québec. Par contre, deux périodes de si courtes durées ne sont sûrement pas suffisantes pour analyser une tendance de stationnarité à long terme, comme mentionné dans leur conclusion. Ils mentionnaient également la possibilité d'utiliser des ensembles multimembres afin d'étudier la sensibilité des estimations aux conditions initiales du modèle utilisé. Le modèle climatique utilisé dans ce projet de maîtrise est

en effet multimembre et il est détaillé dans la section suivante.

Plus récemment, un modèle spatio-temporel a été développé par Jalbert, Favre, Bélisle & Angers [3] en 2017 afin de modéliser les précipitations extrêmes. C'est plus précisément un modèle régional bayésien hiérarchique non stationnaire qui a été construit spécialement pour les simulations climatiques de pluie intense. Une méthode pour obtenir un facteur de correction a été développée pour des données d'événements de pluie intense. Ce facteur de correction pour le niveau de retour associé à la période de retour T entre les années t_1 et t_2 est en fait le rapport des quantiles estimés, soit $z_p(t_2)/z_p(t_1)$. Ce dernier est alors multiplié directement aux niveaux de retour des courbes IDF à actualiser. Cette étude porte sur une seule simulation d'un modèle climatique. Dans le cadre du projet de maîtrise présenté, l'objectif est de généraliser l'aspect non stationnaire du modèle spatio-temporel de valeurs extrêmes pour un ensemble de simulations climatiques.

Un modèle bayésien hiérarchique spatial a été introduit par Sharkey et Winter en 2017 [20]. Dans cet article, la dépendance spatiale est prise en compte dans le calcul des intervalles de confiance et c'est la loi *Generalized Pareto* (GP) qui est utilisée pour l'ajustement. Cette distribution est utilisée lors de la modélisation des excédents au-dessus d'un seuil et fait partie, avec la loi GEV, des distributions importantes de la théorie des valeurs extrêmes. Ce seuil doit être suffisamment élevé afin d'être utilisé dans le cadre formel de la théorie des valeurs extrêmes. Pour revenir à l'article de Sharkey et Winter, une structure spatiale est imposée sur les paramètres de la loi GP, servant à modéliser les données de précipitations observées en Grande-Bretagne excédant un seuil fixé. Les méthodes adoptées par Sharkey et Winter s'apparentent beaucoup à celles utilisées dans ce projet de maîtrise, à l'exception que, dans notre cas, les données analysées proviennent d'un modèle climatique et que la non-stationnarité du jeu de données est prise en compte.

L'initiative de Simonovic, Schardong et Sandink [4] publiée en 2017 ayant pour but d'actualiser les courbes IDF en climat futur a permis de produire des courbes IDF prenant en compte l'effet attendu des changements climatiques pour l'ensemble du Canada. Ils ont utilisé une méthode s'apparentant à celle du plus proche voisin, afin de faire de l'interpolation et de générer des statistiques des précipitations intenses pour n'importe quel emplacement au pays, même ceux sans station météorologique. Notre projet de recherche se distingue des travaux du professeur Simonovic par le fait qu'il est effectué dans le cadre formel de la théorie des valeurs extrêmes.

Une problématique du projet est que les courbes IDF utilisent des données observées de précipitations provenant de stations météorologiques et n'ont pas la même échelle que les données simulées provenant d'un modèle climatique. Afin d'être en mesure de comparer les

données observées avec les données simulées, cela se fait habituellement par le biais d'un facteur de correction qui les mettra à la même échelle. C'est ce qui nous amène à introduire les *Areal Reduction Factors* (ARFs).

Les ARFs sont des facteurs de correction qui prennent des estimations en un point donné et les transforment en estimations sur une région donnée. Par exemple, ils peuvent débiter avec des précipitations extrêmes observées à une station météorologique et transformer cela en estimations de pluie intense sur une cellule d'un modèle climatique, qui représente habituellement une région du domaine d'étude. Pour cela, ils se servent des liens unissant les différentes stations météorologiques entre elles [21]. Les ARFs ont plusieurs applications, comme la définition des zones inondables et l'estimation du risque d'inondation pour des événements rares. Ils servent également à la conception des infrastructures exposées aux aléas du climat. Avec l'amplification des effets du changement climatique, ces facteurs sont très importants pour estimer les événements extrêmes.

Il est dit en théorie que les ARFs sont étroitement liés à la fréquence d'occurrence d'événements associés à de grandes périodes de retour par la structure de dépendance asymptotique des précipitations sur un territoire. Cette relation n'est pas encore tout à fait claire et elle est approfondie dans l'article de Le *et al.* [21]. Il y est mentionné entre autres que les ARFs diminuent avec des événements plus rares. Nombreux articles démontrent à l'aide de différentes méthodes que le niveau de dépendance devient relativement stable avec une période de retour grandissante pour des modèles asymptotiques dépendants. En d'autres mots, les ARFs d'un tel modèle devraient en théorie être stables au-dessus d'un certain point. Pour les modèles asymptotiques indépendants, la théorie stipule que le niveau de dépendance tend à diminuer avec une période de retour grandissante. Les ARFs associés à ce type de modèle devraient donc diminuer avec une période de retour qui augmente [21, p. 712].

L'article de Le, Davison, Engelke, Leonard et Westra publié en 2018 porte effectivement sur les propriétés des précipitations extrêmes et sur les ARFs [21]. Dans cette étude, la méthode de Bell est adoptée pour le calcul des facteurs de correction, où ils sont modélisés comme une fonction de la profondeur et de la moyenne des bassins versants, donnant des estimations spatiales. Les auteurs de l'article prennent des données observées en Australie afin de construire les ARFs et de comprendre leur relation avec la fréquence d'occurrence d'événements associés à de grandes périodes de retour. Cette relation dépend de la structure de dépendance asymptotique des précipitations sur la région concernée et elle est approfondie avec une étude de simulation. Différents types de modèles max-stables sont alors utilisés dans l'article afin de générer des données de précipitations, dans le but d'étudier le comportement asymptotique des ARFs. Cet article de grande envergure serait un bon point de départ si le

projet de maîtrise était étendu à des estimations régionales, soit des estimations prenant en compte l’aspect spatial entre les différents points de grille.

Pour continuer, les modèles max-stables représentent une approche qui est grandement utilisée pour modéliser la dépendance spatiale de valeurs extrêmes. Par contre, ces modèles sont très difficiles à utiliser pour des projets de grande dimension, comme par exemple pour des modèles climatiques. Cette caractéristique empêche l’application des modèles max-stables dans certaines études, comme dans celle de Jalbert *et al.* [3]. Cette approche n’a pas été utilisée pour le projet de maîtrise étant donné l’immense base de données à analyser.

Une étude pour quantifier les changements à venir pour les courbes IDF régionales a été effectuée par Lima, Kwon & Kim en 2018 [9]. Dans cette dernière, les paramètres de la loi GEV sont estimés à l’aide de méthodes bayésiennes. Par contre, la non-stationnarité n’est pas adressée dans cet article. Pour continuer, Reich et Shaby présentent dans un article récent [10] un modèle spatial markovien pour les extrêmes climatiques dans un domaine discret. Leur étude fait l’utilisation de grappes et utilise un modèle spatial, dont la distribution marginale est une loi GEV pour chaque point étudié dans l’espace.

La recherche sortie en 2019 de Ganguli et Coulibaly [22] est une des premières à considérer plus d’un membre d’un modèle climatique pour modéliser des précipitations extrêmes. Ce sont trois modèles régionaux du climat qui ont servi pour la quantification des changements futurs dans les courbes IDF de précipitations sur 8 stations de pluie dans le sud de l’Ontario. La non-stationnarité a également été abordée, en faisant varier le paramètre de position linéairement selon le temps. Comme mentionné dans leur étude, trois membres ne sont peut être pas suffisants pour représenter la variabilité naturelle du système climatique à court terme, d’où la pertinence du projet de maîtrise présenté.

Une revue de la littérature en lien avec ce mémoire vient d’être présentée. Elle a mis en lumière des travaux récents portant sur l’estimation des précipitations extrêmes en présence de changements climatiques. Divers concepts ont été définis, comme la méthode RFA et les ARFs, afin de bonifier la compréhension de ces articles.

3.2 Ensemble de simulations

L’ensemble de simulations utilisé pour ce projet est ClimEx et il est efficacement présenté dans l’article de Leduc *et al* [23]. Pour réaliser les simulations climatiques, ClimEx emploie le modèle climatique MRCC5 [15, 16], développé par le Centre pour l’étude et la simulation du climat à l’échelle régionale. Ces simulations sont à haute résolution spatiale d’environ 12 km sur deux domaines, soit l’Europe centrale et le Nord-Est de l’Amérique du Nord. La

simulation du modèle climatique a été produite et distribuée par Ouranos, un consortium sur la climatologie régionale et l’adaptation aux changements climatiques, dans le but de caractériser « la variabilité naturelle du climat et les événements extrêmes dans un contexte de changements climatiques tout en mettant l’emphase sur les impacts hydrologiques » [24]. Les projections climatiques suivent le scénario de carbone élevé RCP8.5 pour les émissions de gaz à effet de serre et d’aérosols.

Ce scénario fait l’hypothèse que la concentration de gaz à effet de serre va continuer de croître à un taux semblable à celui mesuré actuellement. Le GIEC utilise un total de quatre scénarios d’évolution des gaz à effet de serre. Parmi eux, le RCP8.5 est celui qui suppose les plus grandes émissions de GES, causant le plus alarmant changement climatique [13]. Pour obtenir davantage d’informations concernant ce scénario, se référer au site de l’Atlas climatique du Canada [13].

Au total, ClimEx comprend 50 réalisations équiprobables du climat appelées des membres. Ils sont identifiés par une série de 3 lettres, soit KDA, KDB, . . . , KDX, KEA, . . . , KEX. Chaque membre se distingue par de petites perturbations dans ses conditions initiales de façon à représenter la nature incertaine du système climatique actuel [24]. Il est important d’avoir plusieurs réalisations d’un même modèle climatique afin de caractériser de manière plus robuste l’effet du changement climatique sur les événements extrêmes [23]. Utiliser un ensemble de simulations permet de prédire un plus grand éventail d’événements de précipitations extrêmes. Il est supposé qu’après quelques années suivant l’initialisation des 50 simulations de ClimEx, qui a été effectuée en 1950, elles sont des réalisations indépendantes du système climatique [23]. Plus précisément, après 3 ans, les températures et les précipitations de la Terre et dans l’océan sont considérées indépendantes. Pour les processus dont l’évolution est lente, comme par exemple la circulation dans le fond de l’océan, cela prendrait des centaines, voire des milliers, d’années avant qu’ils soient indépendants [23]. C’est pour cela que Leduc *et al* assument l’indépendance quelques années après 1950, sans toutefois donner un nombre précis. Les données simulées utilisées dans ce projet débutent en 1956 afin de laisser suffisamment de temps pour que l’hypothèse d’indépendance soit vraie tout en conservant suffisamment de données pour avoir un ajustement de qualité.

Tel que mentionné dans l’article de Maillhot *et al.* [14], les ensembles multimembres comme ClimEx sont essentiels pour étudier la relation entre les changements climatiques et la structure des modèles climatiques. En l’occurrence, ils permettent d’analyser les impacts des changements climatiques sur les précipitations extrêmes. La valeur ajoutée d’utiliser un ensemble multimembre ainsi que son accessibilité justifient notre choix d’utiliser ClimEx pour le projet. Également, ClimEx utilise un modèle climatique régional doté d’une résolution spatiale très

fine. Cet aspect nous permettra d'obtenir des estimations de phénomènes de pluies intenses à petite échelle géographique. L'ensemble de simulations ClimEx est très récent et c'est ce qui explique pourquoi aucun article scientifique publié n'en fait l'usage jusqu'à présent. Il sera important dans le futur de trouver des articles qui citent cet ensemble de simulations pour s'assurer de sa performance.

CHAPITRE 4 MÉTHODOLOGIE ET RÉSULTATS

Ce chapitre relate les étapes et les résultats du projet de maîtrise. La méthodologie se décompose en trois étapes. Premièrement, il était nécessaire d'extraire seulement les données essentielles à l'étude parmi un immense bassin de données simulées. Suivait ensuite l'analyse exploratoire de ces données pour la durée 24h, afin de savoir si une quelconque tendance non stationnaire s'en dégageait. Cette seconde étape était primordiale afin d'explorer le contenu du jeu de données et le comportement de ces maxima annuels de précipitations. Puis, une synthèse de la non-stationnarité présente dans les données de ClimEx est présentée. Pour cela, un modèle pour maxima non stationnaires a été développé afin de proposer divers candidats pour la création d'un facteur de correction. Ce dernier pourra servir à la mise à jour des courbes IDF. Les résultats sont présentés pour une cellule en particulier.

4.1 Extraction des données

Les précipitations simulées de 1956 à 2100 pour le nord-est de l'Amérique du Nord ont été fournies par le consortium Ouranos. À partir de cela, l'objectif était d'obtenir les MOAM des précipitations pour chaque cellule du territoire afin d'en estimer la non-stationnarité. La grille étant de taille 280×280 , cela constitue un total de 78 400 tuiles à traiter pour chacun des 50 membres.

Cette étape semble relativement simple en théorie, mais elle fût plus ardue que prévu. La quantité d'information contenue dans les fichiers *Network Common Data Form* (NetCDF) fournis par Ouranos était imposante et occupait un espace de stockage de 7.2 téraoctets (To). Pour situer le lecteur, une quantité de 1 To équivaut à 1000 gigaoctets (Go), et les ordinateurs portables les plus communs de nos jours peuvent avoir une capacité de stockage allant jusqu'à 500 Go [25]. Les fichiers de ClimEx qui nous ont été fournis ne pouvaient même pas être stockés directement sur un ordinateur. Il fallait donc s'assurer d'extraire les informations pertinentes nécessaires au projet.

Notre projet de recherche a utilisé le logiciel NetCDF développé par UCAR/Unidata (voir <http://doi.org/10.5065/D6H70CW6>) afin de suivre le format des fichiers de ClimEx. Ce logiciel constitue un ensemble d'interfaces visant l'accès aux données orienté objets. Il est gratuit, *open source* et a pour but de promouvoir la création, l'accès et le partage de données scientifiques. Les fichiers NetCDF sont grandement utilisés en climatologie, d'où la convention *Climate and Forecast* (CF) (<http://cfconventions.org/>). Cette dernière a été créée afin de

contribuer au traitement et au partage des fichiers NetCDF. Elle fournit une description définitive du contenu de chaque variable ainsi que des propriétés dans le temps et l'espace des données, dans le but que les usagers puissent comparer des données provenant de sources diverses. Ce type de fichier était nouveau et a demandé une bonne adaptation pour en extraire les données.

Pour chacune des cellules des 50 membres, il y avait donc une quantité colossale de données et notre tâche était d'en faire le tri. Tout d'abord, il y avait les données de précipitations à extraire pour le calcul des maxima annuels. Ces précipitations étaient présentées sous la forme de la quantité de pluie tombée à chaque intervalle d'une heure sur une période de plus de 140 ans. Il fallait en second lieu récolter des informations par rapport aux cellules, comme la latitude et la longitude du centre de la cellule, l'élévation moyenne de la cellule et le temps. Ces données étaient cruciales à extraire afin de pouvoir représenter les maxima annuels de façon spatiale. La quantité de données extraites étant très importante, les codes utilisés devaient être efficaces et concis afin de ne pas avoir des temps de calculs gigantesques. L'optimisation des temps de calcul a donc été une préoccupation centrale tout au long du projet de maîtrise.

Les précipitations ont été extraites pour les mois de mai à octobre. En effet, comme ClimEx a des informations sur les précipitations de neige et que cela pourrait influencer les précipitations de pluie, les mois qui correspondent à la période hivernale au Québec ont été retirés. Les données de précipitations de pluie seulement étaient d'intérêt à la recherche. Cette pratique d'exclure des mois pour affiner les estimations de précipitations extrêmes est grandement utilisée (voir par exemple [14, 21]).

Au final, ce sont les données de précipitations du Nord-Est de l'Amérique du Nord de 1956 à 2099 qui ont été extraites pour les mois de mai à octobre. Elles représentaient l'intensité de pluie tombée en une heure exprimée en $\frac{kg}{m^2 \times s}$ pour chaque cellule sur la période d'étude. Avec ces données, les quantités pour chacune des cinq durées étaient ensuite calculées en utilisant des fenêtres glissantes. Cela consiste à additionner les quantités horaires sur une fenêtre de temps de longueur égale à la durée concernée en heures. À titre de rappel, les durées pour le projet sont les suivantes : 1, 2, 6, 12 et 24 heures.

Une fois les quantités de pluie calculées pour chaque durée, les maxima annuels étaient extraits sur ces données. Cela donne une série de 144 maxima annuels par durée par cellule et ce, pour les 50 membres. Les maxima annuels devaient ensuite être convertis en mm/h de façon à pouvoir comparer les quantités de pluie intense entre différentes durées. Le facteur de conversion $f_d = 3600/d$ était multiplié à chaque maximum, pour d correspondant à la durée en heure.

La méthode qui vient d'être énoncée pour obtenir des maxima exprimés en mm/h est expliquée plus en détails pour une durée spécifique. Par exemple, pour la durée 2h, la longueur de la fenêtre était de deux unités donc les quantités horaires étaient additionnées tour à tour dans le but d'obtenir les quantités de précipitations pour la durée 2h, comme illustré à la figure 4.1. Ensuite, pour effectuer le changement d'unité, on multiplie le nombre de secondes dans une heure, soit 3600, aux données. Cela donne donc les quantités de pluie tombées en 2h. Il ne reste plus qu'à diviser les maxima par la durée concernée exprimée en heure pour obtenir la quantité moyenne par heure. Ce facteur est donc résumé par $f_d = 3600/d$.

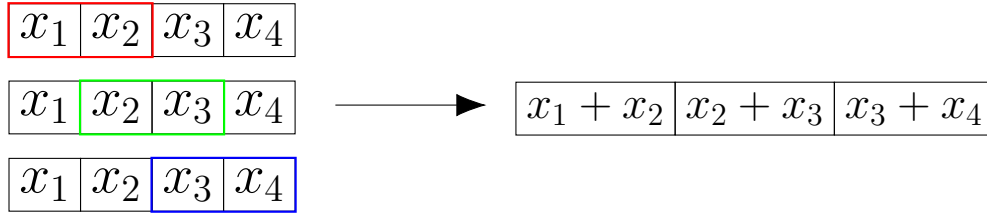


Figure 4.1 Fenêtre glissante permettant d'obtenir la quantité de pluie tombée sur deux heures

Bref, les MOAM utilisés dans le projet s'étendent sur la période de 1956 à 2099 inclusivement et sont disponibles pour chacune des cinq durées suivantes : 1, 2, 6, 12 et 24 heures. Les données simulées sont prises telles quelles à partir de ClimEx et il n'y aura pas d'analyse à savoir si l'ensemble de simulations contient des données aberrantes. Les maxima de précipitations peuvent maintenant être utilisés pour le restant du projet afin d'extraire le maximum d'information concernant la non-stationnarité des événements de pluie intense.

4.2 Analyse exploratoire

La présente section concerne l'analyse exploratoire des données. Cette dernière se penche sur la possible présence de non-stationnarité dans le jeu de données. Comme mentionné dans le chapitre 1, les précipitations extrêmes sont influencées par les changements climatiques et sont considérées dans plusieurs études comme une série non stationnaire. Une analyse exploratoire est utile pour illustrer le comportement de ces maxima annuels de précipitations et guider la modélisation de la série.

Un thème central de la recherche présentée est d'optimiser les temps de calcul. Le projet portant sur les précipitations extrêmes au Québec précisément, une sous-grille a été utilisée pour l'analyse exploratoire ainsi que pour toutes les étapes restantes du projet. La grille utilisée contient le Québec et est de taille 136×154 , réduisant le nombre de cellules à analyser à 20 944. Le fait d'exclure les tuiles situées à l'extérieur de notre région d'intérêt

diminue considérablement le temps de calcul. Même si dans ce mémoire les résultats seront interprétés pour une unique cellule, il était nécessaire de garder toutes les cellules de la grille de taille 136×154 pour effectuer l'analyse exploratoire des données. En effet, cette étape nous permet d'avoir une idée d'ensemble des données et de développer une méthodologie qui sera applicable sur tout le Québec.

L'analyse de stationnarité a été effectuée sur les maxima annuels de la durée 24h contenus dans la grille. Les données ont été utilisées en *mm/jour* dans cette section afin de faciliter l'interprétation des quantités de pluie estimées pour la durée utilisée.

4.2.1 Recherche de non-stationnarité

Pour explorer plus en profondeur un membre de ClimEx, cinq villes ont été choisies arbitrairement puisqu'elles couvrent bien le territoire du Québec. Les villes choisies sont Fermont, Gaspé, Montréal, Radisson et Salluit et sont présentées à la figure 4.2. Elles ont été sélectionnées car elles ont des positions géographiques variées et sont potentiellement exposées à des climats différents, ce qui risque d'être intéressant pour l'analyse.

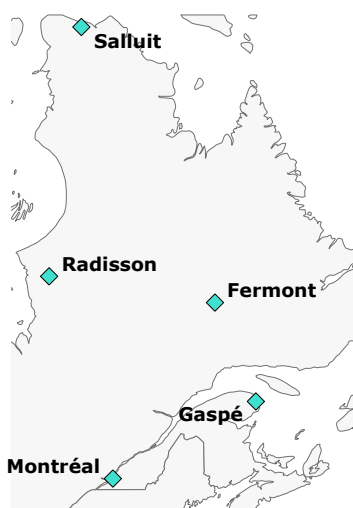
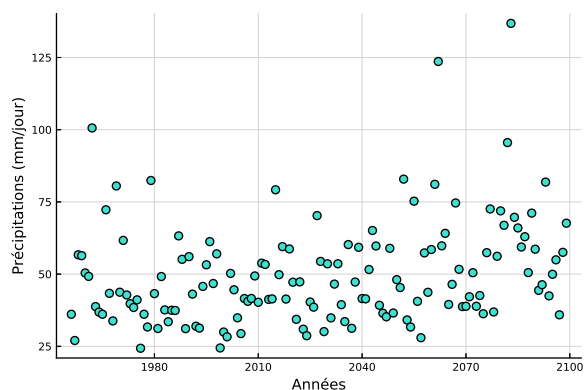


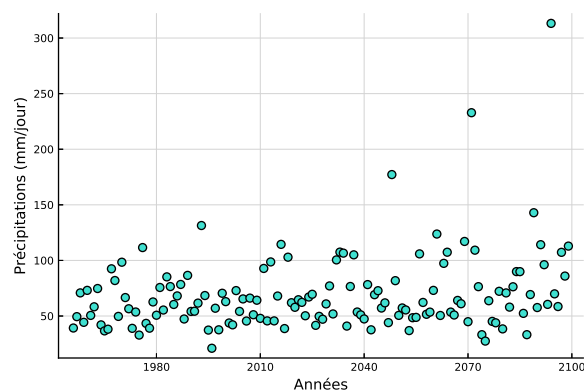
Figure 4.2 Cellules choisies

La figure 4.3 illustrent les maxima annuels pour les cinq cellules correspondant aux villes choisies. Elles correspondent aux cellules du membre KDA dont la latitude et la longitude se rapprochent le plus des coordonnées géographiques des villes ciblées. Les villes de Fermont et Salluit semblent avoir une série de maxima non stationnaire puisque les précipitations augmentent sur la période d'étude. Pour les villes de Gaspé et Radisson, il est plus difficile de dégager une quelconque tendance à partir des données. La série de précipitations de la ville

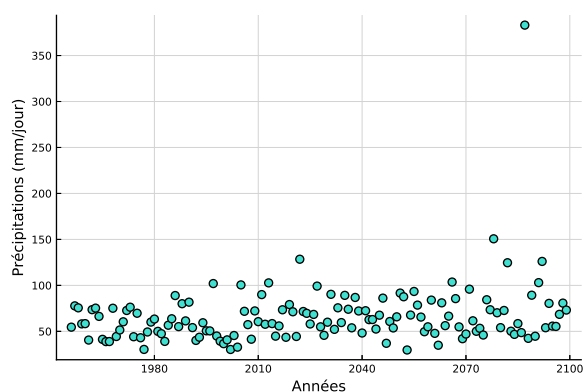
de Montréal semble quant à elle plutôt stationnaire, à l'exception d'une donnée très élevée par rapport aux autres maxima annuels. Ce ne sont que des suppositions suite à l'analyse de graphiques et l'analyse exploratoire des données va donc permettre de savoir s'il y a réellement de la non-stationnarité présente dans les données de ClimEx.



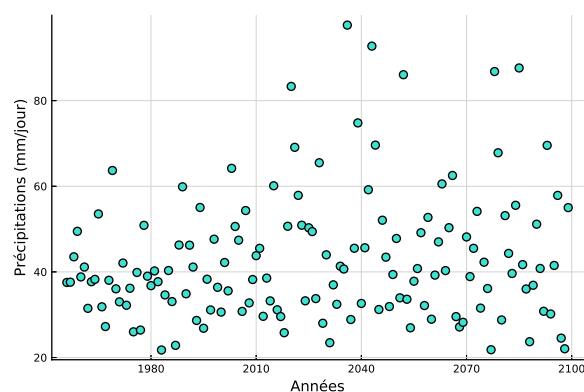
(a) Ville de Fermont



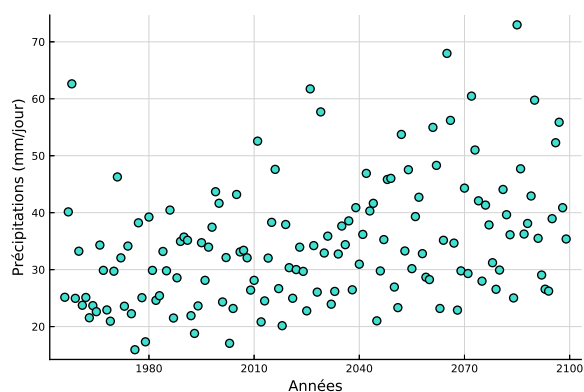
(b) Ville de Gaspé



(c) Ville de Montréal



(d) Ville de Radisson



(e) Ville de Salluit

Figure 4.3 Séries temporelles des maxima pour cinq cellules du membre KDA

Tout d'abord, il fallait déterminer quelle variable explicative était retenue pour l'ajustement.

À titre de rappel, il a été mentionné dans l'introduction que le changement climatique est en grande partie dû aux émissions de GES, où le CO_2 y joue un rôle majeur. Comme mentionné abondamment dans la littérature, les GES ont un effet sur la température, qui elle affecte les précipitations [7]. Même s'il est logique de penser que la température serait une bonne variable explicative potentielle, elle n'a pas été retenue. Cela ajouterait une couche de données simulées supplémentaire et compliquerait le processus si le choix de scénario de GES devait être modifié en cours d'expérience. Il est astucieux de modéliser les précipitations extrêmes avec l'origine même de la chaîne causale, soit les GES. Conséquemment, les deux variables explicatives testées pour ce projet de maîtrise sont les concentrations annuelles de CO_2 et de GES pondérés selon leur pouvoir réchauffant. À la figure 4.4, les valeurs que prennent ces deux variables sont tracées. Elles augmentent toutes les deux de façon continue sur la période d'étude.

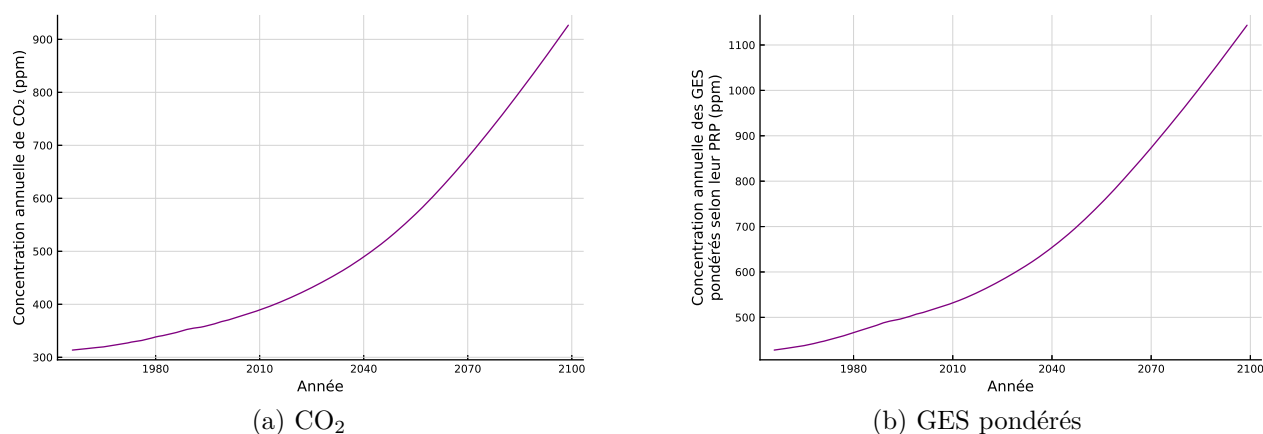


Figure 4.4 Variables explicatives possibles en fonction du temps

La concentration annuelle de CO_2 a été testée comme variable explicative pour le paramètre de position afin de modéliser la non-stationnarité présente dans le jeu de données. En effet, cette quantité a une influence sur les précipitations extrêmes, comme démontré dans l'étude de Zhang, Villarini, Scoccimarro et Vecchi [26]. Étant donné que le CO_2 n'est pas le seul gaz à affecter le réchauffement climatique, la quantité totale des principaux GES pondérés selon leur potentiel de réchauffement planétaire (PRP) a également été testée comme variable explicative. La liste des GES utilisés pour cette variable explicative est la suivante : CO_2 , protoxyde d'azote (N_2O), méthane (CH_4), chlorofluorocarbures (CFC)11 et CFC12. Le PRP d'un GES est une mesure permettant de calculer la capacité de ce gaz à piéger la chaleur dans l'atmosphère par rapport au CO_2 [27]. En effet, chaque GES est différent et possède une durée de vie dans l'atmosphère et un potentiel de rétention de la chaleur qui lui sont spécifiques. Ainsi, chacun des cinq gaz a été pondéré de la manière suivante selon leur PRP [27,28] et les

concentrations sont en parties par million (ppm) :

$$\text{GES}(t) = \text{CO}_2(t) + 298\text{N}_2\text{O}(t) + 23\text{CH}_4(t) + 4000\text{CFC11}(t) + 8500\text{CFC12}(t). \quad (4.1)$$

Pour ce projet, il a été décidé d'incorporer la non-stationnarité dans la paramètre de position via une relation linéaire avec la variable explicative et de garder les paramètres d'échelle et de forme constants. D'autres options pourront être explorées dans le futur, comme incorporer de la non-stationnarité dans le paramètre d'échelle et utiliser d'autres relations, comme une relation quadratique.

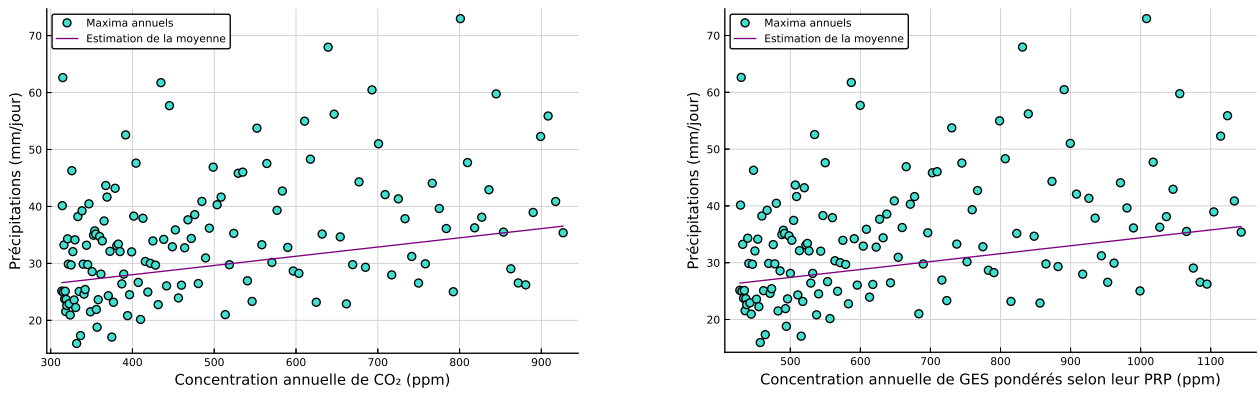


Figure 4.5 Ajustement selon les deux variables explicatives possibles pour la ville de Salluit

Avec la même cellule utilisée pour la ville de Salluit à la figure 4.3, deux ajustements ont été effectués de façon indépendante. Une loi GEV dont le paramètre de position a une relation linéaire avec la variable explicative a été ajustée, donnant des estimations pour les paramètres μ_0 , μ_1 , σ et ξ . Cela a été fait à l'aide de la méthode du maximum de vraisemblance pour les deux variables explicatives possibles. Les séries de maxima sont présentées à la figure 4.5, où la ligne mauve représente l'estimation du paramètre de localisation selon la variable explicative. L'estimation de ce paramètre est présentée pour les deux variables explicatives :

$$\mu(t) = \mu_0 + \mu_1 \times \text{CO}_2(t) \quad \mu(t) = \mu_0 + \mu_1 \times \text{GES}(t)$$

pour une année t en particulier.

D'après la figure 4.5, il n'est pas toujours évident de déterminer laquelle des variables explicatives possibles effectue un meilleur ajustement en regardant les données. Pour faire un choix, les différents ajustements sont comparés à l'aide de l'indice *Bayesian Information Criterion* (BIC). Cet indice développé par Schwarz en 1978 [29] est abondamment utilisé dans

la littérature pour la sélection de modèles. Il est défini dans l'équation suivante [30] :

$$\text{BIC} = -2\ln(\hat{\mathcal{L}}) + k \times \ln(n) \quad (4.2)$$

où $\hat{\mathcal{L}}$ représente la vraisemblance, k le nombre de paramètres du modèle et n la taille de l'échantillon. Le modèle présentant la plus petite valeur de l'indice BIC est celui qui convient le mieux au jeu de données pour cette tuile en particulier. La différence entre les indices BIC des modèles en comparaison peut être interprétée à l'aide du tableau suivant [31] :

Tableau 4.1 Interprétation de l'écart entre deux indices BIC

ΔBIC	Évidence contre le plus grand BIC
0-2	Faible
2-6	Positive
6-10	Forte
>10	Très forte

Dans le cas de la figure 4.5, la concentration annuelle de GES pondérés avait un plus petit BIC et offrait donc un meilleur ajustement. Cela signifie qu'il y a une augmentation significative de l'intensité des précipitations extrêmes sur la période d'étude pour cette cellule. Les détails concernant l'écart entre les différents indices BIC pour cette cellule sont énoncés plus bas.

Afin d'illustrer en détails la procédure qui sera appliquée sur toutes les cellules des 50 membres, nous débutons par ajuster la loi GEV de façon indépendante sur chacune des cellules arbitraires pour trois modèles différents avec la méthode du maximum de vraisemblance. Ce sont le modèle stationnaire, le modèle non stationnaire avec la variable explicative de la concentration annuelle de CO_2 et le modèle non stationnaire avec la concentration annuelle de GES. L'indice BIC servira à trancher quel ajustement est le meilleur. Comme les cellules sont bien étendues sur le territoire québécois, l'ajustement des précipitations sur ces différentes villes sera intéressant puisqu'on peut s'attendre à des résultats différents d'une ville à une autre.

Premièrement, pour la ville de Fermont située dans la région de la Côte-Nord, c'est le modèle non stationnaire avec la concentration de CO_2 qui a le plus petit indice BIC d'une valeur d'environ 1185.2689. L'autre modèle non stationnaire avait un indice très proche, soit de 1185.5125 et le modèle stationnaire était nettement plus haut avec un BIC de 1196.3395. L'écart entre les deux modèles non stationnaires étaient donc très petit dans ce cas. Si on compare plutôt les deux modèles non stationnaires avec le modèle stationnaire, l'écart entre les indices BIC est d'un peu plus de 10 dans les deux cas. Selon le tableau 4.1, cela traduit

une très forte évidence que le modèle stationnaire n'offre pas le meilleur ajustement. Comme illustré à la figure 4.6, le paramètre μ de cette série de précipitations extrêmes augmente légèrement avec la hausse des concentrations de CO_2 et la dispersion des données ne semble pas évoluer sur le graphique.

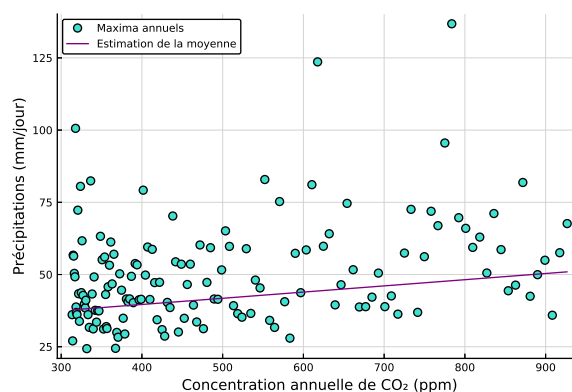


Figure 4.6 Maxima annuels de précipitations pour la durée 24h pour la ville de Fermont

Pour continuer, la ville de Gaspé en Gaspésie avait un meilleur ajustement dans le cas du modèle stationnaire. En effet, l'indice BIC du modèle stationnaire était d'environ 1342.9814 et autour de 1346 pour les deux modèles non stationnaires. Dans ce cas, l'évidence est positive de rejeter les deux modèles non stationnaires. Le modèle stationnaire est celui qui convient le mieux aux données de la ville provenant de ClimEx et les trois paramètres estimés sont donc considérés constants. Les maxima selon les différentes variables explicatives et leur paramètre μ sont présentés à la figure 4.7, afin de montrer que dans les deux cas, le paramètre de position est constant. La dispersion des données de précipitations ne semble pas changer selon la variable explicative.

Pour la ville de Montréal, qui est une île fluviale située dans le sud du Québec, c'est le modèle stationnaire qui offre un meilleur ajustement avec un indice BIC de 1294.6194. Les deux modèles non stationnaires se situent non loin avec des BIC autour de 1295. D'après le tableau 4.1, il n'y a pas d'évidence positive contre les deux modèles non stationnaires. La série de précipitations extrêmes pour Montréal est présentée à la figure 4.8, où le paramètre de position illustré est celui correspondant au modèle stationnaire. Les maxima ne semblent pas avoir un paramètre μ ni une dispersion qui évoluent selon la variable explicative. Par contre, cette série de données simulées est assez particulière, étant donné la valeur de plus de 350mm pour une durée de 24h aux alentours d'une concentration de GES pondérés de 1000 ppm.

La ville de Radisson dans la région du Nord-du-Québec alimente la tendance stationnaire,

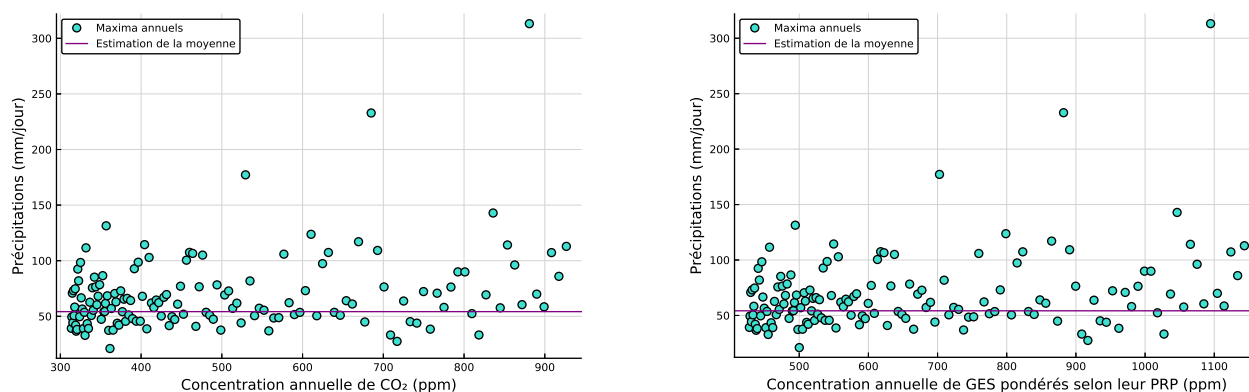


Figure 4.7 Maxima annuels de précipitations pour la durée 24h pour la ville de Gaspé selon les variables explicatives

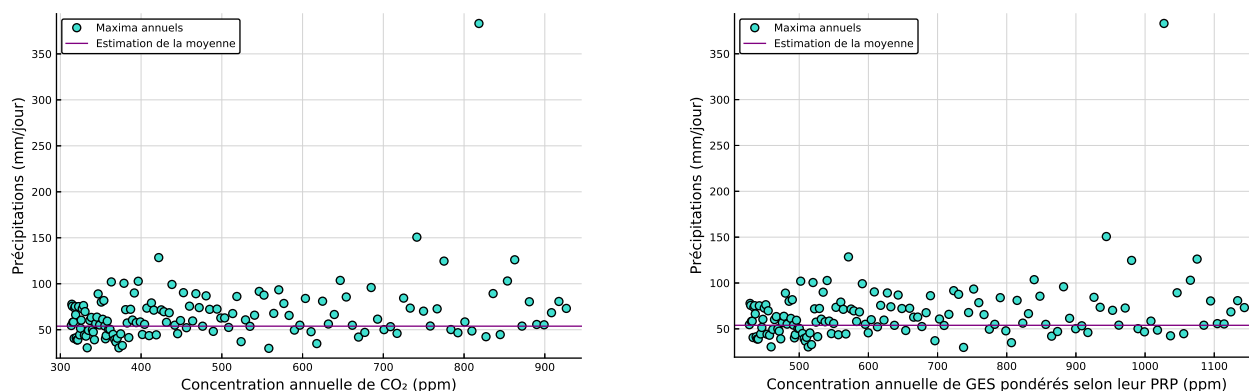


Figure 4.8 Maxima annuels de précipitations pour la durée 24h pour la ville de Montréal selon les variables explicatives

puisque c'est encore le modèle stationnaire qui offre le meilleur ajustement. Ce modèle considérant tous les paramètres constants a obtenu un BIC d'environ 1155.2161, tandis que les modèles non stationnaires ont eu des indices autour de 1159.7. L'écart entre l'indice BIC du modèle stationnaire et celui des deux modèles non stationnaires montre qu'il y a une évidence positive contre les modèles qui incorporent de la non-stationnarité. Les maxima ainsi que leur paramètre de localisation selon la variable explicative sont présentés à la figure 4.9, où la dispersion ne semble pas évoluer.

Puis, pour le village nordique Salluit, le modèle non stationnaire avec la concentration annuelle de GES pondérés selon leur PRP avait le plus petit indice BIC comme mentionné précédemment. Plus précisément, l'indice BIC avait une valeur de 1084.4963 pour le modèle stationnaire, 1070.3019 pour le modèle non stationnaire avec la concentration de CO₂

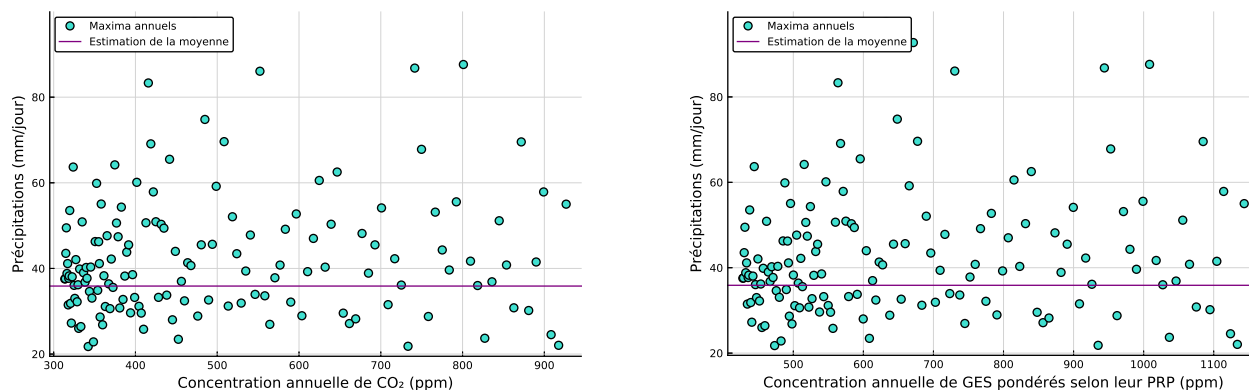


Figure 4.9 Maxima annuels de précipitations pour la durée 24h pour la ville de Radisson selon les variables explicatives

et 1069.9554 pour le modèle avec la concentration de GES pondérés. Il ne semble donc pas avoir de différence significative entre les deux modèles non stationnaires. Par contre, l'évidence contre le modèle stationnaire ici est qualifiée de très forte par rapport aux deux modèles non stationnaires. Les données ont déjà été présentées à la figure 4.5, où l'on peut voir que le paramètre μ évolue avec les deux variables explicatives.

Pour résumer, on remarque que la différence entre les indices BIC des deux modèles non stationnaires était dans tous les cas très petite. Comme attendu, les ajustements étaient différents d'une ville à une autre. Les trois villes choisies qui sont situées plus au sud du Québec avaient toutes un meilleur ajustement avec le modèle stationnaire. Ce résultat est tout de même particulier, puisqu'il induit que les maxima de précipitations de ces villes ne seront pas affectés par l'augmentation du CO_2 et/ou de GES au fil des années. Les indices BIC pour Montréal étaient cependant tous très proches, laissant croire que les données contiennent peut être de la non-stationnarité en faible quantité. Les villes plus au nord, soit Fermont et Salluit, avaient un indice BIC plus petit pour les modèles non stationnaires. Sur les cinq tuiles du membre KDA analysées, une seule était mieux représentée par un modèle non stationnaire ayant comme variable explicative la concentration annuelle des GES pondérés selon leur pouvoir réchauffant. Cette sous-section était pertinente afin de montrer que le modèle offrant le meilleur ajustement varie d'une cellule à une autre. Ces résultats ne peuvent cependant pas être généralisés. Il faut effectuer l'analyse sur l'ensemble des cellules pour les 50 membres.

4.2.2 Choix du modèle d'ajustement

La procédure qui vient d'être utilisée pour les cellules arbitraires est maintenant effectuée en deux étapes sur l'ensemble des cellules de la grille pour les 50 membres. Il est important d'inclure toutes les cellules du Québec dans l'analyse afin de choisir le modèle qui convient le mieux à la majorité.

Pour débiter, les deux modèles non stationnaires sont ajustés de façon indépendante. Par conséquent, la concentration de CO₂ ainsi que la concentration de GES pondérés selon leur PRP ont été ajustées séparément comme variable explicative à relation linéaire avec le paramètre de position de la loi GEV. La comparaison a été effectuée sur l'ensemble des cellules pour chacun des 50 membres à l'aide de l'indice BIC. Pour 48 membres sur un total de 50, incluant le membre KDA, les cellules avaient en moyenne un plus petit BIC lorsque la variable explicative était représentée par la concentration de GES pondérés. Cela se traduit par un meilleur ajustement des données. C'est un résultat assez convaincant, puisque ce sont presque toutes les simulations de ClimEx qui préfèrent la variable explicative des GES à celle du CO₂. Le meilleur modèle non stationnaire est donc celui avec la variable explicative de GES d'après l'indice BIC. Ce résultat est logique puisque cette quantité ne fait qu'ajouter de l'information supplémentaire à la concentration de CO₂.

Le modèle non stationnaire peut maintenant être défini. Il comprend un paramètre de position à relation linéaire avec la variable explicative, et des paramètres d'échelle et de forme constants à travers le temps. Ils sont exprimés de la manière suivante, pour t allant de 1956 à 2099 :

$$\begin{aligned}\mu(t) &= \mu_0 + \mu_1 \times \text{GES}(t) \\ \sigma(t) &= \sigma \\ \xi(t) &= \xi.\end{aligned}\tag{4.3}$$

La variable explicative du paramètre de position représente la concentration annuelle dans l'atmosphère de GES pondérés selon leur PRP en ppm selon le scénario de carbone élevé RCP8.5.

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

$$\begin{aligned}\mu(t) &= \mu \\ \sigma(t) &= \sigma \\ \xi(t) &= \xi.\end{aligned}\tag{4.4}$$

Le principe est le même que précédemment, où les indices BIC des deux modèles sont comparés afin de déterminer le meilleur ajustement. Pour la totalité des 50 membres, le modèle non stationnaire avec la variable explicative de GES est favorable au modèle stationnaire pour la majorité des cellules de la grille. La tendance observée avec les trois cellules arbitraires au sud du Québec n'était donc pas généralisée à toute la province. Ce résultat est crucial, puisqu'il vient confirmer la présence de non-stationnarité dans les précipitations extrêmes. Autrement dit, pour les 50 membres, le paramètre μ_1 est significatif pour la majorité des cellules. Il y a donc une augmentation significative de l'intensité des événements de pluie intense au Québec.

Pour résumer, il a été justifié dans le chapitre 1 que les GES ont une influence sur les précipitations extrêmes. Il a été démontré dans la présente sous-section, à l'aide de l'indice BIC, que le modèle non stationnaire dont le paramètre de position a une relation linéaire avec la concentration annuelle de GES pondérés a un meilleur ajustement sur les maxima annuels.

4.2.3 Membre KDA

Pour continuer, les résultats obtenus avec le modèle non stationnaire pour le membre KDA sont présentés. À titre de rappel, le modèle non stationnaire est présenté à l'équation (4.3) et correspond à celui dont le paramètre de position a une relation linéaire avec la concentration annuelle de GES pondérés selon leur PRP. Les paramètres de la loi GEV ont été estimés par la méthode du maximum de vraisemblance et sont représentés de façon spatiale sur la figure 4.10. Un polygone représentant le Québec a été utilisé pour la création de ces cartes.

L'ordonnée à l'origine du paramètre de position, soit μ_0 , semble être distribuée selon un gradient nord-sud. En effet, la localisation des extrêmes est plus grande dans le sud. Le même constat est à faire pour la carte du paramètre d'échelle, soit σ , qui représente la dispersion de la série de maxima. Cette tendance nord-sud met en lumière un résultat très intéressant, soit que les précipitations extrêmes ont une certaine structure spatiale sur le territoire. Les événements de pluies intenses auraient un paramètre de position plus élevé et également une dispersion plus grande au sud du Québec. Partant de ce fait, les maxima annuels de précipitations extrêmes ont une moyenne plus élevée dans le sud de la province. Toujours sur la figure 4.10, le schéma spatial de la pente du paramètre de position, soit μ_1 , laisse croire qu'il s'agit de bruit spatial. Les données ne semblent pas être distribuées sur le territoire selon une quelconque structure. Un constat semblable est à faire pour le paramètre d'échelle. Il semble y avoir des valeurs plus élevées dans le nord, mais c'est très diffus, et ça ne semble pas correspondre à une structure spatiale précise.

Pour le membre KDA, deux grandes tendances peuvent être déduites des schémas spatiaux présentés. Les paramètres μ_0 et σ suivent une tendance nord-sud sur le territoire, tandis que les paramètres μ_1 et ξ ne varient pas vraiment dans l'espace. Il sera intéressant de comparer ces résultats avec les cartes produites avec les estimations moyennes des 50 membres, puisqu'une seule simulation n'est sûrement pas représentative de l'ensemble ClimEx.

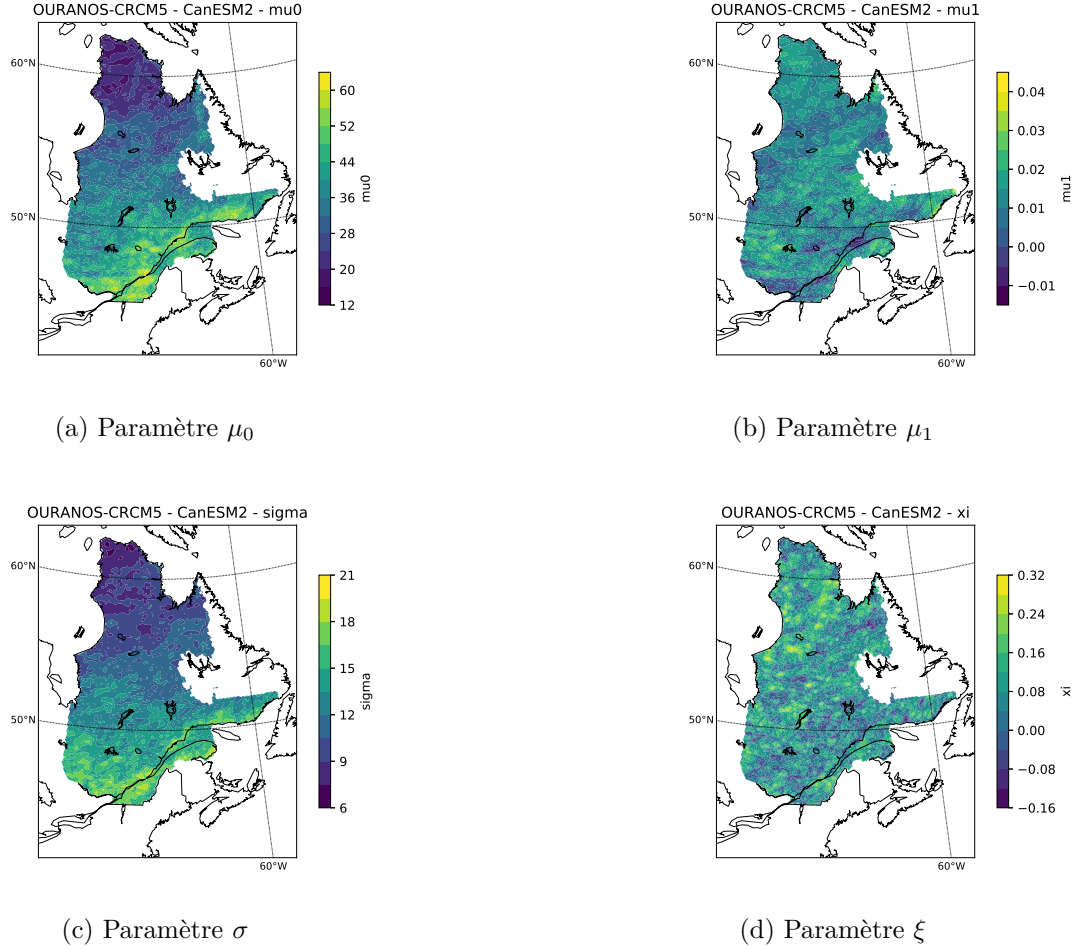


Figure 4.10 Estimations des paramètres de la loi GEV pour le membre KDA et la durée 24h

4.2.4 Moyenne des 50 membres

L'analyse est effectuée en dernier lieu sur l'ensemble des 50 membres de ClimEx. De manière générale, une tendance qui n'est pas nécessairement visible pour une seule simulation ressortira dans l'analyse moyenne sur les 50 membres. Les différentes formules qui seront utilisées pour l'analyse sont tout d'abord présentées. L'estimation moyenne de chaque paramètre de la loi GEV était voulue pour illustrer les résultats de façon spatiale. À titre d'exemple, l'es-

timation moyenne du paramètre μ_0 pour la cellule de coordonnées (i, j) sur l'ensemble des 50 membres a été calculée de la façon suivante :

$$\bar{\mu}_0(i, j) = \frac{1}{50} \sum_{k=1}^{50} \mu_0(i, j, k),$$

où $\mu_0(i, j, k)$ représente l'estimation du paramètre μ_0 pour la cellule (i, j) et le k^e membre. La même démarche a été utilisée pour les paramètres restants (μ_1 , σ et ξ).

Étant donné que le modèle d'ajustement incorpore de la non-stationnarité, nous avons ciblé trois années dans la période pour pouvoir obtenir des estimations du paramètre de position et du quantile associé à une période de retour de 20 ans. Les trois années sont : 1990, 2050 et 2099 et elles ont été choisies arbitrairement puisqu'elles couvrent bien l'ensemble de la période d'étude.

Ainsi, la moyenne sur les 50 membres du paramètre de position μ devait être estimée pour ces trois années. Pour son calcul, les estimations moyennes des paramètres μ_0 et μ_1 ont été utilisées directement. Pour la tuile (i, j) et l'année t , voici le calcul afin d'obtenir l'estimation moyenne de μ :

$$\bar{\mu}(i, j, t) = \bar{\mu}_0(i, j) + \bar{\mu}_1(i, j) \times \text{GES}(t).$$

Puis, les niveaux de retour sont alors estimés pour les trois années énoncées précédemment pour la période de retour $T = 20$ années. Ce calcul est effectué pour chaque membre, sans utiliser les estimations moyennes, puisqu'il s'agit d'une équation non linéaire : Soit $p = 1/T$,

$$z_p(t) = \begin{cases} \mu(t) - \frac{\sigma}{\xi} \times [1 - \{-\log(1 - p)\}^{-\xi}] & \text{si } \xi \neq 0, \\ \mu(t) - \sigma \times \log \{-\log(1 - p)\} & \text{si } \xi = 0. \end{cases} \quad (4.5)$$

On cherche à estimer le niveau de pluie intense qui sera dépassé en moyenne une fois aux 20 ans pour chaque cellule.

Il reste donc le calcul de l'estimation moyenne du quantile pour chacune des trois années à présenter. La notation est modifiée ici pour simplifier la formule. Le quantile dénommé z_t représente en fait le quantile $z_p(t)$ de l'année t avec $p = 1/T$. Pour la cellule (i, j) , le calcul de l'estimation moyennne du quantile de période de retour $T = 20$ de l'année t est le suivant :

$$\bar{z}_t(i, j) = \frac{1}{50} \sum_{k=1}^{50} z_t(i, j, k),$$

où $z_t(i, j, k)$ représente l'estimation du quantile $z_p(t)$ pour la cellule (i, j) et le k^e membre.

La procédure appliquée à chaque membre est maintenant présentée. Tout d'abord, l'estimation des paramètres de la loi GEV calculée plus haut est nécessaire ici. À titre de rappel, ces estimations ont été obtenues en ajustant de façon indépendante la loi GEV sur chacune des cellules de la grille selon le modèle non stationnaire présenté à l'équation (4.3). Pour chacun de ces paramètres, une carte représentant de façon spatiale les valeurs de l'estimation a été créée, comme présenté pour le membre KDA dans la sous-section précédente. Le paramètre de position ainsi que le quantile sont estimés pour les trois années ciblées. Puis, les quantités moyennes ont été calculées avec les démarches présentées en début de sous-section.

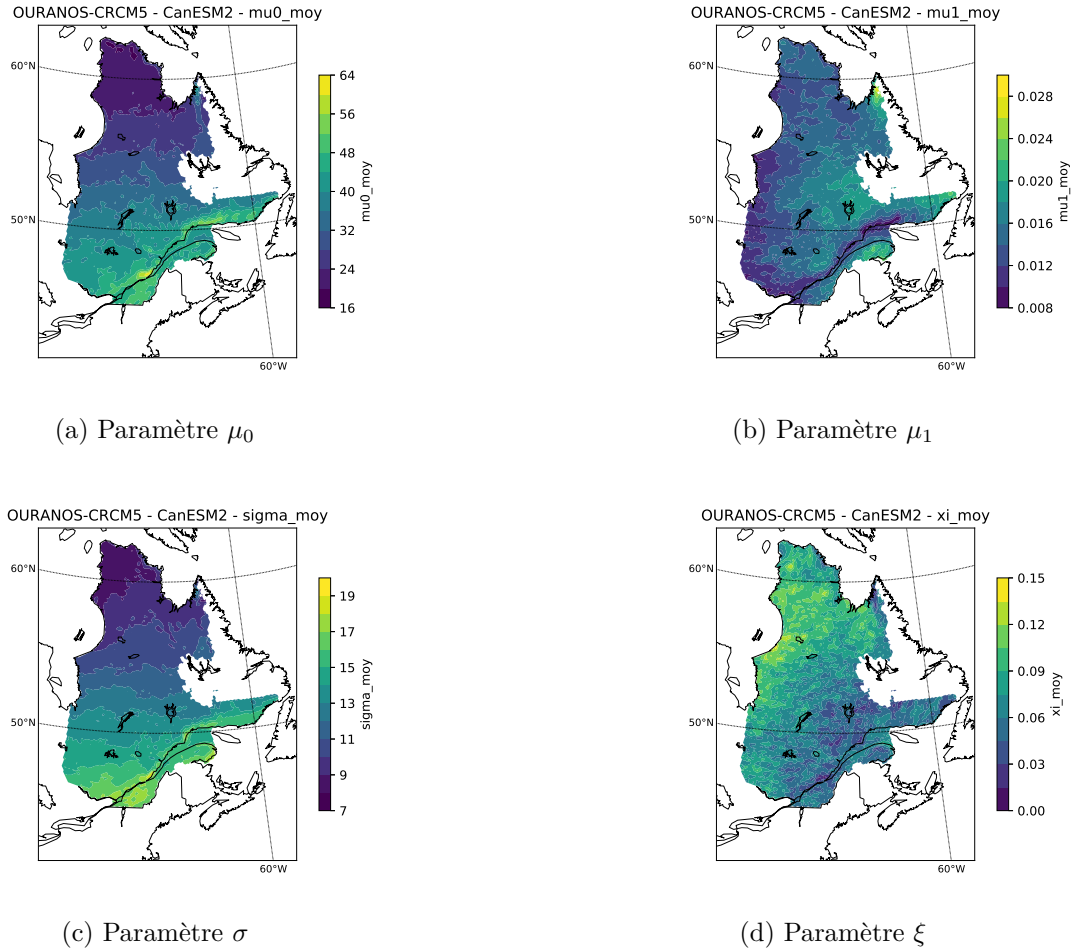


Figure 4.11 Moyenne des estimations des paramètres de la loi GEV sur les 50 membres pour la durée 24h

Avec toutes ces données moyennes, des schémas spatiaux ont pu être créés pour l'analyse. À la figure 4.11, les estimations moyennes des paramètres de la loi GEV sont présentées. Les mêmes tendances semblent se dégager des cartes de μ_0 et σ que lors de l'analyse du membre KDA. En effet, pour le schéma spatial du paramètre μ_0 , l'estimation augmente à mesure que

l'on se déplace vers le sud de la province. Même remarque pour la carte de σ . Ensuite, de légères tendances qui n'étaient pas visibles pour la première simulation semblent ressortir ici. L'estimation moyenne de μ_1 apparaît être plus petite sur les côtes à l'ouest de la province. Comme ce paramètre incorpore la non-stationnarité dans le modèle d'ajustement, on peut dire, grâce au schéma spatial de l'estimation moyenne de μ_1 , que la non-stationnarité des précipitations extrêmes semble plus élevée dans l'est du Québec. Par la suite, l'estimation du paramètre de forme semble quant à elle augmenter vers le nord de la province. Si $\xi > 0$, il est dit en théorie que les maxima n'ont pas de borne supérieure. La queue de la distribution dans ce cas est donc plus épaisse, ce qui veut dire qu'on peut s'attendre à avoir davantage d'événements extrêmes dans le nord du Québec. En résumé, les paramètres μ_0 et σ prennent de plus grandes valeurs au sud de la province et ces tendances étaient visibles même pour une seule simulation. Pour les paramètres μ_1 et ξ , les tendances spatiales ressortaient plutôt lors de l'analyse des estimations moyennes sur les 50 membres. Cela démontre l'importance d'utiliser plusieurs réalisations d'un modèle climatique, afin de calculer des estimations moyennes sur tous les membres et de mieux représenter la nature incertaine du système climatique.

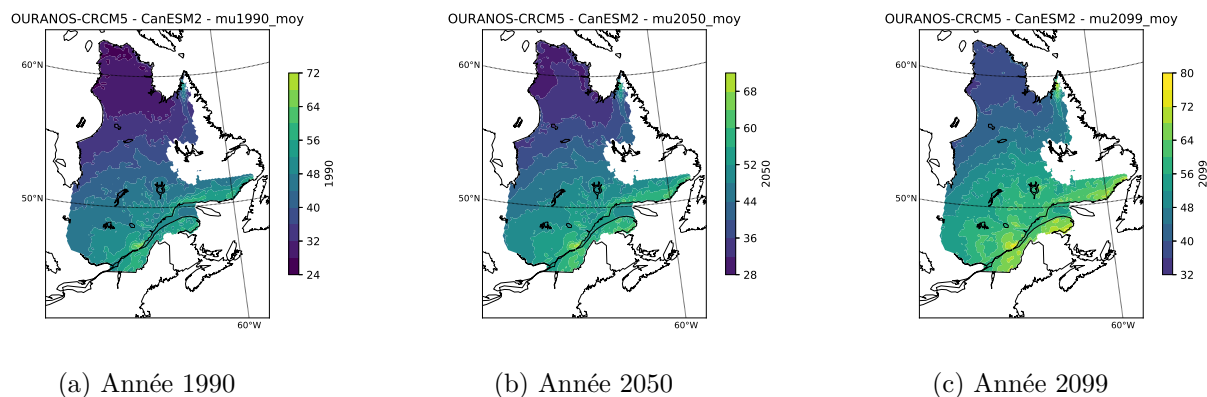


Figure 4.12 Moyenne de l'estimation du paramètre de position de la loi GEV sur les 50 membres pour la durée 24h

L'estimation moyenne du paramètre de position est quant à elle disponible à la figure 4.12 pour les trois années choisies. Il est bien de savoir que les valeurs de l'échelle de couleurs sont comprises dans le même intervalle pour chacune des trois cartes afin de faciliter leur comparaison. L'analyse des deux premières cartes permet de voir qu'il y a une légère modification de μ . L'estimation du paramètre de position semble avoir augmenté faiblement dans le nord du Québec entre 1990 et 2050. Cette augmentation est toutefois très faible, comme en témoignent les intervalles de valeurs. Pour l'estimation de 1990, μ se situe environ entre 24 et 72 tandis que pour l'estimation de 2050, les valeurs sont entre 28 et 72 d'après

la figure 4.12. En comparant les cartes des années 2050 et 2099, on remarque cette fois-ci qu'il y a une nette augmentation sur l'ensemble du territoire. Le Québec en entier apparaît plus pâle sur la carte de 2099 en comparaison avec celle de 2050, ce qui correspond à une augmentation de l'estimation moyenne de μ , où les valeurs se situent maintenant entre 32 et 80. La même tendance nord-sud des précipitations est observée sur les trois schémas spatiaux et la moyenne des précipitations extrêmes est plus élevée dans le sud du Québec.

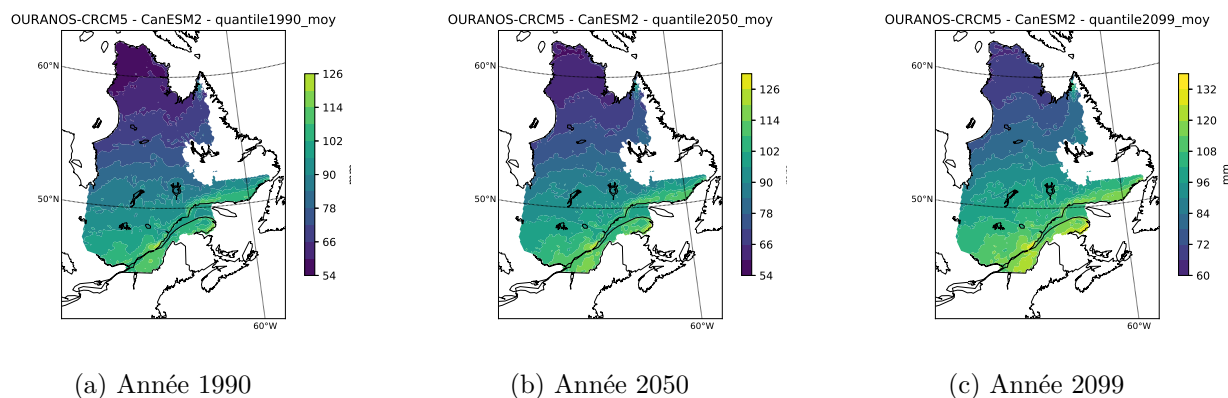


Figure 4.13 Moyenne des estimations des quantiles de la loi GEV pour une période de retour de 20 ans et une durée de 24h

L'estimation moyenne du niveau de retour est présentée à la figure 4.13, avec la même échelle de couleurs pour les trois cartes. Comme dans la figure précédente, une légère différence est observée entre les schémas spatiaux des années 1990 et 2050. Le changement est plus perceptible en comparant les années 2050 et 2099. L'estimation moyenne du quantile semble augmenter sur tout le territoire. De façon générale, cela signifie que le niveau de pluie intense à attendre en moyenne une fois au 20 ans sera plus élevé en 2099 qu'en 2050. La même tendance nord-sud que pour les paramètres μ et σ est observée pour l'estimation des quantiles. Les niveaux de retour estimés prennent de très grandes valeurs, comme par exemple pour l'année 2099, il y a quelques cellules dans le sud du Québec où il est attendu que le niveau de retour de 138 mm par jour soit dépassé en moyenne une fois aux 20 ans. Cela représente une très grande quantité de pluie à tomber en une seule journée.

Dans l'ensemble, la loi GEV a été ajustée à l'aide de la méthode du maximum de vraisemblance sur chacune des tuiles des 50 membres de ClimEx. L'ajustement a été fait pour la durée 24h. Quatre schémas par membre ont été créés afin de visualiser de façon spatiale les estimations. Les niveaux de retour ont été estimés pour les années 1990, 2050 et 2099. Puis, toutes ces estimations ont servies au calcul des estimations moyennes des paramètres de la loi GEV, du paramètre de position μ et du niveau de retour pour les trois années énon-

cées. Les grandes lignes tirées de l'analyse du membre KDA semblent être les mêmes que les conclusions déduites de l'analyse de la moyenne des 50 membres pour les paramètres μ_0 et σ . L'analyse des estimations moyennes a tout de même permis de déceler que les paramètres μ_1 et ξ ont une certaine structure spatiale, ce qui était invisible sur les cartes du membre KDA.

Pour résumer cette section, l'analyse exploratoire des données a permis d'orienter notre recherche. En effet, les données de précipitations fournies par ClimEx semblent, en majorité, convenir à un modèle non stationnaire ayant une variable explicative reliée à la concentration atmosphérique de GES pondérés selon leur PRP. L'ajout de cette variable explicative à notre modèle a permis d'estimer adéquatement les paramètres de position, d'échelle et de forme de la loi GEV et de les illustrer sur des schémas spatiaux à l'allure plutôt lisse. Certaines cellules peuvent néanmoins avoir une série de maxima s'apparentant davantage à un modèle stationnaire. Elles sont en minorité parmi les cellules contenues dans la grille. Également, l'analyse exploratoire a mis en lumière la tendance nord-sud des paramètres μ et σ ainsi que des niveaux de retour. Ce résultat est très intéressant et peut être utilisé dans le futur pour séparer le territoire en différentes zones. À ce stade, les estimations obtenues des paramètres de la loi GEV sont des estimations ponctuelles. Elles ne nous permettent pas de nous prononcer sur les distributions de μ_0 , μ_1 , σ et ξ . Même constat pour les niveaux de retour et le facteur de correction, qui est expliqué dans la section suivante. Pour ces motifs, il faut développer un modèle utilisant des méthodes MCMC afin d'avoir un échantillon de la distribution des paramètres, des quantiles et du facteur de correction.

4.3 Synthèse de la non-stationnarité

Cette dernière section contient la fin de la méthodologie ainsi que la présentation des résultats pour une cellule en particulier. Tout d'abord, le modèle développé dans le cadre du projet est présenté. Ensuite, les différentes alternatives de facteurs de correction sont énoncées puis les résultats pour la cellule associée à Montréal sont interprétés.

4.3.1 Modèle pour maxima non stationnaires

Le modèle développé afin d'estimer les précipitations extrêmes avec l'effet des changements climatiques est présenté. Afin d'extraire l'information et de stocker les données de la manière la plus efficace possible, notre choix s'est arrêté sur un modèle hiérarchique bayésien à trois niveaux. Un tel modèle nous permet d'obtenir des échantillons de la loi *a posteriori* pour chacun des paramètres et des quantités que nous désirons calculer. Cela permet, entre autres, de calculer un intervalle de crédibilité pour le facteur de correction.

Un modèle hiérarchique bayésien structure le problème, comme son nom l'indique, de façon hiérarchique, en liant ses différents niveaux par des relations conditionnelles. Les modèles hiérarchiques sont caractérisés à la fois par leurs hyperparamètres et par leurs paramètres [32]. Dans notre cas, les paramètres sont ceux de la loi GEV ajustée sur les données de ClimEx (soit μ_0 , μ_1 , σ et ξ) et une loi *a priori* leur est associée, dont les paramètres sont appelés hyperparamètres. Les méthodes MCMC seront utilisées pour obtenir un échantillon de la loi *a posteriori*, avec des taux d'acceptation souhaités entre 25 et 50%, comme il est mentionné au chapitre 7 du livre de Robert et Casella [33]. La structure de ce modèle est illustrée pour le paramètre μ_0 de la loi GEV à la figure 4.14, où $\mathbf{X}_k$ et μ_{0_k} représentent respectivement le vecteur des maxima annuels et l'ordonnée à l'origine du paramètre de position pour une cellule du k^e membre.

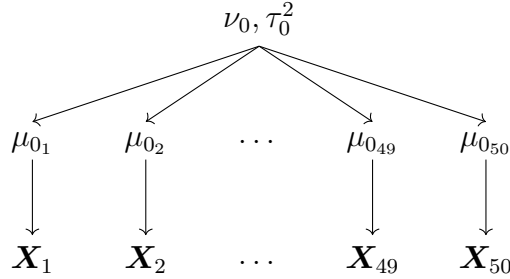


Figure 4.14 Structure du modèle hiérarchique bayésien pour le paramètre μ_0

Cette structure est maintenant expliquée plus en détails. Tout d'abord, il fallait choisir des lois *a priori* pour les paramètres de la loi GEV. Ce choix s'est arrêté sur une loi normale, comme présenté dans l'expression suivante, pour $\phi = \log(\sigma)$:

$$\begin{aligned} f_{\mu_0}(\mu_0) &\propto \mathcal{N}(\mu_0 \mid \nu_0, \tau_0^2) & f_{\phi}(\phi) &\propto \mathcal{N}(\phi \mid \nu_2, \tau_2^2) \\ f_{\mu_1}(\mu_1) &\propto \mathcal{N}(\mu_1 \mid \nu_1, \tau_1^2) & f_{\xi}(\xi) &\propto \mathcal{N}(\xi \mid \nu_3, \tau_3^2). \end{aligned} \quad (4.6)$$

Nous estimons $\phi = \log(\sigma)$ puisque son support est $\mathbb{R}$, tandis que le support de σ est $\mathbb{R}^+$. Les paramètres de la loi GEV suivent donc chacun une loi normale de moyenne ν_i et d'écart-type τ_i . C'est une hypothèse que nous posons, étant donné que la loi normale n'est pas un choix contraignant. Elle est souvent utilisée en pratique. Les lois *a priori* utilisées pour les hyperparamètres sont énoncées pour $i = 0, 1, 2, 3$:

$$f_{\nu_i}(\nu_i) \propto 1 \quad f_{\tau_i^2}(\tau_i^2) \propto \frac{1}{\tau_i^2}. \quad (4.7)$$

En théorie de la décision bayésienne, le fait que la loi *a priori* soit impropre n'est pas problématique si la loi *a posteriori* est propre [18].

Le modèle hiérarchique bayésien développé comporte trois niveaux et ils sont énoncés. Le premier niveau du modèle représente l'ajustement de façon marginale de la loi GEV sur toutes les cellules pour les données provenant de ClimEx, correspondant à la ligne du bas à la figure 4.14. Le second niveau fournit les distributions *a priori* pour les paramètres de la loi GEV. Elles sont présentées à l'équation (4.6). Puis, le troisième et dernier niveau contient les lois *a priori* pour les hyperparamètres, disponibles à l'équation (4.7). Les lois *a priori* qui sont utilisées représentent des choix arbitraires. Le modèle hiérarchique bayésien incorporant ces lois a cependant été testé avec des valeurs simulées et les estimations obtenues étaient très efficaces pour refléter la réalité.

Étant donné que chaque membre se distingue par de faibles perturbations dans ses conditions initiales, chacune des 50 simulations est décrite par l'ajustement d'une loi GEV sur la série de maxima annuels. Les paramètres de la loi GEV sont uniques à chaque membre. L'ajout du 3e niveau vient raffiner notre modèle en modélisant chaque échantillon de 50 observations, soit $(\mu_{01}, \dots, \mu_{050})$ à la figure 4.14, selon une loi normale.

Pour les méthodes MCMC, nous utilisons l'échantillonnage de Gibbs. Cet algorithme consiste à décomposer le problème de départ en sous-problèmes de lois conditionnelles, sans connaître la constante de normalisation. Plus particulièrement, il faut générer des données à partir de la densité $f_{(\Theta|Y=y)}(\theta)$ pour $\theta = (\mu_0, \mu_1, \sigma, \xi, \nu_0, \nu_1, \nu_2, \nu_3, \tau_0^2, \tau_1^2, \tau_2^2, \tau_3^2)$.

Il fallait tout d'abord déterminer ces lois conditionnelles pour les douze paramètres du modèle hiérarchique bayésien pour implémenter l'échantillonnage de Gibbs. En effet, les quatre paramètres de la loi GEV ont chacun deux hyperparamètres, ce qui donne un total de douze paramètres à modéliser à travers notre modèle. Pour commencer, les lois conditionnelles des quatre paramètres de la loi GEV sont énoncées pour une cellule en particulier. Elles sont présentées à l'équation (4.8), où $\mathbf{y} = (y_1, y_2, \dots, y_{144})$ représente la série de maxima de 1956 à 2099 et t les itérations.

$$\begin{aligned}
f_{(\mu_0|-)}(\mu_0) &\propto \prod_{i=1}^{144} \text{GEV} \left(y_i \mid \mu_i^{(1)}, \exp\{\phi^{t-1}\}, \xi^{t-1} \right) \times \mathcal{N} \left(\mu_0 \mid \nu_0^{t-1}, \tau_0^{2t-1} \right) \\
f_{(\mu_1|-)}(\mu_1) &\propto \prod_{i=1}^{144} \text{GEV} \left(y_i \mid \mu_i^{(2)}, \exp\{\phi^{t-1}\}, \xi^{t-1} \right) \times \mathcal{N} \left(\mu_1 \mid \nu_1^{t-1}, \tau_1^{2t-1} \right) \\
f_{(\phi|-)}(\phi) &\propto \prod_{i=1}^{144} \text{GEV} \left(y_i \mid \mu_i^t, \exp\{\phi\}, \xi^{t-1} \right) \times \mathcal{N} \left(\phi \mid \nu_2^{t-1}, \tau_2^{2t-1} \right) \\
f_{(\xi|-)}(\xi) &\propto \prod_{i=1}^{144} \text{GEV} \left(y_i \mid \mu_i^t, \exp\{\phi^t\}, \xi \right) \times \mathcal{N} \left(\xi \mid \nu_3^{t-1}, \tau_3^{2t-1} \right)
\end{aligned} \tag{4.8}$$

où $\mu_i^{(1)} = \mu_0 + \mu_1^{t-1} \times \text{GES}(i)$,
 $\mu_i^{(2)} = \mu_0^t + \mu_1 \times \text{GES}(i)$,
 et $\mu_i^t = \mu_0^t + \mu_1^t \times \text{GES}(i)$.

Nous poursuivons avec la présentation des lois conditionnelles des huit hyperparamètres. Pour les hyperparamètres de μ_0 , on tire des échantillons à partir de ces distributions conditionnelles :

$$\begin{aligned}
f_{(\nu_0|-)}(\nu_0) &\propto \mathcal{N} \left(\nu_0 \mid \bar{\mu}_0^t, \frac{\tau_0^{2t-1}}{50} \right) \\
f_{(\tau_0^2|-)}(\tau_0^2) &\propto \text{InvGamma} \left(\tau_0^2 \mid \frac{50}{2}, \sum_{m=1}^{50} \frac{(\mu_0^t[m] - \nu_0^t)^2}{2} \right),
\end{aligned} \tag{4.9}$$

où $\bar{\mu}_0^t = \frac{1}{50} \sum_{m=1}^{50} \mu_0^t[m]$ et τ_0^2 suit une loi Inverse Gamma. Le même raisonnement a été utilisé pour les autres hyperparamètres et les distributions sont présentées dans l'équation suivante :

$$\begin{aligned}
f_{(\nu_1|-)}(\nu_1) &\propto \mathcal{N} \left(\nu_1 \mid \bar{\mu}_1^t, \frac{\tau_1^{2t-1}}{50} \right) & f_{(\tau_1^2|-)}(\tau_1^2) &\propto \text{InvGamma} \left(\tau_1^2 \mid \frac{50}{2}, \sum_{m=1}^{50} \frac{(\mu_1^t[m] - \nu_1^t)^2}{2} \right) \\
f_{(\nu_2|-)}(\nu_2) &\propto \mathcal{N} \left(\nu_2 \mid \bar{\phi}^t, \frac{\tau_2^{2t-1}}{50} \right) & f_{(\tau_2^2|-)}(\tau_2^2) &\propto \text{InvGamma} \left(\tau_2^2 \mid \frac{50}{2}, \sum_{m=1}^{50} \frac{(\phi^t[m] - \nu_2^t)^2}{2} \right) \\
f_{(\nu_3|-)}(\nu_3) &\propto \mathcal{N} \left(\nu_3 \mid \bar{\xi}^t, \frac{\tau_3^{2t-1}}{50} \right) & f_{(\tau_3^2|-)}(\tau_3^2) &\propto \text{InvGamma} \left(\tau_3^2 \mid \frac{50}{2}, \sum_{m=1}^{50} \frac{(\xi^t[m] - \nu_3^t)^2}{2} \right)
\end{aligned} \tag{4.10}$$

où $\bar{\mu}_1^t = \frac{1}{50} \sum_{m=1}^{50} \mu_1^t[m]$, $\bar{\phi}^t = \frac{1}{50} \sum_{m=1}^{50} \phi^t[m]$ et $\bar{\xi}^t = \frac{1}{50} \sum_{m=1}^{50} \xi^t[m]$.

Dans l'ensemble, l'échantillonnage de Gibbs a donc été utilisé pour décomposer le problème. Les lois pour chaque paramètre sachant tous les autres paramètres sont présentées aux équations (4.8), (4.9) et (4.10). Il y a un total de douze paramètres à estimer qui composent le modèle hiérarchique bayésien, d'où les douze lois conditionnelles présentées. Pour chacun des paramètres, la prochaine étape était de générer un échantillon à partir de sa loi *a posteriori*. Cette étape était effectuée de deux façons différentes selon le groupe de paramètres à estimer.

Le premier groupe englobe les quatre paramètres de la loi GEV à estimer pour chacun des 50 membres. Étant donné que leurs lois conditionnelles présentées à l'équation (4.8) ne correspondent pas à une forme analytique, l'algorithme de Metropolis-Hastings a été utilisé pour générer une réalisation de la loi conditionnelle. Cet algorithme consiste à proposer des candidats qui seront retenus selon une certaine probabilité pour construire un échantillon. On obtient donc 50 échantillons pour chaque paramètre, correspondant aux 50 membres de ClimEx.

Pour les hyperparamètres, représentant le deuxième groupe de paramètres, les lois conditionnelles étaient exprimées sous des formes connues. Les échantillons de la loi *a posteriori* ont été tirés directement des lois présentées aux équations (4.9) et (4.10).

À chaque itération, les paramètres de la loi GEV ainsi que leurs hyperparamètres étaient estimés par les méthodes qui viennent d'être énoncées. Ces estimations étaient stockées afin d'obtenir ultimement un échantillon de la loi *a posteriori*.

Comme l'analyse porte sur une seule cellule de la grille, le nombre total d'itérations choisi est de 50 000 pour une période d'échauffement de 1000 itérations. Le modèle donnait de bons résultats à partir de 20 000 itérations et pourrait être utilisé ultérieurement avec ce nombre d'itérations pour augmenter la rapidité de calcul. Cela serait utile dans le cas où la méthodologie serait appliquée à l'ensemble des cellules de la grille, afin de couvrir le Québec en entier.

4.3.2 Possibilités de facteur de correction

Cette partie de la méthodologie consiste à proposer diverses méthodes à utiliser pour mettre à jour les courbes IDF. Il est pertinent de rappeler que le projet de maîtrise présenté se veut une généralisation de l'aspect non stationnaire de la recherche de Jalbert *et al.* [3]. Pour cette raison, la méthode proposée dans notre recherche pour développer le facteur de correction est tirée de leur article et elle est maintenant détaillée.

Le but du projet étant d'incorporer de la non-stationnarité, le paramètre de position de la loi GEV change selon l'année d'index t et cela implique que les niveaux de retour dépendent également du temps. On dénote $z_p(t)$ le niveau de retour pour la période de retour $T = 1/p$ de l'année indexée t , comme présenté à l'équation (4.5). Le facteur de correction pour le niveau de retour associé à la période de retour $T = 20$ entre les années indexées t_1 et t_2 est dénoté $\delta_p(t_1; t_2)$. Il est défini comme suit pour une seule cellule :

$$\delta_p(t_1; t_2) = \frac{\bar{z}_p(t_2)}{\bar{z}_p(t_1)}, \quad (4.11)$$

où $\bar{z}_p(t)$ correspond à la moyenne des quantiles pour les 50 membres. La notation (i, j) pour désigner une cellule de la grille est omise ici pour simplifier les formules.

Le facteur de correction proposé par Jalbert *et al* afin d'intégrer de la non-stationnarité dans les courbes IDF est donc un ratio des niveaux de retour des deux années concernées. Ce facteur de correction pourra être appliqué simplement en le multipliant aux niveaux de retour, qui eux sont calculés à partir des données observées. Les niveaux de retour mis à jour serviront par la suite au calcul des courbes IDF. Le facteur de correction sera estimé avec le modèle hiérarchique bayésien puisque ce dernier va être en mesure de fournir un intervalle de crédibilité à 95%. Les intervalles de crédibilité sont l'équivalent bayésien des intervalles de confiance en statistique classique. C'est pour cette raison qu'il n'a pas été calculé directement avec les estimations de l'analyse exploratoire.

Pour des raisons de temps, le facteur de correction ne sera pas estimé pour toutes les cellules de la grille dans ce mémoire. Il sera estimé pour la cellule correspondant à la ville de Montréal dans la prochaine sous-section avec $t_1 = 1990$. Sachant cela, il serait tout de même pertinent d'isoler la partie de $\delta_p(t_1; t_2)$ qui est influencée par les paramètres de la loi GEV, afin d'avoir une idée de la tendance spatiale du facteur de correction. Le calcul est présenté pour une seule cellule dans l'équation suivante, où $t_1 = 1990$:

$$\begin{aligned}
 \delta_p(1990; t_2) &= \frac{\bar{z}_p(t_2)}{\bar{z}_p(1990)} \\
 &= \frac{\mu(t_2) - \frac{\sigma}{\xi} [1 - \{-\log(1-p)\}^{-\xi}]}{\bar{z}_p(1990)} \\
 &= \frac{(\mu(1990) - \mu(1990)) + \mu(t_2) - \frac{\sigma}{\xi} [1 - \{-\log(1-p)\}^{-\xi}]}{\bar{z}_p(1990)} \\
 &= 1 + \frac{\mu(t_2) - \mu(1990)}{\bar{z}_p(1990)} \\
 &= 1 + \frac{[\mu_0 + \mu_1 \text{GES}(t_2)] - [\mu_0 + \mu_1 \text{GES}(1990)]}{\bar{z}_p(1990)} \\
 &= 1 + [\text{GES}(t_2) - \text{GES}(1990)] \times \underbrace{\frac{\mu_1}{\bar{z}_p(1990)}}_{\Delta\delta^{(1990)}}.
 \end{aligned} \tag{4.12}$$

La partie désignée par $\Delta\delta^{(1990)}$ correspond à la partie du facteur de correction qui varie selon les paramètres de la loi GEV. On peut se servir des estimations obtenues à l'étape de l'analyse exploratoire pour tenter de déceler une quelconque tendance spatiale sur le territoire. Les estimations obtenues par la méthode du maximum de vraisemblance ont donc été utilisées pour le calcul de $\Delta\delta^{(1990)}(k)$ pour le k^e membre. Ensuite, la moyenne sur les 50 membres a

été calculée pour obtenir $\Delta\delta^{(1990)}$. La carte représentant $\Delta\delta^{(1990)}$ est présentée à la figure 4.15 pour une période de retour de 20 ans. La tendance spatiale illustrée est très semblable à celle observée pour l'estimation moyenne du paramètre μ_1 à la figure 4.11. En effet, l'estimation prend de plus petites valeurs sur les côtes. Également, le facteur de correction sera plus grand dans le nord du Québec d'après le schéma spatial. Ces observations seront utilisées plus tard dans ce mémoire.

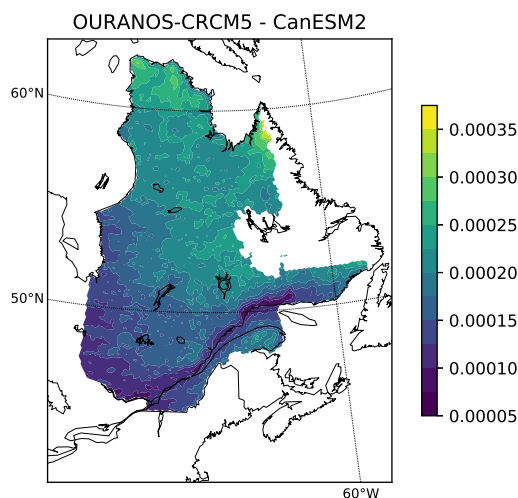


Figure 4.15 Estimation de l'influence du facteur de correction sur le territoire

Comme mentionné précédemment, une problématique du projet de recherche est d'être en mesure de comparer les données de précipitations provenant des stations météorologiques avec les données simulées provenant de ClimEx. Les données simulées sont à l'échelle d'une cellule du modèle, soit une superficie de 12km par 12km. Afin de réaliser cela, il faut déterminer un facteur de correction qui servira à réduire les quantités de précipitations tombées sur la superficie d'une cellule vers des quantités à l'échelle d'un point de grille. Ce facteur est défini comme un ratio qui varie dans le temps et l'espace et il pourra être déterminé à partir des données de ClimEx. Il sera utile afin d'enlever le biais introduit par le modèle climatique utilisé. Pour cela, des méthodes peuvent être proposées pour le développement d'un tel facteur de correction. En effet, la méthode utilisée dans l'article de Mailhot *et al.* [14] qui date de 2007 est toujours d'actualité. Le facteur qu'ils ont développé est un ARF faisant le ratio entre l'intensité de pluie à l'échelle d'une station et l'intensité à l'échelle d'une cellule. Il pourrait être utilisé avec les estimations de notre modèle hiérarchique bayésien. Le modèle développé dans le cadre du projet de maîtrise pourrait également être utilisé pour les ARFs développés dans l'article de Le *et al.* [21]. Il faudrait dans ce cas ajouter un aspect spatial à notre

modèle.

Par conséquent, il y a plusieurs méthodes déjà existantes qui pourraient être utilisées avec notre modèle hiérarchique bayésien pour développer un facteur de correction. Celle proposée par Jalbert *et al* est retenue pour ce mémoire et est appliquée sur une cellule dans la prochaine sous-section.

4.3.3 Résultats pour une cellule

Le but ultime de cette recherche est d’extraire le maximum d’information sur la non-stationnarité des précipitations extrêmes contenue dans ClimEx. Le modèle hiérarchique bayésien qui vient d’être présenté a été utilisé à cette fin pour la série de maxima d’une cellule en particulier. Il s’agit de la cellule contenant la ville de Montréal pour les 50 membres. Elle a été choisie puisque les événements de pluie intense ont habituellement plus de conséquences sur des milieux urbains, car les enjeux économiques sont grands. Tous les résultats énoncés à partir de ce point concernent la série de maxima pour la cellule de Montréal.

Le modèle hiérarchique bayésien a donc permis d’estimer les lois *a posteriori* des paramètres de la loi GEV ainsi que de leurs hyperparamètres. Les chaînes de Markov associées aux hyperparamètres de μ_0 sont illustrées à la figure A.1 en Annexe A. Les distributions marginales des paramètres ν_i sont illustrées à la figure 4.16 et elles ont toutes l’allure d’une cloche. Si on regarde la figure pour le paramètre ν_1 , on voit que sa distribution marginale recoupe la valeur de zéro. Il est pertinent de mentionner ici que, lors de l’analyse exploratoire des données, la cellule de Montréal avait un meilleur ajustement avec le modèle stationnaire pour 32 membres sur 50. Cela justifie donc le fait que le paramètre ν_1 , qui représente la moyenne de la distribution *a posteriori* du paramètre μ_1 , a une distribution marginale qui prend des valeurs autour de zéro. Il y a une cohérence entre les estimations obtenues par maximisation de la vraisemblance et celles obtenues avec le modèle hiérarchique bayésien.

Le modèle développé pour ce projet est donc très polyvalent et permet de prendre en compte les différents types de processus aléatoires possibles. D’un côté, il peut fournir des estimations pour une série de maxima s’apparentant davantage à un modèle stationnaire, en attribuant au paramètre μ_1 une valeur autour de zéro. D’un autre côté, notre modèle peut prendre en compte des série de maxima contenant de la non-stationnarité, grâce au paramètre μ_1 à estimer.

En ayant des échantillons des distributions *a posteriori* des paramètres de la loi GEV, nous pouvons calculer une multitude de quantités. Il est important de cibler les estimations qui seront pertinentes à l’analyse. Tout d’abord, il est judicieux d’estimer les niveaux de retour de

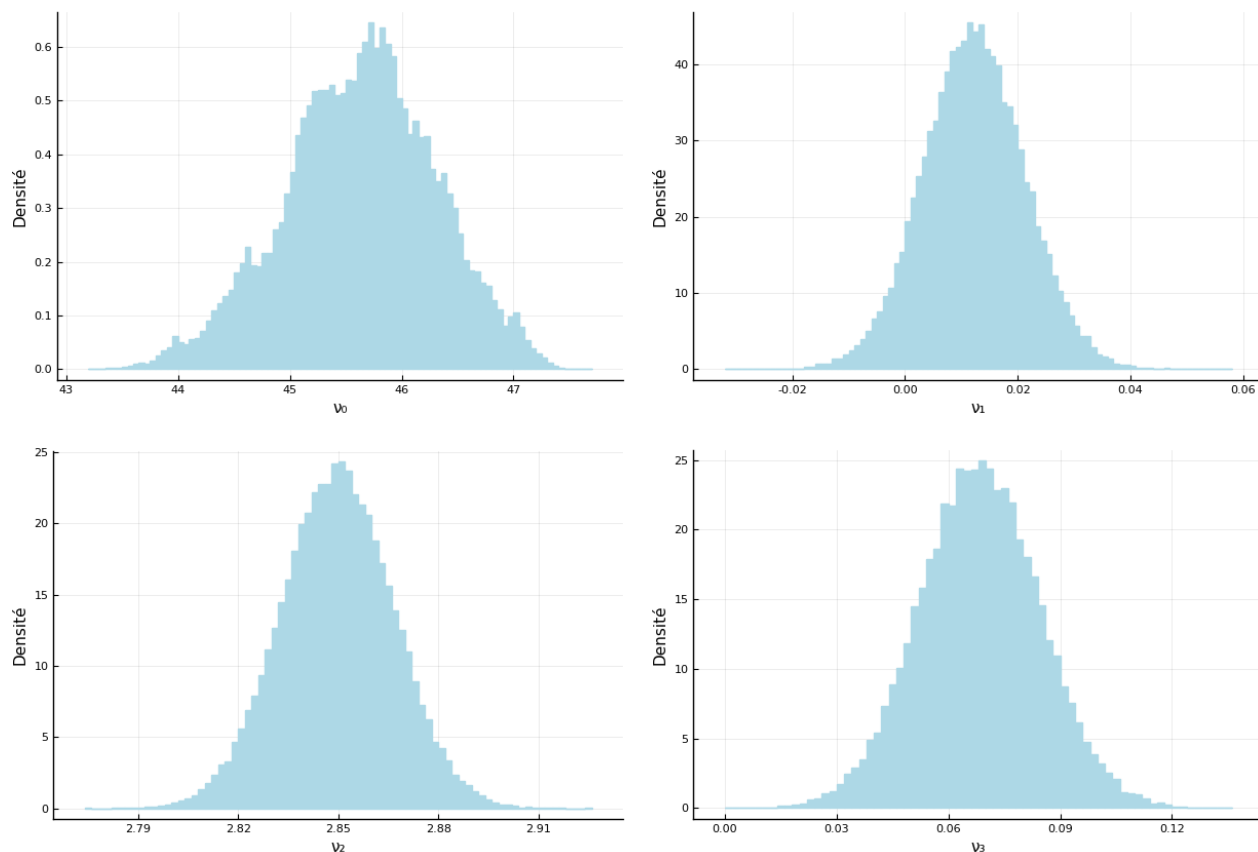


Figure 4.16 Distributions marginales des paramètres ν_0 , ν_1 , ν_2 et ν_3

la loi GEV, puisqu'ils seront nécessaires au calcul du facteur de correction. Ces niveaux avaient déjà été estimés dans l'analyse exploratoire des données, mais grâce au modèle hiérarchique bayésien développé, des intervalles de crédibilité à 95% ont également pu être calculés. Les résultats sont présentés dans le tableau 4.2, où les quantiles sont encore estimés avec une période de retour de 20 ans. Le tableau présente les résultats obtenus avec la méthode MCMC et ceux obtenus dans l'analyse exploratoire avec la méthode du maximum de vraisemblance, notée MV. Les niveaux de retour correspondent à la moyenne des quantiles sur les 50 membres pour la cellule de Montréal.

Tableau 4.2 Estimations des niveaux de retour de la loi GEV selon les deux méthodes

Année	MCMC		MV
	$\bar{z}_p(t)$	IC à 95%	$\bar{z}_p(t)$
1990	108.99	[107.20, 110.88]	107.74
2050	111.73	[110.02, 113.57]	110.62
2099	116.93	[115.04, 118.93]	116.05

Dans l'ensemble, les estimations des deux méthodes sont plutôt proches. On remarque que les estimations ponctuelles du niveau de retour sont plus petites dans le cas de la méthode du maximum de vraisemblance, et ce pour les trois années. Ces dernières estimations sont également incluses dans les intervalles de crédibilité fournis par la méthode MCMC. La méthode MCMC vient donc affiner notre estimation des niveaux de retour en ajoutant un intervalle de crédibilité. Autrement dit, pour une année t en particulier, cela signifie que dans 95% des cas, l'intervalle devrait contenir la vraie valeur du quantile $\bar{z}_p(t)$. Dans les deux méthodes, il est évident que l'estimation du niveau de retour augmente avec les années. Les intervalles de crédibilité obtenus à l'aide des méthodes MCMC permettent de voir que les intervalles des deux premières années se recoupent, tandis que celui pour l'année 2099 ne recoupe pas les intervalles des années 1990 et 2050. Cela signifie que l'intensité d'un événement de précipitations extrêmes attendue en moyenne une fois aux 20 ans augmente entre 1990 et 2099 ainsi qu'entre 2050 et 2099.

On se rappelle que la majorité des membres de la cellule de Montréal avaient un meilleur ajustement pour le modèle stationnaire. C'est donc intéressant de voir que les niveaux de retour attendus augmentent avec les années lorsque l'on considère l'ensemble des simulations avec un modèle non stationnaire, ce qui signifie qu'il y a une possible présence de non-stationnarité. Cela montre donc que si les membres sont analysés de façon catégorique, le modèle stationnaire l'emporte. Par contre, si les simulations sont prises en compte par un modèle non stationnaire, les niveaux de retour estimés augmentent avec les années et cela traduit une présence de non-stationnarité.

Par la suite, on peut calculer le facteur de correction pour la cellule de Montréal avec la méthode de Jalbert *et al.* [3] à partir de l'année 1990. Cette année désigne bien la période avant l'amplification des changements climatiques. Il a été montré dans l'analyse exploratoire des données qu'il y avait peu de différence entre les niveaux de retour des années 1990 et 2050. À titre d'exemple, pour une période de retour de 20 ans, le facteur de correction entre les années 1990 et 2050 est le suivant, pour $p = 1/20$:

$$\delta_{0.05}(1990; 2050) = \frac{\bar{z}_{0.05}(2050)}{\bar{z}_{0.05}(1990)}. \quad (4.13)$$

Le facteur de correction a ainsi été calculé avec l'équation (4.11) pour $t_1 = 1990$ et t_2 allant de 1990 à 2099. Les valeurs obtenues sont tracées à la figure 4.17 avec un intervalle de crédibilité à 95 %, où l'on peut voir que le facteur de correction prend de la valeur à chaque année. En effet, le facteur de correction et son intervalle de crédibilité augmentent de façon continue sur toute la période d'étude. Cet intervalle fût calculé grâce à l'échantillon de la loi *a posteriori*

du facteur de correction fourni par les méthodes MCMC et semble de grandeur acceptable. L'incertitude augmente avec l'amplification des changements climatiques. Par exemple, pour l'année 2000, l'intervalle de crédibilité est de $[1.0018, 1.0025]$, tandis qu'en 2099, l'intervalle de crédibilité est estimé à $[1.0615, 1.0853]$. Cela signifie qu'on estime en 2099 un accroissement des précipitations extrêmes de 6 à 8 % par rapport à 1990.

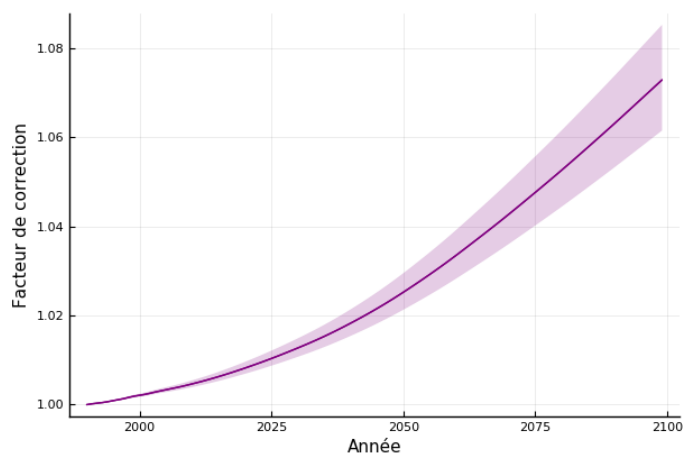


Figure 4.17 Estimation du facteur de correction à partir de 1990 pour une période de retour de 20 ans et une durée de 24h avec un intervalle de crédibilité à 95%

Sur la carte présentée à la figure 4.15, Montréal semble faire partie des cellules dont le facteur de correction a le moins d'effet, tandis que son effet semble être plus grand dans le nord de la province. Il serait intéressant de regarder une cellule dans cette région pour comparer les valeurs du facteur de correction à celles de Montréal. Parmi les cinq cellules arbitraires analysées au début de la section, Salluit était la ville la plus au nord d'après la figure 4.2. Pour cette raison, le facteur de correction pour la ville de Salluit est présenté à la figure B.1 en Annexe. Le facteur de correction augmente sur la période d'étude et prend des valeurs plus élevées si on le compare au facteur de correction de Montréal. En effet, le facteur de correction de Salluit pour l'année 2099 a un intervalle de crédibilité à 95% de $[1.1423, 1.1654]$, traduisant une augmentation estimée entre 14 et 16 %. Il s'agit d'une augmentation représentant le double du maximum du facteur de correction de Montréal. Il sera ainsi primordial de calculer le facteur de correction pour toutes les cellules de la province dans le futur pour mettre à jour les courbes IDF, étant donné que sa valeur varie beaucoup d'une cellule à une autre.

Ce facteur de correction permet donc de quantifier les effets des changements climatiques sur les précipitations extrêmes. Ses valeurs pourront être multipliées directement sur les niveaux de retour estimés à partir de données observées afin de produire des courbes IDF incorporant les effets des changements climatiques. Des courbes IDF actualisées vont permettre d'aug-

menter la résilience face aux changements climatiques des infrastructures dimensionnées selon ces courbes. Ainsi donc, l'adaptation aux changements climatiques doit se faire en utilisant des outils fiables, tels que les courbes IDF qui pourront être produites à l'aide du facteur de correction.

En terminant, les résultats pour la cellule de Montréal ont été présentés. Ils font une synthèse de la non-stationnarité présente dans les maxima annuels de précipitations extrêmes de cette cellule. Le modèle hiérarchique bayésien développé pourra également servir au calcul du facteur de correction pour toutes les cellules de la province.

CHAPITRE 5 CONCLUSION ET RECOMMANDATIONS

Dans ce mémoire, un modèle hiérarchique bayésien a été développé pour les maxima annuels de précipitations d'un ensemble multimembre de simulations climatiques. Les ensembles multimembres, comme ClimEx, sont nécessaires pour étudier la sensibilité aux conditions initiales des modèles et ainsi donner des estimations robustes des précipitations extrêmes.

Pour cela, la théorie des valeurs extrêmes ainsi que la théorie de la décision bayésienne ont d'abord été introduites. Également, un état de l'art concernant les recherches en précipitations extrêmes a été présenté. Les données pertinentes au projet ont été extraites, ce qui a permis de débiter à synthétiser l'information. L'analyse exploratoire a permis d'orienter notre recherche, puisque cette étape a confirmé la présence de non-stationnarité dans les données de ClimEx. Également, cette étape a mis en évidence la distribution spatiale des paramètres de la loi GEV, comme illustré à la figure 4.11, ainsi que de ses niveaux de retour. Puis, le modèle hiérarchique bayésien a été développé dans le but d'extraire le maximum d'information concernant la non-stationnarité dans les séries de précipitations. L'interprétation des résultats a été faite pour la cellule associée à Montréal et a démontré que les niveaux de retour augmentaient sur la période d'étude. Parmi ces résultats, le facteur de correction avec son intervalle de crédibilité à 95% est probablement le plus important. Ce dernier permet de quantifier l'effet des changements climatiques sur l'estimation des précipitations extrêmes en mettant à jour les courbes IDF. Cela va permettre à plusieurs acteurs de la société de travailler avec des outils d'estimation fiables, dont les ingénieurs qui dimensionnent les infrastructures exposées aux aléas du climat. L'outil développé pourra fournir un facteur de correction pour n'importe quel emplacement au Québec, afin de produire des courbes IDF prenant en compte les effets des changements climatiques.

Il est certain que la méthodologie proposée dans ce mémoire comporte certaines limites. En effet, l'incertitude associée au choix de ClimEx est non négligeable. Par exemple, il est dit dans l'article de Mailhot *et al.* [14] que la discrétisation temporelle au pas de temps 15 minutes a pour effet de sous-estimer les vraies valeurs. C'est expliqué par le fait que les maxima extraits à partir des simulations numériques ne concordent pas exactement avec ceux calculés sur des simulations numériques à plus petit pas. Dans notre cas, ClimEx nous a fourni les précipitations pour chaque heure, ce qui signifie que le pas de temps est ici d'une heure pour le MRCC5. L'effet potentiel de sous-estimation mentionné par Mailhot serait donc pire dans notre cas. Cela fait partie des nombreuses restrictions liées à l'utilisation d'un seul modèle climatique et nous devons composer avec.

Pour terminer, quelques améliorations futures sont proposées. Tout d’abord, la non-stationnarité pourrait être incorporée dans le paramètre d’échelle, afin d’étudier l’évolution de la dispersion des précipitations extrêmes avec l’intensification des changements climatiques. Également, d’autres relations pourraient être testées avec la variable explicative, spécialement une relation quadratique.

L’information synthétisée dans ce projet pourra aussi être utilisée directement pour développer un nouveau facteur de correction afin de comparer les données observées avec les données simulées. Les différentes options ont été présentées dans le Chapitre 4.

Concernant le modèle développé, il pourrait être pertinent dans le futur d’introduire des copules pour être en mesure de calculer la probabilité que plusieurs événements aient lieu simultanément. Cela serait utile dans le cas où, par exemple, une tempête a lieu à deux stations météorologiques, ce qui a pour effet que ces deux stations ont leur maxima annuel arrivant la même date.

Une autre potentielle amélioration serait d’introduire un aspect spatial dans les estimations des précipitations extrêmes prenant en compte l’effet des changements climatiques. Une option intéressante serait de développer un modèle spatio-temporel à l’aide de champs de Markov gaussiens intrinsèques. En effet, lors de l’analyse exploratoire des données, il était clair que des tendances spatiales ressortaient pour l’estimation de certains paramètres et des quantiles. Cela reviendrait à combiner le modèle développé par Jalbert *et al* [3] avec celui développé pour ce mémoire. Également, l’ajout d’un aspect spatial permettrait de développer un facteur de correction variant dans l’espace. L’utilisation des ARFs définis dans l’article de Le *et al.* [21] serait possible et bénéfique.

RÉFÉRENCES

- [1] Institut national de santé publique du Québec. Précipitations non hivernales. [En ligne]. Disponible : <http://www.monclimatmasante.qc.ca/public/pr%C3%A9cipitations-non-hivernales.aspx>
- [2] L. Perreault et J. Jalbert, “Estimation des courbes IDF aux postes de transformation d’Hydro-Québec,” Hydro-Québec, Montréal, QC, Rapport synthèse IREQ-2016-0103, 2016.
- [3] J. Jalbert, A.-C. Favre, C. Bélisle et J.-F. Angers, “A spatiotemporal model for extreme precipitation simulated by a climate model, with an application to assessing changes in return levels over North America,” *Journal of the Royal Statistical Society : Series C (Applied Statistics)*, vol. 66, n^o. 5, p. 941–962, nov. 2017. [En ligne]. Disponible : <https://doi.org/10.1111/rssc.12212>
- [4] S. P. Simonovic, A. Schardong et D. Sandink, “Mapping extreme rainfall statistics for canada under climate change using updated intensity-duration-frequency curves,” *Journal of Water Resources Planning and Management*, vol. 143, n^o. 3, 2017. [En ligne]. Disponible : [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000725](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000725)
- [5] IPCC, “Global warming of 1.5 °C. an IPCC special report on the impacts of global warming of 1.5 °C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty [V. Masson-Delmotte, P. Zhai, H. O. Portner, D. Roberts, J. Skea, P.R. Shukla, A. Pirani, Y. Chen, S. Connors, M. Gomis, E. Lonnoy, J. B. R. Matthews, W. Moufouma-Okia, C. Péan, R. Pidcock, N. Reay, M. Tignor, T. Waterfield, X. Zhou (eds.)],” IPCC, Rapport technique SR 1.5, 2018. [En ligne]. Disponible : <https://www.ipcc.ch/sr15/about/content-map/>
- [6] A. Sarhadi et E. D. Soulis, “Time-varying extreme rainfall intensity-duration-frequency curves in a changing climate,” *Geophysical Research Letters*, vol. 44, n^o. 5, p. 2454–2463, 2017. [En ligne]. Disponible : <https://doi.org/10.1002/2016GL072201>
- [7] Institut national de santé publique du Québec. Changements climatiques. [En ligne]. Disponible : <http://www.monclimatmasante.qc.ca/public/changements-climatiques.aspx>
- [8] L’Atlas climatique du Canada. Changement climatique : les principes de base. [En ligne]. Disponible : <https://atlasclimatique.ca/changement-climatique-les-principes-de-base>
- [9] C. H. Lima, H.-H. Kwon et Y.-T. Kim, “A local-regional scaling-invariant bayesian GEV model for estimating rainfall IDF curves in a future climate,”

- Journal of Hydrology*, vol. 566, p. 73–88, nov. 2018. [En ligne]. Disponible : <https://doi.org/10.1016/j.jhydrol.2018.08.075>
- [10] B. J. Reich et B. A. Shaby, “A spatial markov model for climate extremes,” *Journal of Computational and Graphical Statistics*, 2018. [En ligne]. Disponible : <https://doi.org/10.1080/10618600.2018.1482764>
- [11] V. Agilan et N. V. Umamahesh, “Covariate and parameter uncertainty in non-stationary rainfall IDF curve,” *International Journal of Climatology*, vol. 38, n°. 1, p. 365–383, janv. 2018. [En ligne]. Disponible : <https://doi.org/10.1002/joc.5181>
- [12] UPMC- Sorbonne Universités. (2013) Simulations climatiques : entre modèles et modélisations. [En ligne]. Disponible : http://www.upmc.fr/fr/recherche/actualites_de_la_recherche/dossiers_thematiques/les_sciences_du_climat_a_l_upmc/modeler_le_climat/simulations_climatiques_entre_modeles_et_modelisations.html
- [13] L’Atlas climatique du Canada. Glossaire. [En ligne]. Disponible : <https://atlasclimatique.ca/glossaire>
- [14] 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°. 1-2, p. 197–210, déc. 2007. [En ligne]. Disponible : <https://doi.org/10.1016/j.jhydrol.2007.09.019>
- [15] 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, déc. 2013. [En ligne]. Disponible : <https://doi.org/10.1007/s00382-013-1778-9>
- [16] 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, déc. 2013. [En ligne]. Disponible : <https://doi.org/10.1007/s00382-013-1737-5>
- [17] S. Coles, *An Introduction to Statistical Modeling of Extreme Values*. London : Springer, 2001.
- [18] C. P. Robert, *Le choix bayésien*. Paris, France : Springer, 2006. [En ligne]. Disponible : <https://link.springer.com/content/pdf/10.1007%2F2-287-31120-3.pdf>
- [19] J. R. M. Hosking et J. R. Wallis, *Regional Frequency Analysis : An Approach Based on L-Moments*. Cambridge, GB : Cambridge University Press, 1997.

- [20] P. Sharkey et H. C. Winter, “A bayesian spatial hierarchical model for extreme precipitation in great britain,” *Environmetrics*, oct. 2017. [En ligne]. Disponible : <https://arxiv.org/abs/1710.02091>
- [21] P. D. Le, A. C. Davison, S. Engelke, M. Leonard et S. Westra, “Dependence properties of spatial rainfall extremes and areal reduction factors,” *Journal of Hydrology*, vol. 565, p. 711–719, oct. 2018. [En ligne]. Disponible : <https://doi.org/10.1016/j.jhydrol.2018.08.061>
- [22] P. Ganguli et P. Coulibaly, “Assessment of future changes in intensity-duration-frequency curves for Southern Ontario using North American (NA)-CORDEX models with non-stationary methods,” 2019, soumis pour publication. [En ligne]. Disponible : <https://arxiv.org/abs/1706.00122>
- [23] 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, 2019. [En ligne]. Disponible : <https://doi.org/10.1175/JAMC-D-18-0021.1>
- [24] M. Leduc, M. Braun, R. Ludwig, D. Kranzlmüller, F. Brissette, R. Turcotte et H. Komischke, “Projet ClimEx : un grand ensemble de simulations pour explorer la variabilité naturelle et les extrêmes climatiques,” Ouranos, Rapport technique. [En ligne]. Disponible : https://www.ouranos.ca/publication-scientifique/sc_Leduc.pdf
- [25] Fnac. (2019) Bien choisir ordinateur portable. [En ligne]. Disponible : <https://www.fnac.com/Guides/Micro/bien-choisir-ordinateur-portable/choisir-ordinateur-portable.aspx>
- [26] W. Zhang, G. Villarini, E. Scoccimarro et G. A. Vecchi, “Stronger influences of increased CO_2 on subdaily precipitation extremes than at the daily scale,” *Geophysical Research Letters*, vol. 44, n°. 14, p. 7464–7471, 2017. [En ligne]. Disponible : <https://doi.org/10.1002/2017GL074024>
- [27] Gouvernement du Canada. (2019) Potentiels de réchauffement planétaire. [En ligne]. Disponible : <https://www.canada.ca/fr/environnement-changement-climatique/services/changements-climatiques/emissions-gaz-effet-serre/orientation-quantification/potentiels-rechauffement-planetaire.html>
- [28] Environnement et Changement climatique Canada. (2013) Substances appauvrissant la couche d’ozone. [En ligne]. Disponible : <https://www.ec.gc.ca/ozone/default.asp?lang=Fr&n=D57A00&wbdisable=true&fbclid=IwAR27WS36XkdYij6RWfDu3c4kBTFS04dOh0V8S4ukJ0-WREa3cyjPG4ukzL0#ref2>

- [29] G. E. Schwarz, “Estimating the dimension of a model,” *Annals of Statistics*, vol. 6, n°. 2, p. 461–464, 1978. [En ligne]. Disponible : <https://projecteuclid.org/euclid.aos/1176344136>
- [30] Wit, E., Heuvel, E.v.d. and Romeijn, J.-W, “‘All models are wrong...’ : an introduction to model uncertainty,” *Statistica Neerlandica*, vol. 66, n°. 3, p. 217–236, 2012. [En ligne]. Disponible : <https://doi.org/10.1111/j.1467-9574.2012.00530.x>
- [31] Kass, R. E. and Raftery, A. E., “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>
- [32] A. Gelman, J. B. Carlin, H. S. Stern et D. B. Rubin, *Bayesian Data Analysis*, Deuxième éd. Floride, États-Unis : Chapman & Hall/CRC, 2004.
- [33] Robert, C. P. and Casella, G., *Monte Carlo Statistical Methods*, Deuxième éd. New York, États-Unis : Springer, 2004.
- [34] M. Lefebvre, *Processus stochastiques appliqués*, Deuxième éd. Montréal, Québec : Presses internationales Polytechnique, 2014.
- [35] A. C. Davison, S. A. Padoan et M. Ribatet, “Statistical modeling of spatial extremes,” *Statistical Science*, vol. 27, n°. 2, p. 161–186, 2012. [En ligne]. Disponible : <https://projecteuclid.org/euclid.ss/1340110864>

ANNEXE A CHAÎNES DE MARKOV

La figure A.1 illustre les résultats énoncés dans le mémoire des chaînes de Markov obtenues par les procédures MCMC.

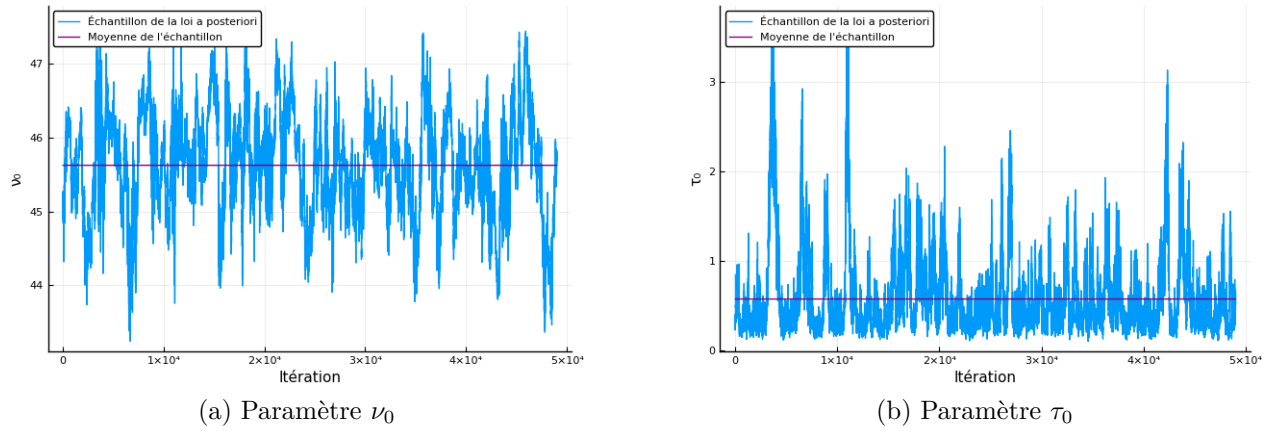


Figure A.1 Estimation de la moyenne et de l'écart-type pour le paramètre μ_0 pour une durée de 24h

ANNEXE B FACTEUR DE CORRECTION

La figure B.1 présente les valeurs du facteur de correction pour la ville de Salluit.

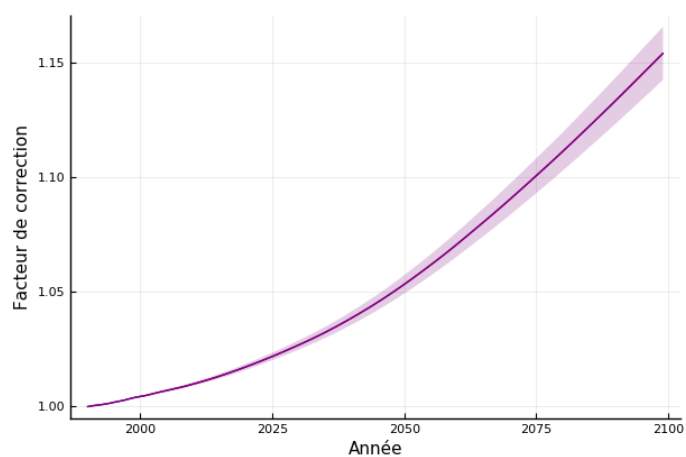


Figure B.1 Estimation du facteur de correction à partir de 1990 pour une période de retour de 20 ans et une durée de 24h avec un intervalle de crédibilité à 95% pour la ville de Salluit