

**Titre:** Conception d'un algorithme variationnel pour l'ajustement d'un modèle hiérarchique bayésien de valeurs extrêmes pour les précipitations  
Title:

**Auteur:** Ferdinand Willemin  
Author:

**Date:** 2025

**Type:** Mémoire ou thèse / Dissertation or Thesis

**Référence:** Willemin, F. (2025). Conception d'un algorithme variationnel pour l'ajustement d'un modèle hiérarchique bayésien de valeurs extrêmes pour les précipitations [Mémoire de maîtrise, Polytechnique Montréal]. PolyPublie.  
Citation: <https://publications.polymtl.ca/67717/>

## Document en libre accès dans PolyPublie

Open Access document in PolyPublie

**URL de PolyPublie:** <https://publications.polymtl.ca/67717/>  
PolyPublie URL:

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

**Programme:** Maîtrise recherche en mathématiques appliquées  
Program:

**POLYTECHNIQUE MONTRÉAL**

affiliée à l'Université de Montréal

**Conception d'un algorithme variationnel pour l'ajustement d'un modèle  
hiérarchique bayésien de valeurs extrêmes pour les précipitations**

**FERDINAND WILLEMIN**

Département de mathématiques et de génie industriel

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

Juin 2025

**POLYTECHNIQUE MONTRÉAL**

affiliée à l'Université de Montréal

Ce mémoire intitulé :

**Conception d'un algorithme variationnel pour l'ajustement d'un modèle hiérarchique bayésien de valeurs extrêmes pour les précipitations**

présenté par **Ferdinand WILLEMIN**

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

**Antoine SAUCIER**, président

**Jonathan JALBERT**, membre et directeur de recherche

**Antoine LEGRAIN**, membre

## DÉDICACE

*À celle dont je tiens tout,  
qui insuffle l'humilité, le respect et la beauté,  
mamie Paulette.*

## REMERCIEMENTS

Je souhaite exprimer ma profonde gratitude à Jonathan, un professeur passionnant et un superviseur bienveillant, qui a su construire un cadre de travail à la fois rigoureux et amusant tout au long du projet.

Un immense merci à Nicolas pour son soutien généreux, ses brillantes idées et ses précieux conseils.

Un grand merci à mes amis et collègues de laboratoire Delhio, Charles, Anne, Gabrielle, Justine et Sami pour l'entraide, l'inspiration et les bons moments passés ensemble.

Un merci tout particulier à mes chers parents, mon frère et ma sœur qui m'ont soutenu inconditionnellement durant ces deux années passées à Montréal.

Enfin, merci à ma coloc Flore pour m'avoir permis d'aller au bout des choses dans la joie et la bonne humeur !

## RÉSUMÉ

L’élévation de température au Québec menace d’impacter la fréquence et l’intensité des précipitations extrêmes, augmentant les risques de catastrophes climatiques telles que les inondations, érosions de sols et débordements d’égouts. Afin d’orienter la stratégie d’adaptation des organisations face à ces nouvelles dynamiques de précipitations, il est primordial de les quantifier. Les modèles statistiques, permettant d’inférer leur distribution à partir de données observées et de simulations climatiques, sont les outils les plus adaptés pour caractériser ces phénomènes aléatoires.

En particulier, un modèle hiérarchique bayésien conçu spécifiquement pour les précipitations extrêmes capture fidèlement la structure de leur comportement. L’ajustement de ce modèle aux données extrêmes consiste à trouver la distribution conditionnelle de ses paramètres sachant les données observées, appelée distribution *a posteriori*. Celle-ci n’étant pas connu analytiquement, elle est estimée *via* un algorithme de type Monte Carlo par Chaîne de Markov (MCMC). Bien que satisfaisant, les algorithmes MCMC peuvent être coûteux en temps et en mémoire et nécessitent des réglages fastidieux pour fonctionner.

Les méthodes variationnelles constituent une alternative prometteuse aux MCMC. L’objectif est de trouver, au sein d’une famille de lois paramétriques simples, une approximation de la loi *a posteriori* en minimisant un coût, nommé divergence de Kullback-Leibler. La famille de lois reposant sur l’hypothèse de champ moyen est usuellement utilisée. Sous cette hypothèse, on peut accéder à l’expression de la loi marginale *a posteriori* optimale d’un paramètre donné vis-à-vis du problème d’optimisation (loi  $q^*$ ), tous les autres paramètres étant fixés. De ce résultat découle l’algorithme Coordinate Ascent for Variational Inference (CAVI) dans lequel les lois  $q^*$  sont mises à jour à tour de rôle. Or, celui-ci n’est pas applicable tel quel au modèle de précipitations extrêmes car les lois  $q^*$  qu’il fait apparaître ne peuvent pas être simplement normalisées.

Dans ce mémoire, l’approximation dite quadratique est utilisée pour approcher les lois  $q^*$  indéterminées de la procédure CAVI par des lois normales. Cette approximation est très précise pour estimer les lois marginales des paramètres de localisation et d’échelle de la loi des valeurs extrêmes généralisée. La combinaison de ces procédés résulte en un algorithme variationnel rapide, donnant accès à des lois marginales *a posteriori* explicites et dont la convergence est facilement vérifiable.

## ABSTRACT

Rising temperatures in Quebec threaten to impact the frequency and intensity of extreme precipitation events, increasing the risk of climate-related disasters such as flooding, soil erosion and sewer overflows. To guide organizations' adaptation strategies in the face of these new precipitation dynamics, it is essential to quantify them. Statistical models, which can infer their distribution from observed data and climate simulations, are the most suitable tools for characterizing these random phenomena.

In particular, a Bayesian hierarchical model designed specifically for extreme precipitation events faithfully captures the structure of their behavior. Fitting this model to extreme data involves finding the conditional distribution of its parameters given the observed data, known as the posterior distribution. As this is not known analytically, it is estimated using a Markov Chain Monte Carlo (MCMC) algorithm. Although satisfactory, MCMC algorithms can be costly in terms of time and memory, and require tedious tuning to operate.

Variational methods are a promising alternative to MCMC. The aim is to find, within a family of simple parametric laws, an approximation to the posterior law by minimizing a cost, called the Kullback-Leibler divergence. The family of laws based on the mean-field hypothesis is usually used. Under this assumption, we can access the expression of the optimal posterior marginal distribution of a given parameter with respect to the optimization problem ( $q^*$  distribution), with all other parameters fixed. From this result derives the Coordinate Ascent for Variational Inference (CAVI) algorithm, in which the  $q^*$  laws are updated in turn. However, this algorithm cannot be applied as it stands to the extreme precipitation model, as the  $q^*$  laws it generates cannot simply be normalized.

In this thesis, the so-called quadratic approximation is used to approximate the indeterminate  $q^*$  laws of the CAVI procedure by normal laws. This approximation is highly accurate for estimating the marginal laws of the location and scale parameters of the generalized extreme value distribution. The combination of these procedures results in a fast variational algorithm, giving access to explicit posterior marginal laws whose convergence is easily verifiable.

## TABLE DES MATIÈRES

DÉDICACE . . . . .	iii
REMERCIEMENTS . . . . .	iv
RÉSUMÉ . . . . .	v
ABSTRACT . . . . .	vi
TABLE DES MATIÈRES . . . . .	vii
LISTE DES TABLEAUX . . . . .	x
LISTE DES FIGURES . . . . .	xi
LISTE DES ANNEXES . . . . .	xiii
LISTE DES SIGLES ET ABRÉVIATIONS . . . . .	xiv
CHAPITRE 1 INTRODUCTION . . . . .	1
1.1 Analyse des valeurs extrêmes . . . . .	1
1.2 Disponibilité des données . . . . .	2
1.3 Problématique . . . . .	2
1.4 Organisation du mémoire . . . . .	3
CHAPITRE 2 REVUE DE LITTÉRATURE . . . . .	4
2.1 Les approches bayésiennes variationnelles . . . . .	4
2.2 L'espérance-propagation . . . . .	5
2.3 L'approximation de Laplace intégrée et imbriquée . . . . .	6
CHAPITRE 3 CADRE THÉORIQUE . . . . .	8
3.1 Inférence bayésienne . . . . .	8
3.2 Les algorithmes Monte-Carlo par Chaîne de Markov . . . . .	9
3.2.1 Metropolis . . . . .	10
3.2.2 Metropolis-Hastings . . . . .	10
3.2.3 Algorithme de Langevin ajusté par Metropolis . . . . .	10
3.2.4 Monte-Carlo Hamiltonien . . . . .	11

3.2.5	Gibbs . . . . .	12
3.2.6	Metropolis-Hastings dans Gibbs . . . . .	12
3.2.7	Réglage des hyperparamètres . . . . .	12
3.3	L'algorithme CAVI . . . . .	14
3.3.1	L'approximation champ moyen . . . . .	14
3.3.2	Densités marginales optimales . . . . .	15
3.3.3	Principe de l'algorithme . . . . .	15
3.4	L'approximation quadratique . . . . .	16
3.5	Théorie des valeurs extrêmes . . . . .	17
3.5.1	Loi des valeurs extrêmes généralisée . . . . .	17
3.5.2	Méthode des maxima par blocs . . . . .	18
3.6	Champs markoviens gaussiens intrinsèques de premier ordre sur une grille réguli��re . . . . .	19
<b>CHAPITRE 4 DONN��ES ET MODLISATION . . . . .</b>		<b>22</b>
4.1	R��analyse Canadienne de Surface . . . . .	22
4.2	Mod��le bay��sien spatial de valeurs extr��mes pour les pr��cipitations sur grille . . . . .	22
<b>CHAPITRE 5 APPROXIMATIONS QUADRATIQUES EN SRIE POUR L'INFRENCE VARIATIONNELLE . . . . .</b>		<b>28</b>
5.1	La famille de lois d'approximation . . . . .	28
5.2	Loi $q_{\{(\mu_i, \phi_i)\} \lambda_i}^*(\mu_i, \phi_i)$ . . . . .	29
5.3	Loi $q_{(\xi \lambda_\xi)}^*(\xi)$ . . . . .	31
5.4	Lois $q_{(\kappa_u \lambda_u)}^*(\kappa_u)$ et $q_{(\kappa_v \lambda_v)}^*(\kappa_v)$ . . . . .	33
5.5	Impl��mentation de l'algorithme . . . . .	33
<b>CHAPITRE 6 R��ULTATS SUR DONN��ES DE SIMULATION . . . . .</b>		<b>35</b>
6.1	Les donn��es . . . . .	35
6.2	Estimations SQAVI . . . . .	35
6.3	Estimations MCMC . . . . .	36
6.4	Comparaison SQAVI et MCMC . . . . .	36
<b>CHAPITRE 7 R��ULTATS SUR DONN��ES DE R��ANALYSE . . . . .</b>		<b>40</b>
7.1	Estimations SQAVI . . . . .	40
7.2	Estimations MCMC . . . . .	41
7.3	Comparaison SQAVI et MCMC . . . . .	41
<b>CHAPITRE 8 DISCUSSION . . . . .</b>		<b>46</b>

CHAPITRE 9 CONCLUSION ET RECOMMANDATIONS . . . . .	48
RÉFÉRENCES . . . . .	50

## LISTE DES TABLEAUX

Tableau 3.1	Taux d'acceptation asymptotiques optimaux des algorithmes MCMC.	13
Tableau 6.1	Valeurs des paramètres du modèle de précipitations extrêmes avec lesquelles les données de simulation ont été générées. . . . .	35
Tableau 6.2	Évaluation et comparaison des estimations SQAVI et MCMC. . . . .	39
Tableau 7.1	Évaluation et comparaison des estimations SQAVI et MCMC. . . . .	44
Tableau A.1	Configuration des paramètres ayant généré les données artificielles de l'analyse. . . . .	54
Tableau A.2	Valeurs moyennes sur 10 exécutions de la corrélation entre $\mu_1$ , $\phi_1$ et $\xi$ pour une grille de taille $30 \times 30$ . . . . .	55
Tableau D.1	Comparaison des algorithmes MCMC. Le meilleur algorithme (Metropolis dans Gibbs optimisé) est mis en évidence en gras dans la colonne grise. . . . .	60

## LISTE DES FIGURES

Figure 3.1	Illustration de l'influence du paramètre de forme $\xi$ sur la loi GEV. . . . .	18
Figure 3.2	Ajustement d'une loi GEV aux maxima annuels de précipitations journalières à l'aéroport Montréal-Trudeau de 1941 à 2016 ( $\mu = 41.58$ , $\sigma = 10.14$ , $\xi = 0.04$ ). . . . .	20
Figure 3.3	Schéma de dépendance de la grille ( $m_1 = m_2 = 3$ ). Les cellules voisines de la cellule 5 sont les cellules 2, 4, 6 et 8. . . . .	21
Figure 4.1	Sous-domaine régulier centré sur le Québec de la grille RCaS. À titre d'illustration, les maxima annuels des précipitations journalières de RCaS pour l'année 2022 sont présentés. . . . .	23
Figure 4.2	Schéma de la modélisation du territoire en grille régulière (à gauche). Maxima de précipitations journalières pour l'année 2022 issus de la réanalyse RCaS (à droite). . . . .	25
Figure 4.3	Densité de la loi a priori du paramètre $\xi$ définie par $f_\xi(\xi) = \text{Beta}(\xi + .5 9, 6)$ . . . . .	27
Figure 4.4	Schéma représentant les relations entre les paramètres du modèle hiérarchique bayésien spatial pour les précipitations extrêmes. . . . .	27
Figure 6.1	Traces de la divergence KL et du mode des lois <i>a posteriori</i> pour chacun des paramètres obtenues par la méthode SQAVI. . . . .	37
Figure 6.2	Approximation des lois <i>a posteriori</i> marginales (courbes rouges) de chacun des paramètres et histogrammes (bandes bleues) des échantillons des lois <i>a posteriori</i> marginales obtenus par MCMC. . . . .	38
Figure 7.1	Localisation de la cellule témoin sur les données. . . . .	40
Figure 7.2	Traces de la divergence et du mode des lois <i>a posteriori</i> pour chacun des paramètres obtenues par la méthode SQAVI. . . . .	42
Figure 7.3	Approximation des lois <i>a posteriori</i> marginales (courbes rouges) de chacun des paramètres et histogrammes (bandes bleues) des échantillons des lois <i>a posteriori</i> marginales obtenus par MCMC pour les précipitations réanalysées de RCaS. . . . .	43
Figure 7.4	Cartes de chaleur des estimations SQAVI (à gauche) et MCMC (à droite) du paramètre de localisation sur le territoire d'intérêt. . . . .	44
Figure 7.5	Cartes de chaleur des estimations SQAVI (à gauche) et MCMC (à droite) du paramètre d'échelle sur le territoire d'intérêt. . . . .	45

Figure 7.6	Cartes de chaleur du niveau de retour associé à la période de retour 100 ans, noté $z_{100}$ , estimé par la méthode SQAVI (à gauche) et MCMC (à droite). . . . .	45
Figure A.1	Évolution des corrélations entre $\xi$ , $\mu_1$ et $\phi_1$ en fonction de la taille de la grille. . . . .	56

**LISTE DES ANNEXES**

Annexe A	CORRÉLATION ENTRE LE PARAMÈTRE DE FORME ET LES PARAMÈTRES INTRA-CELLULAIRES . . . . .	54
Annexe B	HYPOTHÈSE DE LINÉARITÉ DES VARIABLES LATENTES POUR LES PARAMÈTRES INTRA-CELLULAIRES . . . . .	57
Annexe C	CALCUL DES LOIS D'APPROXIMATION OPTIMALES . . . . .	58
Annexe D	ÉTUDE COMPARATIVE DES ALGORITHMES MCMC . . . . .	60

## LISTE DES SIGLES ET ABRÉVIATIONS

MCMC	Markov Chain Monte Carlo
CAVI	Coordinate Ascent for Variational Inference
SQAVI	Serial Quadratic Approximations for Variational Inference
RCaS	Réanalyse Canadienne de Surface
Courbes IDF	Courbes Intensité-Durée-Fréquence
INLA	Integrated Nested Laplace Approximation
Divergence KL	Divergence de Kullback-Liebler
VB	Variational Bayes
EP	Expectation-Propagation
GEV	Generalized Extreme Value
iGMRF	intrinsic Gaussian Markov Random Field

## CHAPITRE 1 INTRODUCTION

Les 9 et 10 août 2024, les vestiges de l'ouragan Debby ont provoqué des records de précipitations dans la région de Montréal et des Laurentides provoquant des inondations aux conséquences bien nombreuses (Lavoie, 2024) : glissements de terrain, débordements d'égoûts, pannes de courant, routes coupées... Avec des dégâts estimés à plus de 2,5 milliards de dollars, c'est l'événement climatique le plus coûteux de l'histoire du Québec selon le Bureau d'assurance du Canada (2024). Or ces épisodes de pluies extrêmes seront vraisemblablement plus fréquents et plus intenses avec l'augmentation des températures due aux changements climatiques (Ouranos, 2025). Prévoir ces aléas avec la plus grande précision possible est donc capital pour préparer nos sociétés à les affronter.

À titre d'exemple, améliorer les prédictions de précipitations extrêmes permettrait de générer des courbes Intensité-Durée-Fréquence (IDF) plus précises. Ces graphiques, fournies par Environnement et Changement Climatique Canada (ECCC), aident à dimensionner les infrastructures humaines (ponts, barrages etc. ...) pour les rendre robustes face aux inondations.

### 1.1 Analyse des valeurs extrêmes

La théorie des valeurs extrêmes (e.g. Coles, 2001) est utilisée pour estimer l'occurrence et l'intensité des précipitations extrêmes à partir des données de précipitations. Les valeurs extrêmes étant rares par définition, les modèles statistiques de valeurs extrêmes souffrent d'une grande variance échantillonnale qui se répercute sur l'estimation de l'occurrence et de la sévérité des événements extrêmes. C'est pourquoi il est important d'exploiter au mieux les informations que l'on possède pour réduire cette variance échantillonnale.

Une façon d'exploiter plus d'information sur les extrêmes est de considérer leur dépendance spatiale. De cette façon, l'information peut être partagée d'un site à l'autre. Pour ce faire, un modèle statistique spatial de valeurs extrêmes doit être utilisé. Pour les précipitations, les modèles hiérarchiques bayésiens sont communément utilisés. Par exemple, dans le modèle proposé par Cooley *et al.* (2007), la dépendance spatiale des précipitations extrêmes enregistrées en plusieurs sites est modélisée par un processus gaussien dans la couche latente d'un modèle hiérarchique bayésien de valeurs extrêmes.

## 1.2 Disponibilité des données

Au Canada, le réseau des stations météorologiques est épars compte tenu de la taille du territoire. L'information sur les précipitations extrêmes n'est pas disponible partout. Il peut alors être avantageux d'utiliser la réanalyse canadienne de surface (RCaS, Gasset *et al.*, 2021). Cet outil simule à haute résolution des variables de surface continues dans l'espace et le temps et en particulier les accumulations de précipitations horaires et journalières. Il se concentre sur la météorologie passée en s'efforçant d'être le plus précis possible afin de fournir des données fiables pour l'analyse et l'inférence. Utiliser ces données simulées pour ajuster un modèle hiérarchique bayésien spatial de valeurs extrêmes (e.g. Jalbert *et al.*, 2017) permet de s'affranchir du problème de manque de données réelles.

## 1.3 Problématique

L'approche bayésienne est très utile pour ajuster des modèles hiérarchiques. En effet, dans ce cadre, l'inférence revient exclusivement à calculer la distribution conditionnelle des paramètres sachant les observations, appelée loi *a posteriori*, qui s'exprime simplement comme produit entre la vraisemblance des observations et les lois *a priori*.

La relation de Bayes donne accès à l'expression non normalisée de la loi *a posteriori* conjointe des paramètres. En revanche il est difficile — voire impossible — de déterminer sa constante de normalisation. Ce problème est courant en statistique bayésienne et de nombreuses méthodes numériques ont été mises au point pour tenter de le surmonter (van de Schoot *et al.*, 2021). Une solution basée sur les algorithmes de type Monte-Carlo par Chaînes de Markov (MCMC) permet d'approcher la loi exacte de manière efficace (Jalbert *et al.*, 2022). Ces algorithmes consistent à générer un échantillon de taille arbitraire issu de la loi cible même sans connaître sa constante de normalisation, donnant accès à des estimations empiriques des paramètres. Les MCMC sont garantis de converger à condition d'être correctement réglés, et leur précision n'est pas altérée en haute dimension.

Malgré leurs performances et leur large acceptation dans la communauté statistique, les MCMC ne sont pas toujours les plus adaptés. Premièrement, ils peuvent être coûteux en temps et en mémoire puisqu'ils nécessitent de générer et stocker des échantillons de grande taille. Ensuite, diagnostiquer leur convergence est un travail fastidieux et exigeant (cf Brooks *et al.* (2011) et la section 3.2). Bien que Roy (2020) souligne de nombreuses avancées dans l'établissement de critères quantifiés pour optimiser les réglages des MCMC, les auteurs précisent que la recherche dans ce domaine reste active et nécessite encore des travaux importants. Quant aux méthodes qualitatives, « elles ne permettent pas d'établir la convergence

[des MCMC] avec certitude ». Enfin, dans notre cas précis, les bases théoriques sur lesquelles les MCMC reposent sont peu accessibles pour les non-initiés, en particulier pour les ingénieurs qui seront amenés à les manipuler.

Le but de ce travail est de proposer un algorithme alternatif aux MCMC pour réaliser l’inférence du modèle de valeurs extrêmes.

Les méthodes variationnelles (Blei *et al.*, 2017), constituent une démarche prometteuse pour atteindre cet objectif. Elles visent à approximer la loi *a posteriori* par une loi explicite plus simple en minimisant une certaine distance. Il s’agit donc d’un problème d’optimisation. Ces techniques présentent un grand potentiel car, en plus de leur rapidité de convergence et leur faible occupation mémoire, leur diagnostic de convergence est simple si bien que très peu d’hyperparamètres nécessitent d’être réglés.

Les objectifs du mémoire consistent à (1) définir une famille de loi d’approximation appropriée pour notre problème ; (2) résoudre le problème d’optimisation *via* un algorithme efficace (3) valider cette méthode sur des données issues de simulations climatiques, en comparant les estimations à celles obtenues par MCMC.

## 1.4 Organisation du mémoire

Dans un premier temps, le chapitre 2 présente les méthodes variationnelles appliquées à l’inférence bayésienne et dresse l’état de l’art des techniques d’approximation privilégiées par la communauté statistique. Le chapitre 3 explore les concepts théoriques de la statistique bayésienne, des algorithmes MCMC, de la méthode Coordinate Ascent for Variational Inference, de la théorie des extrêmes et des champs markoviens gaussiens intrinsèques. Le chapitre 4 décrit le jeu de données de test et introduit le modèle hiérarchique bayésien de précipitation. Le chapitre 5 présente notre procédure variationnelle intitulée *Serial Quadratic Approximations for Variational Inference* (SQAVI) dont les performances sont évaluées aux chapitres 6 et 7 et comparées aux méthodes MCMC. Ces résultats sont discutés au chapitre 8. Enfin, le chapitre 9 établit le bilan de ce travail et mentionne des pistes d’exploration possibles pour de futures recherches.

## CHAPITRE 2 REVUE DE LITTÉRATURE

Le calcul variationnel désigne un ensemble de méthodes visant à maximiser ou minimiser des fonctionnelles, i.e. des fonctions de fonctions, sur un espace donné. Sa première formalisation remonte au XVIII<sup>e</sup> siècle suite aux travaux sur le calcul des variations de Leonhard Euler et Joseph-Louis Lagrange (Fraser, 1994). Son application aux méthodes d'inférence a été introduite à la fin du XX<sup>e</sup> siècle pour optimiser des réseaux de neurones (Frey et Hinton, 1999).

Le problème d'inférence en statistique bayésienne est le suivant. Pour un modèle statistique donné dont la variable aléatoire  $Y$  est caractérisée par la densité de probabilité  $f_{(Y|\theta)}(y)$  où  $\theta$  est le vecteur des paramètres, on cherche la distribution qui décrit  $\theta$  sachant le vecteur (ou la matrice)  $y$  des réalisations de  $Y$ . Cette densité conditionnelle est appelée loi *a posteriori* et est notée  $f_{(\theta|Y=y)}(\theta)$ . L'idée de l'inférence variationnelle est de trouver une loi explicite simple qui s'approche le plus possible de la loi *a posteriori*. En ce sens, elle fait partie des approches d'approximation — pour reprendre les termes de Rue *et al.* (2009) — et équivaut à un problème d'optimisation. Ces stratégies contrastent avec les méthodes exactes telles que les techniques d'échantillonnage parmi lesquelles les algorithmes MCMC (Hastings, 1970) sont les plus populaires.

### 2.1 Les approches bayésiennes variationnelles

Pour approcher la loi *a posteriori*, Hinton et van Camp (1993) et MacKay (1995) proposent de considérer une famille de distributions  $\mathcal{Q}$  et d'en extraire le membre qui minimise l'*énergie variationnelle libre*, correspondant en statistique à la *divergence de Kullback-Leibler* (divergence KL) entre la vraie loi *a posteriori* et la loi d'approximation. La stratégie la plus commune consiste à considérer une classe paramétrique  $\mathcal{Q} = \{q(\theta|\lambda)\}$  avec les paramètres ajustables  $\lambda$ , appelés *paramètres variationnels*. Les valeurs retenues de ces paramètres sont celles qui minimisent la divergence KL.

La flexibilité offerte par ces techniques, dites bayésiennes variationnelles (VB), a encouragé leur application à une vaste gamme de modèles au début des années 2000. Parmi ceux-ci, on retrouve les modèles de Markov cachés (MacKay, 1997), les modèles de mélange (Humphreys et Titterington, 2000), les modèles graphiques (Jordan *et al.*, 1999) et les modèles d'espace-état (Beal, 2003). Pour chacune de ces applications, des algorithmes spécifiques ont été proposés.

À la suite de ces études, un cadre général de résolution a pu être établi. Ce cadre repose sur une hypothèse d'indépendance *a posteriori* des paramètres, appelée *hypothèse de champ moyen*. L'algorithme Coordinate Ascent for Variational Inference (CAVI, Bishop, 2006) est le plus populaire. Inspiré de la méthode d'Espérance-Maximisation (EM, Dempster *et al.*, 1977), il estime alternativement chaque paramètre variationnel en fonction des autres. Il est particulièrement efficace pour les modèles où les lois conditionnelles complètes appartiennent à la famille exponentielle (Blei *et al.*, 2017), ce qui n'est malheureusement pas le cas dans cette étude.

Des travaux récents ont permis d'améliorer les algorithmes bayésiens variationnels afin de les adapter aux problèmes impliquant des données massives, d'éviter les minima locaux, d'intégrer de la dépendance entre les paramètres et d'augmenter la précision de l'approximation (Blei *et al.*, 2017). L'approche la plus prometteuse est celle du *Black Box Variational Inference* (Ranganath *et al.*, 2013). Elle exploite les techniques relatives à l'inférence variationnelle stochastique pour minimiser la divergence KL. L'idée est d'écrire le gradient de cette divergence comme une espérance et de l'estimer par la méthode Monte-Carlo. Une fois le gradient connu, des méthodes de descente de gradient peuvent être employées. En plus de pouvoir s'appliquer sur n'importe quelle classe de loi d'approximation  $\mathcal{Q}$ , cet algorithme peut profiter des méthodes stochastiques pour réduire son temps de calcul, notamment *via* un sous-échantillonnage des observations. Cependant une telle démarche nécessite de contrôler la variance du gradient estimé, requérant ainsi l'utilisation de techniques complexes.

À ce jour, aucun ouvrage ne se concentre spécifiquement sur l'application de telles méthodes à des modèles de valeurs extrêmes. Or, elles pourraient être d'une grande utilité, notamment grâce à leur fort pouvoir d'interprétabilité qui permet l'extrapolation, essentielle dans l'étude des extrêmes. Ce travail propose donc d'explorer l'utilisation des méthodes bayésiennes variationnelles à un modèle de valeurs extrêmes.

## 2.2 L'espérance-propagation

Parallèlement aux méthodes bayésiennes variationnelles sont apparues des méthodes dites séquentielles, notamment l'*Assumed-Density Filtering* (ADF) qui tire son nom de la théorie du contrôle (Maybeck, 1982). Tout comme les méthodes VB, l'ADF requiert de fixer au préalable une famille de distributions d'approximation  $\{q\}_{q \in \mathcal{Q}}$ .

Le principe est de traiter les vecteurs d'observations un par un. À chaque itération  $i$ , une distribution *a posteriori* partielle est construite à partir de la densité d'approximation courante et de l'information portée par le  $i$ -ème vecteur de données. La nouvelle fonction d'approximation

mation mise à jour est celle qui minimise la divergence KL avec la loi *a posteriori* partielle. La nature séquentielle de ce processus en fait sa principale faiblesse : l'information contenue dans les observations qui sont traitées au début est perdue au fil des itérations. L'ADF est également sensible à l'ordre des itérations, ce qui empêche d'accélérer l'algorithme en traitant les données par paquets.

Pour dépasser ces limites, l'algorithme d'Espérance-Propagation (EP, Minka, 2001, 2013) utilise le même principe mais en adoptant un point de vue différent. Plutôt que d'approcher séquentiellement la loi *a posteriori* en parcourant les vecteurs d'observations, il approche l'information portée par les vecteurs d'observations, avant de construire une loi *a posteriori* exacte basée sur ces approximations. En approchant l'information contenue dans les données plutôt que la loi finale, il peut traiter simultanément tous les vecteurs d'observations à la fois.

Cette méthode s'apparente aux approches bayésiennes variationnelles, dans la mesure où elle raffine itérativement une loi d'approximation afin de la rapprocher au plus près de la distribution *a posteriori* en minimisant la divergence KL. Néanmoins plusieurs différences sont à souligner. Premièrement, pour une densité *a posteriori*  $p$  et une densité d'approximation  $q$ , l'algorithme EP vise à minimiser  $KL(p||q) = \int p \log \frac{p}{q}$ , tandis que les méthodes bayésiennes variationnelles minimisent  $KL(q||p)$ . Par conséquent, l'algorithme EP a tendance à surestimer la variance globale de la distribution *a posteriori*, tandis que les méthodes bayésiennes variationnelles ont tendance à sous-estimer la variance de la loi *a posteriori* (voir par exemple Bishop, 2006, chapitre 10.1). Deuxièmement, la convergence de la loi estimée par l'algorithme EP n'est pas assurée, contrairement à celle estimée par les approches bayésiennes variationnelles (Sun, 2013). C'est pourquoi on priviliege ces dernières dans notre étude.

### 2.3 L'approximation de Laplace intégrée et imbriquée

Plus récemment, des techniques d'approximation spécifiques aux modèles gaussiens latents ont vu le jour. Il s'agit de modèles hiérarchiques pour lesquels les lois *a priori* correspondent à des champs gaussiens, ce qui est le cas dans notre étude. L'*Integrated Nested Laplace Approximation* (INLA, Rue *et al.*, 2009) fait partie des approches les plus communes. Dans la suite de cette section, on note  $d$  la dimension du vecteur  $\boldsymbol{\theta}$ ,  $\theta_i$  ( $1 \leq i \leq d$ ) ses composantes et  $\boldsymbol{\kappa}$  le vecteur des hyperparamètres qui régissent les lois *a priori* des  $\theta_i$ .

La méthode repose sur la possibilité d'exprimer les lois marginales *a posteriori* des  $\theta_i$  comme une intégrale du produit de la loi  $f_{(\theta_i|Y=\mathbf{y},\boldsymbol{\kappa})}(\theta_i)$  et de la loi  $f_{(\boldsymbol{\kappa}|Y=\mathbf{y})}(\boldsymbol{\kappa})$ . Le but de cet algorithme est de calculer cette intégrale numériquement en parcourant le support de la loi

$f_{(\kappa|Y=y)}(\kappa)$ . Pour rendre cette intégration numérique possible (*i.e.* transformer l'intégrale en une somme finie), une procédure sur mesure permet de générer un nombre raisonnable de valeurs de  $\kappa$  qui explorent suffisamment bien le domaine de leur loi marginale *a posteriori*. La loi marginale  $f_{(\kappa|Y=y)}(\kappa)$  est quant à elle approchée par l'approximation de Laplace telle que définie par Tierney *et al.* (1986). En outre, les lois conditionnelles  $f_{(\theta_i|Y=y, \kappa)}(\theta_i)$  doivent être approximées à faible coût car  $d$  peut être très grand. Pour cela, une approximation dite de Laplace simplifiée est utilisée.

Plus récemment, INLA a été appliqué sur un modèle statistique pour l'analyse des précipitations extrêmes (Opitz *et al.*, 2018). Cette étude a permis à l'algorithme de prouver sa capacité à prendre en compte les lois extrêmes telle que la loi de Pareto Généralisée (GPD).

Bien que très répandu dans la communauté statistique, INLA présente certains inconvénients au regard de notre application. Tout d'abord, l'algorithme requiert de régler des hyperparamètres qui interviennent dans la procédure d'exploration de l'espace de  $\kappa$ . De plus, il ne fournit pas de loi explicite mais seulement des estimations numériques. L'approche bayésienne variationnelle, ne nécessitant pas d'hyperparamètres à régler, donnant accès à des lois explicites et s'interprétant simplement comme un problème de minimisation, semble être une meilleure alternative à explorer. Néanmoins, il convient d'étudier dans quelles mesures les simplifications qu'elle engendre peuvent détériorer les estimations dans le cas d'un modèle de valeurs extrêmes.

## CHAPITRE 3 CADRE THÉORIQUE

Ce chapitre présente les principaux concepts théoriques sur lesquels repose notre travail. Il débute par une série de définitions importantes de la statistique bayésienne, dont celle de loi *a posteriori*, un élément central de notre étude. Puis, il détaille le fonctionnement des algorithmes Monte-Carlo par chaîne de Markov qui sont aujourd’hui privilégiés pour l’inférence bayésienne. Il explique ensuite la procédure Coordinate Ascent for Variational Inference, une méthode variationnelle qui constitue la base de notre solution. Enfin, il introduit la loi des valeurs extrêmes généralisée et les champs markoviens gaussiens intrinsèques qui composent le modèle hiérarchique bayésien pour les précipitations extrêmes. Tout au long de ce chapitre, on note  $\boldsymbol{\theta}$  le vecteur des paramètres d’intérêt à estimer,  $Y$  la variable aléatoire représentant les observations (scalaire, vecteur ou matrice) et  $y$  une réalisation de  $Y$ .

### 3.1 Inférence bayésienne

En statistique bayésienne, les paramètres sont considérés comme des variables aléatoires. Le théorème de Bayes permet de calculer leur densité de probabilité sachant les observations :

$$f_{(\boldsymbol{\theta}|Y=y)}(\boldsymbol{\theta}) = \frac{f_{(Y|\boldsymbol{\theta})}(y)f_{\boldsymbol{\theta}}(\boldsymbol{\theta})}{f_Y(y)}. \quad (3.1)$$

La loi  $f_{(\boldsymbol{\theta}|Y=y)}(\boldsymbol{\theta})$  est appelée loi *a posteriori* et le but de l’inférence bayésienne est de l’estimer. La loi  $f_{(Y|\boldsymbol{\theta})}(y)$  correspond à la vraisemblance des observations. Elle est généralement facile à calculer dans le cas où les observations sont considérées indépendantes conditionnellement aux paramètres  $\boldsymbol{\theta}$  et au modèle statistique. La quantité  $f_{\boldsymbol{\theta}}(\boldsymbol{\theta})$  est appelée loi *a priori*. C’est une densité qui peut contenir de l’information que l’analyste possède sur les paramètres avant d’avoir vu les données. S’il n’y a pas d’information *a priori*, la densité peut être non informative. La quantité  $f_Y(y)$ , souvent appelée l’évidence, est particulièrement difficile à calculer puisqu’elle correspond à l’intégrale :

$$f_Y(y) = \int_{\Theta} f_{(Y|\boldsymbol{\theta})}(y)f_{\boldsymbol{\theta}}(\boldsymbol{\theta})d\boldsymbol{\theta}.$$

Dans la majorité des cas, les quantités  $f_{(Y|\boldsymbol{\theta})}(y)$  et  $f_{\boldsymbol{\theta}}(\boldsymbol{\theta})$  sont connues car déterminées par le modèle, tandis que  $f_Y(y)$  est inconnue. Des méthodes sont donc utilisées pour estimer la loi *a posteriori*, parmi lesquelles l’échantillonnage Monte-Carlo par chaîne de Markov et les méthodes variationnelles. L’équation (3.1) s’écrit souvent

$$f_{(\boldsymbol{\theta}|Y=y)}(\boldsymbol{\theta}) \propto f_{(Y|\boldsymbol{\theta})}(y) f_{\boldsymbol{\theta}}(\boldsymbol{\theta}) \quad (3.2)$$

pour mettre en évidence la forme fonctionnelle en  $\boldsymbol{\theta}$  de la loi *a posteriori*, i.e. la loi sans constante de normalisation.

L'équation (3.2) souligne l'importance du choix de la loi *a priori*, qui se combine avec l'information portée par les observations pour réaliser l'inférence. Moins les données sont nombreuses, plus l'influence de la loi *a priori* devient prépondérante dans l'estimation des paramètres, et inversement. Les lois informatives peuvent donc aussi bien améliorer l'estimation des paramètres que la déteriorer si elles sont choisies à tort. *A contrario*, les lois *a priori* non informatives évitent de biaiser l'inférence mais la rendent très sensible aux données, augmentant le risque de surajustement du modèle.

Les algorithmes d'inférence bayésienne ont souvent recourt aux *lois conditionnelles complètes* des paramètres. Soit  $\theta_i$  le  $i$ -ème paramètre du modèle, alors sa loi conditionnelle complète se note  $f_{(\theta_i|y, \boldsymbol{\theta}_{-i})}$ , où  $\boldsymbol{\theta}_{-i}$  est le vecteur contenant tous les paramètres sauf le  $i$ -ème. Elle est parfois notée  $f_{(\theta_i|Y=y, \dots)}$  pour alléger le texte. Cette loi ne doit pas être confondue avec  $f_{(\theta_i|Y=y)}$ , la loi marginale *a posteriori* du  $i$ -ème paramètre.

### 3.2 Les algorithmes Monte-Carlo par Chaîne de Markov

Les algorithmes Monte-Carlo par Chaîne de Markov (MCMC, Hastings, 1970) désignent un ensemble de méthodes servant à échantillonner des lois de probabilité sans nécessairement connaître les constantes de normalisation. En statistique bayésienne, les méthodes MCMC sont largement utilisées pour générer un échantillon de la loi *a posteriori* des paramètres car sa constante de normalisation est souvent inconnue. Les méthodes MCMC fonctionnent selon un processus itératif d'acceptation-rejet. Le principe d'une itération est le suivant. À chaque paramètre, on propose un candidat — i.e. une valeur — généré aléatoirement *via* une *loi instrumentale*. Ce candidat est accepté avec une certaine *probabilité d'acceptation*. S'il est accepté, alors il devient la nouvelle valeur du paramètre. S'il est rejeté, alors le paramètre garde sa valeur précédente. Le ratio du nombre de candidats acceptés sur la taille de la chaîne s'appelle le *taux d'acceptation*, il est un indicateur de la qualité de l'échantillonnage (cf Section 3.2.7). L'évolution des valeurs prises par un paramètre au cours des itérations s'appelle la *trace* du paramètre. À partir d'un certain nombre d'itérations, et si la probabilité d'acceptation est bien choisie, on entre dans la *phase d'échantillonnage* : les échantillons générés se comportent selon la loi *a posteriori* normalisée. Bien que les algorithmes MCMC partagent cette base de fonctionnement, ils diffèrent selon le processus de génération de candidats et la probabilité

d'acceptation. Dans la suite de cette section, on présente les principales variantes de ces algorithmes. Soient  $\theta$  une composante quelconque de  $\boldsymbol{\theta}$  et  $d$  sa dimension.

### 3.2.1 Metropolis

L'algorithme de *Metropolis* (Metropolis *et al.*, 1953) fonctionne en utilisant une marche aléatoire symétrique centrée sur l'état présent, *i.e.* la valeur courante du paramètre, pour générer un candidat. La variance du pas de la marche aléatoire est un hyperparamètre appelé *variance instrumentale* qu'il convient de régler au préalable. Si  $\tilde{\theta}$  est le candidat proposé au paramètre  $\theta$ , alors le candidat est accepté avec la probabilité :

$$\alpha(\tilde{\theta}, \theta) = \min \left\{ 1, \frac{\tilde{f}_{(\tilde{\theta}|y)}(\tilde{\theta})}{\tilde{f}_{(\theta|y)}(\theta)} \right\} \quad (3.3)$$

où  $\tilde{f}_{(\theta|Y=y)}(\theta)$  désigne la densité *a posteriori* non normalisée (3.2).

### 3.2.2 Metropolis-Hastings

L'algorithme de *Metropolis-Hastings* (Hastings, 1970) vient relaxer la contrainte de symétrie de la marche aléatoire en proposant une autre probabilité d'acceptation. Si  $\pi_{(\tilde{\theta}|\theta)}(\tilde{\theta})$  désigne la densité instrumentale — non forcément symétrique — du paramètre  $\theta$ , alors la probabilité d'acceptation s'écrit

$$\alpha(\tilde{\theta}, \theta) = \min \left\{ 1, \frac{\tilde{f}_{(\tilde{\theta}|y)}(\tilde{\theta})\pi_{(\theta|\tilde{\theta})}(\theta)}{\tilde{f}_{(\theta|y)}(\theta)\pi_{(\tilde{\theta}|\theta)}(\tilde{\theta})} \right\}. \quad (3.4)$$

En plus de maximiser l'exploration de l'espace de  $\theta$  (Peskun, 1973), l'équation (3.4) donne un plus large choix de probabilités instrumentales à utiliser, permettant dans certains cas d'accélérer la convergence de l'algorithme, notamment en guidant la génération de candidats vers les régions où la densité *a posteriori* est plus élevée. C'est ce qui est notamment exploité dans l'*Algorithme de Langevin ajusté par Metropolis* (MALA, Roberts et Stramer, 2002).

### 3.2.3 Algorithme de Langevin ajusté par Metropolis

L'algorithme MALA provient de la discréétisation d'un processus de diffusion Langevin défini par l'équation différentielle stochastique

$$d\boldsymbol{\theta}_t = d\mathbf{b}_t + \frac{1}{2}\nabla \log \tilde{f}_{(\boldsymbol{\theta}|y)}(\boldsymbol{\theta}_t)dt \quad (3.5)$$

où  $\mathbf{b}_t$  dénote un mouvement brownien de dimension  $d$  au temps  $t$ .

L'avantage de ce processus est qu'il est stationnaire vis-à-vis de la distribution cible  $\tilde{f}_{(\boldsymbol{\theta}|y)}(\boldsymbol{\theta})$ . Ainsi, les candidats générés par ce processus vont explorer en priorité les zones à forte densité de la loi *a posteriori*. Par conséquent ils seront plus souvent acceptés et la chaîne va converger plus rapidement vers la loi cible.

Pour pouvoir utiliser l'équation (3.5), il faut la discréteriser. La discréterisation d'Euler-Maruyama (Maruyama, 1955) avec un pas  $h > 0$  nous donne

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t + \frac{h}{2} \nabla \log \tilde{f}_{(\boldsymbol{\theta}|y)}(\boldsymbol{\theta}_t) + \sqrt{h} Z_{t+1}$$

où  $\{Z_t\}_{t \geq 1}$  sont indépendants identiquement distribuées (i.i.d.) de loi  $\mathcal{N}_d(\mathbf{0}, I_d)$ , avec  $\mathcal{N}_d$  la loi normale multidimensionnelle de dimension  $d$ . En découle la loi instrumentale

$$\tilde{\boldsymbol{\theta}} \sim \mathcal{N}_d \left( \boldsymbol{\theta}_t + \frac{h}{2} \nabla \log \tilde{f}_{(\boldsymbol{\theta}|y)}(\boldsymbol{\theta}_t), hI_d \right).$$

Celle-ci étant non symétrique, elle s'insère dans un algorithme de Metropolis-Hastings pour former l'algorithme MCMC correspondant. Le *paramètre d'échelonnage*  $h > 0$ , réglé par l'utilisateur, joue le rôle de la variance instrumentale de l'algorithme de Metropolis.

Bien que plus efficace que l'algorithme de Metropolis grâce au choix plus précis des candidats, le MALA souffre d'un coût d'exécution élevé notamment à cause du calcul du terme  $\nabla \log \tilde{f}_{(\boldsymbol{\theta}|y)}(\boldsymbol{\theta}_t)$  à chaque itération. Ce compromis entre efficacité et coût computationnel explique pourquoi il n'est pas systématiquement adopté.

### 3.2.4 Monte-Carlo Hamiltonien

Également dans le but de proposer des candidats de manière plus efficace, l'algorithme de Monte-Carlo Hamiltonien (HMC, Duane *et al.*, 1987; Neal, 2011) introduit une variable auxiliaire indépendante, le momentum  $\boldsymbol{\phi}$ , de loi  $\mathcal{N}_d(\mathbf{0}, M)$  où  $M$  est généralement une matrice diagonale. A chaque itération du HMC, on génère une réalisation de  $\boldsymbol{\phi}$ , puis on propose simultanément un candidat  $(\tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\phi}})$  via une procédure gouvernée par un *paramètre d'échelonnage* et un *nombre de sauts*, qui imite la dynamique hamiltonienne en physique. Les candidats étant générés de manière asymétrique, cet algorithme est également un cas particulier de Metropolis-Hastings dont la probabilité d'acceptation est

$$\alpha\{(\tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\phi}}), (\boldsymbol{\theta}, \boldsymbol{\phi})\} = \min \left\{ 1, \frac{\tilde{f}_{(\boldsymbol{\theta}|y)}(\tilde{\boldsymbol{\theta}}) \mathcal{N}_d(\tilde{\boldsymbol{\phi}}|\mathbf{0}, M)}{\tilde{f}_{(\boldsymbol{\theta}|y)}(\boldsymbol{\theta}) \mathcal{N}_d(\boldsymbol{\phi}|\mathbf{0}, M)} \right\}. \quad (3.6)$$

Une variante du HMC, le *no-U-turn sampler* (NUTS, Homan et Gelman, 2014), est particuli-

lièrement appréciée car elle détermine le nombre de sauts optimal à chaque itération.

### 3.2.5 Gibbs

L'échantillonneur de Gibbs (Geman et Geman, 1984) se base sur l'utilisation des lois conditionnelles complètes afin d'étudier l'espace des paramètres. Bien qu'il ne fasse pas intervenir de probabilité d'acceptation, il fait partie des algorithmes MCMC car il est un cas particulier de l'algorithme de Metropolis-Hastings, pour lequel les candidats sont toujours acceptés.

Dans l'échantillonneur de Gibbs, on détermine pour chaque paramètre  $\theta$  sa loi conditionnelle complète (cf Section 3.1) en fonction des autres paramètres. Le plus souvent, la loi conditionnelle complète possède une forme explicite facile à échantillonner. On génère donc la nouvelle valeur de  $\theta$  à partir de sa loi conditionnelle complète sans passer par l'étape d'acceptation, puis on fait de même pour le paramètre suivant et ainsi de suite. Parcourir tous les paramètres correspond à une itération. Au fil des itérations, les lois conditionnelles complètes sont affinées. À partir d'un certain nombre d'itérations, les échantillons sont générés par la loi cible.

### 3.2.6 Metropolis-Hastings dans Gibbs

Lorsque, dans l'échantillonneur de Gibbs, la loi conditionnelle complète de  $\theta$  n'est pas une loi explicite, il est néanmoins possible de générer une réalisation de ce paramètre grâce à une *étape Metropolis-Hastings* à une itération. Dans cette étape, un candidat est généré *via* une distribution instrumentale et est accepté avec la probabilité (3.4), dans laquelle la loi cible n'est pas la loi *a posteriori* mais la loi conditionnelle complète. Pour un nombre d'itérations assez élevé et moyennant un pas instrumental adéquat, cette méthode équivaut à échantillonner  $\theta$  selon sa loi conditionnelle complète, permettant de se ramener à l'algorithme de Gibbs.

L'avantage de cette méthode est de pouvoir appliquer l'algorithme de Gibbs à des modèles où seules certaines lois conditionnelles complètes sont explicites. Il existe autant de variantes de cet algorithme que de variantes de l'algorithme Metropolis-Hastings : Metropolis dans Gibbs, MALA dans Gibbs, NUTS dans Gibbs etc...

### 3.2.7 Réglage des hyperparamètres

Les algorithmes MCMC requièrent de choisir avec précaution un certain nombre d'hyperparamètres pour fonctionner correctement :

- La *taille de la chaîne* : nombre d’itérations de l’algorithme. Plus elle est longue, plus on a de réalisations pour procéder à l’inférence. En revanche le coût de simulation associé est plus élevé.
- Le *temps de chauffe* : quantité de réalisations à supprimer au début de la chaîne avant que celle-ci entre dans sa phase d’échantillonnage. Parfois, ce seuil est évident. Dans certains cas, en particulier si la chaîne converge lentement, il faut faire attention à ne pas le choisir trop court, sans quoi des réalisations absurdes pour la loi cible s’introduisent dans le calcul des estimations. *A contrario*, le choisir trop grand implique de supprimer des réalisations de la phase d’échantillonnage et donc de perdre en précision.
- Le *pas instrumental* : désigne aussi bien la variance instrumentale de l’algorithme de Metropolis que le paramètre d’échelonnage du MALA et HMC. Ce paramètre joue sur l’amplitude de la loi instrumentale et est réglé grâce aux taux d’acceptation du Tableau 3.1. Plus il est élevé, plus l’écart de valeurs entre les candidats successifs d’une même chaîne sont importants. Il dicte donc la vitesse à laquelle la chaîne va parcourir l’espace du paramètre. S’il est trop faible, alors la chaîne converge trop lentement, impliquant un taux d’acceptation trop élevé. S’il est trop élevé, alors les candidats risquent de sortir trop souvent des zones à forte densité et donc d’être rejeter trop souvent, impliquant un taux d’acceptation trop faible.
- L’*amincissement* (en anglais *thinning*) : nombre de réalisations consécutives à ignorer dans une chaîne avant de conserver la suivante. Par exemple, si l’amincissement est de 5, cela signifie que seul un échantillon toutes les 5 itérations de la chaîne est conservé. Cela permet d’améliorer la qualité des échantillons en réduisant l’autocorrélation entre les réalisations, tout en réduisant la taille de la chaîne. L’autocorrélation au sein d’une chaîne peut être mesurée grâce à la *taille d’échantillon effective* (ESS, Gelman *et al.*, 2013).

TABLEAU 3.1 Taux d’acceptation asymptotiques optimaux des algorithmes MCMC.

Algorithme	Metropolis-Hastings	MALA	HMC
Taux d’acceptation optimal (%)	23.4	57.4	65

Le calibrage des hyperparamètres est une étape critique dans le fonctionnement d’un algorithme MCMC. Si un seul hyperparamètre est mal ajusté, l’échantillon de la loi *a posteriori* est erroné. De plus, cette étape ne peut être entièrement automatisée, car elle exige une analyse manuelle des traces. L’expertise nécessaire peut représenter un obstacle, limitant ainsi la diffusion de ces algorithmes parmi les groupes de recherche et d’analyse. Ces défis justifient

l'exploration d'autres méthodes, telles que les approches bayésiennes variationnelles. Parmi elles, l'algorithme Coordinate Ascent for Variational Inference (CAVI, Blei *et al.*, 2017) constitue une approche prometteuse, notamment grâce à sa capacité d'adaptation et la sobriété de son diagnostic de convergence.

### 3.3 L'algorithme CAVI

En tant que méthode bayésienne variationnelle (VB, Section 2.1), l'algorithme CAVI cherche à trouver le membre  $\boldsymbol{\theta} \mapsto q^{min}(\boldsymbol{\theta})$  de la famille de densités  $\mathcal{Q}$  qui minimise la divergence de Kullback-Leibler définie par :

$$KL(q||f_{(\boldsymbol{\theta}|Y=y)}) = \int q(\boldsymbol{\theta}) \log \frac{q(\boldsymbol{\theta})}{f_{(\boldsymbol{\theta}|Y=y)}(\boldsymbol{\theta})} d\boldsymbol{\theta}. \quad (3.7)$$

Cette quantité est positive et nulle si et seulement si  $q = f_{(\boldsymbol{\theta}|Y=y)}$ . Notons qu'il s'agit d'une divergence et non d'une distance puisqu'elle n'est pas symétrique :  $KL(q||f_{(\boldsymbol{\theta}|Y=y)}) \neq KL(f_{(\boldsymbol{\theta}|Y=y)}||q)$ .

Pour se ramener aux méthodes d'optimisation classiques, on considère une famille de densités d'approximation paramétriques  $\mathcal{Q} = \{q_{(\boldsymbol{\theta}|\boldsymbol{\lambda})}, \boldsymbol{\lambda} \in \Lambda\}$  où  $\boldsymbol{\lambda}$  est le vecteur des paramètres variationnels dont l'ensemble de définition est noté  $\Lambda$ . Ainsi on cherche

$$\boldsymbol{\lambda}^{min} = \underset{\boldsymbol{\lambda} \in \Lambda}{\operatorname{argmin}} \, KL(q_{(\boldsymbol{\theta}|\boldsymbol{\lambda})} || f_{(\boldsymbol{\theta}|Y=y)}). \quad (3.8)$$

#### 3.3.1 L'approximation champ moyen

L'approximation champ moyen tient son nom d'un cadre de résolution de problèmes établi en physique (Parisi, 1988). Il consiste à considérer comme famille  $\mathcal{Q}$  l'ensemble des fonctions de la forme

$$q_{(\boldsymbol{\theta}|\boldsymbol{\lambda})}(\boldsymbol{\theta}) = \prod_i q_{(\theta_i|\boldsymbol{\lambda}_i)}(\theta_i) \quad (3.9)$$

traduisant une indépendance *a posteriori* des lois marginales d'approximation.

Cette forme présente plusieurs avantages. Premièrement, elle permet d'obtenir les lois marginales optimales de chaque paramètre sachant les autres. Ensuite, moyennant des lois  $q_{(\theta_i|\boldsymbol{\lambda}_i)}$  assez simples, elle permet de générer facilement des échantillons *a posteriori* puisque chaque composante  $\theta_i$  peut être générée indépendamment des autres paramètres puis concaténée pour former un échantillon conjoint. Enfin, elle autorise l'ajout de contraintes supplémentaires aux  $q_{(\theta_i|\boldsymbol{\lambda}_i)}$  pour s'adapter à la nature des paramètres, ce qui est exploité pour construire

la procédure décrite au Chapitre 5.

### 3.3.2 Densités marginales optimales

Dans le cadre de l'approximation champ-moyen, la loi marginale  $q_{(\theta_i|\lambda_i)}^*$  qui minimise la divergence KL lorsque tous les indices  $j \neq i$  sont fixés est égale, à une constante de normalisation près, à (Gelman *et al.*, 2013)

$$q_{(\theta_i|\lambda_i)}^*(\theta_i) \propto \exp\{\mathbb{E}_{-i}[\log f_{(\theta_i|y,\boldsymbol{\theta}_{-i})}(\theta_i)]\}, \quad (3.10)$$

où

- $\mathbb{E}_{-i}[\cdot]$  correspond à l'espérance selon la loi d'approximation de tous les paramètres sauf celui d'indice  $i$ , *i.e.*  $\prod_{j \neq i} q_{(\theta_j|\lambda_j)}$  ;
- $\boldsymbol{\theta}_{-i}$  correspond au vecteur de tous les paramètres sauf celui d'indice  $i$  ;
- $f_{(\theta_i|y,\boldsymbol{\theta}_{-i})}(\theta_i)$  correspond à la loi conditionnelle complète du paramètre  $\theta_i$ .

On note  $q_{(\boldsymbol{\theta}|\lambda)}^*$  la loi d'approximation conjointe donnée par

$$q_{(\boldsymbol{\theta}|\lambda)}^*(\boldsymbol{\theta}) = \prod_i q_{(\theta_i|\lambda_i)}^*(\theta_i) \quad (3.11)$$

et on appelle loi  $q^*$  conjointe le produit des lois marginales optimales.

L'algorithme CAVI met à jour successivement les lois  $q_{(\theta_i|\lambda_i)}^*$ , permettant d'obtenir une approximation satisfaisante de  $q^{min}$ . Autrement dit,  $q_{(\boldsymbol{\theta}|\lambda)}^* \approx q^{min}$ . En pratique, seuls les paramètres variationnels optimaux  $\lambda_i^*$  nécessitent d'être calculés. Ils sont obtenus par l'identité

$$q_{(\theta_i|\lambda_i)}^*(\theta_i) = q_{(\theta_i|\lambda_i^*)}(\theta_i). \quad (3.12)$$

### 3.3.3 Principe de l'algorithme

Pour initialiser l'algorithme, les paramètres  $\lambda_i$  sont fixés arbitrairement. A chaque itération, on parcourt les indices  $i$  et on calcule les paramètres variationnels optimaux  $\lambda_i^*$  en utilisant successivement les équations (3.10) et (3.12) ; les paramètres variationnels  $\{\lambda_j\}_{j \neq i}$  des autres cellules étant fixés. Au fil des itérations et de la mise à jour des paramètres variationnels, la divergence KL entre la loi  $q_{(\boldsymbol{\theta}|\lambda)}^*$  (3.11) et la loi *a posteriori* diminue. Le processus s'arrête lorsqu'un minimum est atteint. Parfois, l'identification des paramètres  $\lambda_i^*$  décrite par l'équation (3.12) n'est pas possible. Une solution consiste alors à approcher les lois  $q_{(\theta_i|\lambda_i)}^*(\theta_i)$  par des lois plus simples, ce qui sera le cas au chapitre 5.

### 3.4 L'approximation quadratique

L'approximation quadratique est une technique permettant d'approcher une densité unimodale (ou multimodale mais pour laquelle un mode domine largement les autres) par une loi normale centrée en son mode.

Plus précisément, soit  $f$  une densité unimodale définie sur un support quelconque et  $x_0$  son mode. Soit  $g = \log f$ . Un développement de Taylor à l'ordre 2 de  $g$  en  $x_0$  s'écrit

$$g(x) \approx g(x_0) + g'(x_0)(x_0 - x) + \frac{1}{2}g''(x_0)(x_0 - x), \quad x \in \mathcal{X}.$$

Or, puisque  $x_0$  correspond au mode,  $g'(x_0) = 0$ . Par conséquent, l'approximation rapportée à  $f$  s'exprime :

$$\begin{aligned} f(x) &\approx \exp\left(g(x_0) + \frac{1}{2}g''(x_0)(x_0 - x)\right), \\ &\propto \exp\left(\frac{1}{2}g''(x_0)(x_0 - x)\right). \end{aligned}$$

On reconnaît la forme fonctionnelle de  $\mathcal{N}\left(x \mid x_0, -\frac{1}{g''(x_0)}\right)$ , où  $\mathcal{N}(\cdot \mid \eta, \nu^2)$  correspond à la densité de la loi normale de moyenne  $\eta$  et de variance  $\nu^2$ .

Cette méthode s'applique également aux densités multidimensionnelles. Soient  $\mathbf{x}$  et  $\mathbf{x}_0$  les vecteurs équivalents appartenant à l'espace vectoriel quelconque de dimension  $k \geq 2$ . Supposons  $f$  (et donc  $g$ ) multidimensionnelles. Le développement de Taylor s'écrit :

$$g(\mathbf{x}) \approx g(\mathbf{x}_0) + \nabla g(\mathbf{x}_0)^T(\mathbf{x}_0 - \mathbf{x}) + \frac{1}{2}(\mathbf{x}_0 - \mathbf{x})^T H_g(\mathbf{x}_0)(\mathbf{x}_0 - \mathbf{x}), \quad \mathbf{x} \in \mathcal{X}$$

où  $\nabla g(\mathbf{x}_0)$  désigne le gradient de  $g$  au point  $\mathbf{x}_0$  et  $H_g(\mathbf{x}_0)$  désigne la matrice hessienne de  $g$  évaluée en  $\mathbf{x}_0$ .

L'approximation résultante s'écrit alors

$$f(\mathbf{x}) \approx \mathcal{N}_k \left\{ \mathbf{x} \mid \mathbf{x}_0, -H_g(\mathbf{x}_0)^{-1} \right\} \tag{3.13}$$

Cette méthode prend parfois le nom d'*approximation de Laplace* ou d'*approximation normale* dans la littérature. Cependant, afin de ne pas la confondre avec la méthode développée par Tierney *et al.* (1986) — notamment utilisée dans l'algorithme INLA — le terme d'*approximation quadratique* est utilisé dans ce mémoire. De plus, ce nom a l'avantage de

décrire explicitement la procédure employée.

### 3.5 Théorie des valeurs extrêmes

La théorie des valeurs extrêmes est une discipline de la statistique qui étudie le comportement des réalisations rares d'une variable aléatoire. Les premières formalisations de ces techniques remontent au début du XXe siècle (Fisher et Tippett, 1928).

#### 3.5.1 Loi des valeurs extrêmes généralisée

Soit  $X_1, X_2, \dots$  une suite de variables aléatoires indépendantes et identiquement distribuées selon la fonction de répartition inconnue  $F$ . Soit  $Y_n = \max\{X_1, \dots, X_n\}$  le maximum des  $n$  premières variables. La fonction de répartition de  $Y_n$  peut être obtenue par :

$$\begin{aligned}\mathbb{P}\{Y_n \leq z\} &= \mathbb{P}\{X_1 \leq z, \dots, X_n \leq z\} \\ &= \mathbb{P}\{X_1 \leq z\} \times \dots \times \mathbb{P}\{X_n \leq z\} \\ &= \{F(z)\}^n.\end{aligned}$$

Une première idée pour estimer  $\{F(z)\}^n$  est d'estimer  $F$  et de l'élever à la puissance  $n$ . Or cette approche est très sensible aux erreurs d'estimation de  $F$  puisque celles-ci sont exacerbées lors du calcul de  $\{F(z)\}^n$ . Une alternative, préconisée par la théorie des valeurs extrêmes, consiste à approcher directement la loi de  $Y_n$  par une famille de loi qui prend exclusivement en compte les réalisations extrêmes. Dans le cas asymptotique ( $n \rightarrow \infty$ ), le Théorème 3.5.1 exhibe la seule forme de loi possible pour  $Y_n$ . Ce résultat est utilisé pour approcher  $\{F(z)\}^n$  lorsque  $n$  est assez grand.

**Théorème 3.5.1** ((Fisher et Tippett, 1928; Gnedenko, 1943; Jenkinson, 1955)). *S'il existe des suites de constantes  $\{a_n > 0\}$  et  $\{b_n\}$  telles que*

$$\lim_{n \rightarrow \infty} \mathbb{P}\{(Y_n - b_n)/a_n \leq z\} = G(z) \quad (3.14)$$

*pour une fonction de répartition non dégénérée  $G$ , alors  $G$  est nécessairement de la forme suivante :*

$$G(z) = \exp \left\{ - (1 + \xi z)^{-1/\xi} \right\}, \quad \xi \in \mathbb{R}$$

*définie sur  $\{z : 1 + \xi z > 0\}$ .*

On dit que  $G$  appartient à la famille des lois des valeurs extrêmes généralisées (GEV, Jenkin-

son, 1955). Selon la valeur du paramètre  $\xi$ , appelé paramètre de forme, la nature des extrêmes est différente (Figure 3.1) :

1. Si  $\xi > 0$ , la queue de la distribution décroît de façon polynomiale (queue lourde).
2. Si  $\xi = 0$ , la queue de la distribution décroît exponentiellement (queue légère).
3. Si  $\xi < 0$ , la queue de la distribution est bornée à droite (queue finie).

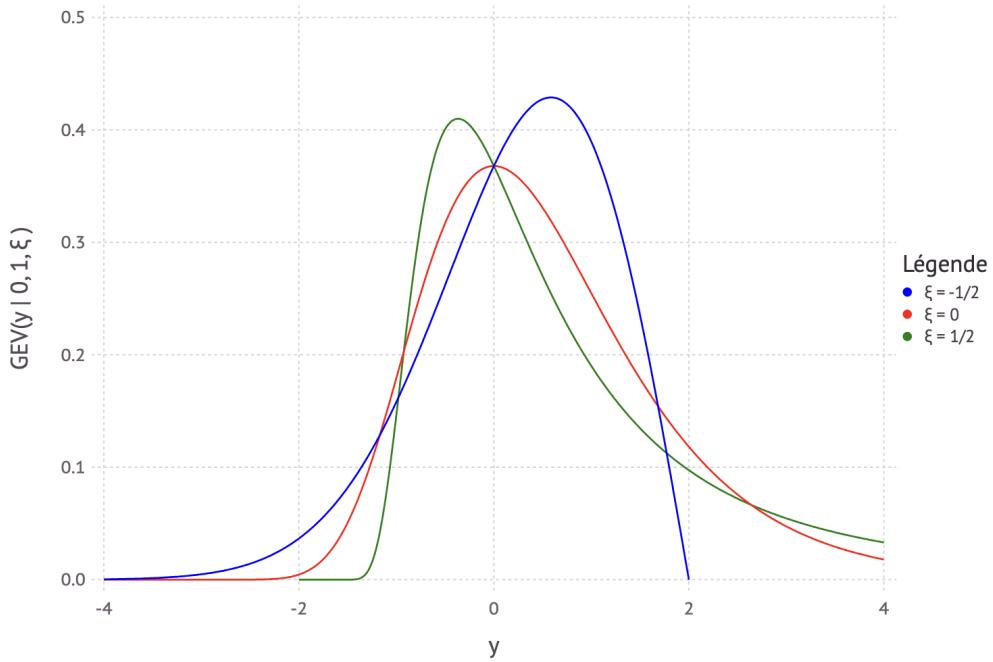


FIGURE 3.1 Illustration de l'influence du paramètre  $\xi$  sur la loi GEV.

### 3.5.2 Méthode des maxima par blocs

Pour une valeur de  $n$  assez grande, on peut, en utilisant l'équation (3.14), obtenir l'approximation suivante

$$\mathbb{P}\{(Y_n - b_n)/a_n \leq z\} \approx G(z) \quad (3.15)$$

donc de manière équivalente

$$\begin{aligned} \mathbb{P}\{Y_n \leq z\} &\approx G\{(z - b_n)/a_n\} \\ &= G^*(z) \end{aligned}$$

où  $G^*$  est un membre dit non standardisé de la famille GEV, dont la forme est la suivante (Coles, 2001)

$$G^*(z) = \exp \left\{ - \left[ 1 + \xi \left( \frac{z - \mu}{\sigma} \right) \right]^{-1/\xi} \right\}, \quad \mu \in \mathbb{R}, \sigma \in \mathbb{R}_+, \xi \in \mathbb{R},$$

définie sur  $\{z : 1 + \xi(z - \mu)/\sigma > 0\}$ .

On dit alors que  $Y_n$  suit une loi GEV de paramètres  $\mu$ ,  $\sigma$ , et  $\xi$  et on note  $Y_n \sim \text{GEV}(\mu, \sigma, \xi)$ . Le paramètre de localisation  $\mu$  représente la valeur typique des extrêmes observés. Plus  $\mu$  augmente, plus les extrêmes ont tendance à être intenses. Le paramètre d'échelle  $\sigma$  se comporte de manière similaire à une variance : il mesure la dispersion des extrêmes autour du paramètre de localisation.

Pour effectuer l'inférence des paramètres d'une loi GEV à partir d'une réalisation de  $n$  variables aléatoires i.i.d.  $x_1, \dots, x_n$ , une méthode consiste à partitionner les observations en  $m$  groupes de taille égale et d'en extraire pour chacun la valeur maximale. La série de maxima résultante est alors utilisée comme échantillon pour estimer  $\mu$ ,  $\sigma$  et  $\xi$ . C'est la méthode des maxima par blocs (Coles, 2001).

La taille des blocs joue un rôle important dans l'inférence. Si elle est trop grande, alors trop peu de données sont disponibles pour l'ajustement, rendant la variance de l'estimation très élevée. Inversement, si elle est trop faible, alors les valeurs utilisées ne sont pas suffisamment extrêmes si bien que la loi GEV n'est pas appropriée pour décrire leur répartition, et l'estimation est donc biaisée. Il s'agit alors de trouver un compromis entre biais et variance.

Dans l'étude des précipitations journalières extrêmes, comme pour les extrêmes climatiques en général, Coles (2001, Section 3.3.1) recommande de choisir des blocs d'un an. En effet, des blocs de taille inférieure fourniraient des variables aléatoires soumises à l'influence de la saisonnalité, violant l'hypothèse de variables i.i.d., tandis qu'une taille plus élevée (plusieurs années à la fois) diminuerait considérablement le nombre de blocs. La Figure 3.2 montre un exemple d'inférence sur des données réelles.

### 3.6 Champs markoviens gaussiens intrinsèques de premier ordre sur une grille régulière

Soit une grille 2D régulière finie de taille  $m_1 \times m_2 = m$  pour laquelle chaque cellule  $i \in \{1, \dots, m\}$  renferme une variable aléatoire  $\theta_i$ . On s'intéresse à la loi du vecteur aléatoire  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)$ .

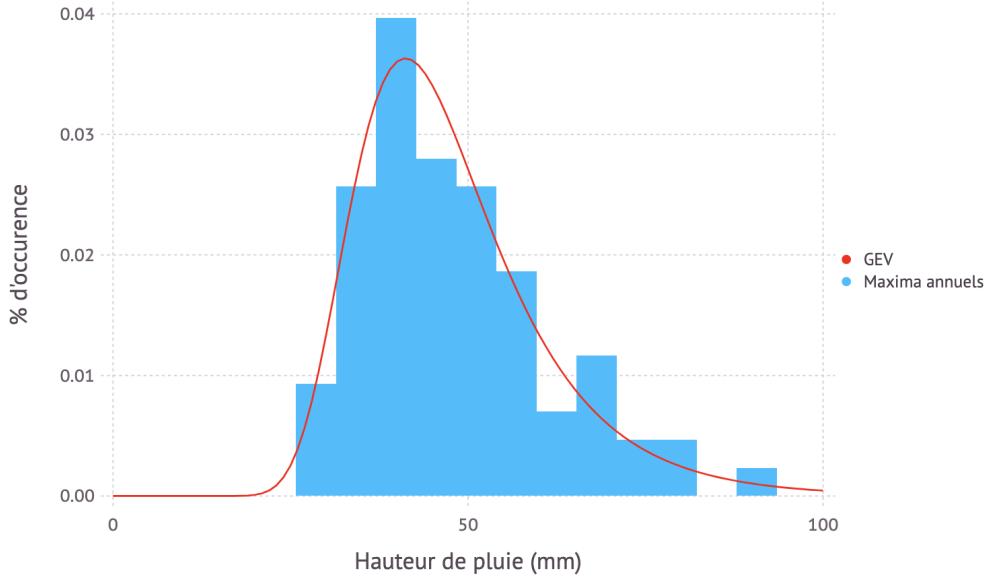


FIGURE 3.2 Ajustement d'une loi GEV aux maxima annuels de précipitations journalières à l'aéroport Montréal-Trudeau de 1941 à 2016 ( $\mu = 41.58$ ,  $\sigma = 10.14$ ,  $\xi = 0.04$ ).

Dans cette grille, la loi conditionnelle d'une variable sachant toutes les autres ne dépend que des cellules voisines. Cette hypothèse dite de Markov s'écrit :

$$\theta_i \mid \boldsymbol{\theta}_{-i} \sim \theta_i \mid (\theta_j)_{j:j \sim i},$$

où la notation  $j \sim i$  désigne le fait que les cellules  $i$  et  $j$  sont voisines. Les cellules voisines correspondent ici aux cellules immédiatement adjacentes dans les directions des points cardinaux (Figure 3.3).

Le champ markovien gaussien intrinsèque (iGMRF, Rue et Held, 2005) de premier ordre respecte cette structure. Dans ce modèle, la loi conditionnelle de la cellule  $i$  sachant ses voisines est la suivante :

$$f_{(\theta_i|\boldsymbol{\theta}_{-i})}(\theta_i) = \mathcal{N}\left(\theta_i \mid \bar{\theta}_i, \frac{1}{n_i \kappa}\right);$$

où  $\mathcal{N}(\cdot \mid \eta, \nu^2)$  correspond à la densité de la loi normale de moyenne  $\eta$  et de variance  $\nu^2$ ,  $\bar{\theta}_i$  correspond à la moyenne des valeurs des cellules adjacentes,  $n_i$  correspond au nombre de cellules adjacentes en considérant les bordures de la grille et  $\kappa$  correspond à la précision du champ. Plus  $\kappa$  est grand, plus le champ est lisse. La variance de la loi conditionnelle dépend donc de la précision et du nombre de cellules adjacentes.

Cet ensemble de lois conditionnelles découle d'une loi conjointe impropre pour l'ensemble des cellules dont la forme fonctionnelle peut s'écrire :

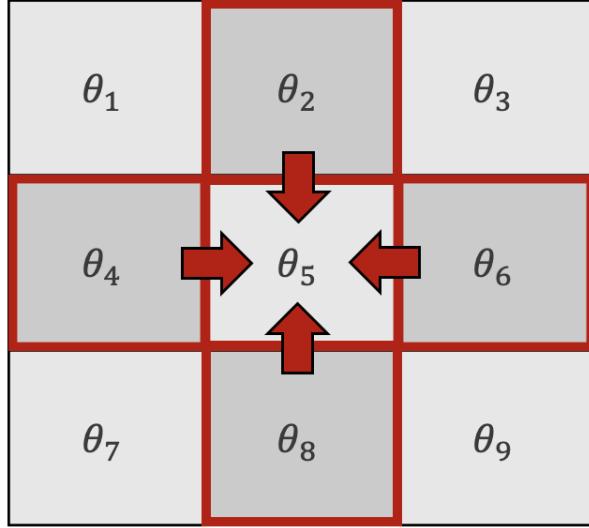


FIGURE 3.3 Schéma de dépendance de la grille ( $m_1 = m_2 = 3$ ). Les cellules voisines de la cellule 5 sont les cellules 2, 4, 6 et 8.

$$f_{\boldsymbol{\theta}}(\boldsymbol{\theta}) \propto \kappa^{\frac{m-1}{2}} \exp \left\{ -\frac{\kappa}{2} \boldsymbol{\theta}^T W \boldsymbol{\theta} \right\}; \quad (3.16)$$

où  $W$  est appelée matrice de structure du iGMRF et est définie par :

$$W_{ij} = \begin{cases} n_i & \text{si } i = j \\ -1 & \text{si } j \sim i \\ 0 & \text{sinon.} \end{cases}$$

On note  $\boldsymbol{\theta} \sim \text{iGMRF}(\kappa)$  la densité d'une telle loi conjointe. L'expression (3.16) est une conséquence du théorème d'Hammersley-Clifford (1971) présenté par Besag (1974).

En outre, la matrice de structure est une matrice creuse puisqu'une cellule donnée ne peut avoir au plus que quatre voisins. Cette propriété permet d'exploiter les calculs numériques peu coûteux spécifiques aux matrices creuses. Pour plus de détails concernant les iGMRFs et leur implémentation numérique, le livre de Rue et Held (2005) peut être consulté.

Un résultat utile pour la suite du mémoire est l'égalité suivante

$$\boldsymbol{\theta}^T W \boldsymbol{\theta} = \sum_{i=1}^m \sum_{j:j \sim i} \theta_i^2 - \theta_i \theta_j. \quad (3.17)$$

## CHAPITRE 4 DONNÉES ET MODÉLISATION

Ce chapitre présente la Réanalyse Canadienne de Surface, pour laquelle les extrêmes des précipitations réanalysées seront étudiés, ainsi que le modèle spatial bayésien hiérarchique proposé par Jalbert *et al.* (2017) pour l'analyse des valeurs extrêmes.

### 4.1 Réanalyse Canadienne de Surface

La Réanalyse Canadienne de Surface (RCaS, Gasset *et al.*, 2021) est une réanalyse nord-américaine des variables météorologiques de surface, qui assimile les observations aux stations de mesure dans les prévisions météorologiques obtenues par le modèle de prévision d'Environnement et Changement climatique Canada. Les précipitations réanalysées sont disponibles à l'échelle horaire sur une grille régulière couvrant l'Amérique du Nord. La résolution horizontale de la grille est d'environ 10 km pour les latitudes canadiennes. Les précipitations ont été réanalysées du 1er janvier 1980 au 31 décembre 2023.

Pour la présente étude, l'analyse se concentre exclusivement sur les précipitations d'un sous-domaine de la grille centré sur la région du Québec (voir figure 4.1). Ce sous-domaine est représenté par une grille de dimensions  $m_1 \times m_2$ , comprenant un total de  $m = m_1 m_2$  cellules. Comme l'analyse porte sur des données journalières, il est nécessaire d'agréger les données horaires sur une période de 24 heures. À partir de ces données journalières, les maxima annuels sont extraits afin de constituer le jeu de données final. En somme, chaque cellule dispose de  $n_{obs} = 44$  maxima annuels. Le maximum annuel pour l'année  $1 \leq j \leq n_{obs}$  et pour la cellule  $1 \leq i \leq m$  est désigné par  $y_{ij}$ .

### 4.2 Modèle bayésien spatial de valeurs extrêmes pour les précipitations sur grille

Un modèle bayésien spatial de valeurs extrêmes pour des données disposées sur une grille régulière a été développé par Jalbert *et al.* (2017). L'utilité du modèle spatial est de partager les informations sur les extrêmes entre les cellules de façon à réduire l'incertitude des paramètres estimés. La modélisation de la dépendance spatiale s'effectue dans la couche latente d'un modèle hiérarchique bayésien. Puisque les données sont sur grille, la couche latente est constituée d'un champ markovien gaussien, qui permet des calculs rapides et précis.

Au niveau de la couche des données du modèle hiérarchique, il est supposé que la variable aléatoire  $Y_{ij}$  décrivant le maximum annuel de l'année  $j$  à la cellule  $i$  est distribuée selon la

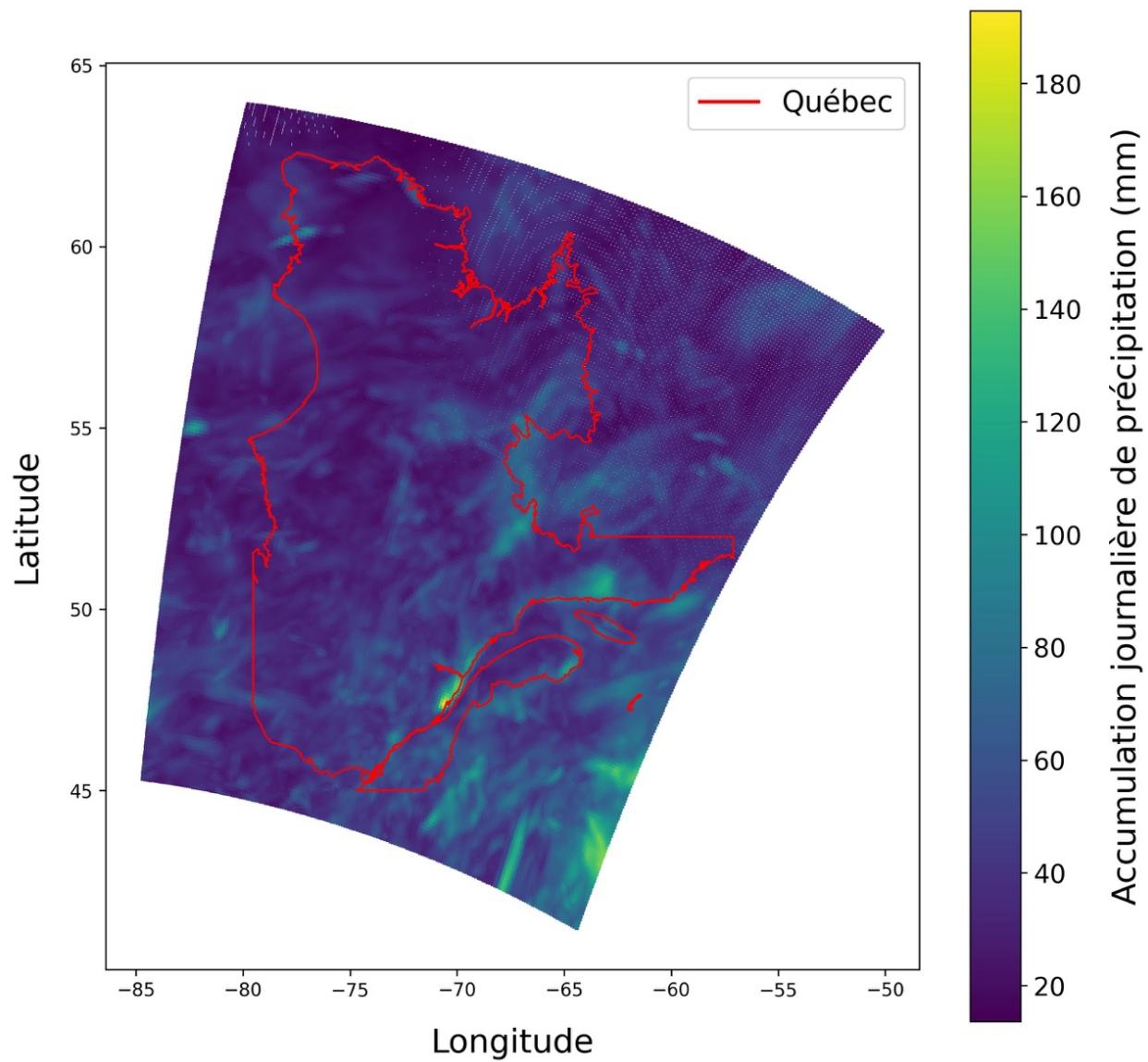


FIGURE 4.1 Sous-domaine régulier centré sur le Québec de la grille RCaS. À titre d'illustration, les maxima annuels des précipitations journalières de RCaS pour l'année 2022 sont présentés.

loi GEV de paramètres  $(\mu_i, \sigma_i, \xi)$  :

$$f_{(Y_{ij}|\mu_i, \sigma_i, \xi)}(y_{ij}) = \mathcal{G}EV(y_{ij} | \mu_i, \sigma_i, \xi),$$

où  $\mathcal{G}EV(y | \mu, \sigma, \xi)$  dénote la densité de la loi GEV de paramètres  $(\mu, \sigma, \xi)$  évaluée à  $y$ . Il convient de noter que les paramètres de localisation et d'échelle de la loi GEV sont dépendants de la cellule  $i$ , pour s'ajuster aux différences météorologiques de chaque cellule, alors que le paramètre de forme demeure le même pour toute la grille. Cette hypothèse imposée par Jalbert *et al.* (2017) suggère que la nature des extrêmes est la même partout sur le territoire (à queue lourde, légère ou bornée). Elle permet d'exploiter l'information de toutes les cellules pour estimer ce paramètre dont l'estimation est particulièrement difficile en pratique.

Pour le vecteur  $\mathbf{Y}_i = (Y_{ij} : 1 \leq j \leq n)$  les  $n$  maxima annuels de la cellule  $i$  sont supposés indépendants :

$$f_{(\mathbf{Y}_i|\mu_i, \sigma_i, \xi)}(\mathbf{y}_i) = \prod_{j=1}^n \text{GEV}(y_{ij} | \mu_i, \sigma_i, \xi).$$

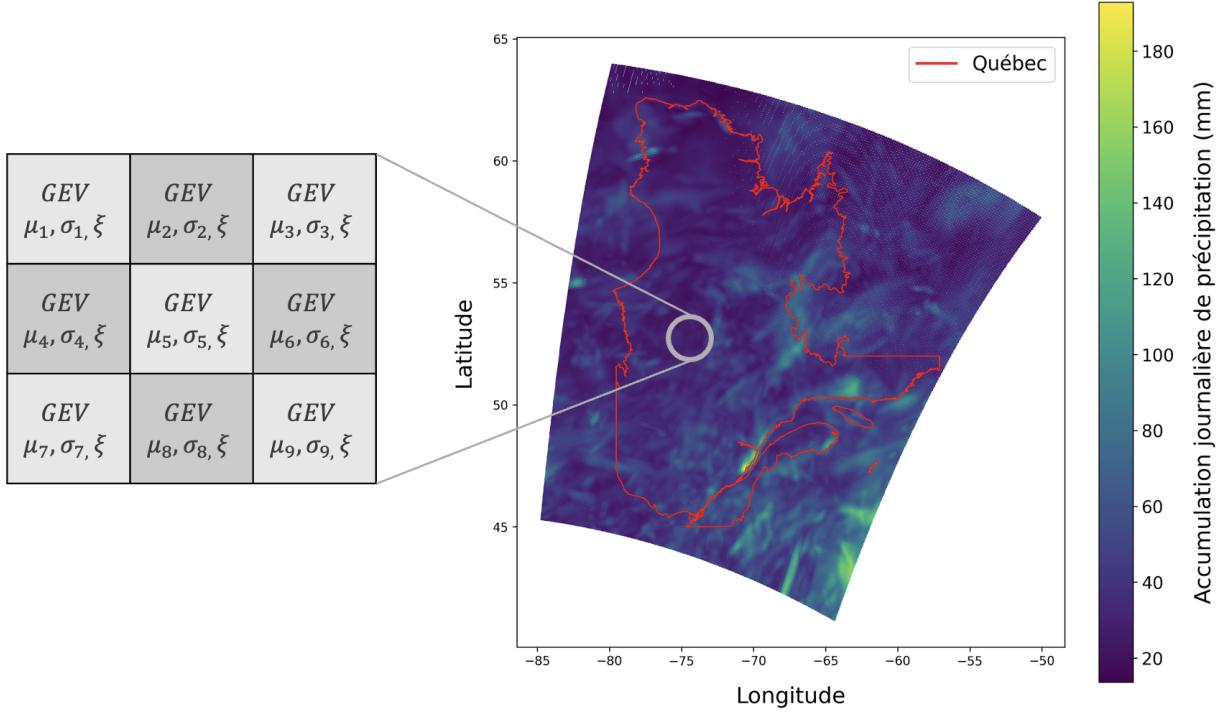


FIGURE 4.2 Schéma de la modélisation du territoire en grille régulière (à gauche). Maxima de précipitations journalières pour l'année 2022 issus de la réanalyse RCaS (à droite).

La dépendance spatiale du vecteur  $\boldsymbol{\mu} = (\mu_i : 1 \leq i \leq m)$  des paramètres de localisation de la loi GEV est modélisée dans la couche latente par un champ markovien gaussien de premier ordre :

$$\boldsymbol{\mu} | \kappa_u \sim \text{iGMRF}(\kappa_u);$$

où  $\kappa_u > 0$  correspond à la précision du champ. De façon équivalente, la dépendance spatiale du vecteur  $\boldsymbol{\phi} = (\phi_i : 1 \leq i \leq m)$ , où  $\phi_i = \log \sigma_i$  correspond au logarithme du paramètre d'échelle de la loi GEV, est aussi modélisée par un champ markovien gaussien de premier ordre :

$$\boldsymbol{\phi} | \kappa_v \sim \text{iGMRF}(\kappa_v);$$

où  $\kappa_v > 0$ .

Finalement, la couche des lois *a priori* du modèle hiérarchique bayésien modélise la connais-

sance *a priori* des paramètres  $(\kappa_u, \kappa_v, \xi)$  du modèle. Pour les précisions  $\kappa_u$  et  $\kappa_v$  des champs markoviens, des lois faiblement informatives  $Gamma(1, 1/100)$  sont utilisées. Celles-ci portent de l'information (elles ne sont pas uniformes sur le support de loi) mais cette information n'est pas significative sur la loi *a posteriori*. Une loi informative pour le paramètre de forme de la loi GEV pour les précipitations extrêmes proposée par Martins et Stedinger (2000) est utilisée (figure 4.3). La densité de la loi *a priori* est donc la suivante :

$$f_{(\kappa_u, \kappa_v, \xi)}(\kappa_u, \kappa_v, \xi) = Gamma\left(\kappa_u \middle| 1, \frac{1}{100}\right) \times Gamma\left(\kappa_v \middle| 1, \frac{1}{100}\right) \times Beta\left(\xi + \frac{1}{2} \middle| 9, 6\right);$$

où  $Gamma(y | \alpha, \beta)$  et  $Beta(y | \alpha, \beta)$  dénotent respectivement la densité de la loi gamma et de la loi bêta de paramètres  $(\alpha, \beta)$  évaluée à  $y$ .

En assemblant la couche des données, la couche latente et la couche des lois *a priori* (figure 4.4), la forme fonctionnelle de la loi *a posteriori* du modèle s'écrit ainsi :

$$\begin{aligned} f_{(\mu, \phi, \xi, \kappa_u, \kappa_v) | \mathbf{Y} = \mathbf{y}}(\boldsymbol{\mu}, \boldsymbol{\phi}, \xi, \kappa_u, \kappa_v) & \quad (4.1) \\ & \propto \left\{ \prod_{i=1}^m \prod_{j=1}^n \mathcal{GEV}(y_{ij} | \mu_i, \exp \phi_i, \xi) \right\} \\ & \quad \times \kappa_u^{\frac{m-1}{2}} \exp\left(-\kappa_u \frac{\boldsymbol{\mu}^\top W \boldsymbol{\mu}}{2}\right) \times \kappa_v^{\frac{m-1}{2}} \exp\left(-\kappa_v \frac{\boldsymbol{\phi}^\top W \boldsymbol{\phi}}{2}\right) \\ & \quad \times Gamma\left(\kappa_u \middle| 1, \frac{1}{100}\right) \times Gamma\left(\kappa_v \middle| 1, \frac{1}{100}\right) \times Beta\left(\xi + \frac{1}{2} \middle| 9, 6\right), \end{aligned}$$

où  $W$  désigne la matrice de structure des iGMRFs relatifs à  $\boldsymbol{\mu}$  et  $\boldsymbol{\phi}$ . Celle-ci possède la même forme pour les deux vecteurs de paramètres.

Pour alléger le texte, on écrit parfois  $\mathcal{GEV}(y_{ij} | \mu_i, \phi_i, \xi)$  pour désigner  $\mathcal{GEV}(y_{ij} | \mu_i, \exp \phi_i, \xi)$ .

La constante de normalisation de la loi (4.1) ne s'exprime pas sous une forme analytique et est très difficile, voire impossible, à estimer. C'est pourquoi Jalbert *et al.* (2017) ont utilisé une méthode de Monte-Carlo par chaînes de Markov pour tirer un échantillon aléatoire de cette loi. Ce travail propose une approche différente. Au lieu d'avoir recours au MCMC, une famille d'approximations analytiques plus simples de la loi *a posteriori* est développée, et le membre de cette famille qui l'approxime le mieux est identifié en minimisant la divergence de Kullback-Leibler.

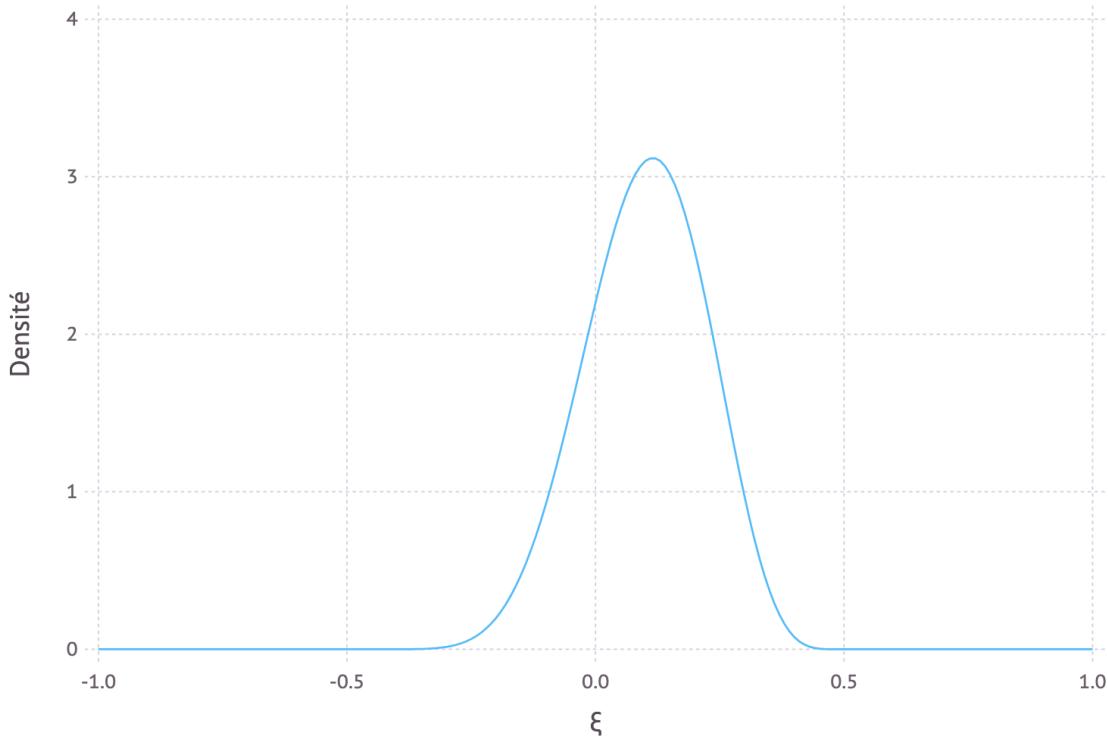


FIGURE 4.3 Densité de la loi a priori du paramètre  $\xi$  définie par  $f_\xi(\xi) = \text{Beta}(\xi + .5|9, 6)$ .

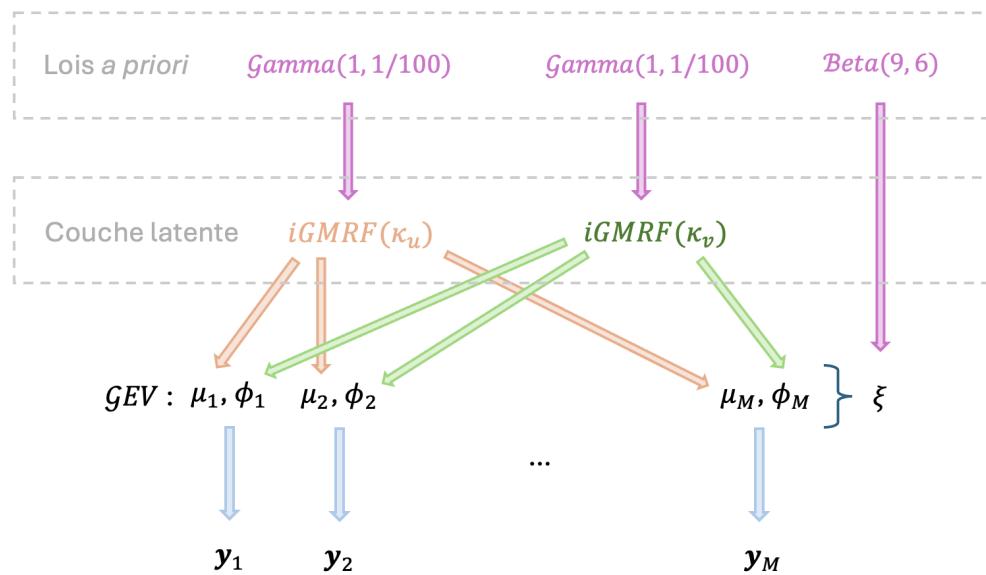


FIGURE 4.4 Schéma représentant les relations entre les paramètres du modèle hiérarchique bayésien spatial pour les précipitations extrêmes.

## CHAPITRE 5 APPROXIMATIONS QUADRATIQUES EN SÉRIE POUR L'INFÉRENCE VARIATIONNELLE

Ce chapitre présente la procédure variationnelle nommée *Serial Quadratic Approximations for Variational Inference* (SQAVI), développée pour approximer la loi *a posteriori* du modèle spatial bayésien de valeurs extrêmes présenté au chapitre 4. Elle consiste en un algorithme CAVI (c.f. section 3.3) dont les lois  $q^*$  indéterminées sont approchées par des lois normales. De cette façon, elle se sert du caractère quasi-normal imposé par les lois *a priori* iGMRFs pour contourner la difficulté d'estimer les termes de vraisemblance de la loi *a posteriori*. De plus, la famille de lois d'approximation est conçue de sorte à conserver la structure de dépendance entre les paramètres de la loi GEV dans chaque cellule, essentielle à l'estimation des quantiles extrêmes.

### 5.1 La famille de lois d'approximation

Tel qu'expliqué en section 2.1, la conception d'une méthode variationnelle débute avec le choix d'une famille de lois d'approximation  $\mathcal{Q}$ . Ensuite, on identifie le meilleur membre de cette famille au sens du problème d'optimisation décrit par l'équation (3.8). La famille  $\mathcal{Q}$  doit être suffisamment simple pour faciliter les calculs et l'interprétation, mais elle doit être suffisamment complexe pour conserver les éléments structurels importants de la loi cible. Sa forme finale est donc le fruit de compromis entre simplicité et précision.

Pour le modèle considéré (section 4.2), plusieurs éléments structurels sont identifiés :

1. Les paramètres de localisation  $\mu_i$  et de log-échelle  $\phi_i$  de la loi GEV au sein d'une même cellule  $i$  sont dépendants et doivent le demeurer dans la loi d'approximation. On impose donc une loi conjointe pour ces deux paramètres  $q_{\{(\mu_i, \phi_i)|\lambda_i\}}(\mu_i, \phi_i)$ , qui permet de maintenir cette dépendance.
2. Le traitement du paramètre de forme  $\xi$  de la loi GEV est plus délicat puisqu'il agit à la fois à l'échelle locale au niveau des données et à l'échelle globale comme paramètre commun à toutes les cellules de la grille. Or, à mesure que la taille de grille augmente, la corrélation de  $\xi$  avec les paramètres de cellule  $(\mu_i, \phi_i)_{i=1,\dots,M}$  se réduit (Annexe A). Ce comportement suggère qu'il devrait être considéré au même titre que les paramètres de précision, *i.e.* ayant une loi marginale d'approximation indépendante des autres paramètres.
3. Les paramètres de précisions  $\kappa_u$  et  $\kappa_v$ , qui représentent l'intensité de la dépendance

dans la structure iGMRF, sont globaux et mutuellement indépendants des autres paramètres. Ils possèdent donc chacun leur propre loi d'approximation, respectivement  $q_{(\kappa_u|\lambda_u)}(\kappa_u)$  et  $q_{(\kappa_v|\lambda_v)}(\kappa_v)$ .

En somme, la loi d'approximation  $q \in \mathcal{Q}$  est de la forme suivante :

$$q_{(\boldsymbol{\theta}|\boldsymbol{\lambda})}(\boldsymbol{\theta}) = \prod_{i=1}^m q_{\{(\mu_i, \phi_i)|\boldsymbol{\lambda}_i\}}(\mu_i, \phi_i) q_{(\xi|\boldsymbol{\lambda}_\xi)}(\xi) q_{(\kappa_u|\boldsymbol{\lambda}_u)}(\kappa_u) q_{(\kappa_v|\boldsymbol{\lambda}_v)}(\kappa_v), \quad (5.1)$$

avec les paramètres  $\boldsymbol{\theta} = (\mu_1, \phi_1, \dots, \mu_m, \phi_m, \xi, \kappa_u, \kappa_v)$  et les paramètres variationnels  $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_m, \boldsymbol{\lambda}_\xi, \boldsymbol{\lambda}_u, \boldsymbol{\lambda}_v)$ .

La loi d'approximation conjointe qui minimise la divergence KL avec la loi *a posteriori*, notée  $q_{(\boldsymbol{\theta}|\boldsymbol{\lambda})}^{min}$ , s'écrit donc

$$q_{(\boldsymbol{\theta}|\boldsymbol{\lambda})}^{min}(\boldsymbol{\theta}) = q_{(\boldsymbol{\theta}|\boldsymbol{\lambda}^{min})}(\boldsymbol{\theta}) = \prod_{i=1}^m q_{\{(\mu_i, \phi_i)|\boldsymbol{\lambda}_i^{min}\}}(\mu_i, \phi_i) q_{(\xi|\boldsymbol{\lambda}_\xi^{min})}(\xi) q_{(\kappa_u|\boldsymbol{\lambda}_u^{min})}(\kappa_u) q_{(\kappa_v|\boldsymbol{\lambda}_v^{min})}(\kappa_v). \quad (5.2)$$

Chacun des facteurs de ce produit peut être approché par une mise à jour successive des lois marginales  $q^*$  (3.10). Cependant, les lois  $q^*$  ne correspondent pas toujours à des lois explicites. Des approximations supplémentaires sont donc nécessaires. La loi décrite dans l'équation (5.2) correspondant à un produit de lois marginales, on peut traiter chacun des facteurs de manière indépendante. Les sections suivantes détaillent l'identification des membres qui minimisent la divergence KL pour chacune des marginales.

## 5.2 Loi $q_{\{(\mu_i, \phi_i)|\boldsymbol{\lambda}_i\}}^*(\mu_i, \phi_i)$

Dans l'approche CAVI, l'équation (3.10) permet d'obtenir la loi conjointe optimale des paramètres de localisation et d'échelle de la loi GEV à la cellule  $i$  :

$$q_{\{(\mu_i, \phi_i)|\boldsymbol{\lambda}_i\}}^*(\mu_i, \phi_i) \propto \exp \left[ \mathbb{E}_{\setminus(\mu_i, \phi_i)} \left\{ \log f_{\{(\mu_i, \phi_i)|Y=y, \boldsymbol{\mu}_{-i}, \boldsymbol{\phi}_{-i}, \xi, \kappa_u, \kappa_v\}}(\mu_i, \phi_i) \right\} \right] \quad (5.3)$$

où pour tout vecteur  $\boldsymbol{\theta}$ ,  $\boldsymbol{\theta}_{-i}$  désigne le vecteur  $\theta$  privé de sa composante  $i$  et  $\mathbb{E}_{\setminus(\mu_i, \phi_i)}$  désigne l'espérance selon la loi d'approximation de tous les paramètres excepté  $(\mu_i, \phi_i)$ , *i.e.* selon

$$\prod_{l \neq i} q_{\{(\mu_l, \phi_l)|\hat{\boldsymbol{\lambda}}_l\}}^*(\mu_l, \phi_l) q_{(\xi|\hat{\boldsymbol{\lambda}}_\xi)}^*(\xi) q_{(\kappa_u|\hat{\boldsymbol{\lambda}}_u)}^*(\kappa_u) q_{(\kappa_v|\hat{\boldsymbol{\lambda}}_v)}^*(\kappa_v).$$

L'expression de la loi conditionnelle complète relative à  $(\mu_i, \phi_i)$  peut être obtenue à partir de l'équation (4.1) en ne gardant que les termes faisant intervenir ces deux variables

$$\begin{aligned} f_{\{(\mu_i, \phi_i) | Y=y, \boldsymbol{\mu}_{-i}, \boldsymbol{\phi}_{-i}, \xi, \kappa_u, \kappa_v\}}(\mu_i, \phi_i) \\ \propto \prod_{j=1}^{n_{obs}} \mathcal{GEV}(y_{ij} | \mu_i, \phi_i, \xi) \mathcal{N}(\mu_i | \bar{\mu}_i, 1/\kappa_u n_i) \mathcal{N}(\phi_i | \bar{\phi}_i, 1/\kappa_v n_i), \end{aligned} \quad (5.4)$$

où

$$\begin{aligned} \bar{\mu}_i &= \frac{1}{n_i} \sum_{j:j \sim i} \mu_j, \\ \bar{\phi}_i &= \frac{1}{n_i} \sum_{j:j \sim i} \phi_j. \end{aligned}$$

Ainsi, en injectant (5.4) dans (5.3),

$$\begin{aligned} q_{\{(\mu_i, \phi_i) | \boldsymbol{\lambda}_i\}}^*(\mu_i, \phi_i) \propto \exp \left[ \mathbb{E}_{\setminus(\mu_i, \phi_i)} \left\{ \sum_{j=1}^{n_{obs}} \log \mathcal{GEV}(y_{ij} | \mu_i, \phi_i, \xi) \right. \right. \\ \left. \left. + \log \mathcal{N}(\mu_i | \bar{\mu}_i, 1/\kappa_u n_i) + \log \mathcal{N}(\phi_i | \bar{\phi}_i, 1/\kappa_v n_i) \right\} \right]. \end{aligned} \quad (5.5)$$

La forme (5.5) ne fait pas apparaître de loi paramétrique simple, en particulier à cause de l'opérateur  $\mathbb{E}_{\setminus(\mu_i, \phi_i)}$ . Pour simplifier, l'approximation suivante est utilisée :

$$\begin{aligned} \mathbb{E}_{\setminus(\mu_i, \phi_i)} [\log f_{\{(\mu_i, \phi_i) | Y=y, \boldsymbol{\mu}_{-i}, \boldsymbol{\phi}_{-i}, \xi, \kappa_u, \kappa_v\}}(\mu_i, \phi_i)] &\approx \log f_{\{(\mu_i, \phi_i) | y, \hat{\boldsymbol{\mu}}_{-i}, \hat{\boldsymbol{\phi}}_{-i}, \hat{\xi}, \hat{\kappa}_u, \hat{\kappa}_v\}}(\mu_i, \phi_i) \\ &\approx \log \hat{f}_{\{(\mu_i, \phi_i) | Y=y, -\}}, \end{aligned} \quad (5.6)$$

où pour tout paramètre  $\theta$ , vecteur ou scalaire,  $\hat{\theta} = \mathbb{E}_\theta[\theta]$  désigne l'espérance de  $\theta$  selon sa loi marginale d'approximation.

L'équation (5.6) est exacte et vérifiée si  $\log f_{\{(\mu_i, \phi_i) | \boldsymbol{Y}=y, \boldsymbol{\mu}_{-i}, \boldsymbol{\phi}_{-i}, \xi, \kappa_u, \kappa_v\}}(\mu_i, \phi_i)$  est une fonction linéaire en  $(\boldsymbol{\mu}_{-i}, \boldsymbol{\phi}_{-i}, \xi, \kappa_u, \kappa_v)$ , c'est pourquoi elle est nommée *hypothèse de linéarité des variables latentes*. Dans le modèle considéré, lorsque  $\xi = 0$ , cette hypothèse de linéarité est satisfaite. Pour  $\xi \approx 0$ , ce qui est souvent le cas en extrêmes climatiques, cette hypothèse est raisonnable (voir Annexe B pour une justification théorique).

En utilisant cette hypothèse de linéarité dans l'équation (5.5), on obtient

$$q_{\{(\mu_i, \phi_i) | \boldsymbol{\lambda}_i\}}^*(\mu_i, \phi_i) \propto \mathcal{N}(\mu_i | \hat{\mu}_i, 1/\hat{\kappa}_u n_i) \times \mathcal{N}(\phi_i | \hat{\phi}_i, 1/\hat{\kappa}_v n_i) \times \prod_{j=1}^{n_{obs}} \text{GEV}(y_{ij} | \mu_i, \phi_i, \hat{\xi}).$$

Cette expression correspond à un produit entre une loi normale pour chacun des paramètres de localisation et de log-échelle de la cellule et la vraisemblance GEV. Ce dernier terme constitue la source de la dépendance intra-cellule entre les deux paramètres  $\mu_i$  et  $\phi_i$ . Sa formulation matricielle s'écrit

$$q_{\{(\mu_i, \phi_i) | \boldsymbol{\lambda}_i\}}^*(\mu_i, \phi_i) \propto \exp \left\{ -\frac{1}{2} \begin{bmatrix} \mu_i - \hat{\mu}_i \\ \phi_i - \hat{\phi}_i \end{bmatrix}^T A_i \begin{bmatrix} \mu_i - \hat{\mu}_i \\ \phi_i - \hat{\phi}_i \end{bmatrix} + \sum_{j=1}^{n_{obs}} \log \text{GEV}(y_{ij} | \mu_i, \phi_i, \hat{\xi}) \right\}, \quad (5.7)$$

et permet de reconnaître dans l'exponentielle un terme de loi normale bidimensionnelle de variance  $A_i^{-1}$  avec  $A_i = \text{diag}(n_i \hat{\kappa}_u, n_i \hat{\kappa}_v)$  et un terme de perturbation issu de la vraisemblance des données.

La forme fonctionnelle de  $q_{\{(\mu_i, \phi_i) | \boldsymbol{\lambda}_i\}}^*(\mu_i, \phi_i)$  ne s'exprime pas sous une forme analytique. Alors, nous proposons d'utiliser l'approximation quadratique afin d'approcher cette loi par une loi normale bidimensionnelle de paramètres  $(\boldsymbol{\eta}_i, \Sigma_i)$ . Ces paramètres sont trouvés numériquement en utilisant l'équation (3.13). On obtient donc :

$$q_{\{(\mu_i, \phi_i) | \boldsymbol{\lambda}_i\}}^*(\mu_i, \phi_i) \approx \mathcal{N}_2 \{(\mu_i, \sigma_i) | \boldsymbol{\eta}_i, \Sigma_i\}.$$

Cette approximation suit la forme quasi-normale révélée par l'expression (5.7) en intégrant le terme de perturbation dans le vecteur moyenne  $\boldsymbol{\eta}_i$  et dans la matrice de covariance  $\Sigma_i$ . De cette manière, la loi conditionnelle complète de chaque paramètre pris séparément reste normale, et la relation de dépendance entre les deux est contenue dans la matrice  $\Sigma_i$ .

Finalement,  $\boldsymbol{\lambda}_i^* = (\boldsymbol{\eta}_i, \Sigma_i)$ .

### 5.3 Loi $q_{(\xi | \boldsymbol{\lambda}_\xi)}^*(\xi)$

Pour étudier la loi  $q_\xi^*$ , il est commode de considérer le paramètre d'échelle  $\sigma_i$  plutôt que  $\phi_i$ . Pour rappel,  $\sigma_i = \exp \phi_i$ . Or pour  $1 \leq i \leq m$ ,  $\phi_i \sim \mathcal{N}(\eta_{\phi_i}, \nu_{\phi_i}^2)$  ce qui implique  $\sigma_i \sim \log \mathcal{N}(\eta_{\phi_i}, \nu_{\phi_i}^2)$ . Donc, l'espérance d'une loi log-normale étant connue,

$$\hat{\sigma}_i = \exp \left( \mathbb{E}_{\phi_i}[\phi_i] + \frac{\mathbb{V}_{\phi_i}[\phi_i]}{2} \right).$$

L'équation (3.10) conduit à

$$q_{(\xi|\lambda_\xi)}^*(\xi) \propto \exp [\mathbb{E}_{\lambda_\xi} \{\log f_{(\xi|Y=y,\mu,\phi,\kappa_u,\kappa_v)}(\xi)\}], \quad (5.8)$$

où la forme fonctionnelle de la loi conditionnelle complète de  $\xi$  peut être obtenue via l'équation (4.1) en ne gardant que les termes faisant intervenir la variable  $\xi$

$$f_{(\xi|Y=y,\mu,\phi,\kappa_u,\kappa_v)}(\xi) \propto \prod_{i=1}^m \prod_{j=1}^{n_{obs}} \text{GEV}(y_{ij}|\mu_i, \sigma_i, \xi) \text{Beta}(\xi + .5|9, 6). \quad (5.9)$$

Ainsi, en injectant (5.9) dans (5.8),

$$q_{(\xi|\lambda_\xi)}^*(\xi) \propto \exp \left[ \mathbb{E}_{\lambda_\xi} \left\{ \sum_{i=1}^m \sum_{j=1}^{n_{obs}} \log \text{GEV}(y_{ij}|\mu_i, \sigma_i, \xi) + \log \text{Beta}(\xi + .5|9, 6) \right\} \right].$$

Comme pour  $q_{\{(\mu_i, \phi_i)|\lambda_i\}}^*$ , cette forme ne fait pas apparaître de loi paramétrique simple. On applique donc l'hypothèse de linéarité des variables latentes présentée en section 5.2

$$\begin{aligned} \mathbb{E}_{\lambda_\xi} [\log f_{(\xi|y,\mu,\phi,\kappa_u,\kappa_v)}(\xi)] &\approx \log f_{(\xi|y,\hat{\mu},\hat{\phi},\hat{\kappa}_u,\hat{\kappa}_v)}(\xi) \\ &\approx \log \hat{f}_{\{\xi|Y=y,\dots\}}(\xi). \end{aligned} \quad (5.10)$$

Plus précisément, on peut écrire

$$\begin{aligned} \log \hat{f}_{\{\xi|Y=y,\dots\}}(\xi) &= \sum_{i=1}^m \sum_{j=1}^{n_{obs}} \log \left\{ \left( 1 + \xi \frac{y_{ij} - \hat{\mu}_i}{\hat{\sigma}_i} \right)^{-1-1/\xi} \right\} - \left( 1 + \xi \frac{y_{ij} - \hat{\mu}_i}{\hat{\sigma}_i} \right)^{-1/\xi} \\ &\quad + \log \{(\xi + .5)^5 (.5 - \xi)^8\} + cste. \end{aligned}$$

La fonction  $\hat{f}_{\{\xi|y,\dots\}}(\xi)$  ne faisant pas apparaître de loi connue, on lui applique l'approximation quadratique de la Section 3.4 i.e.

$$\hat{f}_{\{\xi|Y=y,\dots\}}(\xi) \approx \mathcal{N}_2 \{\xi|\eta_e, \sigma_e^2\}.$$

Finalement,

$$q_{(\xi|\lambda_\xi)}^*(\xi) \approx \mathcal{N}_2 \{\xi|\eta_e, \sigma_e^2\}, \quad (5.11)$$

et par identification  $\lambda_\xi^* = (\eta_e, \sigma_e^2)$ .

Contrairement à  $(\mu_i, \phi_i)$ , l'application de l'approximation de l'équation (5.10) n'est pas justifiée théoriquement. Elle pourrait donc conduire à des imprécisions dans l'estimation de la

loi *a posteriori* marginale pour  $\xi$ .

#### 5.4 Lois $q_{(\kappa_u|\boldsymbol{\lambda}_u)}^*(\kappa_u)$ et $q_{(\kappa_v|\boldsymbol{\lambda}_v)}^*(\kappa_v)$

L'équation (3.10) fait directement apparaître des lois explicites simples, à savoir

$$q_{(\kappa_u|\boldsymbol{\lambda}_u)}^*(\kappa_u) = \text{Gamma}(\kappa_u|a, b_u) \quad (5.12)$$

$$q_{(\kappa_v|\boldsymbol{\lambda}_v)}^*(\kappa_v) = \text{Gamma}(\kappa_v|a, b_v). \quad (5.13)$$

Donc par identification  $\boldsymbol{\lambda}_u^* = (a, b_u)$  et  $\boldsymbol{\lambda}_v^* = (a, b_v)$  avec

$$a = \frac{m-1}{2} + 1, \quad (5.14)$$

$$b_u = \frac{\mathbb{E}_{\boldsymbol{\mu}}[\boldsymbol{\mu}^T W \boldsymbol{\mu}]}{2} + \frac{1}{100}, \quad (5.14)$$

$$b_v = \frac{\mathbb{E}_{\boldsymbol{\phi}}[\boldsymbol{\phi}^T W \boldsymbol{\phi}]}{2} + \frac{1}{100}. \quad (5.15)$$

Le détail des calculs est donné en Annexe C.

#### 5.5 Implémentation de l'algorithme

L'ensemble des approximations précédentes donne la loi  $q^*$  suivante

$$q_{(\boldsymbol{\theta}|\boldsymbol{\lambda})}^*(\boldsymbol{\theta}) = \text{Gamma}(\kappa_u|a, b_u) \text{Gamma}(\kappa_v|a, b_v) \mathcal{N}(\xi|\eta_e, \sigma_e^2) \prod_{i=1}^m \mathcal{N}_2((\mu_i, \phi_i)|\boldsymbol{\eta}_i, \Sigma_i). \quad (5.16)$$

#### Initialisation

Soit  $\boldsymbol{\lambda}^{(0)} = [\boldsymbol{\eta}_1^{(0)}, \dots, \boldsymbol{\eta}_m^{(0)}, \eta_e^{(0)}, b_u^{(0)}, b_v^{(0)}]^\top$  le vecteur des valeurs initiales des paramètres variationnels. Comme tout algorithme d'optimisation, ces valeurs peuvent avoir une influence sur la convergence de l'algorithme.

Typiquement, les  $\boldsymbol{\eta}_i^{(0)}$  ( $1 \leq i \leq m$ ) et  $\eta_e^{(0)}$  sont déterminés par les estimations de maximum de vraisemblance des paramètres qui leur sont associées, tandis que  $b_u^{(0)}$  et  $b_v^{(0)}$  sont fixés de sorte que  $\mathbb{E}_{\{\kappa_u|a, b_u^{(0)}\}} [\kappa_u] = \mathbb{E}_{\{\kappa_v|a, b_v^{(0)}\}} [\kappa_v] = 1$ .

#### Itérations

Pour chaque itération  $1 \leq \ell \leq L$ , les étapes suivantes sont effectuées :

1. Pour chaque cellule  $i$ , on calcule  $\hat{f}_{\{(\mu_i, \phi_i)|y, \dots\}}$  à partir de  $\eta_e^{(\ell-1)}$ ,  $b_u^{(\ell-1)}$ ,  $b_v^{(\ell-1)}$  et  $\boldsymbol{\eta}_j^{(\ell-1)}$  où  $j : j \sim i$ . Puis on en déduit  $\boldsymbol{\eta}_i^{(\ell)}$  et  $\Sigma_i^{(\ell)}$ , les paramètres variationnels relatifs à  $(\mu_i, \phi_i)$ .
2. On utilise les  $\boldsymbol{\eta}_i^{(\ell)}$ ,  $1 \leq i \leq m$  pour calculer  $\hat{f}_{(\xi| \dots)}$  et en déduire  $\eta_e^{(\ell)}$  et  $\sigma_e^{2(\ell)}$  via l'approximation quadratique.
3. On utilise  $\boldsymbol{\eta}_i^{(\ell)}$ ,  $1 \leq i \leq m$  pour calculer  $b_u^{(\ell)}$  via (5.14) et  $b_v^{(\ell)}$  via (5.15).

Ces mises à jour des paramètres variationnels modifient successivement les lois marginales d'approximation, pour la rapprocher de la vraie loi *a posteriori*.

### Critère de convergence :

Toutes les  $\tau$  itérations, on effectue un calcul de la divergence de Kullback-Leibler entre l'approximation et la loi cible, en considérant que  $\boldsymbol{\lambda}^* = \boldsymbol{\lambda}^{(\ell)}$  où  $\ell$  est la dernière itération effectuée. Puisque la divergence ne se calcule pas de façon analytique dans notre cas, on l'estime par une méthode Monte-Carlo. Pour ce faire, on génère  $K$  échantillons de la loi d'approximation  $q_{(\boldsymbol{\theta}|\boldsymbol{\lambda}^{(\ell)})}$ , puis on estime la divergence :

$$KL(q_{(\boldsymbol{\theta}|\boldsymbol{\lambda}^{(\ell)})} || f_{(\boldsymbol{\theta}|Y=y)}) = \int q_{(\boldsymbol{\theta}|\boldsymbol{\lambda}^{(\ell)})}(\boldsymbol{\theta}) \log \frac{q_{(\boldsymbol{\theta}|\boldsymbol{\lambda}^{(\ell)})}(\boldsymbol{\theta})}{f_{(\boldsymbol{\theta}|Y=y)}(\boldsymbol{\theta})} d\boldsymbol{\theta} \approx \sum_{k=1}^K \log \frac{q_{(\boldsymbol{\theta}|\boldsymbol{\lambda}^{(\ell)})}(\boldsymbol{\theta}_k)}{f_{(\boldsymbol{\theta}|Y=y)}(\boldsymbol{\theta}_k)}$$

où  $\boldsymbol{\theta}_k \sim q_{(\boldsymbol{\theta}|\boldsymbol{\lambda}^{(\ell)})}$ ,  $1 \leq k \leq K$ .

Les itérations sont interrompues lorsque les variations de la divergence KL estimées sont suffisamment faibles entre deux évaluations successives. Un ensemble de  $\tau$  itérations s'appelle une *époque*. L'organisation en époques permet d'éviter le calcul de la divergence KL à chaque itération, qui peut être coûteux et est exclusivement effectué pour attester la convergence.

## CHAPITRE 6 RÉSULTATS SUR DONNÉES DE SIMULATION

Une première étape consiste à tester l'algorithme SQAVI développé sur un jeu de données simulé. Ces données sont produites à partir du modèle bayésien hiérarchique de valeurs extrêmes pour les précipitations décrit à la section 4.2. Les valeurs des paramètres ayant générées ces données sont appelées les vraies valeurs. Le but de ce chapitre est d'étudier la capacité de l'algorithme SQAVI à estimer précisément la loi *a posteriori* de ces paramètres. Pour cela, la loi d'approximation est comparée à un échantillon de la loi *a posteriori* estimée par MCMC.

### 6.1 Les données

Le jeu de données de simulation est une grille de taille  $m_1 = m_2 = 200$  pour laquelle chaque cellule contient  $n_{obs} = 100$  données. En tout, la grille possède  $m = 40\,000$  cellules. La cellule médiane, i.e. la cellule d'indice 20 000, est utilisée comme référence dans l'étude des paramètres de localisation et d'échelle. Les données sont générées à partir du modèle présenté en section 4.2 avec la configuration donnée dans le tableau 6.1 pour la cellule de référence.

TABLEAU 6.1 Valeurs des paramètres du modèle de précipitations extrêmes avec lesquelles les données de simulation ont été générées.

Paramètre	$\mu_{20\,000}$	$\phi_{20\,000}$	$\xi$	$\kappa_u$	$\kappa_v$
Vraie valeur	40.0668	0.8663	0.05	1	10

### 6.2 Estimations SQAVI

Pour les besoins de la simulation, l'algorithme est exécuté sur 5 époques de 5 itérations chacune sans fixer de critère d'arrêt. Les paramètres variationnels sont initialisés avec des valeurs arbitraires pour tester la robustesse de l'algorithme.

La figure 6.1a montre la décroissance de la divergence KL au fil des itérations. Cette décroissance n'est pas monotone en raison de son estimation par méthode Monte-Carlo. Les traces du mode des approximations des lois *a posteriori* marginales sont présentées aux figures 6.1b, 6.1c, 6.1d, 6.1e et 6.1f. Elles montrent une convergence du mode des lois, dont les valeurs finales correspondent aux estimations SQAVI des paramètres. Ces estimations semblent proches des vraies valeurs du tableau 6.1. Comme mesure d'évaluation, la distance

euclidienne relative aux vraies valeurs est donnée au tableau 6.2. Cette distance s'écrit

$$d(\hat{\theta}_{SQAVI}, \theta) = \frac{|\hat{\theta}_{SQAVI} - \theta|}{|\theta|}$$

où  $\theta$  est la vraie valeur du paramètre  $\theta$  et  $\hat{\theta}_{SQAVI}$  est l'estimation SQAVI obtenue.

Les lignes rouges de la figure 6.2 illustre les approximations obtenues par la méthode SQAVI des lois *a posteriori* marginales des paramètres du modèle.

### 6.3 Estimations MCMC

Suite à une étude préliminaire comparant différentes méthodes MCMC (annexe D), l'algorithme Metropolis dans Gibbs optimisé de Jalbert *et al.* (2022) est retenu comme algorithme de référence pour évaluer la solution variationnelle. Un échantillon de taille 9000 de la loi *a posteriori* a été extrait d'une chaîne de 10 000 itérations, où les 1000 premières itérations ont été rejetées en tant qu'itérations de la phase de chauffe.

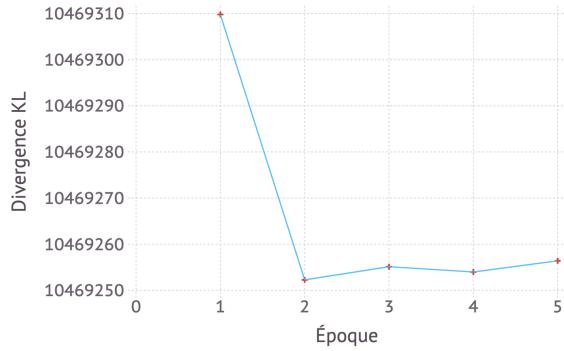
Les histogrammes bleus de la figure 6.2 illustrent les échantillons obtenus par MCMC des lois *a posteriori* marginales des paramètres du modèle.

### 6.4 Comparaison SQAVI et MCMC

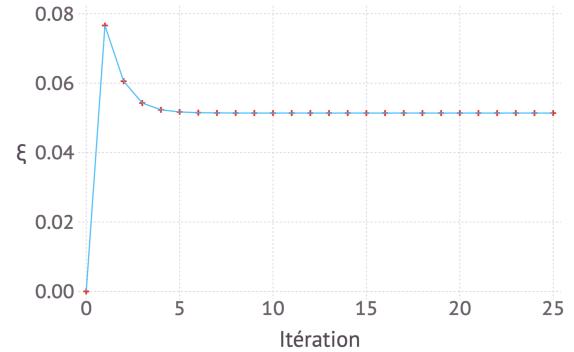
La figure 6.2 permet de comparer les approximations des lois *a posteriori* marginales  $q^*$  des paramètres obtenues par l'approche variationnelle SQAVI aux échantillons des lois *a posteriori* marginales obtenus par MCMC. Les valeurs du mode des deux densités pour chaque paramètre ainsi que les mesures d'évaluation associées sont présentées au tables 6.2. Ces résultats conduisent à plusieurs observations.

Premièrement, l'approximation de la loi marginale *a posteriori* de  $(\mu_{20\,000}, \phi_{20\,000})$  semble particulièrement précise : les lois MCMC et SQAVI coïncident, aussi bien du point de vue de la position que de la dispersion. Présenté seulement pour  $i = 20\,000$  par soucis de lisibilité, ce résultat s'étend à toutes les cellules de la grille. Une telle précision peut s'expliquer par la cohérence des approximations vis-à-vis de la forme des lois  $q^*_{\{(\mu_i, \phi_i)\} \rightarrow \cdot}$ . En effet, l'hypothèse de linéarité des variables latentes et l'approximation quadratique sont justifiées (voir l'annexe B et de même que l'expression (5.7)).

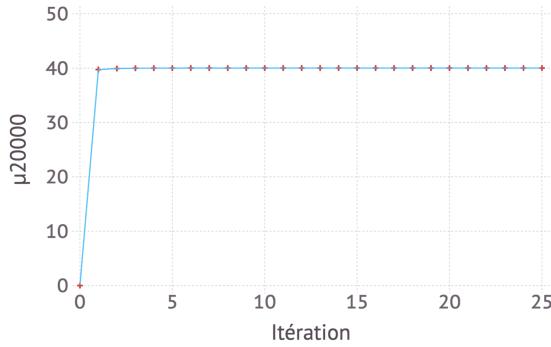
Dans le cas du paramètre  $\xi$ , on constate la présence d'un décalage entre les modes des deux densités marginales : SQAVI la surestime la vraie valeur. Cette différence peut être due à l'hypothèse de linéarité des variables latentes et à l'approximation quadratique qui,



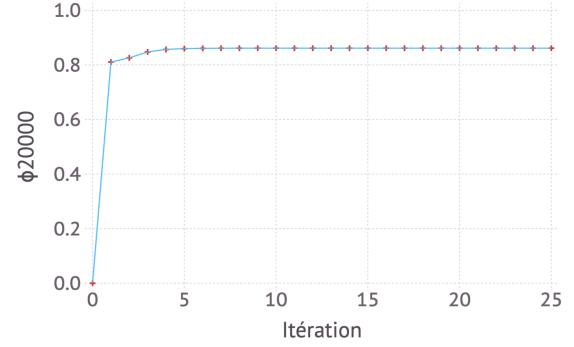
(a) Évolution de la divergence KL en fonction des époques.



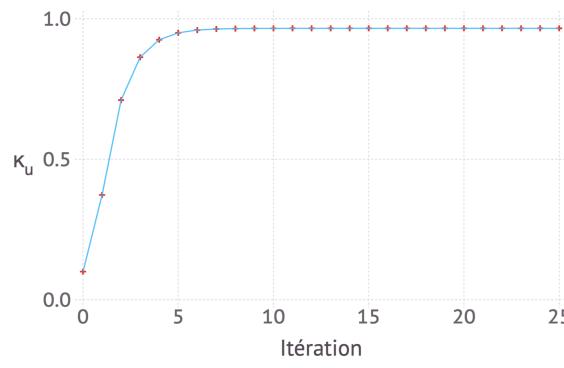
(b) Trace SQAVI du paramètre  $\xi$ .



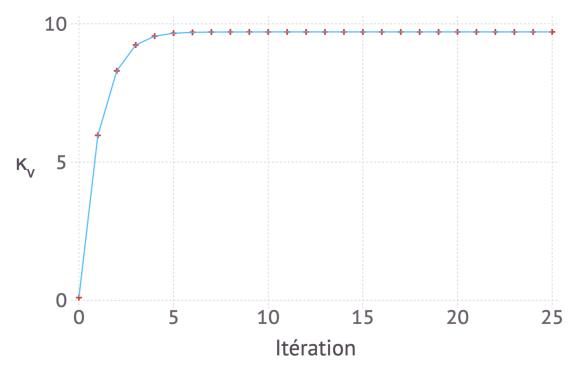
(c) Trace SQAVI du paramètre  $\mu_{20000}$ .



(d) Trace SQAVI du paramètre  $\phi_{20000}$ .



(e) Trace SQAVI du paramètre  $\kappa_u$ .



(f) Trace SQAVI du paramètre  $\kappa_v$ .

FIGURE 6.1 Traces de la divergence KL et du mode des lois *a posteriori* pour chacun des paramètres obtenues par la méthode SQAVI.

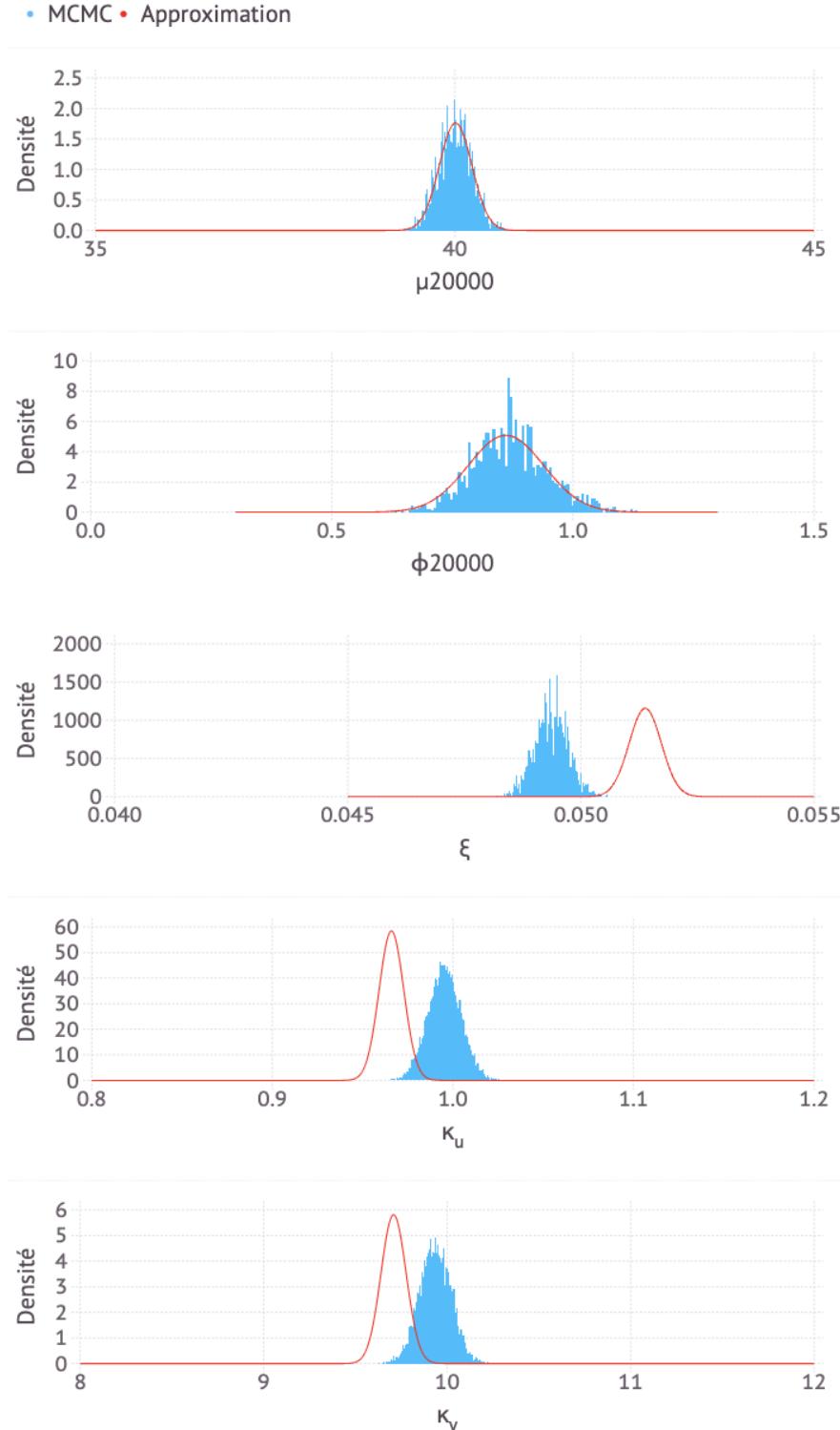


FIGURE 6.2 Approximation des lois *a posteriori* marginales (courbes rouges) de chacun des paramètres et histogrammes (bandes bleues) des échantillons des lois *a posteriori* marginales obtenus par MCMC.

contrairement aux paramètres  $(\mu_i, \phi_i)$ , n'ont pas de fondement théorique solide. Malgré tout, la différence entre les estimations est très faible en valeur absolue et la distance relative aux vraies valeurs est du même ordre de grandeur pour les deux algorithmes.

Un biais est également observé pour les paramètres  $\kappa_u$  et  $\kappa_v$ . Pour ces paramètres, la méthode SQAVI sous-estime les modes. Cette erreur peut être causée par un phénomène de propagation d'erreur. En effet, le calcul des paramètres  $b_u$  et  $b_v$ , permettant de déterminer les estimations de  $\kappa_u$  et  $\kappa_v$ , utilise la variance et l'espérance au carré de tous les paramètres de localisation et de log-échelle respectivement (annexe C). De petites erreurs sur ces quantités peuvent s'accumuler et perturber le résultat. La comparaison des distances relatives montre que le MCMC est plus précis pour ces paramètres. Néanmoins, leur estimation n'est pas aussi exigeante que pour les paramètres de la loi GEV. Les paramètres de précision sont essentiellement un outil pour quantifier la dépendance entre les paramètres des lois GEV. Ils ne sont pas utilisés directement pour estimer les quantiles extrêmes. Par conséquent, pour les applications courantes, ces imprécisions ont un impact restreint. Elles ne limitent donc pas l'utilisation de SQAVI, qui procure de nombreux avantages.

TABLEAU 6.2 Évaluation et comparaison des estimations SQAVI et MCMC.

Paramètre	$\mu_{20\ 000}$	$\phi_{20\ 000}$	$\xi$	$\kappa_u$	$\kappa_v$
Vraie valeur	40.0668	0.8663	0.05	1	10
Estimation SQAVI	40.0102	0.8611	0.0514	0.9660	9.7084
Estimation MCMC	40.0076	0.8703	0.04936	0.9954	9.9359
$d(\hat{\theta}_{SQAVI}, \theta)$	0.001412	0.006002	0.02800	0.03400	0.02916
$d(\hat{\theta}_{MCMC}, \theta)$	0.001477	0.004547	0.01271	0.004644	0.006413
$ \hat{\theta}_{MCMC} - \hat{\theta}_{SQAVI} $	0.002667	0.009109	0.002015	0.02934	0.2275

## CHAPITRE 7 RÉSULTATS SUR DONNÉES DE RÉANALYSE

L'algorithme SQAVI est exécuté sur les précipitations journalières de la réanalyse RCaS. Les estimations et les lois *a posteriori* des paramètres de localisation et d'échelle seront présentées pour une cellule témoin dont les coordonnées correspondent à celles de la ville de Montréal (figure 7.1). On note les paramètres de localisation et de log-échelle associés à cette cellule  $\mu_{MTL}$  et  $\phi_{MTL}$ . L'approximation de la loi *a posteriori* sera comparée à un échantillon de celle-ci obtenu par MCMC.

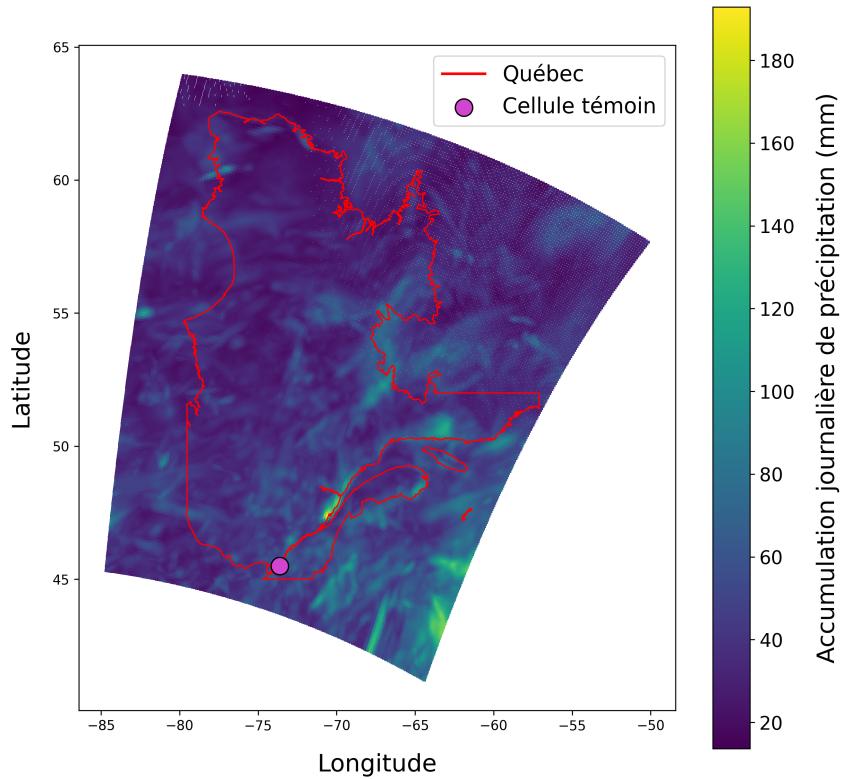


FIGURE 7.1 Localisation de la cellule témoin sur les données.

### 7.1 Estimations SQAVI

L'algorithme est exécuté sur 15 époques de 1 itération chacune afin d'observer la valeur de la divergence KL après chaque itération. L'itération 0 correspond à l'initialisation. Les modes des paramètres de localisation et de log-échelle ( $\eta_i, 1 \leq i \leq m$ ) sont initialisés grâce aux valeurs de leurs estimations par maximum de vraisemblance. Le mode du paramètre de

forme ( $\eta_e$ ) est initialisé à la valeur moyenne des estimations des paramètres de forme obtenues par maximum de vraisemblance en chaque cellule. Les paramètres variationnels  $b_u$  et  $b_v$  sont arbitrairement initialisés à de telle sorte que  $\mathbb{E}_{\{\kappa_u|a,b_u\}}[\kappa_u] = \mathbb{E}_{\{\kappa_v|a,b_v\}}[\kappa_v] = 1$ .

La figure 7.2 illustre les traces de la divergence KL et du mode des lois *a posteriori* marginales pour chacun des paramètres du modèle. Ces traces suggèrent qu'un minimum de la divergence est atteint avant la fin des 15 itérations. La figure 7.3 illustre les lois *a posteriori* marginales approximées (en rouge) par la méthode SQAVI.

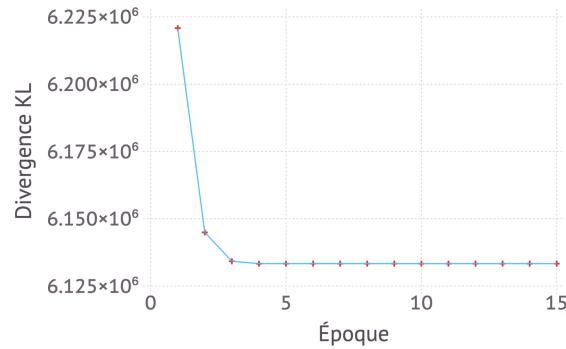
## 7.2 Estimations MCMC

Un échantillon de la loi *a posteriori* du modèle hiérarchique de valeurs extrêmes ajusté sur les précipitations réanalysées a été généré par méthode MCMC (Jalbert *et al.*, 2022). La figure 7.3 illustre les histogrammes (bandes bleues) des échantillons des lois *a posteriori* marginales pour chacun des paramètres.

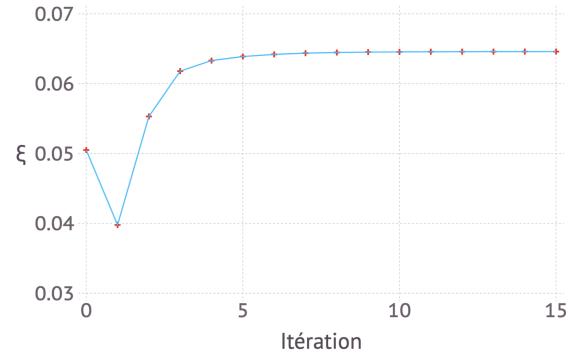
## 7.3 Comparaison SQAVI et MCMC

Les lois marginales  $q^*$  obtenues à l'aide des dernières mises à jour des paramètres variationnels de l'algorithme SQAVI sont superposées aux histogrammes normalisés des échantillons MCMC dans la figure 7.3. Comme pour les données simulées, (chapitre 6) les approximations des paramètres de localisation et de log-échelle sont excellentes. Les cartes de chaleur des estimations obtenues par SQAVI et MCMC sur le territoire d'étude (figures 7.4 et 7.5) révèlent des estimations très proches et une structure de dépendance spatiale conservée entre les deux algorithmes. C'est ce qu'on souhaitait obtenir : retrouver les lois MCMC via notre algorithme variationnel.

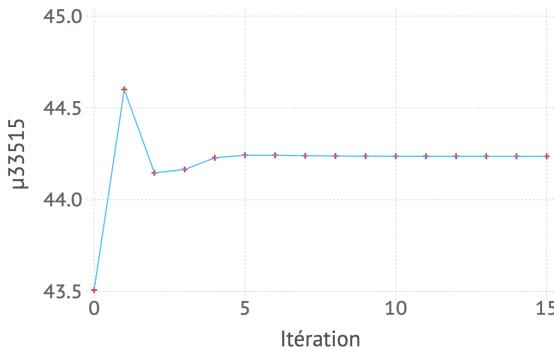
En revanche, l'erreur sur l'estimation du paramètre de forme  $\xi$  est plus élevée que pour les données simulées (tableau 7.1). Les conséquences de cet écart peuvent être analysées *via* les cartes de chaleurs des niveaux de retour associés à la période de retour 100 ans (figure 7.6). Ces cartes sont importantes car elles matérialisent une extrapolation typique sur les extrêmes qui peut être réalisée par les spécialistes œuvrant en sciences du climat. Elles représentent la quantité de précipitation journalière susceptible d'être observée en moyenne tous les 100 ans à chaque point de grille. Le calcul du niveau de retour exploite les estimations de  $\mu$ ,  $\sigma$  et  $\xi$  (cf *e.g.* Coles (2001)), permettant de se rendre compte de l'effet de la différence des estimations du paramètre de forme sur les quantiles extrêmes. Aucune différence majeure sur les niveaux de retour étant observée, cette différence est jugée négligeable d'un point de vue pratique.



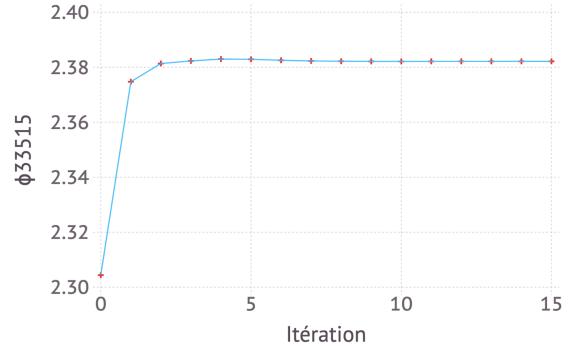
(a) Évolution de la divergence KL en fonction des époques.



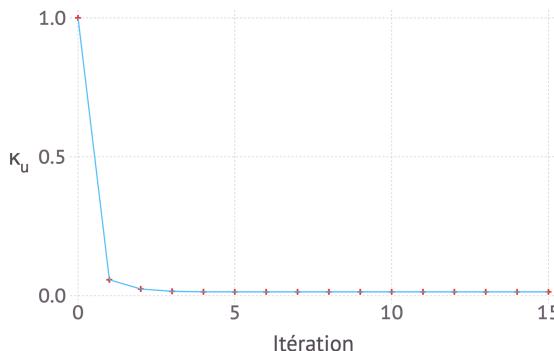
(b) Trace SQAVI du paramètre  $\xi$ .



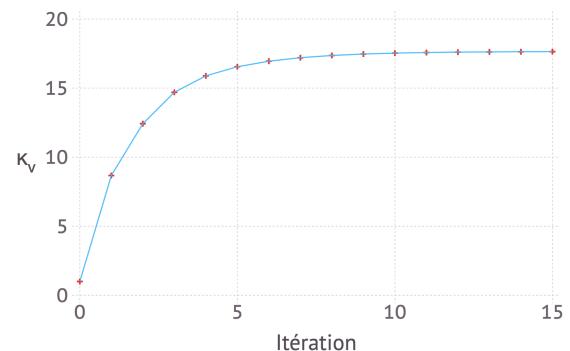
(c) Trace SQAVI du paramètre  $\mu_{MTL}$ .



(d) Trace SQAVI du paramètre  $\phi_{MTL}$ .



(e) Trace SQAVI du paramètre  $\kappa_u$ .



(f) Trace SQAVI du paramètre  $\kappa_v$ .

FIGURE 7.2 Traces de la divergence et du mode des lois *a posteriori* pour chacun des paramètres obtenues par la méthode SQAVI.

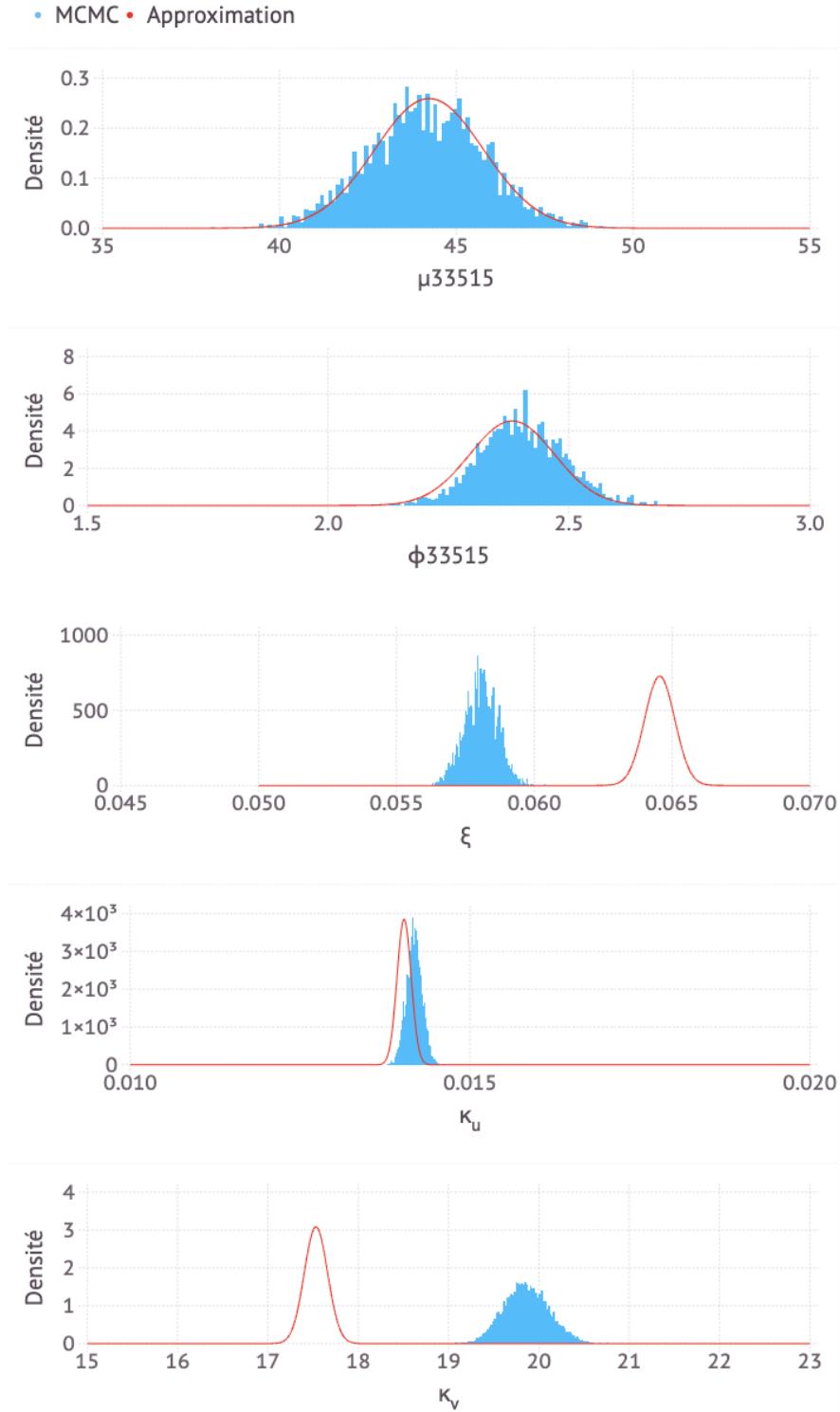


FIGURE 7.3 Approximation des lois *a posteriori* marginales (courbes rouges) de chacun des paramètres et histogrammes (bandes bleues) des échantillons des lois *a posteriori* marginales obtenus par MCMC pour les précipitations réanalysées de RCaS.

Enfin, la figure 7.3 montre que l'algorithme SQAVI est plus précis pour estimer  $\kappa_u$  que  $\kappa_v$ . De fait, l'espérance de  $\kappa_u | Y = y$  étant visiblement plus faible que celle de  $\kappa_v | Y = y$ , elle suggère une dépendance plus faible entre les paramètres de localisation, qui s'aligne davantage avec l'hypothèse de champ moyen (section 3.3.1).

TABLEAU 7.1 Évaluation et comparaison des estimations SQAVI et MCMC.

Paramètre	$\mu_{MTL}$	$\phi_{MTL}$	$\xi$	$\kappa_u$	$\kappa_v$
Estimation SQAVI	44.2362	2.3821	0.06459	0.01404	17.6408
Estimation MCMC	44.1560	2.4030	0.05806	0.01418	19.8619
$ \hat{\theta}_{MCMC} - \hat{\theta}_{SQAVI} $	0.08019	0.02084	0.006536	0.0001387	2.2211

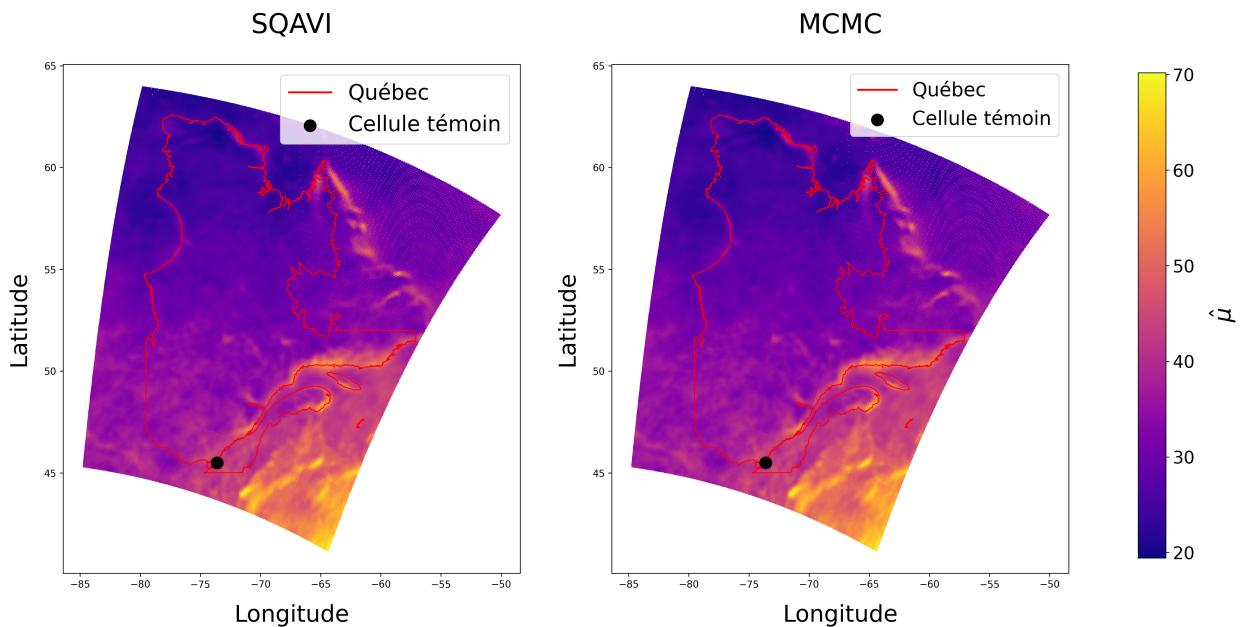


FIGURE 7.4 Cartes de chaleur des estimations SQAVI (à gauche) et MCMC (à droite) du paramètre de localisation sur le territoire d'intérêt.

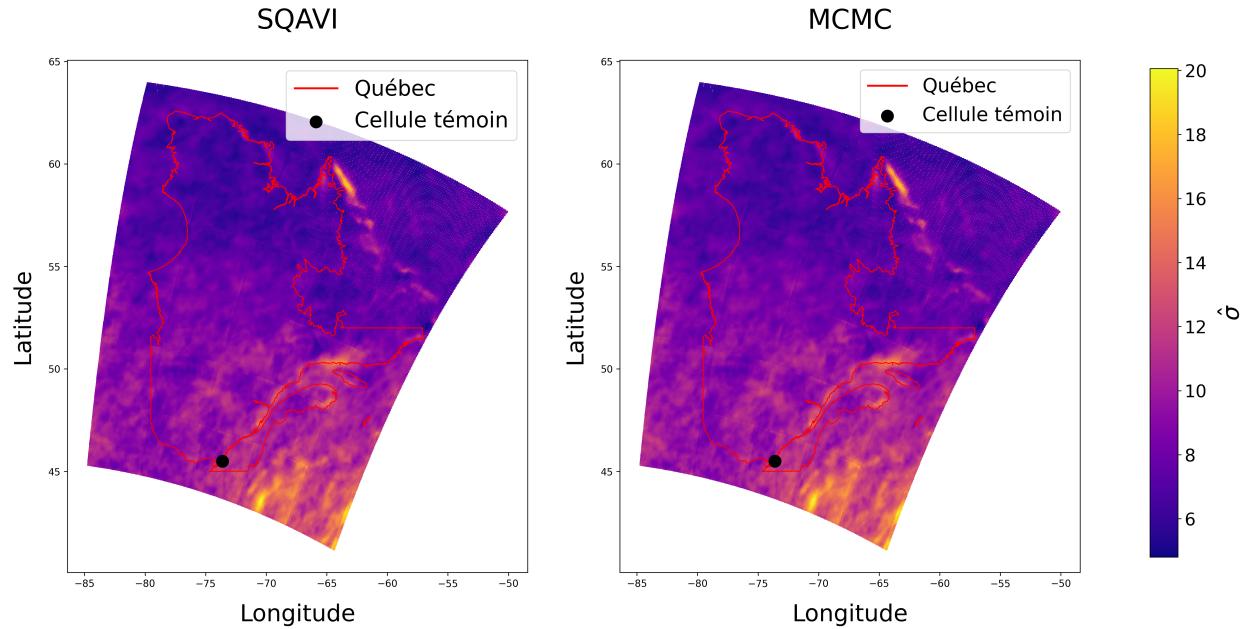


FIGURE 7.5 Cartes de chaleur des estimations SQAVI (à gauche) et MCMC (à droite) du paramètre d'échelle sur le territoire d'intérêt.

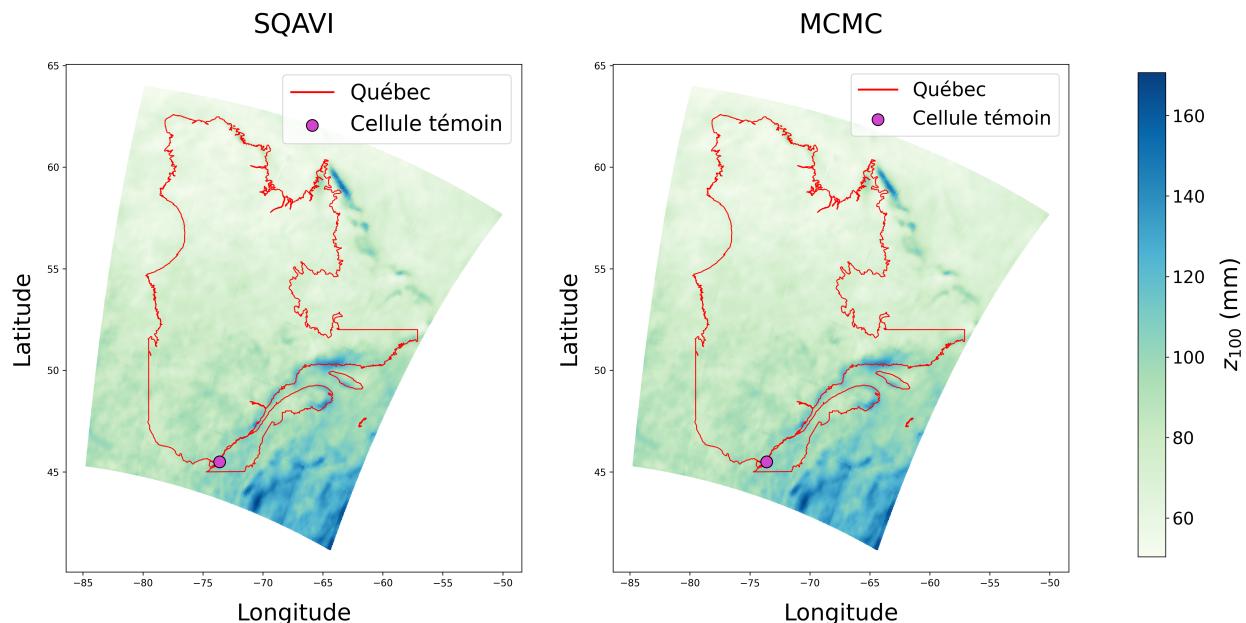


FIGURE 7.6 Cartes de chaleur du niveau de retour associé à la période de retour 100 ans, noté  $z_{100}$ , estimé par la méthode SQAVI (à gauche) et MCMC (à droite).

## CHAPITRE 8 DISCUSSION

Les résultats des chapitres 6 et 7 illustrent clairement le potentiel de la méthode variationnelle proposée. Elle donne accès à une loi d'approximation explicite, proche de la vraie loi *a posteriori*, via une procédure ne nécessitant pas de réglage au préalable. Néanmoins, l'algorithme n'est pas parfaitement précis pour les paramètres globaux du modèle. Concernant les paramètres de précision des lois *a priori* iGMRF, les écarts entre les modes de la loi d'approximation et la loi cible sont faibles étant donné que ces variables n'interviennent pas directement dans l'extrapolation des extrêmes. Cette conclusion est moins évidente pour le paramètre de forme de la loi GEV, dont la surestimation peut engendrer des erreurs importantes dans le calcul des niveaux de retour associés aux périodes de retours élevées (environ 10 000 ans). Certaines hypothèses imposées dans cette étude permettent d'expliquer ces altérations, donnant des pistes de réflexion pour les corriger.

Au niveau du modèle hiérarchique bayésien (section 4.2), le choix d'un paramètre de forme constant sur toute la grille relève de considérations pratiques et n'est pas rigoureusement justifié. En effet, La complexité inhérente à l'estimation de ce paramètre conduit à mobiliser l'ensemble des informations disponibles, à savoir l'intégralité de la grille. Cette décision entraîne une faible incertitude sur l'estimation qui amplifie le biais de la méthode variationnelle. Certains tests non présentés dans ce mémoire semblent indiquer que l'écart d'estimation avec l'algorithme MCMC se réduit pour des tailles de grille plus faibles. Un modèle doté de paramètres de forme régionaux, *i.e.* regroupant des sous-ensembles de la grille, pourrait constituer un compromis efficace pour atténuer l'erreur observée.

L'approximation champ moyen (section 3.3.1) constitue également une hypothèse incertaine, qui donne l'impression d'un choix de facilité. Compte tenu de la tolérance relativement élevée dans l'estimation des paramètres de précision, l'indépendance de leurs lois marginales *a posteriori* est un postulat acceptable. En revanche, l'indépendance *a posteriori* du paramètre de forme est plus délicate à justifier. Bien que l'annexe A corroborre ce choix, son fondement théorique reste à établir. Par ailleurs, elle serait bien plus fragile dans un modèle à paramètres de forme régionaux.

L'hypothèse de linéarité des variables latentes appliquée au paramètre de forme en (5.10) est très certainement la plus discutable. Elle modifie la structure de la loi marginale optimale de  $\xi$  de manière expédiente, alors même que les variables qui apparaissent dans son expression sont fortement intriquées les unes aux autres. Comme la loi  $q_\xi^*$  s'écrit sous la forme d'une espérance, une utilisation judicieuse de la méthode Monte-Carlo permettrait possiblement

d'obtenir une approximation plus fine. Une telle étude n'a pas été explorée dans le cadre de ce travail.

L'approximation quadratique est également source de questionnement. Rien n'indique que le paramètre de forme se comporte selon une loi quasi-normale et d'autres approches plus fines peuvent être essayées. Entre autre, la méthode de Laplace de Tierney *et al.* (1986), notamment utilisée dans l'application de la méthode INLA à un modèle d'extrêmes de précipitation (Opitz *et al.*, 2018), pourrait être une bonne alternative.

Malgré ces simplifications, l'algorithme fournit des résultats très prometteurs, si bien que l'extrapolation des extrêmes n'est pratiquement pas modifiée par rapport à celle obtenue par méthode MCMC. Non seulement la procédure proposée est aisément interprétable et dotée d'un diagnostic de convergence sobre, mais elle dispose d'une grande marge d'amélioration qui pourrait la rendre incontournable dans l'étude des événements climatiques extrêmes.

Notons enfin que ce travail ne fournit pas d'estimation du coût asymptotique de l'algorithme SQAVI. Il ne permet donc pas de comparer rigoureusement son temps d'exécution avec celui de la méthode MCMC. Cependant, le faible nombre d'itérations qu'il requiert pour atteindre un minimum laisse entendre qu'il pourrait présenter un gain de temps significatif sur la méthode MCMC. Les exécutions des algorithmes ayant permis d'obtenir les résultats des chapitres 6 et 7 vont également dans ce sens.

## CHAPITRE 9 CONCLUSION ET RECOMMANDATIONS

L'étude des précipitations extrêmes est essentielle pour prévenir nos sociétés face à l'augmentation de la fréquence et de l'intensité des inondations. Une telle étude est permise par l'ajustement du modèle de Jalbert *et al.* (2017), un modèle hiérarchique bayésien de valeurs extrêmes qui modélise les maxima annuels de précipitations organisés sur une grille régulière. À ce jour, l'ajustement de ce modèle était réalisé *via* des méthodes MCMC qui, bien qu'efficaces, requièrent une expertise pour être configurée, limitant leur diffusion à des non statisticiens au sein des institutions de recherche.

L'algorithme SQAVI développé dans ce travail constitue une alternative fiable aux méthodes MCMC pour l'estimation des paramètres du modèle des précipitations extrêmes. Son objectif est de déterminer, parmi une famille de lois d'approximation, la loi la plus proche de la distribution *a posteriori* selon une certaine distance : la divergence de Kullback-Leibler. La famille de lois proposée s'appuie sur l'hypothèse de champ moyen dans laquelle les paramètres sont supposés indépendants *a posteriori*, exceptés les paramètres de localisation et de log-échelle d'une même cellule qui sont modélisés par une loi conjointe. Avec une telle famille, la forme des lois marginales d'approximation optimales est déterminée. Lorsque ces lois ont des formes indéterminées, elles sont approchées par des lois normales *via* une hypothèse de linéarité et une approximation dite quadratique. La solution consiste alors à mettre à jour itérativement les lois marginales optimales pour s'approcher de la vraie distribution *a posteriori*. Elle s'inspire de la procédure CAVI de Bishop (2006).

Pour tester sa capacité à approcher efficacement la distribution *a posteriori*, l'algorithme SQAVI est appliqué d'une part à un jeu de données simulé et d'autre part à des données de réanalyse issues d'un modèle climatique couvrant le territoire du Québec. Ses performances sont comparées à celle d'un algorithme MCMC de type Metropolis dans Gibbs, défini comme algorithme de référence.

Les résultats sont très positifs au regard de la simplicité des approximations effectuées, en particulier pour les paramètres de localisation et d'échelle, dont les lois marginales d'approximation coïncident avec les échantillons obtenus par MCMC. Cependant, la méthode est moins précise pour estimer le paramètre de forme, essentiel pour l'extrapolation des extrêmes. De nombreuses améliorations peuvent encore être réalisées, notamment en affinant la famille de distribution d'approximation.

Globalement, l'évaluation permet de valider l'utilisation de l'algorithme proposé pour l'inférence du modèle hiérarchique bayésien. Une suite logique serait d'adapter l'algorithme à

d'autres modèles tel que celui de Jalbert *et al.* (2022), conçus pour les jeux de données réelles intégrant des variables explicatives. Des recherches ultérieures pourraient également étudier quelles approximations permettraient d'appliquer la méthode variationnelle à des modèles basés sur la loi de Pareto Généralisée (GPD, Coles, 2001) voire à la loi de Pareto Généralisée étendue (EGPD, Gamet et Jalbert, 2022).

En outre, il serait intéressant d'appliquer l'algorithme *Black Box Variational Inference* (Ranganath *et al.*, 2013) au modèle hiérarchique bayésien. Son caractère très général laisse supposer qu'il peut fournir des lois d'approximations plus complexes et donc plus fines que la méthode construite dans ce rapport. Cependant, cette méthode nécessite de relever certaines difficultés d'implémentation qui pourraient inclure l'utilisation d'hyperparamètres. Il s'agirait alors de trouver un juste compromis entre précision et facilité d'utilisation.

## RÉFÉRENCES

- BEAL, M. J. (2003). *Variational algorithms for approximate Bayesian inference*. Thèse de doctorat, UCL (University College London).
- BESAG, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society : Series B (Methodological)*, 36(2):192–225.
- BISHOP, C. M. (2006). *Pattern recognition and machine learning*. Information science and statistics. Springer, New York.
- BLEI, D. M., KUCUKELBIR, A. et McAULIFFE, J. D. (2017). Variational inference : A review for statisticians. *Journal of the American Statistical Association*, 112(518):859–877.
- BROOKS, S., GELMAN, A., JONES, G. et MENG, X.-L. (2011). *Handbook of Markov Chain Monte Carlo*. CRC Press LLC, Milton, UNITED KINGDOM.
- BUREAU D'ASSURANCE DU CANADA (2024). L'événement météorologique le plus coûteux de l'histoire du Québec – Les inondations du mois d'août ont causé près de 2.5 milliards de dollars de dommages assurés. <https://fr.ibc.ca/news-insights/news/the-costliest-severe-weather-event-in-quebec-s-history-august-flooding-caused-nearly-2-5-billion-in-insured-damage>. Consulté le 22 mai 2025.
- COLES, S. (2001). *An introduction to statistical modeling of extreme values*. Springer Series in Statistics. Springer, London.
- COOLEY, D., NYCHKA, D., et NAVEAU, P. (2007). Bayesian spatial modeling of extreme precipitation return levels. *Journal of the American Statistical Association*, 102(479):824–840.
- DEMPSTER, A. P., LAIRD, N. M. et RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society : Series B (Methodological)*, 39(1):1–22.
- DUANE, S., KENNEDY, A. D., PENDLETON, B. J. et ROWETH, D. (1987). Hybrid Monte Carlo. *Physics Letters B*, 195(2):216–222.
- FISHER, R. A. et TIPPETT, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24(2):180–190.

- FRASER, C. G. (1994). The origins of Euler's variational calculus. *Archive for History of Exact Sciences*, 47(2):103–141.
- FREY, B. J. et HINTON, G. E. (1999). Variational learning in nonlinear gaussian belief networks. *Neural Computation*, 11(1):193–213.
- GAMET, P. et JALBERT, J. (2022). A flexible extended generalized Pareto distribution for tail estimation. *Environmetrics*, 33(6):e2744.
- GASSET, N., FORTIN, V., DIMITRIJEVIC, M., CARRERA, M., BILODEAU, B., MUNCASTER, R., GABORIT, É., ROY, G., PENTCHEVA, N., BULAT, M., WANG, X., PAVLOVIC, R., LESPINAS, F., KHEDHAOUIRIA, D. et MAI, J. (2021). A 10km North American precipitation and land-surface reanalysis based on the GEM atmospheric model. *Hydrology and Earth System Sciences*, 25(9):4917–4945.
- GELMAN, A., CARLIN, J. B., STERN, H. S., DUNSON, D. B., VEHTARI, A. et RUBIN, D. B. (2013). *Modal and distributional approximations*. Chapman and Hall/CRC, 3 édition.
- GEMAN, S. et GEMAN, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6):721–741.
- GNEDENKO, B. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Annals of Mathematics*, 44(3):423–453.
- HASTINGS, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109.
- HINTON, G. E. et van CAMP, D. (1993). Keeping the neural networks simple by minimizing the description length of the weights. In *Proceedings of the sixth annual conference on Computational learning theory*, COLT '93, page 5–13, New York, NY, USA. Association for Computing Machinery.
- HOMAN, M. D. et GELMAN, A. (2014). The No-U-turn sampler : adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.*, 15(1):1593–1623.
- HUMPHREYS, K. et TITTERINGTON, D. M. (2000). *Approximate Bayesian inference for simple mixtures*, page 331–336. Physica, Heidelberg.
- JALBERT, J., FAVRE, A.-C., BÉLISLE, C. et ANGERS, J.-F. (2017). 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*, 66(5):941–962.
- JALBERT, J., GENEST, C. et PERREAULT, L. (2022). Interpolation of precipitation extremes on a large domain toward IDF curve construction at unmonitored locations. *Journal of Agricultural, Biological and Environmental Statistics*, 27(3):461–486.
- JENKINSON, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, 81(348):158–171.
- JORDAN, M. I., GHAHRAMANI, Z., JAAKKOLA, T. S. et SAUL, L. K. (1999). An introduction to variational methods for graphical models. *Machine Learning*, 37(2):183–233.
- LAVOIE, L. (2024). La tempête de vendredi au Québec entre dans l'histoire : quantité de pluie «sans précédent» à Montréal et ailleurs.
- MACKAY, D. J. C. (1995). Ensemble learning and evidence maximization. *Technical report*. Cavendish Laboratory, University of Cambridge, Cambridge.
- MACKAY, D. J. C. (1997). Ensemble learning for hidden markov models. Cavendish Laboratory, University of Cambridge, Cambridge.
- MARTINS, E. S. et STEDINGER, J. R. (2000). Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resources Research*, 36(3):737–744.
- MARUYAMA, G. (1955). Continuous Markov processes and stochastic equations. *Rendiconti del Circolo Matematico di Palermo*, 4(1):48–90.
- MAYBECK, P. S. (1982). *Stochastic models, estimation, and control*, volume 3. Academic Press, 1st édition.
- METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H. et TELLER, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092.
- MINKA, T. P. (2001). *A family of algorithms for approximate Bayesian inference*. Thèse de doctorat, Massachusetts Institute of Technology.
- MINKA, T. P. (2013). Expectation Propagation for approximate Bayesian inference. (arXiv :1301.2294).

- NEAL, R. M. (2011). *MCMC using Hamiltonian dynamics*. arXiv :1206.1901.
- OPITZ, T., HUSER, R., BAKKA, H. et RUE, H. (2018). INLA goes extreme : Bayesian tail regression for the estimation of high spatio-temporal quantiles. *Extremes*, 21(3):441–462.
- OURANOS (2025). Crues et inondations. <https://www.ouranos.ca/fr/phenomenes-climatiques/crues-inondations-changements-projetes>. Consulté le 22 mai 2025.
- PARISI, G. (1988). *Statistical field theory*. Redwood City, CA. : Addison-Wesley Pub. Co.
- PESKUN, P. H. (1973). Optimum Monte-Carlo sampling using Markov chains. *Biometrika*, 60(3):607–612.
- RANGANATH, R., GERRISH, S. et BLEI, D. M. (2013). Black Box Variational Inference. (arXiv :1401.0118).
- ROBERTS, G. O. et STRAMER, O. (2002). Langevin diffusions and metropolis-hastings algorithms. *Methodology And Computing In Applied Probability*, 4(4):337–357.
- ROY, V. (2020). Convergence diagnostics for Markov Chain Monte Carlo. *Annual Review of Statistics and Its Application*, 7(Volume 7, 2020):387–412.
- RUE, H. et HELD, L. (2005). *Gaussian Markov Random Fields : theory and applications*. Chapman and Hall/CRC, New York.
- RUE, H., MARTINO, S. et CHOPIN, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 71(2):319–392.
- SUN, S. (2013). A review of deterministic approximate inference techniques for bayesian machine learning. *Neural Computing and Applications*, 23(7):2039–2050.
- TIERNEY, L., et KADANE, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, 81(393):82–86.
- van de SCHOOT, R., DEPAOLI, S., KING, R., KRAMER, B., MÄRTENS, K., TADESSE, M. G., VANNUCCI, M., GELMAN, A., VEEN, D., WILLEMSSEN, J. et YAU, C. (2021). Bayesian statistics and modelling. *Nature Reviews Methods Primers*, 1(1):1–26.

## ANNEXE A CORRÉLATION ENTRE LE PARAMÈTRE DE FORME ET LES PARAMÈTRES INTRA-CELLULAIRES

Le but de cette analyse est de comparer la corrélation entre  $\xi$  et  $\mu_1$  et entre  $\xi$  et  $\phi_1$  à celle entre  $\mu_1$  et  $\phi_1$  au fur et à mesure que la taille de la grille augmente.

Pour cela, on considère une grille régulière carrée ( $m_1 = m_2$ ) pour laquelle on génère des données artificielles à partir du modèle bayésien hiérarchique de valeurs extrêmes pour les précipitations décrit à la section 4.2. Les vraies valeurs des paramètres du modèle permettant de générer les données sont fournies au tableau A.1.

TABLEAU A.1 Configuration des paramètres ayant générée les données artificielles de l’analyse.

Paramètre	$\mu$	$\phi$	$\xi$	$\kappa_u$	$\kappa_v$
Valeur	$\approx 40$	$\approx 1$	0.05	1	10

La notation  $\approx 40$  signifie que les paramètres  $\mu_i$  ont été générés par un iGMRF centré en 40, *i.e.*  $\mu_i \sim iGMRF(\kappa_u) + 40$ ,  $\forall i \in \{1, \dots, m\}$ . La même notation vaut pour  $\phi$ .

Pour simplifier l’étude, on fixe les valeurs de  $\kappa_u$  et  $\kappa_v$  du modèle d’inférence. La loi conjointe *a posteriori* s’écrit donc

$$\begin{aligned}
 f_{(\boldsymbol{\mu}, \boldsymbol{\phi}, \xi | Y)}(\boldsymbol{\mu}, \boldsymbol{\phi}, \xi) &\propto \left\{ \prod_{i=1}^m \prod_{j=1}^{n_{obs}} \mathcal{GEV}(y_{ij} | \mu_i, \phi_i, \xi) \right\} \\
 &\quad \times 1^{\frac{m-r}{2}} \exp \left( -1 \frac{\boldsymbol{\mu}^\top W \boldsymbol{\mu}}{2} \right) \\
 &\quad \times 10^{\frac{m-r}{2}} \exp \left( -10 \frac{\boldsymbol{\phi}^\top W \boldsymbol{\phi}}{2} \right) \\
 &\quad \times \mathcal{Beta}(\xi + .5 | 9, 6)
 \end{aligned} \tag{A.1}$$

La méthode consiste à calculer la matrice de covariance approximative des paramètres  $\mu_1$ ,  $\phi_1$  et  $\xi$  pour des tailles de grille croissantes.

Ainsi, pour chaque taille de grille, les paramètres  $\mu_i$  et  $\phi_i$ ,  $i \neq 1$  sont estimés *via* l’algorithme SQAVI (section 5.5), permettant de calculer la loi conditionnelle complète  $f_{\{(\mu_1, \phi_1, \xi), \dots\}}$ . Une fois obtenue, l’approximation quadratique appliquée à cette densité fournit la matrice de covariance des trois paramètres.

La corrélation est une version normalisée de la covariance. Elle permet donc une meilleure comparaison de l'influence mutuelle entre les paramètres. La relation entre ces deux quantités s'écrit, pour  $\mu_1$  et  $\phi_1$ ,

$$\text{Corr}(\mu_1, \phi_1) = \frac{\text{Cov}(\mu_1, \phi_1)}{\sigma_{\mu_1} \sigma_{\phi_1}}$$

où  $\text{Corr}(\mu_1, \phi_1)$  désigne la corrélation entre  $\mu_1$  et  $\phi_1$ ,  $\text{Cov}(\mu_1, \phi_1)$  la covariance entre  $\mu_1$  et  $\phi_1$ ,  $\sigma_{\mu_1}$  l'écart-type de  $\mu_1$  et  $\sigma_{\phi_1}$  l'écart-type de  $\phi_1$ .

Pour chaque taille de grille, on répète la procédure 10 fois en enregistrant les valeurs absolues des corrélations, puis on garde la moyenne des valeurs obtenues. Ainsi, les résultats sont moins sensibles à la génération aléatoire des données artificielles. L'évolution résultante est présentée Figure A.1 et les valeurs pour une grille de 30 cellules de côté sont dévoilés au Tableau A.2.

TABLEAU A.2 Valeurs moyennes sur 10 exécutions de la corrélation entre  $\mu_1$ ,  $\phi_1$  et  $\xi$  pour une grille de taille  $30 \times 30$ .

$\text{Corr}(\mu_1, \phi_1)$	$\text{Corr}(\xi, \mu_1)$	$\text{Corr}(\xi, \phi_1)$
0.3512	0.0091	0.0039

La Figure A.1 suggère que  $\text{Corr}(\xi, \mu_1)$  et  $\text{Corr}(\xi, \phi_1)$  tendent vers 0 à mesure que la taille de grille augmente. En comparaison, ce n'est pas le cas de  $\text{Corr}(\mu_1, \phi_1)$  qui semble fluctuer autour d'une valeur constante. Un tel phénomène est aisément compréhensible. Plus il y a de cellules dans la grille, plus l'information pour estimer  $\xi$  est riche et diversifiée. L'information provenant de la première cellule est donc diluée parmi celle provenant de toute la grille, donc l'influence de  $\mu_1$  et  $\phi_1$  sur  $\xi$  diminue. A contrario,  $\mu_1$  et  $\phi_1$  ne sont pas impactés par l'augmentation de la taille de la grille. Leur influence mutuelle n'a donc pas de raison de varier en moyenne.

Pour une taille de grille  $30 \times 30$ , le Tableau A.2 révèle que la corrélation entre  $\mu_1$  et  $\phi_1$  est plus de 38 fois supérieure à celle entre  $\xi$  et  $\mu_1$  et plus de 90 fois supérieure à celle entre  $\xi$  et  $\phi_1$ . Étant donné que les jeux de données usuellement utilisés comprennent des grilles de taille entre 35 et 200 cellules de côté, ces résultats permettent de négliger la dépendance entre  $\xi$  et  $(\mu_1, \phi_1)$ . Cette conclusion s'étend naturellement à toutes les cellules de la grille. Par conséquent, elle justifie l'utilisation d'une loi marginale *a posteriori* indépendante pour  $\xi$ .

Notons que pour chaque exécution de la procédure, les estimations des paramètres  $\mu_i$  et  $\phi_i$ ,  $i \neq 1$ , auraient pu être obtenues par MCMC. L'algorithme SQAVI étant très précis pour ces

paramètres (cf chapitre 8), les résultats auraient été les mêmes.

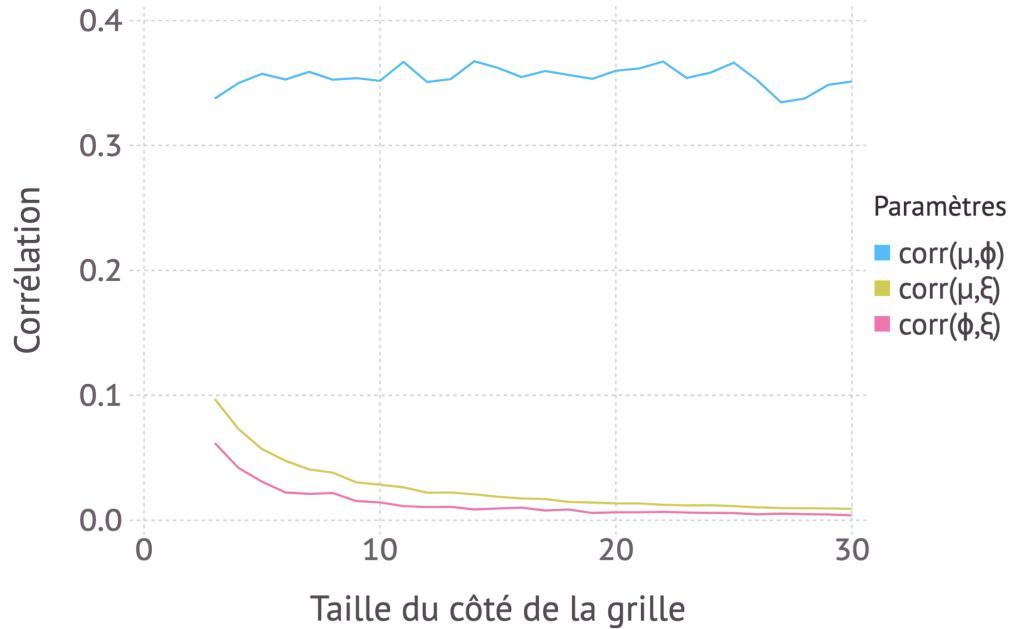


FIGURE A.1 Évolution des corrélations entre  $\xi$ ,  $\mu_1$  et  $\phi_1$  en fonction de la taille de la grille.

## ANNEXE B HYPOTHÈSE DE LINÉARITÉ DES VARIABLES LATENTES POUR LES PARAMÈTRES INTRA-CELLULAIRES

Dans la méthode développée, nous avons utilisé une hypothèse de linéarité des variables latentes pour les paramètres intra-cellulaires. Cette annexe montre pourquoi cette hypothèse est justifiée pour le modèle statistique utilisée.

$$GEV(\mu_i, \phi_i, \xi) \xrightarrow{\xi \rightarrow 0} \frac{1}{\exp \phi_i} \exp \left( -\frac{y_{ij} - \mu_i}{\exp \phi_i} \right) \exp \left\{ -\exp \left( -\frac{y_{ij} - \mu_i}{\exp \phi_i} \right) \right\}$$

donc, à l'aide de (3.17),

$$\begin{aligned} \log f_{\{(\mu_i, \phi_i) | y, \mu_{-i}, \phi_{-i}, \xi, \kappa_u, \kappa_v\}}(\mu_i, \phi_i) &\underset{\xi \rightarrow 0}{\approx} \sum_{j=1}^{n_{obs}} -\phi_i - \frac{y_{ij} - \mu_i}{\exp \phi_i} - \exp \left( -\frac{y_{ij} - \mu_i}{\exp \phi_i} \right) \\ &\quad - \frac{\kappa_u}{2} \left( n_i \mu_i^2 - 2\mu_i \sum_{j:j \sim i} \mu_j \right) \\ &\quad - \frac{\kappa_v}{2} \left( n_i \phi_i^2 - 2\phi_i \sum_{j:j \sim i} \phi_j \right). \end{aligned}$$

Cette expression est linéaire en  $\kappa_u$ ,  $\kappa_v$  et  $\mu_j$ ,  $\phi_j$ ,  $j : j \sim i$ . Elle est donc linéaire en tous les paramètres exceptés  $(\mu_i, \phi_i)$  et respecte donc (5.6). On peut alors extrapoler ce résultat pour  $\xi \approx 0$ .

## ANNEXE C CALCUL DES LOIS D'APPROXIMATION OPTIMALES

Pour  $\kappa_u$

On rappelle que la loi conditionnelle complète de  $\kappa_u$  s'écrit

$$f_{(\kappa_u|y,\mu,\phi,\xi,\kappa_v)}(\kappa_u) \propto \kappa_u^{\frac{m-1}{2}} \exp\left(\frac{-\kappa_u}{2}\boldsymbol{\mu}^T W \boldsymbol{\mu}\right) \exp\left(\frac{-\kappa_u}{100}\right)$$

donc

$$\begin{aligned} \log f_{(\kappa_u|y,\mu,\phi,\xi,\kappa_v)}(\kappa_u) &= \frac{m-1}{2} \log \kappa_u - \kappa_u \frac{\boldsymbol{\mu}^T W \boldsymbol{\mu}}{2} - \frac{\kappa_u}{100} + cste \\ &= \frac{m-1}{2} \log \kappa_u - \kappa_u \left( \frac{\boldsymbol{\mu}^T W \boldsymbol{\mu}}{2} + \frac{1}{100} \right) + cste \end{aligned}$$

Pour obtenir  $q_u^*$  on applique l'opérateur  $E_{\setminus \kappa_u}[\cdot]$  donc

$$\begin{aligned} \log q_u^*(\kappa_u) &= \frac{m-1}{2} \log \kappa_u - \kappa_u \left( \frac{1}{2} E_{\setminus \kappa_u}[\boldsymbol{\mu}^T W \boldsymbol{\mu}] + \frac{1}{100} \right) + cste \\ &= \frac{m-1}{2} \log \kappa_u - \kappa_u \left( \frac{1}{2} E_{\boldsymbol{\mu}}[\boldsymbol{\mu}^T W \boldsymbol{\mu}] + \frac{1}{100} \right) + cste \end{aligned} \quad (\text{C.1})$$

On reconnaît alors la forme fonctionnelle d'une loi  $\Gamma\left(\frac{m-1}{2} + 1, \frac{E_{\boldsymbol{\mu}}[\boldsymbol{\mu}^T W \boldsymbol{\mu}]}{2} + \frac{1}{100}\right)$ .

Calcul de  $E_{\boldsymbol{\mu}}[\boldsymbol{\mu}^T W \boldsymbol{\mu}]$

D'après (3.17)

$$\boldsymbol{\mu}^T W \boldsymbol{\mu} = \sum_{i=1}^m \left( n_i \mu_i^2 - \sum_{j:j \sim i} \mu_i \mu_j \right)$$

Or d'une part la famille d'approximation (5.1) suggère que pour  $i, j \in \mathbb{N}$ , si  $i \neq j$  alors  $\mu_i|y \perp \mu_j|y$ , ce qui implique

$$E_{\boldsymbol{\mu}}[\mu_i \mu_j] = E_{\mu_i}[\mu_i] E_{\mu_j}[\mu_j]$$

et d'autre part

$$E_{\mu_i}[\mu_i^2] = \text{Var}[\mu_i] + E_{\mu_i}[\mu_i]^2, \quad i \in \mathbb{N}$$

donc finalement

$$\begin{aligned} \text{E}_{\boldsymbol{\mu}}[\boldsymbol{\mu}^T W \boldsymbol{\mu}] &= \sum_{i=1}^m \left\{ n_i (\text{Var}[\mu_i] + \text{E}_{\mu_i}[\mu_i]^2) - \text{E}_{\mu_i}[\mu_i] \sum_{j:j \sim i} \text{E}_{\mu_j}[\mu_j] \right\} \\ &= \sum_{i=1}^m n_i (\text{Var}[\mu_i] + \text{E}_{\mu_i}[\mu_i]^2) - \sum_{i=1}^m \left\{ \text{E}_{\mu_i}[\mu_i] \sum_{j:j \sim i} \text{E}_{\mu_j}[\mu_j] \right\} \end{aligned} \quad (\text{C.2})$$

Pour  $\kappa_v$

La méthode est exactement la même. Les résultats s'obtiennent donc en remplaçant  $\boldsymbol{\mu}$  par  $\boldsymbol{\phi}$  dans les expressions (C.1) et (C.2) :

$$\log q_v^*(\kappa_v) = \frac{m-1}{2} \log \kappa_v - \kappa_v \left( \frac{1}{2} \text{E}_{\boldsymbol{\phi}}[\boldsymbol{\phi}^T W \boldsymbol{\phi}] + \frac{1}{100} \right) + cste \quad (\text{C.3})$$

$$\text{E}_{\boldsymbol{\phi}}[\boldsymbol{\phi}^T W \boldsymbol{\phi}] = \sum_{i=1}^m \left\{ n_i (\text{Var}[\phi_i] + \text{E}_{\phi_i}[\phi_i]^2) - \text{E}_{\phi_i}[\phi_i] \sum_{j:j \sim i} \text{E}_{\phi_j}[\phi_j] \right\} \quad (\text{C.4})$$

## ANNEXE D ÉTUDE COMPARATIVE DES ALGORITHMES MCMC

Une étude comparative entre différentes méthodes MCMC permet d'identifier celle qui servira de référence pour évaluer la méthode variationnelle proposée. Quatre algorithmes différents de type Metropolis-Hastings dans Gibbs (section 3.2.6) sont testés :

1. Metropolis dans Gibbs de marche aléatoire gaussienne.
2. MALA dans Gibbs.
3. NUTS dans Gibbs.
4. Metropolis dans Gibbs optimisé via l'hypothèse de Markov (Jalbert *et al.*, 2022).

Pour les comparer, les algorithmes sont appliqués à un même jeu de données comportant des données de précipitations extrêmes simulées sur une grille de taille  $3 \times 3$ . La métrique permettant de les discriminer est la taille d'échantillon effective (ESS, Gelman *et al.*, 2013) par seconde, qui mesure la qualité d'un échantillon généré. Plus elle est élevée, moins les valeurs au sein d'un même échantillon sont corrélées, donc meilleure est l'estimation de la loi *a posteriori*. Pour l'obtenir, l'ESS par seconde de la chaîne de chaque  $\mu_i$  ( $i = 1, \dots, 9$ ) est calculée, puis moyennée sur les 9 chaînes.

Le tableau D.1 présente les configurations globales de chaque algorithme testé et la valeur de leur ESS par seconde moyenne sur les 9 chaînes des  $\{\mu_i\}_{i=1,\dots,9}$ . Il montre que l'algorithme Metropolis dans Gibbs optimisé de Jalbert *et al.* (2022) est le plus efficace. C'est donc lui qui sera retenu comme algorithme de référence pour évaluer la solution variationnelle.

TABLEAU D.1 Comparaison des algorithmes MCMC. Le meilleur algorithme (Metropolis dans Gibbs optimisé) est mis en évidence en gras dans la colonne grise.

Algorithme	Metropolis dans Gibbs	MALA dans Gibbs	NUTS dans Gibbs	Metropolis dans Gibbs optimisé
Nombre d'itérations	10000	10000	1000	<b>10000</b>
Nombre d'itérations de chauffe	2000	2000	200	<b>2000</b>
Temps d'exécution (s)	0.737	0.646	12.8	<b>0.253</b>
ESS/s	$759 \pm 141$	$1019 \pm 281$	$53 \pm 5$	<b><math>2332 \pm 406</math></b>