UNIVERSITÉ DU QUÉBEC
INSTITUT NATIONAL DE LA RECHERCHE SCIENTIFIQUE
CENTRE EAU TERRE ENVIRONNEMENT

SÉRIES DE DURÉES PARTIELLES : APPLICATION EN CONTEXTE NON HOMOGÈNE ET NON STATIONNAIRE

Par

Simon LACHANCE-CLOUTIER

Mémoire présenté pour l'obtention du grade de *Magister Scientiae*, M.Sc.
en sciences de l'eau

Jury d'évaluation

Président du jury et examinateur externe	
Examinateur externe	Théo Vischel Université Joseph Fourier
Examinateur interne	David Huard Consortium Ouranos
Directeur de recherche	Alain Mailhot INRS, Centre-Eau, Terre et Environnement
Codirectrice de recherche	Anne-Catherine Favre Université Laval

© Droits réservés de Simon LACHANCE-CLOUTIER, 2011

RÉSUMÉ

Les extrêmes de précipitations peuvent entraîner des dégâts importants lorsque les infrastructures ne sont pas conçues en conséquence. L'analyse de ces extrêmes est cependant difficile puisqu'ils sont associés à des événements rares et que les séries de précipitations disponibles couvrent souvent de courtes périodes. De plus, dans un contexte de changements climatiques, on anticipe une augmentation de l'amplitude et de la fréquence de ces extrêmes. On doit alors obligatoirement considérer la non stationnarité des séries. Deux approches sont généralement considérées pour l'analyse fréquentielle statistique des extrêmes de précipitations : la méthode des séries de maxima annuels (SMA) et la méthode des séries de durées partielles (SDP). La première approche consiste à construire les séries des maxima annuels de précipitations pour une durée donnée. L'approche SDP se présente comme une alternative à l'approche SMA et permet de considérer plusieurs évènements par année en retenant tous ceux excédant un certain seuil. Comparativement à l'approche SMA, l'utilisation de cette méthode permet d'accroître la taille des séries et de bénéficier ainsi de plus d'informations.

Le présent mémoire vise à développer une méthodologie pour l'utilisation des SDP dans un contexte non stationnaire. Les variabilités intra-annuelle et interannuelle des extrêmes sont considérées. L'approche développée sera dans un premier temps appliquée à l'analyse des extrêmes en climat futur tels que simulés par le Modèle canadien Couplé de Climat Global de troisième génération (MCCG3) et, dans un deuxième temps, à l'analyse de données de stations pluviométriques réparties au sud du Québec. Dans le cas des données du MCCG3, l'application de la méthode conclut à une augmentation des extrêmes de précipitations en climat futur pour l'ensemble du territoire. L'analyse des extrêmes saisonniers révèle que pour la saison estivale (mois de juin, juillet et août), la valeur du total journalier de précipitation, dépassé en moyenne une fois tous les 20 ans, diminuera en climat futur pour les régions situées au sud du 50^{e} parallèle. L'hiver (mois de décembre, janvier et février) est la saison pour laquelle l'augmentation relative des extrêmes sera la plus importante.

Pour l'application aux données de séries observées, l'approche SDP a été régionalisée afin d'améliorer l'inférence de certains paramètres. Plusieurs modèles ont été proposés, considérant

des valeurs locales ou régionales de chacun des paramètres de la distribution de Pareto généralisée (GPD). Différentes covariables spatiales ont également été introduites dans ces modèles afin de tenir compte d'une possible structure spatiale des paramètres. Les différents modèles ont été classés en vertu du critère d'Akaike corrigé (AIC_c). Les résultats obtenus montrent, à l'instar des résultats du MCCG3, la présence d'une variabilité intra-annuelle très significative des paramètres. Les modèles comportant une variabilité intra-annuelle des valeurs régionales des paramètres d'échelle et de forme représentent de loin les meilleurs candidats selon le critère AIC_c. L'introduction de covariables spatiales dans le modèle révèle une structure spatiale au niveau des paramètres et améliore encore davantage la qualité des ajustements. Des trois covariables spatiales considérées, à savoir la latitude, la longitude et l'élévation, la latitude est généralement celle qui améliore le plus l'inférence. Les cartes de cumuls de précipitations pour le temps de retour 10 ans pour différentes durées montrent l'existence d'un gradient nord-sud très évident avec des valeurs plus élevées au sud du territoire.

AVANT-PROPOS

"It is not the strongest of the species that survives, nor the most intelligent that survives. It is the one that is the most adaptable to change."

Charles Darwin

Cette citation du célèbre naturaliste anglais et auteur de la théorie de l'évolution nous rappelle à quel point l'adaptabilité est fondamentale. Dans un contexte de changements climatiques, cette affirmation de Darwin prend une saveur toute particulière. Tandis qu'il serait sans doute exagéré et alarmiste de dire que la survie de l'homme, en tant qu'espèce, est menacée par ces changements, force est d'admettre que les sociétés qui subiront le moins de contrecoups seront celles ayant les moyens, l'expertise pour mettre en place les mesures d'adaptation appropriées. Mais avant d'entreprendre, il faut comprendre. En ce domaine, il reste beaucoup de travail à faire afin de bien cerner l'enjeu et les conséquences probables de ces changements sur notre environnement. J'espère que ce mémoire saura y contribuer ne serait-ce que d'une infime façon.

REMERCIEMENTS

Dans un premier temps j'aimerais remercier ma famille immédiate, mon père Gilles, ma mère Anne ainsi que ma sœur cadette, Stéphanie. Sans leur encadrement et l'environnement familial qu'ils m'ont offert, je ne serais sans doute pas où j'en suis aujourd'hui. Il y a une part d'eux dans tout ce que je fais et je leur dois bien.

Merci également à tous mes amis et collègues pour leur support, les débats animés et les bons moments passés à refaire le monde autour d'une chope. Sans eux, ces deux années n'auraient pas filé à la vitesse d'une comète.

Un merci particulier à Alain, mon directeur, qui a fait office de mentor au cours de ces travaux. Son franc-parler et son sens critique m'ont permis de m'améliorer à plusieurs égards. J'ai beaucoup appris sous son égide, non seulement sur les statistiques des extrêmes et sur le travail en milieu de recherche, mais sur moi-même également. On dit souvent que dans la vie il faut savoir s'entourer de gens qui nous font grandir, à ce chapitre je crois être très bien tombé avec Alain. Merci pour tout.

Finalement j'aimerais remercier ma codirectrice, Anne-Catherine Favre, pour ses bons conseils et tous ceux que j'ai omis de nommer et qui m'ont accompagné de près ou de loin lors de mes années à la maîtrise. Que ce soit par leur aide, leurs commentaires, leurs suggestions ou leur encadrement.

À la mémoire de mon père.

TABLE DES MATIÈRES

RÉSUMÉ III
AVANT-PROPOS V
REMERCIEMENTS VII
TABLE DES MATIÈRES IX
LISTE DES TABLEAUX XI
LISTE DES FIGURES XIII
LISTE DES ABRÉVIATIONS ET DES SIGLES XVII
1. INTRODUCTION 1
2. MÉTHODE DES SÉRIES DE DURÉES PARTIELLES 5
2.1 Distribution de Pareto généralisée 5
2.2 Estimation des paramètres 7
2.3 Sélection du seuil 9
2.4 Declustering 11
3. MÉTHODE DES SÉRIES DE DURÉES PARTIELLES EN RÉGIME NON STATIONNAIRE 13
3.1 Estimation des paramètres en régime non stationnaire 13
3.2 Variabilité intra-annuelle 14
3.2.1 Séries non homogènes de Pareto généralisées 15
4. APPLICATION AUX SÉRIES DES PRÉCIPITATIONS INTENSES SIMULÉES PAR LE MODÈLE CANADIEN COUPLÉ DE CLIMAT GLOBAL (MCCG3) 17
4.1. Données utilisées 17
4.2 Estimation du seuil en régime non stationnaire 19
4.3 Modélisation de la dépendance temporelle des paramètres 23
4.4 Stabilisation de la fréquence de dépassement du seuil 26
4.5 CJP extrêmes en climat futur 28
4.6 Variabilité intra-annuelle 31
4.7 Sensibilité au taux annuel moyen utilisé pour fixer le seuil 37
4.8 Comparaison avec la littérature 41

5. APPLICATION AUX SÉRIES OBSERVÉES 45
5.1 Données utilisées 45
5.2 Estimation du seuil 47
5.3 Modélisation de la variation temporelle des paramètres 50
5.3.1 Représentation en séries de Fourier 51
5.3.2 Séries mensuelles 52
5.4 Analyse locale versus régionale 52
5.4.1 Régionalisation 53
5.5 Critère d'Akaike (Akaike's Information Criterion) 56
5.6 Modèles locaux et régionaux 59
5.6.1 Proposition de modèles 59
5.6.2 Classement AIC_c des modèles 63
5.7 Modèles régionaux avec covariables spatiales 67
5.7.1 Modélisation de la dépendance spatiale des paramètres 68
5.7.2 Proposition de modèles régionaux avec covariables spatiales 69
5.7.3 Classement AIC_c des modèles régionaux avec dépendance spatiale 70
5.7.4 Variation spatiale et intra-annuelle des paramètres 74
5.7.5 Estimations des cumuls de précipitations extrêmes 75
5.7.6 Sensibilité des résultats au modèle retenu 77
5.7.7 Sensibilité au taux annuel moyen retenu pour définir le seuil 79
6. CONCLUSION 83
BIBLIOGRAPHIE 89
ANNEXE A : DÉTAILS SUR L'INTERPOLATION DES PALIERS DES VALEURS MENSUELLES DES PARAMÈTRES DE LA GPD. 93
ANNEXE B : VALEUR DE CJP DÉPASSÉE EN MOYENNE N FOIS EN 20 ANS POUR DIFFÉRENTES PÉRIODES 97
ANNEXE C : AUGMENTATION RELATIVE DES VALEURS DE CJP DÉPASSÉES EN MOYENNE N FOIS EN 20 ANS EN CLIMAT FUTUR COMPARATIVEMENT À LA PÉRIODE 1850-1950 101
ANNEXE D : MOYENNE SUR LES TUILES D'UNE LATITUDE DONNÉE DE LA VARIATION RELATIVE DU CJP DÉPASSÉ EN MOYENNE N FOIS EN 20 ANS AU COURS DE DIFFÉRENTES SAISONS 105
ANNEXE E : DÉTAILS DE LA MÉTHODE D'OBTENTION DU SEUIL POUR LES SÉRIES DE STATIONS MÉTÉOROLOGIQUES 107
ANNEXE F : CARTES POUR LE SUD DU QUÉBEC DU CUMUL DE PRÉCIPITATION DE TEMPS DE RETOUR 10 ANS POUR DIFFÉRENTES DURÉES. 109

LISTE DES TABLEAUX

Tableau 3.1 : Paramètres utilisés afin de générer deux échantillons issus d'une GPD et paramètres obtenus à la suite de la calibration d'une GPD sur la série non homogène. 15

Tableau 4.1 : Pourcentage des tuiles dont les SDP sont considérées stationnaires par le test de Mann-Kendall avec un seuil α = 0.05. 22

Tableau 4.2 : Écarts relatifs par rapport à la période 1980-2000 du CJP dépassé en moyenne une fois en 20 ans (P_{20}). 41

Tableau 4.3 : Anomalies de température et taux d'augmentation par degré d'anomalie des CJP par rapport à la période 1980-2000. 42

Tableau 4.4 : Nombre moyen de dépassements de P_{20} (période 1980-2000) durant les diverses périodes considérées 42

Tableau 5.1 : Subdivision du territoire en trois sous-régions 48

Tableau 5.2 : Fraction des séries SDP considérées stationnaires par le test de Mann-Kendall avec un seuil α =0.05. 50

Tableau 5.3 : Nomenclature des modèles présentant un paramètre de forme régional 61

Tableau 5.4 : Nomenclature des modèles présentant un paramètre de forme local. 62

Tableau 5.5 : Nomenclature des modèles construits en blocs mensuels 62

Tableau 5.6 : Classe des modèles qui présentent au moins une covariable spatiale. 69

Tableau 5.7 : Nomenclature des modèles régionaux avec covariables spatiales proposés. 70

LISTE DES FIGURES

Figure 2.1 : Moyenne des écarts au seuil en fonction du seuil pour une série de 7300 valeurs issues d'une distribution normale de paramètres $\mu = 0$ et $\sigma = 3$ tronquée en zéro 10

Figure 2.2 : Exemple d'une série avant (a) et après (b) l'exercice de *declustering*........ 12

Figure 3.1 : Graphique quantile-quantile (qqplot) montrant la comparaison entre les valeurs associées aux quantiles d'une GPD ajustée (GPD-fit) à une série non homogène issue d'une combinaison de GPD et les valeurs associées aux quantiles empiriques de la série non homogène (courbe continue). 16

Figure 4.1 : Territoire couvert par l'étude et discrétisation de l'espace en 420 tuiles par le MCCG3. 18

Figure 4.2 : Division de la période 1850-2100 en 48 séries pseudo-homogènes 19

Figure 4.3 : Valeurs des seuils correspondant à une probabilité quotidienne de dépassement p = 12/365 ≈ 0.0329 pour le mois de décembre (en moyenne 1 dépassement par mois de décembre) pour la tuile en encadré rouge de la Figure 4.1. 20

Figure 4.4 : Variabilités inter et intra-annuelle du seuil pour la période 1950-1952 pour un taux moyen de 12 dépassements par année, soit environ un par mois, pour la tuile de l'encadré rouge de la Figure 4.1.. 21

Figure 4.5 : Évolution du paramètre d'échelle pour le mois de décembre pour la tuile de l'encadré rouge de la Figure 4.1. 24

Figure 4.6 : Variabilités inter et intra-annuelle du paramètre d'échelle pour la période 1950-1952 pour la tuile de l'encadré rouge de la Figure 4.1 25

Figure 4.7 : Probabilité mensuelle de dépassements du seuil pour différentes périodes pour un taux moyen de 12 dépassements par an (un dépassement par mois en moyenne) 26

Figure 4.8 : Distribution du nombre annuel de dépassements du seuil pour un seuil correspondant à un taux moyen fixé à 12 dépassements par an. 27

Figure 4.9 : Hauteur (mm) du CJP dépassé en moyenne une fois en 20 ans sur chaque période considérée. 29

Figure 4.10 : Rapports des CJP dépassés en moyenne une fois chaque 20 ans pour les périodes 2020-2040 et 2060-2080 comparativement à la période de référence 1850-1950. 30

Figure 4.11 : Variation relative du CJP dépassé en moyenne une fois chaque 20 ans au cours des différentes saisons de la période 2060-2080 relativement à leurs homologues de la période 1850-1950. 32

Figure 4.12 : Moyenne sur les tuiles d'une latitude donnée de la variation relative du CJP dépassé en moyenne une fois chaque 20 ans au cours des différentes saisons de la période 2060-2080 et leurs homologues de la période 1850-1950 33

Figure 4.13 : Les six tuiles retenues (encadrées en rouge) pour l'inspection mensuelle des CJP extrêmes. 34

Figure 4.14 : CJP (mm) dépassé en moyenne une fois chaque 20 ans au cours de chacun des mois pour les tuiles retenues (Figure 4.13) 35

Figure 4.15 : Ratio des résultats des taux moyens $\lambda = 3$ et $\lambda = 6$ dépassements par rapport à $\lambda = 12$ des variations relatives du CJP dépassé en moyenne une fois chaque 20 ans au cours de la période 2060-2080 et de la période 1850-1950. 38

Figure 4.16 : Ratio des résultats pour $\lambda = 3$ par rapport à $\lambda = 12$ des variations relatives du CJP dépassé en moyenne une fois chaque 20 ans au cours des différentes saisons de la période 2060-2080 et de la période 1850-1950. 39

Figure 5.1 : Années pour lesquelles des mesures de mai à octobre sont disponibles pour les 109 stations.. 46

Figure 5.2 : Cartes des stations météorologiques retenues pour analyse.. 47

Figure 5.3 : Moyenne des seuils pour la durée 30 minutes pour un taux moyen de 6 dépassements MAO en fonction du jour julien pour les stations des trois zones 49

Figure 5.4 : Cumul (mm) sur 30 minutes de temps de retour 10 ans estimé selon le modèle 2L0L (voir Tableau 5.4). L'interpolation est effectuée par la méthode des polygones de Thiessen. 53

Figure 5.5 : Schéma du processus d'optimisation des paramètres pour l'analyse régionale. 55

Figure 5.6 : Arborescence des modèles possibles. 59

Figure 5.7 : Exemple pour le modèle 1L2R de la nomenclature utilisée. 60

Figure 5.8 : Classement des cinq meilleurs modèles sur les 40 modèles proposés 65

Figure 5.9 : Classement AIC_c des 40 modèles proposées pour une durée 30 minutes 66

Figure 5.10 : Exemple de nomenclature pour le modèle 2-1-LGE. 69

Figure 5.11 : Classements pour les huit durées des cinq meilleurs modèles sur les 36 modèles proposés. 71

Figure 5.12 : Classement AIC_c des 40 modèles proposés pour la durée 30 minutes. 73

Figure 5.13 : Valeur du paramètre d'échelle et du paramètre de forme en fonction du jour julien en différents points du territoire. 75

Figure 5.14 : Cumul de précipitation sur 30 minutes (mm) de temps de retour 10 ans selon le modèle 2-1-LG. 76

Figure 5.15 : Écart relatif, par rapport au modèle 2-1-LG, des valeurs de cumul de précipitation sur 30 minutes de temps de retour 10 ans prédites par le modèle 2-2-LG et le modèle 2-2-L pour chacune des 107 stations. 78

Figure 5.16 : Classement des cinq meilleurs modèles pour la durée 30 minutes en fonction des différents taux annuels moyens de dépassement. 79

Figure 5.17 : Écarts relatifs (%) des prédictions entre le taux considéré et le taux de 6 dépassements pour chaque station du cumul de précipitation sur 30 minutes de temps de retour 10 ans. 81

Figure A-1 : Valeur du seuil pour le mois d'août de la tuile en encadré rouge de la Figure 4.1 93

Figure A-2 : Positionnements des nœuds pour la période 1950-2000 de la Figure A-1 94

Figure A-3 : Courbe obtenue à la suite du lissage des droites en tirets noirs reliant les différents nœuds par la méthode LOESS avec une portée de 75 ans pour le seuil du mois d'août de la tuile en encadré rouge de la Figure 4.1. 96

Figure B-1 : CJP (mm) dépassé en moyenne deux fois en 20 ans sur chaque période considérée. 97

Figure B-2 : CJP (mm) dépassé en moyenne quatre fois en 20 ans sur chaque période considérée. 98

Figure B-3 : CJP (mm) dépassé en moyenne 20 fois en 20 ans sur chaque période considérée. 99

Figure C-1 : Rapports des CJP dépassés en moyenne deux fois en 20 ans pour les périodes 2020-2040 et 2060-2080 comparativement à la période de référence 1850-1950. 101

Figure C-2 : Rapports des CJP dépassés en moyenne quatre fois en 20 ans pour les périodes 2020-2040 et 2060-2080 comparativement à la période de référence 1850-1950. 102

Figure C-3 : Rapports des CJP dépassés en moyenne 20 fois en 20 ans pour les périodes 2020-2040 et 2060-2080 comparativement à la période de référence 1850-1950. 103

Figure D-1 : Moyenne sur les tuiles d'une latitude donnée de la variation relative du CJP dépassé en moyenne 0.5 fois en 20 ans ($N = 0.5$) au cours des différentes saisons de la période 2060-2080 et leurs homologues de la période 1850-1950 105

Figure D-2 : Moyenne sur les tuiles d'une latitude donnée de la variation relative du CJP dépassé en moyenne deux fois en 20 ans ($N = 2$) au cours des différentes saisons de la période 2060-2080 et leurs homologues de la période 1850-1950. 106

Figure E-1 : Exemple de la sélection du seuil pour la station de l'aéroport Pierre Elliott Trudeau pour la durée 30 minutes.......... 108

Figure F-1 : Carte du cumul de précipitation (mm) sur 5 minutes de temps de retour 10 ans selon le modèle 2-1-LGE. 109

Figure F-2 : Carte du cumul de précipitation (mm) sur 10 minutes de temps de retour 10 ans selon le modèle 2-1-LGE. 109

Figure F-3 : Carte du cumul de précipitation (mm) sur 15 minutes de temps de retour 10 ans selon le modèle 2-2-LGE. 110

Figure F-4 : Carte du cumul de précipitation sur 1 heure (mm) de temps de retour 10 ans selon le modèle 2-1-LG.......... 110

Figure F-5 : Carte du cumul de précipitation (mm) sur 2 heures de temps de retour 10 ans selon le modèle 2-1-L. 111

Figure F-6 : Carte du cumul de précipitation (mm) sur 6 heures de temps de retour 10 ans pour le modèle 2-0-LGE.......... 111

Figure F-7 : Carte du cumul de précipitation (mm) sur 12 heures de temps de retour 10 ans selon le modèle 1-1-LGE. 112

LISTE DES ABRÉVIATIONS ET DES SIGLES

AIC	*Akaike Information Criterion*
CC	Changements climatiques
CJP	Cumul journalier de précipitation
DJF	Décembre Janvier Février
GES	Gaz à effet de serre
GEV	*Generalized Extreme Value*
GIEC	Groupe d'experts Intergouvernemental sur l'Évolution du Climat
GPD	*Generalized Pareto Distribution*
i.i.d.	Indépendamment et identiquement distribués
JJA	Juin Juillet Août
MAM	Mars Avril Mai
MAO	Mai à Octobre
MCCG3	Modèle canadien couplé de climat global de troisième génération
MCG	Modèle de circulation générale
MRC	Modèle régional du climat
NARCCAP	North American Regional Climate Change Assessment Program
SDP	Séries de durées partielles
SMA	Séries de maxima annuels
SON	Septembre Octobre Novembre
SRES	*Special report on emission scenarios*

1. INTRODUCTION

La question des changements climatiques (CC) anime plus que jamais les tribunes, tant au niveau sociétal, politique que scientifique. Pour cette dernière communauté, l'origine des changements fait presque l'unanimité. Des rapports tels que celui du GIEC (2007) et des études comme celle de Zhang *et al.* (2007) démontrent que l'activité humaine, par l'émission de GES (gaz à effet de serre) et aérosols, est à l'origine de ces chambardements. L'augmentation de la concentration des GES et aérosols dans l'atmosphère modifie le bilan radiatif terrestre. Cela a pour conséquence de réduire les pertes d'énergie par radiation infrarouge de la surface terrestre. Il en résulte une augmentation graduelle de la température moyenne de surface ainsi qu'une multitude de bouleversements étant donnée le rôle clé que le bilan radiatif joue dans les processus climatiques. On notera, entre autres, la hausse du niveau des océans (Titus et Narayanan, 1996), la fonte des glaces (Gregory et Oerlemans, 1998) ainsi qu'une modification des régimes de précipitations (Dai *et al.,* 1998). Ces faits établis, il s'agit désormais d'anticiper adéquatement les impacts des CC. Pour ce faire, il importe d'abord d'établir des projections du climat futur.

Un outil incontournable à cet effet réside dans les modèles de circulation générale (MCG). Ces derniers intègrent une représentation mathématique des principaux processus régissant le climat terrestre et permettent ainsi de simuler l'évolution du climat. Ces modèles requièrent de définir des scénarios d'émission de GES et aérosols afin de décrire l'évolution des concentrations et des émissions de GES en climat futur. Le rapport spécial sur les scénarios d'émission de GES (Nakicenovic *et al.*, 2000) présente plusieurs scénarios susceptibles de représenter la réalité future. Puisque celle-ci est *de facto* incertaine, il importe en effet de pouvoir explorer différents scénarios possibles, des plus pessimistes aux plus optimistes. Il va de soi qu'une part de l'incertitude associée aux impacts des CC est inhérente au scénario d'émission de GES retenu.

L'étude des précipitations, et particulièrement des épisodes pluvieux extrêmes, est importante à plusieurs égards. Ce type d'information est primordial pour la conception de plusieurs ouvrages hydrauliques (par exemple réseaux de collecte des eaux pluviales, bassins de rétention, ponceaux...). On a alors recours à l'intensité d'une pluie de temps de retour donné pour leur dimensionnement, fixant ainsi le niveau de service jugé adéquat (Mailhot *et al.*, 2007). La description adéquate des évènements extrêmes est donc cruciale dans un contexte d'analyse du risque. On anticipe par ailleurs que l'intensité des extrêmes de précipitation augmentera de manière plus substantielle que les précipitations moyennes (Emori et Brown, 2005). En effet, l'intensité des événements extrêmes est liée à la capacité de l'atmosphère à emmagasiner de l'humidité alors que les précipitations de plus faible intensité sont régies principalement par le bilan d'énergie. Dans cette optique, le GIEC (2007), dans son rapport de synthèse, conclut à une augmentation très probable de la fréquence d'occurrence des événements extrêmes pour la plupart des régions. De même, Min *et al.* (2011) montrent que l'émission anthropique de gaz à effet de serre a contribué en partie à l'augmentation de l'intensité des précipitations extrêmes observée au cours de la seconde moitié du XX^{e} siècle.

Différentes méthodes statistiques sont utilisées pour extraire l'information contenue dans les séries hydroclimatiques. L'analyse statistique des séries issues des modèles climatiques dans un contexte de CC pose cependant un défi additionnel puisque ces séries sont non stationnaires. La non stationnarité d'une série signifie que les moments de la loi statistique qui régissent le phénomène varient en fonction du temps (Meylan *et al.*, 2008). En ce qui a trait à l'analyse fréquentielle des extrêmes de précipitations en contexte stationnaire, deux approches sont généralement considérées, basées sur : 1) les séries des maxima annuels, et 2) les séries de durées partielles. La méthode des séries de maxima annuels (SMA ou AMS pour l'acronyme anglais), que l'on appelle également « block maxima approach » (Coles, 2001), consiste à construire des séries en ne conservant que le cumul de précipitations maximal enregistré pendant une durée donnée (par exemple une heure) pour une année donnée. Cette approche jouit, comparativement à sa rivale, d'une littérature plus abondante quant à son utilisation en régime non stationnaire. L'approche par séries de durées partielles (SDP ou POT en anglais) se présente comme une alternative à l'approche SMA. Plutôt que de considérer seulement le cumul annuel maximal, la méthode SDP permet de considérer plusieurs cumuls de précipitation en retenant tous ceux qui excèdent un certain seuil. L'avantage de pouvoir considérer un plus grand nombre de données impose cependant de sélectionner un seuil adéquat. L'usage des SDP requiert également de s'intéresser à la question de l'autocorrélation afin de retenir des cumuls indépendants (Madsen *et al.*, 1997a). Ce sont principalement les difficultés associées à ces deux étapes, l'indépendance des cumuls de précipitation et la sélection d'un seuil adéquat, qui posent un frein à un usage plus répandu de l'approche SDP.

Ce mémoire se donne comme objectif principal d'adapter l'approche SDP pour une utilisation en régime non stationnaire. La variabilité intra-annuelle des occurrences sera également prise en compte en permettant aux paramètres de la distribution de varier en fonction du moment dans l'année, permettant ainsi de reproduire le patron saisonnier d'occurrence des extrêmes de précipitations. Dans la seconde partie du mémoire, l'approche sera régionalisée et appliquée à l'analyse de séries observées. Le document est organisé comme suit. Le chapitre 2 présente la méthode des SDP en régime stationnaire. Il y est question de la loi statistique qui régit la distribution des cumuls de précipitation au-dessus d'un seuil, des principales méthodes d'estimation des paramètres, de la sélection du seuil et de la notion de *declustering*. Le chapitre 3 traite de l'emploi des SDP en régime non stationnaire. Les questions d'estimation des paramètres et de variabilité intra-annuelle y sont abordées. Le chapitre 4 présente une application aux séries de précipitations intenses simulées par le modèle climatique couplé global de troisième génération (MCCG3) d'Environnement Canada. Le domaine couvert comprend l'ensemble du Canada ainsi que la partie septentrionale des États-Unis. Les cumuls journaliers de précipitations (CJP) extrêmes sont étudiés. La méthodologie employée afin d'estimer le seuil et de représenter la dépendance temporelle des paramètres est exposée ainsi que les changements anticipés dans les valeurs extrêmes de CJP en climat futur. L'importance des changements sur chacune des saisons est étudiée. Par la suite, la sensibilité de l'approche en regard du seuil servant à construire les SDP est évaluée. Finalement, une section du chapitre est consacrée à la comparaison des résultats avec ceux mentionnés dans la littérature.

Le chapitre 5 porte sur l'utilisation de l'approche SDP dans un cadre de régionalisation et sur la sélection d'un modèle adéquat. Les données analysées sont des cumuls de précipitations sur différents pas de temps provenant de stations jaugées réparties au sud de la province de Québec. La stationnarité de ces séries historiques a d'abord été vérifiée. L'analyse se concentre sur la variation intra-annuelle des paramètres. La régionalisation est employée dans le but d'améliorer l'inférence statistique. L'objectif recherché dans le cadre de cette partie est de développer une méthodologie permettant la sélection d'un modèle optimal à l'échelle régionale. Finalement, le chapitre 6 synthétise les principaux constats et résultats de ce travail et énonce des perspectives.

2. MÉTHODE DES SÉRIES DE DURÉES PARTIELLES

L'approche par séries de durées partielles est une méthode d'analyse des extrêmes qui consiste à ne retenir que les valeurs supérieures à un seuil donné. La distribution binomiale $B(n,p)$ permet de modéliser l'occurrence de ces dépassements. Lorsque la probabilité p d'observer un dépassement est suffisamment faible (p < 0.1) et que le nombre d'essais n est suffisamment grand (n > 50), la distribution du nombre annuel de dépassements peut être estimée par une distribution de Poisson $P(\lambda)$ d'intensité $\lambda = n\,p$ (Hines et Montgomery, 1990).

La distribution de l'intensité des dépassements est quant à elle décrite par une Distribution de Pareto Généralisée (GPD) (Pickands, 1975). Le seuil doit cependant être suffisamment élevé afin de ne pas violer la théorie asymptotique sur laquelle s'appuie le développement de la GPD (Coles, 2001). Pour que ces distributions soient valides, il importe également que les échantillons soient indépendants et identiquement distribués (i.i.d.). Cela implique que chaque terme de la série soit issue de la même distribution et que ces valeurs soient indépendantes.

2.1 Distribution de Pareto généralisée

La densité de probabilité de la GPD s'écrit sous la forme :

$$f\left(x\middle|k,\alpha,\xi\right)=\begin{cases}\dfrac{1}{\alpha}\left[1-k\dfrac{(x-\xi)}{\alpha}\right]^{\frac{1}{k}-1} & k\neq 0\\[2ex] \dfrac{1}{\alpha}\exp\left(-\dfrac{x-\xi}{\alpha}\right) & k=0\end{cases} \tag{2.1}$$

avec les paramètres de position (ξ) (seuil), d'échelle (α) et de forme (k). Elle est définie pour $\xi \leq x \leq \xi + \dfrac{\alpha}{k}$ si $k > 0$ et $\xi \leq x < \infty$ si $k \leq 0$. La fonction est bornée lorsque $k > 0$ et non bornée lorsque $k \leq 0$; on parle alors d'une distribution de type « heavy-tailed ».

La fonction de répartition prend la forme :

$$F(x \mid k,\alpha,\xi) = \begin{cases} 1-\left[1-\dfrac{k(x-\xi)}{\alpha}\right]^{\frac{1}{k}} & k \neq 0 \\ \\ 1-\exp\left(\dfrac{-(x-\xi)}{\alpha}\right) & k=0 \end{cases} \tag{2.2}$$

Le cas spécial $k=0$ correspond à une distribution exponentielle de moyenne α. La notation employée ici est celle de Hosking et Wallis (1987). Il est cependant important de noter que certains auteurs utilisent la notation $k=-k$.[1]

La GPD possède une particularité intéressante lorsque l'on augmente le seuil. On peut montrer, dans ce cas, que la distribution demeure une GPD dont les paramètres peuvent s'exprimer en fonction des valeurs de ces derniers avant changement du seuil (Madsen *et al.*, 1997a). Ainsi, pour une série GPD décrite par $f(x \mid k,\alpha,\xi)$, pour tout $\xi' > \xi$, la série demeurera une GPD décrite par $f(x \mid k',\alpha',\xi')$ où $k'=k$ et $\alpha' = \alpha - k\,(\xi' - \xi)$. Pour une série homogène, ce résultat implique qu'au-delà d'un certain seuil minimal, les valeurs seront assurément distribuées selon une GPD.

---

[1] Lorsque l'on travaille avec les fonctions intégrées de Matlab par exemple, il faut prendre garde d'inverser le signe de k.

Il existe un lien entre les paramètres de la GPD et ceux de la distribution GEV servant à décrire les SMA. Pour ces distributions, le paramètre de forme (k) est le même et les paramètres de la GEV peuvent être dérivés de ceux de la GPD (Madsen *et al.,* 1997a) :

$$\mu = \xi + \alpha \ln(\lambda) \qquad k = 0$$
$$\mu = \xi + \frac{\alpha}{k}\ln(1-\lambda^{-k}) \qquad k \neq 0 \tag{2.3}$$

$$\sigma = \alpha\lambda^{-k} \tag{2.4}$$

où μ et σ sont respectivement les paramètres de position et d'échelle de la GEV, ξ et α le seuil et le paramètre d'échelle de la GPD, k le paramètre de forme et λ est l'intensité du processus de Poisson décrivant le processus d'occurrence de dépassement du seuil.

La raison de cette relation est assez simple. La probabilité qu'un maximum annuel soit plus petit ou égal à x tel que décrit par la GEV est donnée par la probabilité de ne pas excéder x, probabilité que permet d'évaluer la GPD. Coles (2001) discute plus amplement de la relation entre la GPD et la GEV.

2.2 Estimation des paramètres

Des trois paramètres de la GPD, seuls les paramètres d'échelle et de forme nécessitent d'être estimés à partir des observations, le seuil (position) devant être préalablement sélectionné afin de construire les SDP. Différentes méthodes peuvent être employées afin d'obtenir une estimation des paramètres, notamment la méthode des moments et du maximum de vraisemblance. Pour une description plus exhaustive des différentes méthodes d'estimation des paramètres de la GPD, le lecteur peut se référer à Bermudez et Kotz (2009).

La méthode des moments (MOM) permet d'estimer les paramètres à partir des deux premiers moments empiriques, soit la moyenne et l'écart-type de l'échantillon. Les estimateurs des paramètres α et k de la GPD pour la méthode des moments sont donnés par (Madsen *et al.*, 1997a) :

$$\hat{\alpha}_{\text{mom}} = \frac{1}{2}\hat{\mu}\left(\frac{\hat{\mu}^2}{\hat{\sigma}^2}+1\right) \tag{2.5}$$

$$\hat{k}_{\text{mom}} = \frac{1}{2}\left(\frac{\hat{\mu}^2}{\hat{\sigma}^2}-1\right) \tag{2.6}$$

où $\hat{\mu}$ et $\hat{\sigma}$ correspondent respectivement à la moyenne et à l'écart-type des écarts au seuil ($x-\xi$).

La méthode du maximum de vraisemblance consiste de son côté à trouver le couple de paramètres (α,k) qui maximise le logarithme de la vraisemblance. Les estimateurs des moments peuvent très bien servir de point de départ à l'optimisation de cette fonction.

Soit un échantillon de n valeurs dénotées $\{x_i\}$ avec $i=1,\ldots,n$. Le logarithme de la fonction de vraisemblance pour cette série s'écrit alors :

$$\ln\left\{\prod_{i=1}^{n} f(x_i)\right\} = \begin{cases} -n\ln(\alpha)+\sum_{i=1}^{n}\left\{\left(\frac{1}{k}-1\right)\ln\left(1-\frac{k(x_i-\xi)}{\alpha}\right)\right\} & k\neq 0 \\ -n\ln(\alpha)-\frac{1}{\alpha}\sum_{i=1}^{n}(x_i-\xi) & k=0 \end{cases} \tag{2.7}$$

Bien que l'estimateur du maximum de vraisemblance soit asymptotiquement non biaisé, il peut être biaisé pour des échantillons de petites tailles ($n < 100$). Il convient alors d'utiliser une autre méthode d'estimation, par exemple la méthode des moments. Hosking et Wallis (1987) montrent que, de manière générale, pour des échantillons de taille inférieure à 500, les méthodes des moments et des moments pondérés sont plus fiables que la méthode du maximum de vraisemblance. Les résultats de Hosking et Wallis (1987) et de Madsen *et al.* (1997a) montrent également que les performances relatives des méthodes d'estimation dépendent de la taille de l'échantillon, du quantile que l'on cherche à évaluer et de la valeur du paramètre de forme k.

2.3 Sélection du seuil

La principale difficulté de la méthode des SDP réside dans la détermination du seuil. Cette étape constitue un frein majeur à son utilisation au profit de méthodes plus simples d'applications telles que les SMA. Le seuil doit être suffisamment élevé de sorte que les valeurs retenues correspondent à la GPD. Un seuil trop bas se traduira par un biais dans l'estimation des paramètres de la distribution et par conséquent dans l'inférence des quantiles. À l'opposé, il faut éviter de sélectionner un seuil inutilement élevé puisque cela a pour effet de diminuer la taille des séries et d'augmenter ainsi les incertitudes d'échantillonnage. Le choix du seuil revient à trouver un équilibre entre le gain en information et l'obtention de séries indépendantes. Le seuil optimal est le plus bas pour lequel la GPD soit valable. Il n'existe cependant aucun critère formel permettant d'identifier une telle valeur. Par contre, certaines méthodes ont été proposées qui, moyennant un apport manuel qui laisse place à un certain niveau de subjectivité, permettent de guider l'utilisateur dans le choix de ce paramètre. Lang *et al.* (1999) présentent une revue des différents tests et méthodes permettant de guider le choix du seuil. Parmi ces méthodes on retrouve celle proposée par Davison et Smith (1990) et reprise par Coles (2001) qui reste sans doute l'une des plus utilisées lorsque vient le temps de faire ce choix délicat.

Cette méthode s'appuie sur la propriété de troncature présentée à la section précédente. De ce résultat, on peut montrer que, pour une GPD, la moyenne des écarts au seuil varie linéairement avec l'augmentation du seuil. La méthode consiste essentiellement à dessiner le graphique de l'écart moyen au seuil en fonction du seuil sélectionné. Théoriquement, on s'attend à observer un comportement linéaire à partir d'une certaine valeur de seuil.

La Figure 2.1 présente le résultat obtenu à la suite de l'application de la méthode sur une série de 7300 valeurs générées à partir d'une distribution normale $\mathcal{N}(0,3)$ tronquée en zéro.

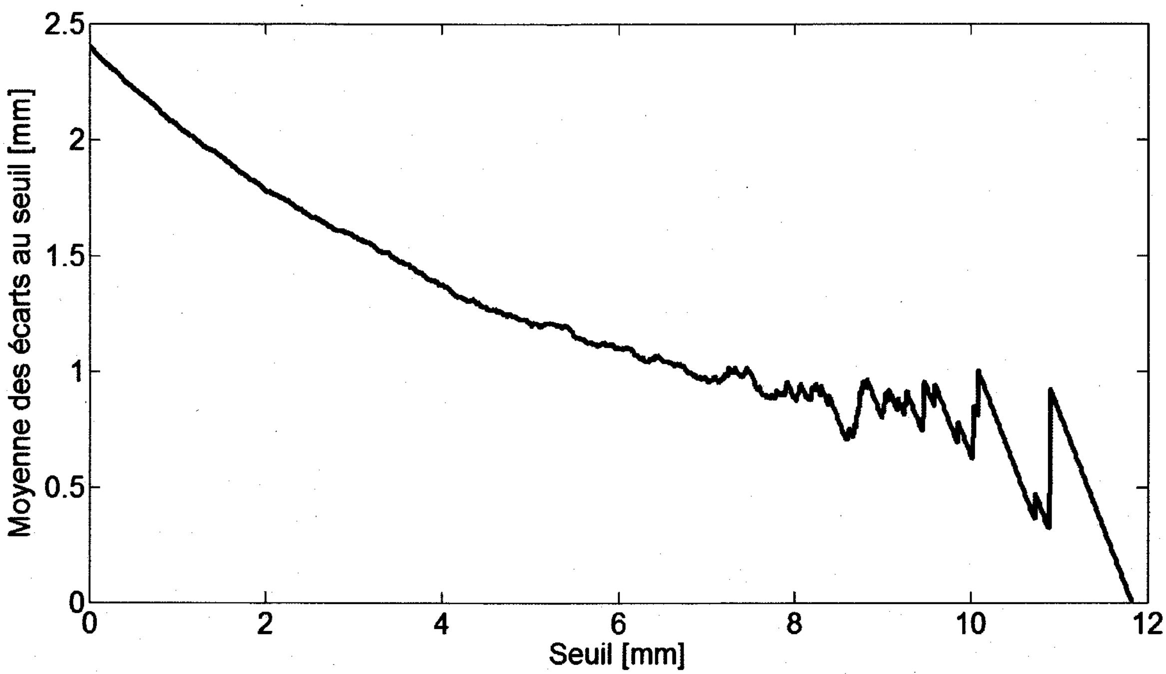

Figure 2.1 : Moyenne des écarts au seuil en fonction du seuil pour une série de 7300 valeurs issues d'une distribution normale de paramètres $\mu = 0$ et $\sigma = 3$ tronquée en zéro

Une analyse visuelle de la Figure 2.1 suggère un comportement approximativement linéaire pour les valeurs de seuil supérieures à 6 mm. Inférer une valeur de seuil à partir de ce résultat demeure cependant un exercice subjectif. Il est en effet difficile d'automatiser cette approche. Une approche « manuelle » demande quant à elle que le travail d'identification du seuil soit refait pour chaque série étudiée, ce qui peut s'avérer particulièrement fastidieux lorsque le volume de séries à analyser est important.

Une alternative intéressante pouvant servir de guide lors de la sélection du seuil consiste à considérer le taux de dépassement annuel ou encore le quantile correspondant. De nombreux auteurs ont choisi cette manière de procéder (Dierckx et Teugels, 2010; Kysely *et al.,* 2010; Coelho *et al.,* 2008; Mendez *et al.,* 2008; Smith, 1989). Il est parfois plus pratique de travailler à partir d'un taux de dépassement fixe, par exemple, retenir en moyenne trois valeurs par année (Zhang *et al.*, 2001). Le seuil adéquat entre deux régions aux régimes pluviométriques différents risque de varier grandement. Par exemple, le seuil dépassé en moyenne deux fois par année dans une région humide sera bien différent de celui dépassé en moyenne deux fois par année dans une région sèche ou nordique. Ces dépassements de seuil sont cependant aussi rares dans l'une ou l'autre de ces régions et correspondent *grosso modo* au même quantile de leur distribution respective. Pour des territoires qui comportent des régions aux pluviométries variables, le taux de dépassement est un guide particulièrement intéressant. À cet effet, Mekis et Hogg (1999) soulignent que, pour le Canada par exemple, un seuil fixe n'est pas approprié pour toutes les stations en raison des variations substantielles des régimes de précipitations. Ils ont, pour cette raison, choisi de travailler avec un seuil basé sur le quantile de l'événement, soit l'équivalent de travailler avec un taux de dépassement.

2.4 Declustering

Conjointement aux efforts déployés dans la sélection adéquate d'un taux de dépassement, il faut considérer le critère d'indépendance statistique des séries. Dans le cas de phénomènes météorologiques tels que les précipitations, il apparaît évident que des journées consécutives de précipitations risquent d'appartenir au même événement et donc de violer le critère d'indépendance. Physiquement, l'indépendance statistique se traduit par le fait qu'observer un certain cumul de précipitations au cours d'un jour donné n'a pas d'incidence sur la probabilité d'observer un cumul donné au cours des jours précédents ou subséquents. Or, en pratique, de par leur nature, les événements sont souvent corrélés temporellement. C'est problématique dans le cas de la méthode des séries de durées partielles, particulièrement lorsque le seuil est bas et que par conséquent le nombre de valeurs retenues est important.

Afin de pallier ce problème, on a recourt au *declustering* qui consiste à ne retenir que la plus élevée des valeurs d'une séquence successive de valeurs supérieures au seuil. La Figure 2.2 montre un exemple de *declustering*.

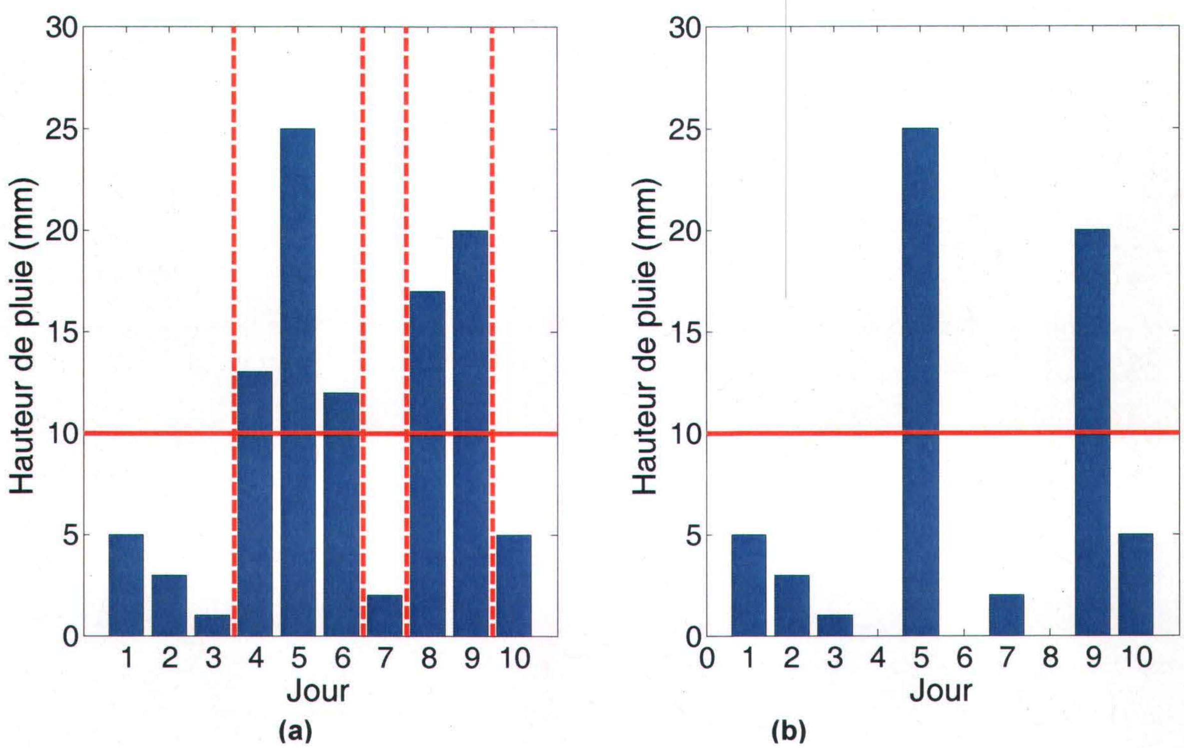

Figure 2.2 : Exemple d'une série avant (a) et après (b) l'exercice de *declustering*. Les courbes en rouge pointillées délimitent les « *clusters* » identifiés et le trait rouge continu la valeur de seuil utilisée.

Il est bon de noter que plusieurs approches ont été proposées pour procéder au *declustering* des séries. Les critères de *declustering* diffèrent selon le type de données analysées. Par exemple, dans le cas de températures journalières, Kysely *et al.* (2010) imposent une période minimale de quatre jours entre deux valeurs. La méthode sélectionnée et illustrée à la Figure 2.2 est sans doute l'une des plus simples et plusieurs études ont déjà attesté de son efficacité (Sugahara *et al.*, 2009; Khaliq *et al.*, 2006). La qualité du *declustering* peut par ailleurs être vérifiée à l'aide de tests d'indépendance. Un des inconvénients du *declustering* est qu'il peut conduire à une sous-estimation de la probabilité réelle d'occurrence particulièrement pour les évènements près du seuil qui sont davantage corrélés. La distribution indiquera par exemple qu'un évènement n'arrivera en moyenne qu'une fois par mois alors qu'en réalité, il pourrait survenir plus fréquemment. L'interprétation des résultats est délicate et correspond en fait à la probabilité que le maximum d'un *cluster* excède une certaine valeur.

3. MÉTHODE DES SÉRIES DE DURÉES PARTIELLES EN RÉGIME NON STATIONNAIRE

Dans un contexte de changements climatiques, il est essentiel de pouvoir prendre en compte le caractère non stationnaire des extrêmes de précipitations (Mailhot *et al.*, 2007). La non stationnarité invalide la condition de variables identiquement distribuées nécessaire à une utilisation conventionnelle de l'approche SDP. Pour pallier cette lacune, l'hypothèse de base consiste à supposer que la non stationnarité des séries aura pour effet de faire varier un ou plusieurs des paramètres de la GPD en fonction du temps (Kysely *et al.*, 2010; Coelho *et al.*, 2008; Mendez *et al.*, 2008; Katz *et al.*, 2002; Strupczweski *et al.*, 2001). Ce chapitre porte, dans un premier temps (section 3.1), sur l'approche « classique » utilisée pour l'estimation des paramètres en régime non stationnaire. La section 3.2 aborde la question de la variabilité intra-annuelle des séries qui invalide également le critère de variables aléatoires identiquement distribuées par la présence d'une non homogénéité.

3.1 Estimation des paramètres en régime non stationnaire

L'approche la plus classique consiste à proposer différents modèles de dépendance temporelle des paramètres. Cette technique n'est pas exclusive à la GPD et peut être mise en œuvre pour modéliser toute distribution en régime non stationnaire (El Aldouni *et al.*, 2007; Khaliq *et al.*, 2006; Katz *et al.*, 2002; Coles, 2001). La relative simplicité et la grande souplesse offertes par cette approche expliquent sans doute qu'elle soit si souvent utilisée. Différentes hypothèses, quant au modèle qui permet de décrire l'évolution des paramètres d'échelle et de forme, doivent d'abord être énoncées.

Le modèle linéaire constitue l'approche la plus simple et consiste à décrire les paramètres de la façon suivante :

$$k(t) = k_0 + b(t - t_0) \tag{3.1}$$

$$\alpha(t) = \alpha_0 + d(t - t_0) \tag{3.2}$$

où k_0 et α_0 sont respectivement les valeurs de k et α au temps $t = t_0$, b et d les taux de variation de k et α en fonction du temps et t_0 le temps initial.

Plusieurs modèles différents sont possibles et l'approche par maximum de vraisemblance est la plus souvent utilisée pour estimer les paramètres de chacun d'entre eux. En comparant par la suite ces différents modèles, il est possible de sélectionner le « meilleur » modèle sur la base du critère d'information d'Akaike (*Akaike information criterion*, AIC; voir section 5.4 pour une description de ce score).

Cette manière de procéder comporte toutefois des limites. Certaines difficultés sont inhérentes à cette approche, notamment lorsque vient le moment de dresser la liste des modèles possibles. *A priori*, on n'a souvent aucune idée de la forme des dépendances temporelles à considérer. Les modèles deviennent d'autant plus complexes à élaborer lorsque l'on souhaite considérer en plus la variabilité intra-annuelle des paramètres. Cette procédure permet par ailleurs de retenir uniquement le meilleur modèle parmi la liste de modèles proposés. Dans le cas où tous les modèles proposés sont « médiocres », on ne retiendra alors que le « moins mauvais modèle » (Burnham et Anderson, 2002).

3.2 Variabilité intra-annuelle

La non stationnarité se manifeste également à l'échelle intra-annuelle. Bien qu'il soit étymologiquement correct de parler de non stationnarité (Meylan *et al.,* 2008, section 3.4) induite par la non homogénéité, on parlera dans ce cas de « variabilité intra-annuelle ». Ce terme permettra de spécifier l'origine de la variation des paramètres et d'éviter toute confusion. Cette variabilité intra-annuelle est attribuable aux cycles saisonniers. En effet, les processus physiques générateurs de pluies et, par extension, d'évènements extrêmes, diffèrent selon la période de l'année considérée. De ce fait, il est raisonnable de penser que l'amplitude de dépassement du seuil et la fréquence de dépassement de ce dernier dépendront de la période de l'année considérée, le corollaire étant que les paramètres de la GPD varieront également au cours de l'année.

Bon nombre d'auteurs ont considéré ou ont au moins abordé la question de la variabilité intra-annuelle dans leurs travaux (Sugahara *et al.,* 2009; Coelho *et al.*, 2008; Coles *et al.*, 2003). La prise en compte explicite de la variabilité intra-annuelle constitue une approche d'autant plus rigoureuse qu'elle tend à nous rapprocher des hypothèses de séries i.i.d. (sans toutefois nous assurer que ces conditions sont entièrement respectées). Il peut être hasardeux de tenter de calibrer une GPD avec des paramètres constants, sachant que la distribution d'un mélange non homogène de GPD n'est pas une GPD (voir section 3.2.1).

3.2.1 Séries non homogènes de Pareto généralisées

Pour se convaincre de la nécessité d'examiner la variabilité intra-annuelle des paramètres, considérons l'exemple suivant. Deux échantillons de N variables sont générés à partir de deux GPD possédant des paramètres différents. La disparité entre ces paramètres s'apparente, par exemple, aux différences observables entre deux périodes de l'année. Les paramètres, associés aux distributions GPD-1 et GPD-2 sont choisis de façon arbitraire. Ils sont présentés au Tableau 3.1.

Tableau 3.1 : Paramètres utilisés afin de générer deux échantillons issus d'une GPD et paramètres obtenus à la suite de la calibration d'une GPD sur la série non homogène.

Distribution	k	α	ξ	N
GPD 1	-0,30	10,0	0	5 000
GPD 2	-0,10	30,0	0	5 000
GPD-fit	-0,29	17,1	0	10 000

On combine ensuite ces deux séries homogènes en une seule série hétérogène. La série obtenue présente des variables non identiquement distribuées. Si l'on choisit d'ignorer la possible variabilité des paramètres et que l'on ajuste une GPD à l'aide de la méthode du maximum de vraisemblance, on obtiendra les valeurs de paramètres présentés au Tableau 3.1 à la ligne DPG-fit.

Il est alors possible de comparer la valeur des quantiles modélisés par une GPD et les quantiles empiriques (Figure 3.1). Les quantiles empiriques ($\hat{F}(x_{[r]})$) sont évalués à partir de l'équation suivante (Cunnane, 1978) :

$$\hat{F}(x_{[r]}) = \frac{r - 0.44}{N + 0.12} \tag{3.3}$$

où $x_{[r]}$ désigne la valeur de rang r et N à la taille de l'échantillon.

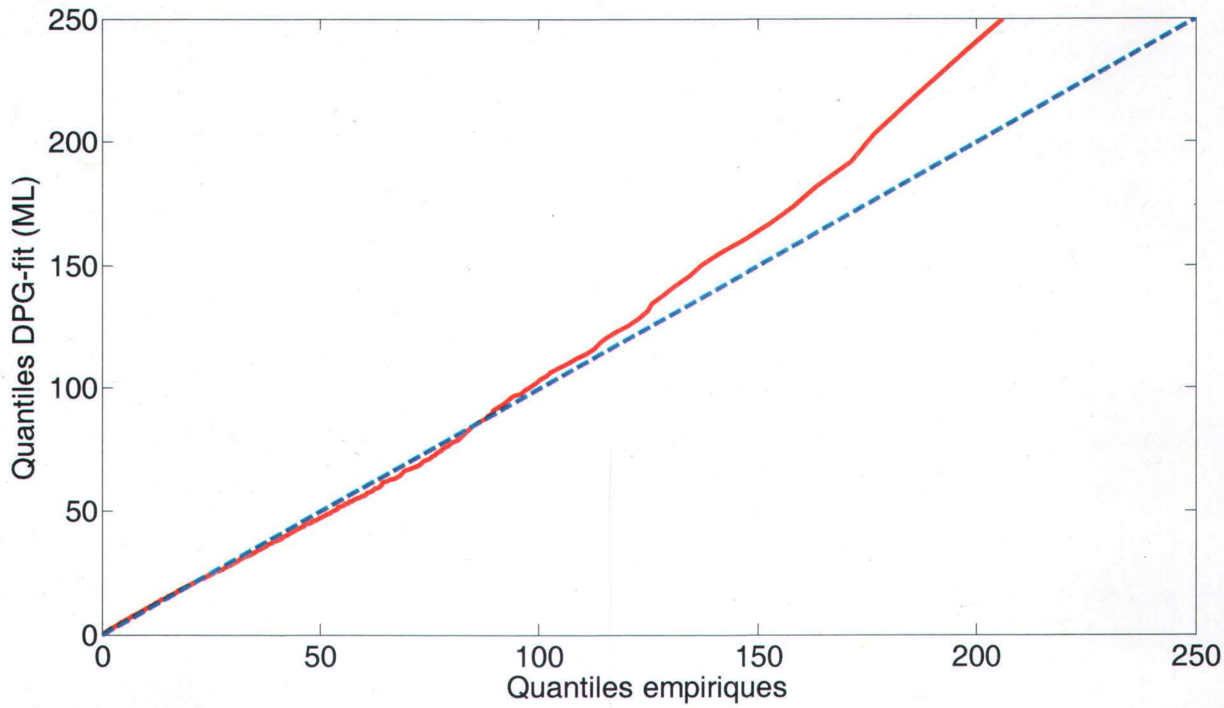

Figure 3.1 : Graphique quantile-quantile (qqplot) montrant la comparaison entre les valeurs associées aux quantiles d'une GPD ajustée (GPD-fit) à une série non homogène issue d'une combinaison de GPD et les valeurs associées aux quantiles empiriques de la série non homogène (courbe continue). La courbe en tirets correspond à la droite de pente 1.

Cet exemple illustre le danger potentiel d'ignorer la variabilité intra-annuelle. La Figure 3.1 permet de constater qu'une série de valeurs non identiquement distribuées ne peut être décrite dans ce cas par une GPD homogène. Des différences importantes au niveau des valeurs prédites et théoriques pour les grands quantiles sont observables.

4. APPLICATION AUX SÉRIES DES PRÉCIPITATIONS INTENSES SIMULÉES PAR LE MODÈLE CANADIEN COUPLÉ DE CLIMAT GLOBAL (MCCG3)

Ce chapitre présente l'application de la méthode SDP aux séries de précipitations intenses simulées par le modèle canadien couplé de climat global de troisième génération (MCCG3). Le but consiste à utiliser l'approche SDP en régime non stationnaire afin d'évaluer l'impact potentiel des CC sur les extrêmes de précipitations. Le chapitre se compose de huit sections. La section 4.1 présente les données utilisées et le territoire couvert par l'analyse. Les deux sections subséquentes visent à décrire la méthodologie développée afin d'adapter l'approche SDP au régime non stationnaire. Plus spécifiquement la section 4.2 s'intéresse à l'estimation du seuil alors que la section 4.3 aborde la question de la modélisation de la dépendance temporelle des paramètres de la GPD. La section 4.4 traite de la stabilisation de la fréquence de dépassement et la section 4.5 présente les résultats obtenus en climat futur. La section 4.6 résume les résultats quant à l'impact des CC sur la variabilité intra-annuelle. La section 4.7 discute de la sensibilité des résultats au taux moyen de dépassement utilisé pour fixer le seuil. Enfin la section 4.8 compare les résultats obtenus à ceux présentés dans la littérature.

4.1. Données utilisées

Les données utilisées comprennent les séries de CJP simulées par le modèle canadien couplé de climat global de troisième génération (MCCG3 version T47) (Flato et Boer, 2001) pour la période 1850 à 2100. Cette troisième version du modèle comporte plusieurs améliorations par rapport au modèle de deuxième génération décrit par McFarlane *et al.* (1992). Le modèle simule le climat terrestre à l'aide d'un maillage de surface d'approximativement 3,75 degrés en latitude et en longitude, ce qui correspond grossièrement à des tuiles de 390 x 260 km à une latitude de 47°.

L'analyse porte sur 420 tuiles (12 x 35) couvrant le Canada ainsi qu'une partie du nord des États-Unis (voir Figure 4.1). Les précipitations sont disponibles sous forme de kg $m^{-2}s^{-1}$ et sont ensuite converties en mm/jour. À partir de l'an 2000, le modèle utilise des scénarios d'émission de GES et aérosols afin d'établir l'évolution future des émissions et des concentrations de ces derniers. Le scénario considéré dans la présente étude est le SRES A2 (Nakicenovic *et al.*, 2000), qui prévoit une augmentation de la concentration en GES de plus de 100 % d'ici 2100 par rapport à la concentration en 2000.

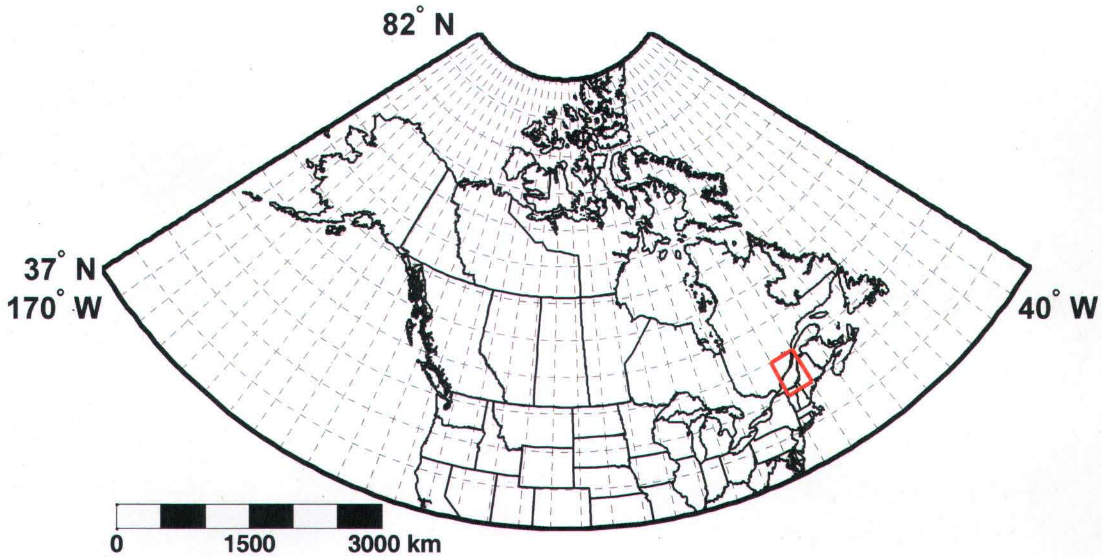

Figure 4.1 : Territoire couvert par l'étude et discrétisation de l'espace en 420 tuiles par le MCCG3. La tuile en encadré rouge correspond à la tuile utilisée dans les prochains exemples.

Cinq simulations utilisant des conditions initialement différentes, que l'on appelle des membres, sont disponibles. Une manière de considérer l'information des cinq membres consiste à supposer qu'ils sont tous issus d'une même distribution statistique et que les différences inter-membres sont uniquement attribuables à la variabilité interne du climat (de Elia *et al.*, 2008). Cette hypothèse d'ergodicité permet de regrouper les séries des cinq membres et d'ainsi améliorer l'échantillonnage. Mailhot *et al.* (2010) ont montré que ces cinq membres peuvent en effet être considérés comme issus de la même distribution en régime stationnaire. Cette approche a été employée dans le cadre de la présente étude.

4.2 Estimation du seuil en régime non stationnaire

Plusieurs auteurs ont proposé de choisir le seuil en fixant un quantile ou, de manière équivalente, en fixant un taux moyen de dépassements (voir section 2.3). L'approche développée s'appuie sur l'estimation du seuil à partir de la sélection d'un taux moyen de dépassements fixe. Cette approche présente, *a priori*, certains avantages. Dans un premier temps, elle permet de représenter chaque période avec un nombre de valeurs au *prorata* de sa durée. On évite ainsi de surreprésenter une période au détriment d'une autre puisqu'on s'assure que le taux moyen de dépassements est constant dans le temps. De plus, on évite de devoir définir « manuellement » le seuil comme ce serait le cas pour la méthode décrite par Coles *et al.* (2003). Puisque l'on désire considérer à la fois la non stationnarité sur de longues périodes induite par les changements climatiques et la variabilité intra-annuelle mensuelle, il est indispensable de faire varier le seuil à la fois de manière inter et intra-annuelle. Pour ce faire, la période 1850-2100 a été subdivisée en quatre sous-périodes. Le découpage des sous-périodes est arbitraire, cependant ces dernières doivent être relativement stationnaires du point de vue des extrêmes. Pour la présente étude, une période plus longue a été utilisée pour la période 1850-1950 faisant office de période de référence. Les 150 années restantes ont été divisées en trois sous-périodes de 50 ans chacune de sorte à pouvoir détecter le signal des CC. Les quatre sous-périodes ont, à leur tour, été divisées en 12 séries représentant chacune un mois de l'année afin de permettre de tenir compte de la variabilité intra-annuelle. On obtient un total de 48 séries (4 sous-périodes x 12 séries mensuelles = 48 séries) tel qu'illustré à la Figure 4.2. La fragmentation de la série vise à obtenir des SDP pseudo-homogènes, c'est-à-dire approximativement stationnaires. La stationnarité des SDP sera ultérieurement vérifiée à l'aide du test non paramétrique de Mann-Kendall (Hamed et Rao, 1998).

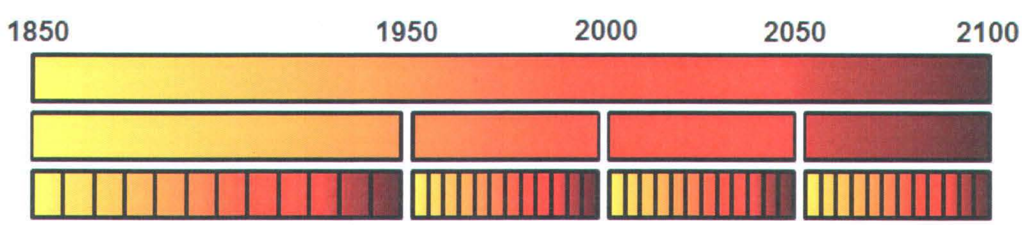

Figure 4.2 : Division de la période 1850-2100 en 48 séries pseudo-homogènes

Il importe désormais d'identifier le seuil qui correspond au taux de dépassement fixé pour chacune de ces séries. On teste différentes valeurs de seuil jusqu'à identifier celle qui permet de retenir le nombre désiré de 12 valeurs par année en moyenne. On a recours au *declustering* présenté à la section 1.3 afin d'assurer de l'indépendance statistique des valeurs retenues. Une fois cette étape complétée, on se retrouve avec 48 seuils (un par série) qui garantissent un taux de dépassement unique sur chacune des subdivisions de la période 1850-2100. Pour un mois donné, décembre par exemple, on peut tracer les valeurs de seuil correspondant aux différentes périodes (Figure 4.3).

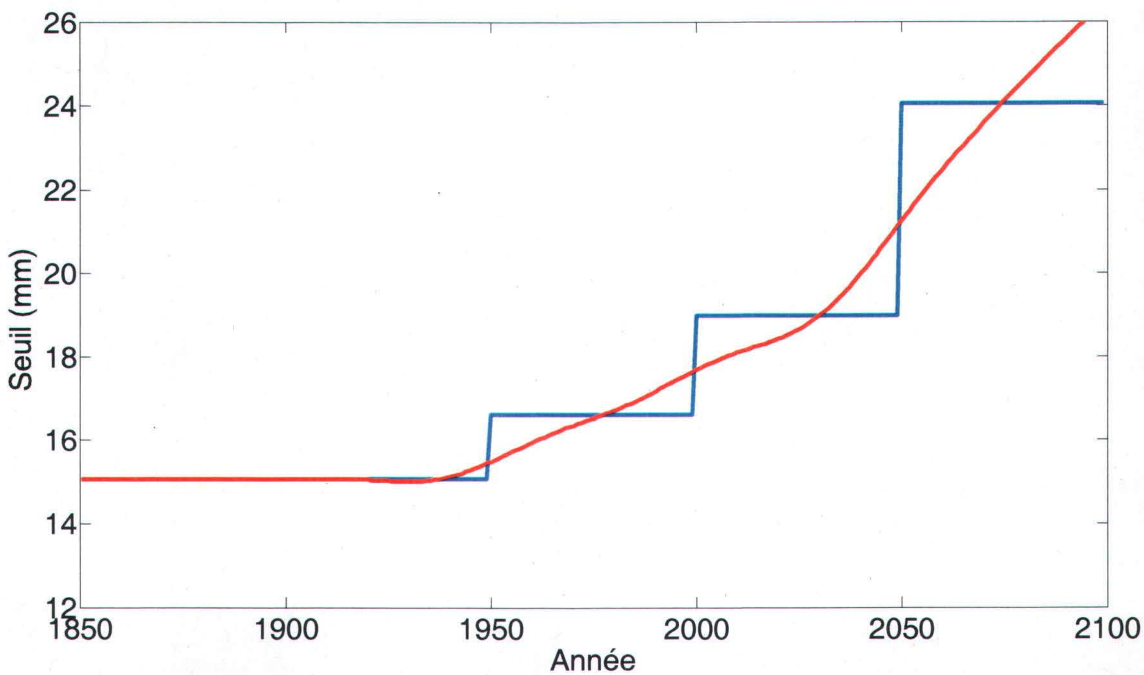

Figure 4.3 : Valeurs des seuils correspondant à une probabilité quotidienne de dépassement p = 12/365 ≈ 0.0329 pour le mois de décembre (en moyenne 1 dépassement par mois de décembre) pour la tuile en encadré rouge de la Figure 4.1. Les plateaux en trait continu bleu sont le résultat de l'algorithme d'identification du seuil et le trait rouge est une courbe lissée afin de représenter le comportement présumé continu de l'évolution du seuil.

La probabilité au dépassement étant constante, l'augmentation du seuil, qui correspond à une élévation des plateaux, suggère une augmentation de l'intensité des pluies extrêmes en climat futur. La représentation par plateaux est un artefact résultant de l'analyse. Intuitivement, on s'attend à observer une variation continue du seuil. Une courbe continue, représentée par le trait rouge à la Figure 4.3, a donc été ajustée afin que, pour chaque plateau, l'aire au-dessus et sous les plateaux soient approximativement équivalentes (voir l'annexe A pour les détails). La valeur moyenne sur la période correspond ainsi à la valeur du seuil pour la période discrète considérée. En répétant le processus pour les 12 mois de l'année, on obtient pour chaque mois, à l'instar de la Figure 4.3, des courbes qui modélisent la variation interannuelle du seuil. On peut reconstruire à partir de ces courbes une série modélisant à la fois les variations intra-annuelle et interannuelle du paramètre de position à partir des valeurs de seuils mensuels pour chacune des 250 années de la série. On obtient alors une courbe constituée d'une succession de plateaux. La Figure 4.4 présente en trait bleu ce résultat en faisant un zoom sur la période 1950-1952. Le seuil est lissé cette fois à l'aide de la méthode LOESS (*Locally Weighted Scatterplot Smoothing*) (Cleveland et Devlin, 1988). La courbe rouge de la Figure 4.4 présente le résultat de ce lissage. Ce seuil, variable en fonction du temps, est utilisé pour construire les SDP.

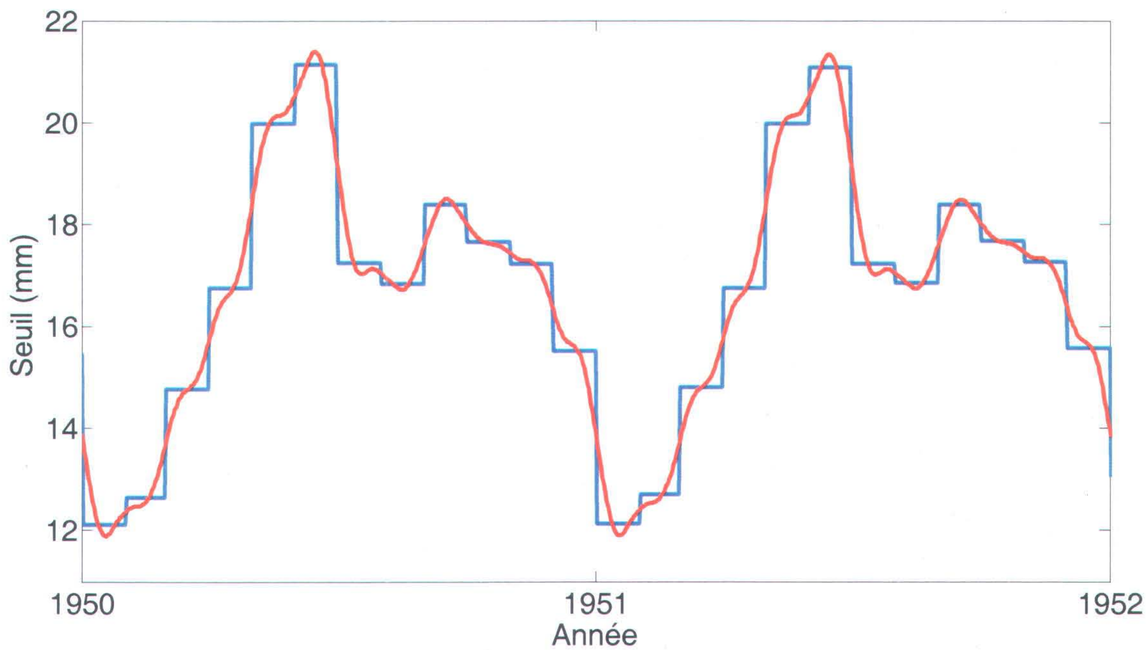

Figure 4.4 : Variabilités inter et intra-annuelle du seuil pour la période 1950-1952 pour un taux moyen de 12 dépassements par année, soit environ un par mois, pour la tuile de l'encadré rouge de la Figure 4.1. Le trait bleu est obtenu en juxtaposant les valeurs interpolées pour chacun des mois. Le trait rouge est le résultat du lissage par la méthode LOESS.

On peut vérifier la stationnarité des SDP obtenues à l'aide du test de Mann-Kendall. Le Tableau 4.1 présente pour chacun des blocs, la fraction de tuiles considérées stationnaires par le test avec un seuil α = 0.05.

Tableau 4.1 : Pourcentage des tuiles dont les SDP sont considérées stationnaires par le test de Mann-Kendall avec un seuil α = 0.05.

Mois	Périodes			
	1850-1950	1950-2000	2000-2050	2050-2100
1	96	90	95	89
2	95	94	93	88
3	96	93	94	89
4	95	96	92	86
5	96	92	93	94
6	93	94	94	93
7	95	95	94	87
8	94	94	93	90
9	95	94	94	92
10	94	93	95	93
11	97	95	93	91
12	94	94	96	82

Le Tableau 4.1 montre que la majorité des SDP peuvent être considérées stationnaires selon le test de Mann-Kendall. Le taux de non rejet de l'hypothèse nulle est d'environ 95 %, soit celui auquel on s'attend pour un seuil α = 0.05. Les pourcentages obtenus pour la période 2050-2100 sont légèrement plus faibles et la moyenne est inférieure à 95%, ce qui suggère une plus grande proportion de séries non stationnaires pour cette période. Ce résultat est conforme avec l'idée que les changements climatiques se feront progressivement sentir et que la période susceptible de connaître les plus grands changements est la période 2050-2100.

Des tests autres que le test de Mann-Kendall auraient pu être utilisés afin de vérifier la stationnarité des séries. Celui-ci a d'abord été retenu pour sa simplicité. De plus, l'objectif était uniquement de pouvoir statuer sur la stationnarité des séries sans chercher à déterminer le type de non-stationnarité comme le permettent les tests plus élaborés (Quessy *et al,.* 2011). Des résultats très similaires (non présentés) à ceux du Tableau 4.1 ont par ailleurs été obtenus à partir d'une méthodologie différente consistant à vérifier si le paramètre de pente d'une régression linéaire des SDP avec le temps comme variable explicative est significatif.

4.3 Modélisation de la dépendance temporelle des paramètres

Une fois les SDP construites, il importe de déterminer les paramètres de forme et d'échelle de la GPD. La modélisation en régime non stationnaire des paramètres peut s'effectuer de diverses manières. La section 3.1 décrit l'approche classique, une méthode paramétrique qui consiste à proposer différents modèles de dépendance temporelle des paramètres. Devant la difficulté d'identifier des modèles réalistes, une approche non paramétrique (dans le sens qu'aucune relation fonctionnelle n'est définie *a priori*) est développée.

La méthode proposée permet de nous guider dans la définition des modèles de dépendance temporelle à considérer. Elle permet de visualiser l'évolution temporelle des paramètres en n'imposant pas de cadre rigide à la variation de ces paramètres. Cette procédure sera utile pour confirmer ou infirmer le choix initial des modèles à retenir dans le cadre d'une approche paramétrique.

Pour obtenir les paramètres d'échelle et de forme de la GPD, une approche similaire à la procédure décrite précédemment pour le seuil a été considérée. La période d'étude est subdivisée en 48 séries qui correspondent aux 12 mois de l'année et aux quatre périodes définies à la section 4.2 (voir Figure 4.1). Les SDP sont obtenues à partir du seuil déterminé antérieurement (voir section 4.2) et selon la méthode établie. Un certain nombre de valeurs est retenu pour chacune des séries en fonction du seuil, formant ainsi les SDP. Le Tableau 4.1 montre que ces SDP sont en grande majorité stationnaires. Il est alors possible d'estimer les valeurs des paramètres de la DPG pour chacune des SDP à l'aide d'une approche classique, en l'occurrence la méthode des moments (voir section 2.1). À l'instar de ce qui a été obtenu dans le cas du paramètre de position, on obtient pour chaque mois une succession de quatre plateaux retraçant la valeur des paramètres de forme et d'échelle. Pour les mêmes raisons et par des méthodes similaires, des courbes continues ont été ajustées à ces plateaux afin d'éviter une représentation discontinue non réaliste de l'évolution temporelle des paramètres. La Figure 4.5 présente, à titre d'exemple, l'évolution du paramètre d'échelle pour le mois de décembre.

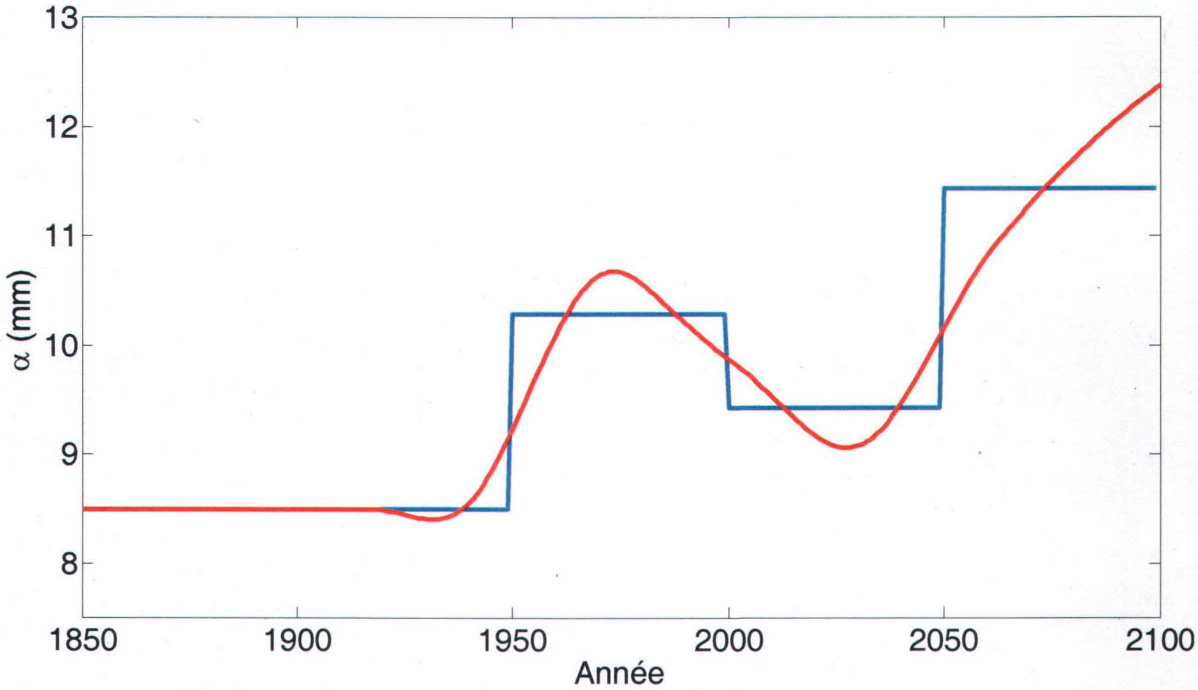

Figure 4.5 : Évolution du paramètre d'échelle pour le mois de décembre pour la tuile de l'encadré rouge de la Figure 4.1. Le trait bleu correspond à la valeur estimée pour chaque période et le trait rouge est obtenu après lissage

Des courbes semblables à celle de la Figure 4.5 ont été obtenues pour chaque mois pour les paramètres de forme et d'échelle. À partir de ces courbes, en appliquant la procédure décrite à la section 4.2 pour le seuil, on construit une courbe qui modélise les variations intra et interannuelle des paramètres. Un exemple de résultat obtenu pour le paramètre d'échelle est présenté à la Figure 4.6.

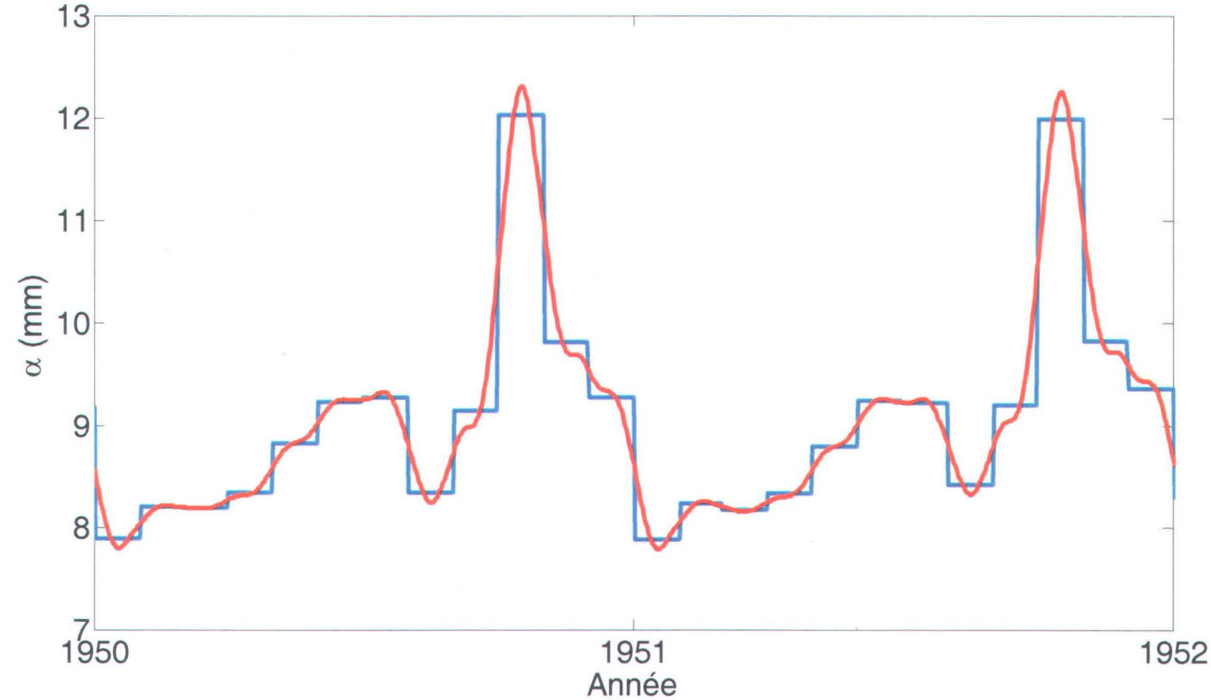

Figure 4.6 : Variabilités inter et intra-annuelle du paramètre d'échelle pour la période 1950-1952 pour la tuile de l'encadré rouge de la Figure 4.1. Le trait bleu est obtenu en juxtaposant les valeurs interpolées à chaque mois et le trait rouge est obtenu après lissage à l'aide de la méthode LOESS.

Une courbe similaire à celle présentée à la Figure 4.6 est obtenue pour le paramètre de forme. On obtient ainsi des courbes continues sur la période 1850-2100 pour les trois paramètres de la GPD. L'interpolation des paramètres de forme et d'échelle est faite de manière indépendante. Cela peut, dans certains cas, poser problème puisque l'interdépendance des paramètres n'est pas considérée. Il y a alors perte d'information puisque cette corrélation n'est pas prise en compte dans le modèle. Une alternative consisterait à interpoler la moyenne et l'écart-type pour ensuite recalculer les valeurs de paramètres à partir de la méthode des moments. Il serait également possible de ne pas interpoler les plateaux. On observerait cependant des discontinuités inter-mensuelles importantes et aucune variation intra-mensuelle des valeurs de paramètres. À partir des courbes de paramètres, on peut estimer la probabilité au dépassement d'une intensité donnée ou encore le nombre de dépassements moyen d'un certain seuil pendant une période donnée.

4.4 Stabilisation de la fréquence de dépassement du seuil

Dans un premier temps, la méthodologie développée doit permettre de stabiliser la fréquence d'occurrence des événements au-dessus du seuil. L'atteinte de cet objectif est évaluée en s'intéressant à la distribution du nombre de dépassements mensuels et annuels pour les différentes périodes considérées. La Figure 4.7 montre la probabilité mensuelle de dépassement du seuil pour les différentes périodes.

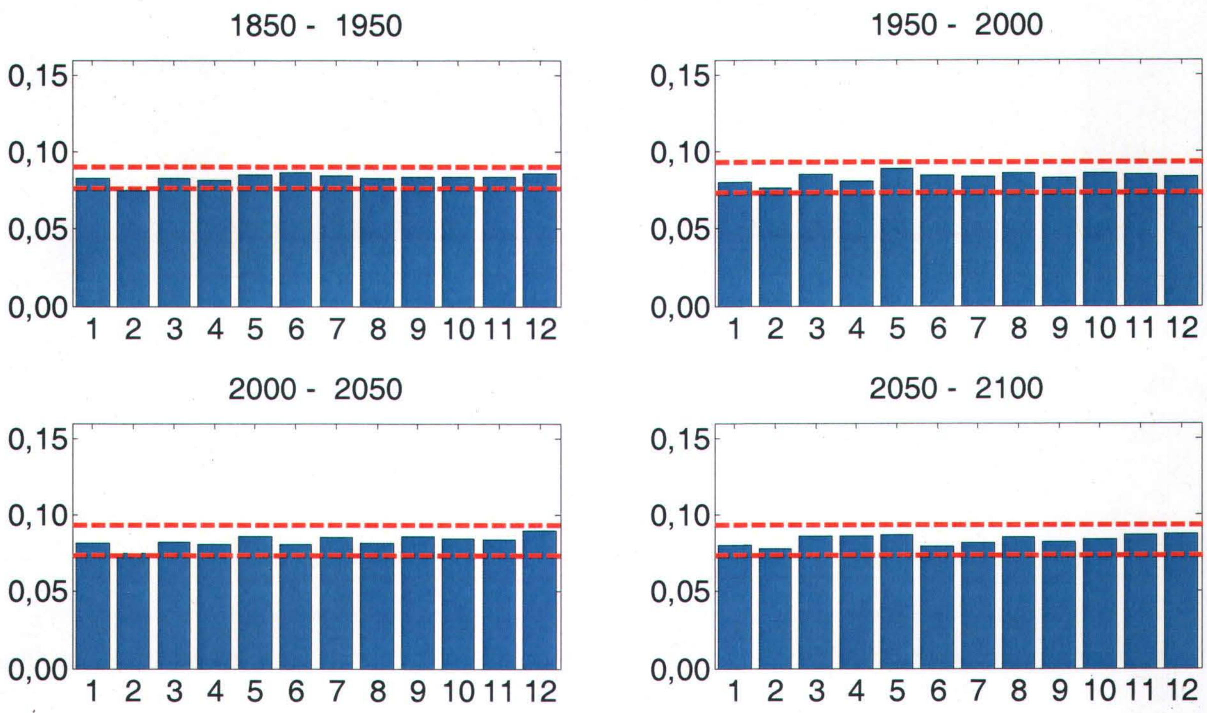

Figure 4.7 : Probabilité mensuelle de dépassements du seuil pour différentes périodes pour un taux moyen de 12 dépassements par an (un dépassement par mois en moyenne). Les courbes en tirets rouges correspondent à l'intervalle de confiance théorique à 95 % basé sur la loi de Poisson.

La Figure 4.7 permet de constater que la probabilité de valeurs supérieures au seuil est constante tous les mois. On remarque par ailleurs que, pour chaque mois, la probabilité est comprise dans l'intervalle de confiance à 95 % obtenu à partir de la distribution de Poisson pour la durée et l'intensité du processus considéré. La Figure 4.8 présente, quant à elle, la distribution du nombre annuel de dépassements du seuil pour l'ensemble de la période 1850-2100.

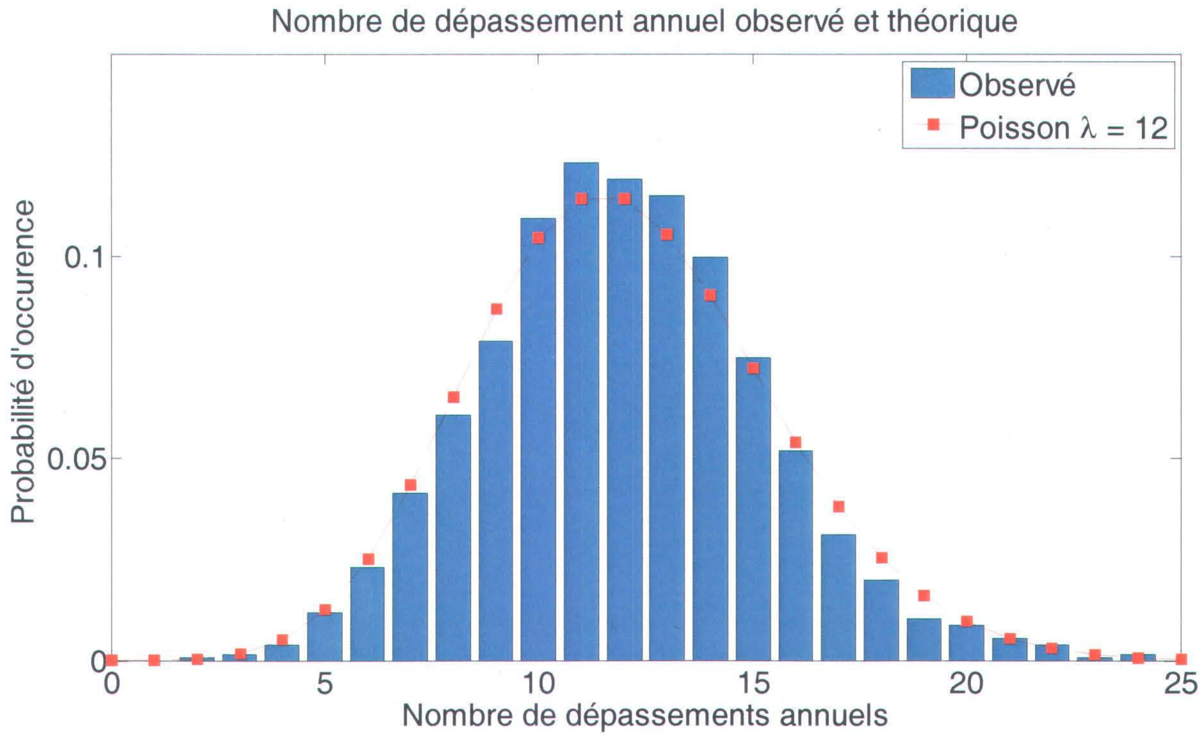

Figure 4.8 : Distribution du nombre annuel de dépassements du seuil pour un seuil correspondant à un taux moyen fixé à 12 dépassements par an. Les marqueurs rouges correspondent à la distribution de Poisson théorique d'intensité 12 dépassements/an.

Cette figure montre que la distribution du nombre annuel de dépassements correspond bien à une distribution de Poisson d'intensité 12 dépassements/an. Ce résultat est important puisqu'il suggère, d'une part, que le taux de dépassement annuel est stationnaire et, d'autre part, que les variables de la série sont indépendantes puisque ces deux critères sont nécessaires pour qu'une distribution binomiale soit approchée par un processus de Poisson (Hines et Montgomery, 1990). La Figure 4.8 permet d'attester, dans une certaine mesure, de l'efficacité de la méthode de *declustering* et de l'aptitude de la méthodologie à stabiliser le taux de dépassements.

4.5 CJP extrêmes en climat futur

Afin de pouvoir comparer les CJP extrêmes entre deux périodes distinctes, on peut s'intéresser au CJP qui, en moyenne, sera dépassé N fois au cours de la période d'intérêt. Ce calcul s'effectue à partir de l'ensemble des paramètres. La probabilité de dépasser un CJP de hauteur x au jour j est donnée par :

$$p_j = \frac{\lambda}{365}\left[1 - F(x \mid k_j, \alpha_j, \xi_j)\right] \tag{4.1}$$

où F est la fonction cumulative de la GPD (éq. 2.2) c'est-à-dire la probabilité de non dépassement de x conditionnellement au dépassement du seuil. La variable λ est le taux de dépassement annuel moyen du seuil et le facteur $\lambda / 365$ correspond à la probabilité de dépasser le seuil au jour j.

Pour une période donnée, en additionnant les probabilités journalières de dépassement d'un CJP de hauteur x pour chacun des n jours de la fenêtre, on obtient le nombre de dépassement moyen $\bar{N}$ de ce CJP pendant cette période de n jours.

$$\bar{N} = \sum_{j=1}^{n} \frac{\lambda}{365}\left[1 - F\left(x \mid k_j, \alpha_j, \xi_j\right)\right] \tag{4.2}$$

Cette formule assume l'indépendance des dépassements à l'échelle quotidienne, hypothèse plausible en vertu du processus de *declustering*. L'équation 4.2 permet de déterminer le CJP x qui sera dépassé en moyenne N fois au cours de la période d'intérêt. La probabilité de dépasser x peut, quant à elle, varier substantiellement en fonction des saisons et de l'année. Ce calcul se veut une alternative à la détermination du temps de retour qui, en régime non stationnaire, perd son sens et qui, pour être évalué, nécessite une connaissance de la valeur des paramètres jusqu'à un temps $t \to \infty$ (Olsen *et al.*, 1998). Un parallèle peut cependant être fait entre ces deux résultats dans un cadre stationnaire. En effet, en régime stationnaire, une intensité qui correspond à une période de retour T sera dépassée en moyenne une fois chaque T année; on a alors $T = \frac{n}{N}\frac{1}{365}$. En régime non stationnaire, le parallèle ne tient plus car la probabilité de dépassement varie en fonction du temps.

La Figure 4.9 présente le CJP qui, sur chacune des périodes, est dépassé en moyenne une fois tous les 20 ans (N = 1/20 ans) obtenu en fixant λ= 12. La période 1850-1950 est choisie comme période de référence car elle est considérée représentative du climat historique et stationnaire (voir Tableau 4.1). La figure présente les résultats pour l'ensemble des tuiles terrestres du territoire à l'étude.

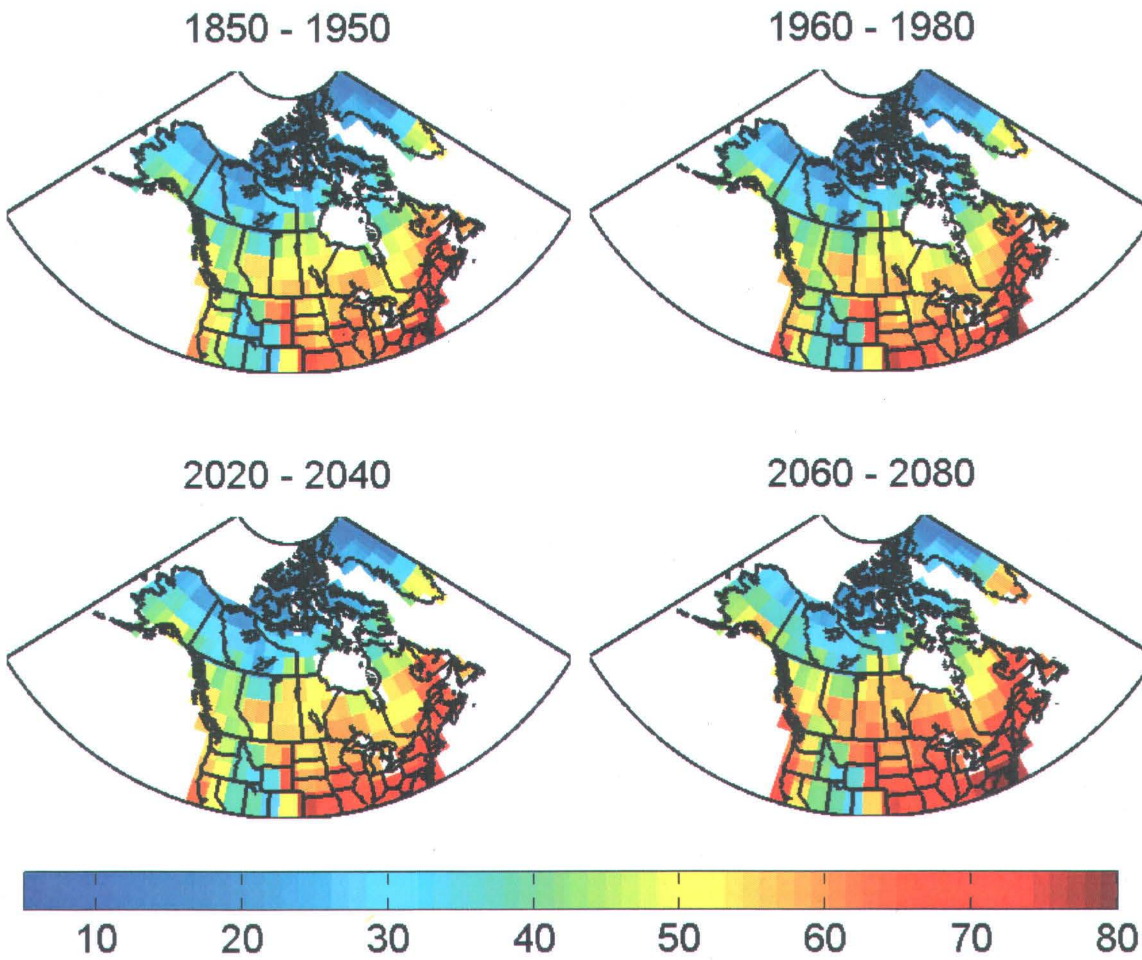

Figure 4.9 : Hauteur (mm) du CJP dépassé en moyenne une fois en 20 ans sur chaque période considérée.

La Figure 4.9 montre une certaine cohérence spatiale, les tuiles voisines présentent des valeurs semblables, et indique une tendance à l'augmentation des extrêmes de précipitation en climat futur. L'annexe B regroupe les résultats pour les hauteurs de CJP dépassées en moyenne deux, quatre et 20 fois sur une période de 20 ans. Bien que les valeurs varient, on retrouve pour ces figures des patrons spatiaux très similaires.

Afin de mieux visualiser l'importance de ces augmentations, la Figure 4.10 présente le rapport des CJP pour les deux périodes futures 2020-2040 et 2060-2080, comparativement à la période de référence 1850-1950.

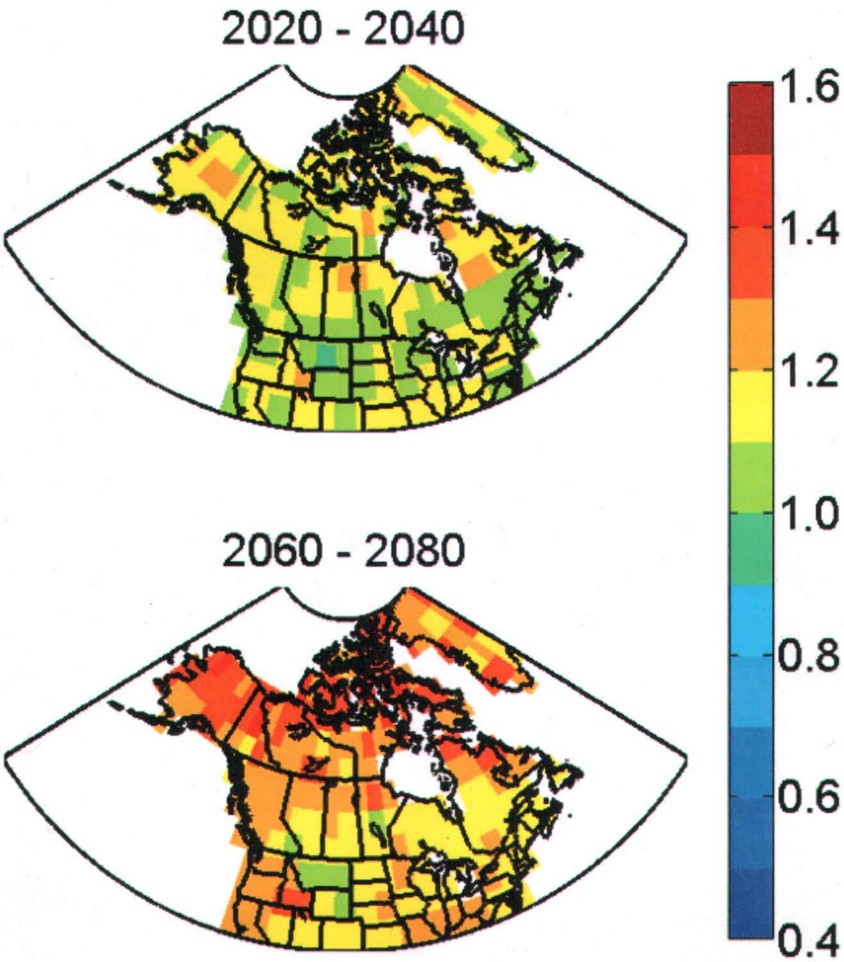

Figure 4.10 : Rapports des CJP dépassés en moyenne une fois chaque 20 ans pour les périodes 2020-2040 et 2060-2080 comparativement à la période de référence 1850-1950.

La Figure 4.10 permet de conclure à une augmentation à grande échelle des CJP dépassés en moyenne une fois chaque 20 ans. Pour la période 2020-2040 et l'ensemble du territoire à l'étude, le facteur d'augmentation varie en moyenne entre 1 et 1.2 comparativement à la valeur de référence. Pour la période 2060-2080, les hausses sont plus marquées, oscillant plutôt entre 10 et 40 %. Certains patrons régionaux sont également observables pour cette période. Pour les régions plus au nord, l'augmentation relative est généralement plus élevée, de 20 à 50 %. Il importe cependant de souligner que, pour ces régions, les valeurs de référence sont généralement plus faibles (Figure 4.9). Au niveau de la côte ouest, l'augmentation est de 20 à 30 % alors que pour l'est du Canada et des États-Unis elle se chiffre entre 10 et 20 %. Au sud des Grands Lacs, l'augmentation relative est légèrement plus élevée que sur la côte, de 20 à 30 %, alors que pour la région des Prairies et au sud de celles-ci, elle est comprise entre 0 et 20 %. Ces résultats n'illustrent qu'un exemple de ce qu'il est possible d'extraire à partir des analyses effectuées et ne concernent qu'une fréquence de dépassement donnée. L'annexe C présente le rapport des CJP dépassés en moyenne deux, quatre et 20 fois sur une période de 20 ans. Ces résultats sont très similaires à ceux de la Figure 4.10.

4.6 Variabilité intra-annuelle

Une des particularités intéressantes de l'approche développée réside dans le fait qu'elle permet de considérer la variabilité intra-annuelle des extrêmes. Il est ainsi possible d'observer comment les changements climatiques affecteront les probabilités saisonnières et mensuelles des extrêmes. Le CJP associé à un nombre moyen de dépassements N sur une période fixée est considéré dans ce qui suit. Par exemple, si on s'intéresse au mois de décembre sur la période 2060-2080, le CJP dépassé en moyenne une fois chaque 20 ans sera dépassé en moyenne une fois au cours des 20 mois de décembre de la période. Les variations relatives sont exprimées par le rapport entre le CJP de la période d'intérêt et la période de référence 1850-1950. La Figure 4.11 présente la variation relative du CJP (mm) dépassé en moyenne une fois chaque 20 ans au cours des différentes saisons de la période 2060-2080 relativement à leurs homologues de 1850-1950.

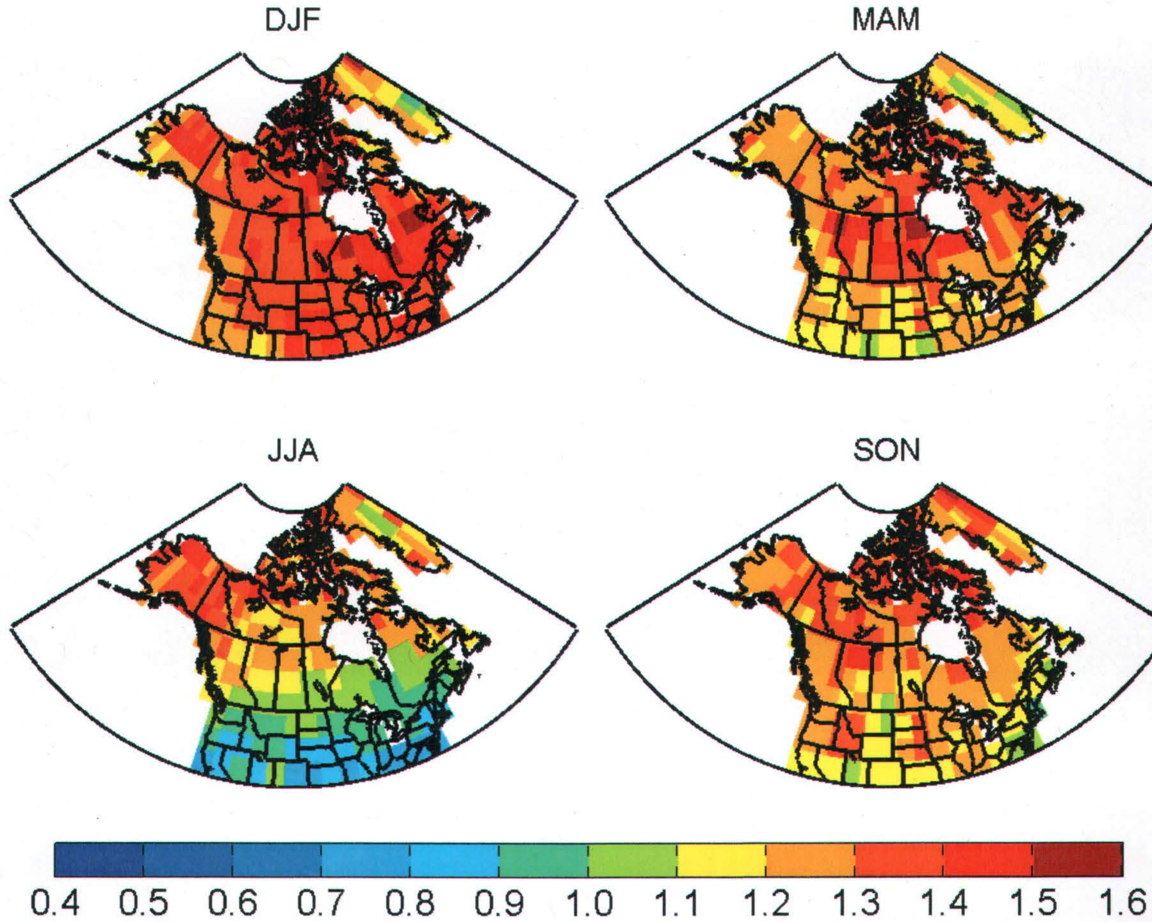

Figure 4.11 : Variation relative du CJP dépassé en moyenne une fois chaque 20 ans au cours des différentes saisons de la période 2060-2080 relativement à leurs homologues de la période 1850-1950.

La Figure 4.11 montre que, généralement, pour l'hiver (décembre, janvier, février), le printemps (mars, avril, mai) et l'automne (septembre, octobre, novembre), le CJP qui correspond au nombre de dépassements retenu augmente en climat futur pour le territoire considéré. Les hausses les plus prononcées sont associées à la saison hivernale avec des valeurs relatives comprises entre 20 et 60 %. Pour les mois d'été, l'analyse prévoit une diminution assez importante des valeurs au sud du territoire. La partie nord du territoire verra une augmentation alors que les régions plus au sud subiront une diminution de l'ordre de 10 à 30 %. Le résultat est sensiblement le même si l'on considère des CJP plus rares (nombre de dépassements moyen plus petit) ou plus fréquents (nombre de dépassements moyen plus élevé).

Afin de résumer l'information obtenue, la moyenne de variation relative pour chacune des 12 latitudes du territoire couvert a été estimée (Figure 4.12). Ces moyennes régionales permettent d'illustrer certains comportements généraux.

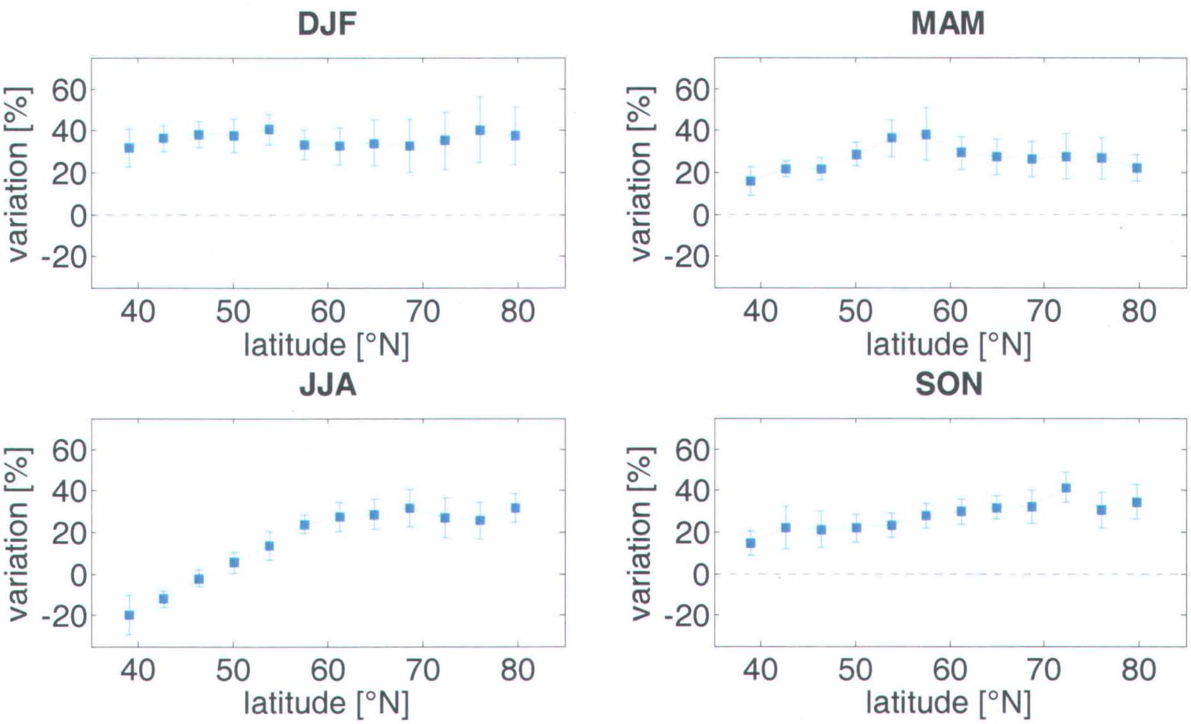

Figure 4.12 : Moyenne sur les tuiles d'une latitude donnée de la variation relative du CJP dépassé en moyenne une fois chaque 20 ans au cours des différentes saisons de la période 2060-2080 et leurs homologues de la période 1850-1950. Les barres d'erreur représentent l'écart-type sur les valeurs des tuiles pour une latitude donnée.

La Figure 4.12 met en évidence des augmentations plus importantes à toutes les latitudes pour les mois hivernaux (DJF), ces valeurs oscillant autour de 35 %. On remarque également un gradient marqué de la variation en fonction de la latitude pour les mois d'été. En moyenne, on observe une diminution des moyennes estivales des CJP pour les tuiles en deçà du 50^e parallèle. À cet effet, on notera que Mailhot *et al.* (2010) observaient une tendance similaire en s'intéressant au mois d'occurrence de l'événement annuel maximal. La saison automnale semble également présenter un léger gradient positif en fonction de la latitude. Celui-ci est cependant moins prononcé qu'en été et les augmentations moyennes demeurent positives à toutes les latitudes. Ces résultats ne signifient pas nécessairement que les événements les plus intenses arriveront en hiver en climat futur mais plutôt que, pour une fréquence d'occurrence donnée, ces événements seront relativement plus importants. Il importe de considérer la valeur absolue afin de déterminer quels seront les principaux mois ou saisons générateurs d'extrêmes en climat futur. La Figure 4.12 montre par ailleurs que pour les latitudes inférieures à 60°N, les comportements observés sont très différents selon les saisons particulièrement pour la saison estivale. Ce constat ne semble pas spécifique au nombre moyen de dépassements. Des résultats similaires ont été obtenus en considérant les CJP dépassés en moyenne 2 fois et 0.5 fois chaque 20 ans. Ces courbes sont présentées à l'annexe D.

Afin d'analyser la variation intra-annuelle (à l'échelle mensuelle) des extrêmes en climat futur, six tuiles susceptibles de représenter globalement le comportement de différentes régions d'intérêt ont été sélectionnées. L'emplacement de ces tuiles est illustré à la Figure 4.13.

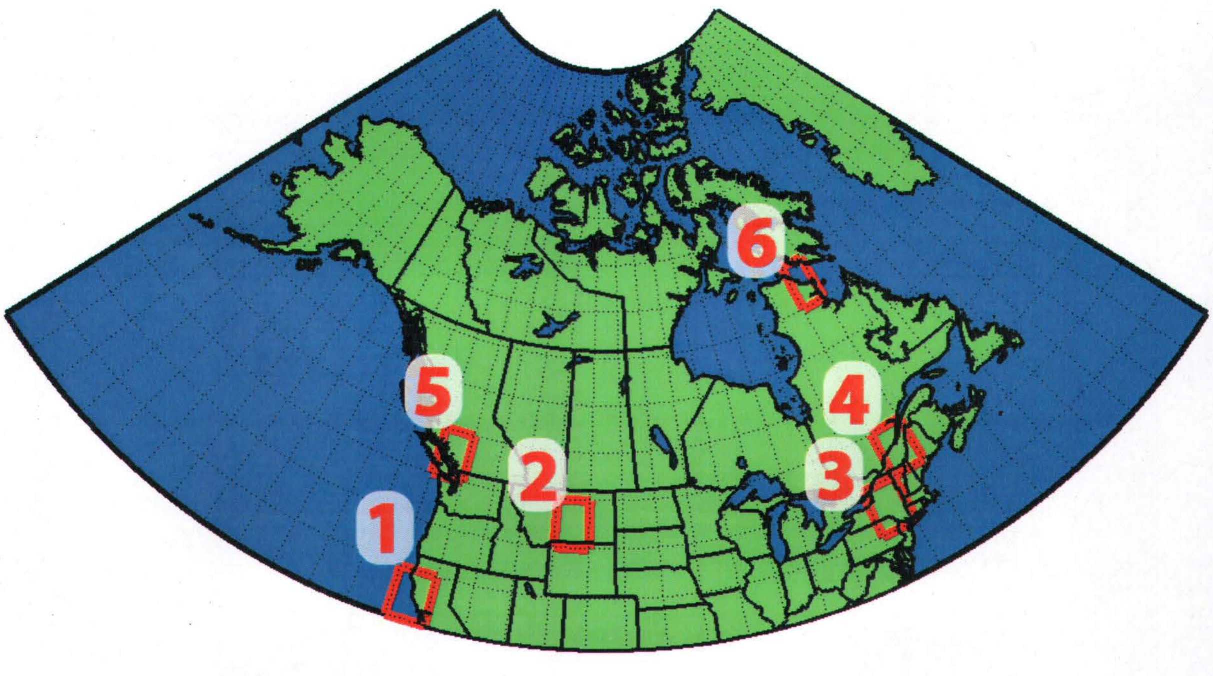

Figure 4.13 : Les six tuiles retenues (encadrées en rouge) pour l'inspection mensuelle des CJP extrêmes.

Les six tuiles ont été choisies de sorte qu'elles représentent *grosso modo* différentes régions de la Figure 4.11. La tuile 1 représente la côte ouest américaine; c'est également la tuile terrestre la plus au sud et la plus à l'ouest du territoire couvert. La tuile 2, qui couvre le Montana, représente le centre des États-Unis et est également typique des Prairies canadiennes. La tuile 3 englobe une partie de l'État de New York au nord des États-Unis. La tuile 4 couvre le sud du Québec et inclut les villes de Québec et de Montréal. La tuile 5 contient le sud-ouest de la

Colombie-Britannique et la ville de Vancouver. La tuile 6 se situe au nord du Québec mais est également représentative du comportement général du Nord Canadien.

Le choix de ces six tuiles permet un bref aperçu des différents types d'évolution des patrons mensuels. Le CJP, dépassé en moyenne une fois chaque 20 ans au cours des différents mois pour les périodes 1850-1950 et 2060-2080, a été calculé pour chacune de ces tuiles. Les résultats obtenus sont présentés à la Figure 4.14.

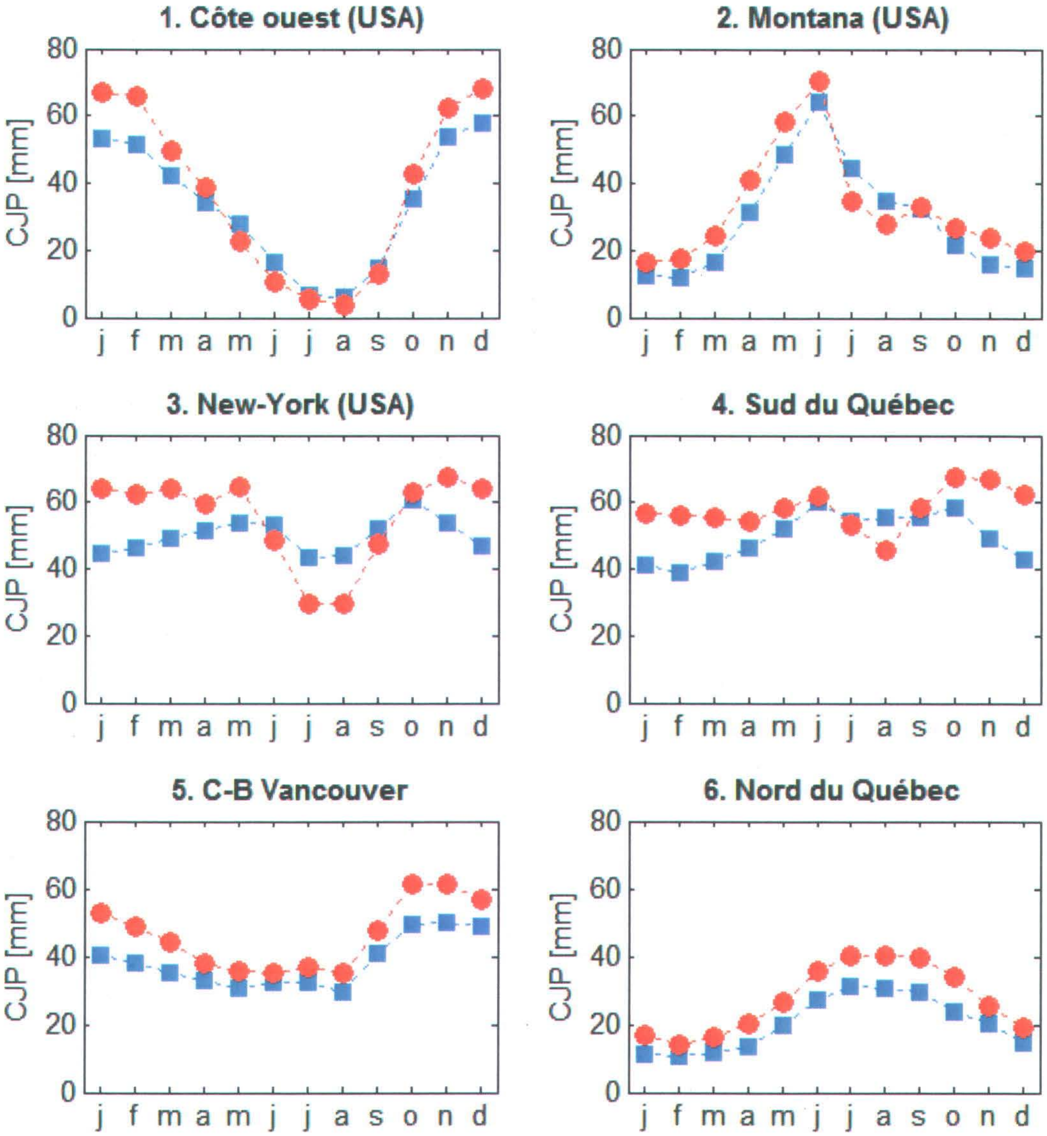

Figure 4.14 : CJP (mm) dépassé en moyenne une fois chaque 20 ans au cours de chacun des mois pour les tuiles retenues (Figure 4.13). Les cercles rouges représentent la période future 2060-2080 et les carrés bleus indiquent la période de référence 1850-1950.

Dans un premier temps, la Figure 4.14 démontre une certaine cohérence inter-mensuelle des résultats obtenus. Il apparaît également que les changements vont se manifester différemment selon les régions géographiques. Le fait que pour un nombre moyen de dépassements donné, le CJP augmente signifie en retour que pour une valeur de CJP fixe, le nombre moyen de dépassements de ce CJP augmente. Ainsi en climat futur, certains mois verront une augmentation substantielle de la probabilité de dépassement d'un CJP donné (mois d'hiver) alors que d'autres verront une baisse (mois d'été).

La tuile qui couvre la côte ouest américaine est caractérisée par une importante variabilité intra-annuelle des extrêmes. Pour les mois estivaux, les valeurs de CJP extrêmes sont très faibles. Ce sont essentiellement les mois d'hiver qui sont propices à la génération de CJP de valeur élevée. En climat futur, une hausse de la valeur de CJP est anticipée pour les mois hivernaux alors qu'il y aura très peu de changements, sinon une légère diminution pour les mois estivaux.

La tuile qui couvre le Montana est typique des régions centrales et se caractérise par une variabilité intra-annuelle importante au niveau des extrêmes. Contrairement à la tuile décrite précédemment, ce sont cependant les mois de mai à juillet, avec un pic centré sur juin, qui sont les plus propices à générer des CJP de valeur élevée. En climat futur, une hausse des CJP est prévue pour les mois de septembre à juin et une légère baisse pour les mois de juillet et d'août.

Pour les tuiles de l'État de New York et du sud du Québec, on constate une diminution des valeurs en été. Cette diminution est plus marquée pour la tuile de l'État de New York, située légèrement plus au sud. Cette diminution est cohérente avec les résultats de la Figure 4.11 et de la Figure 4.12. Pour ces deux tuiles, les moyennes mensuelles de CJP sont maximales au printemps et à l'automne, en climat présent comme en climat futur. Les mois d'hiver sont ceux pour lesquels l'augmentation des CJP est maximale.

Pour la tuile située sur le sud de la Colombie-Britannique, on remarque qu'en climat présent comme en climat futur, ce sont les mois d'automne qui présentent les CJP les plus élevés. Pour tous les mois, on remarque une augmentation en climat futur, apparemment moins marquée en été.

Pour la tuile située au nord du Québec, la courbe est essentiellement décalée vers le haut en climat futur. Les valeurs les plus élevées de CJP se situent durant la saison estivale. L'augmentation appréhendée semble également un peu plus importante en été.

4.7 Sensibilité au taux annuel moyen utilisé pour fixer le seuil

Un élément important à considérer est le choix du taux de dépassement annuel moyen (λ) et son impact sur les résultats. Sans être complètement arbitraire, le choix de ce taux ne se base sur aucune méthode objective. Il est important de regarder la sensibilité des résultats vis-à-vis du taux utilisé pour fixer le seuil. La variation relative des CJP extrêmes en climat futur étant *a priori* la variable la plus intéressante, c'est la sensibilité à l'égard de ce résultat qui sera présentée. Dans un premier temps, on s'intéresse à la variation du CJP dépassé en moyenne une fois sur la fenêtre 2060-2080. À partir de deux taux différents, soit $\lambda = 3$ et $\lambda = 6$, on obtient des cartes similaires à la Figure 4.10 pour la période 2060-2080. Cette carte présente la variation relative du CJP dépassé en moyenne une fois sur la fenêtre 2060-2080 par rapport à celui dépassé en moyenne une fois chaque 20 ans pour la période 1850-1950. Afin d'évaluer la sensibilité des résultats en fonction du taux retenu, on divise les résultats obtenus à partir des taux $\lambda = 3$ et $\lambda = 6$ par ceux obtenus en considérant $\lambda = 12$. Des valeurs inférieures au taux initialement choisi sont retenues afin de se rapprocher des valeurs de taux généralement rapportées dans la littérature (≈ 3 à 4 dépassements par année). On obtient alors la Figure 4.15, qui présente la variation relative des résultats en fonction du taux retenu.

Puisque le taux a une influence directe sur le nombre de dépassements retenus, il va de soi qu'une modification de celui-ci entraînera une certaine variabilité des résultats. Cependant l'important est de pouvoir distinguer si le choix du taux introduit un biais au niveau des estimations. On peut sans trop de risque émettre l'hypothèse que si une différence est observable pour une tuile isolée, cette variation sera principalement attribuable à la variabilité introduite par la modification du taux. En l'absence de biais, on devrait observer une mosaïque de valeurs aléatoires situées de part et d'autre de la valeur 1. Par contre, si on remarque une certaine cohérence spatiale du ratio relatif entre deux taux, cela laisse supposer que le taux retenu était initialement inadéquat pour la région donnée.

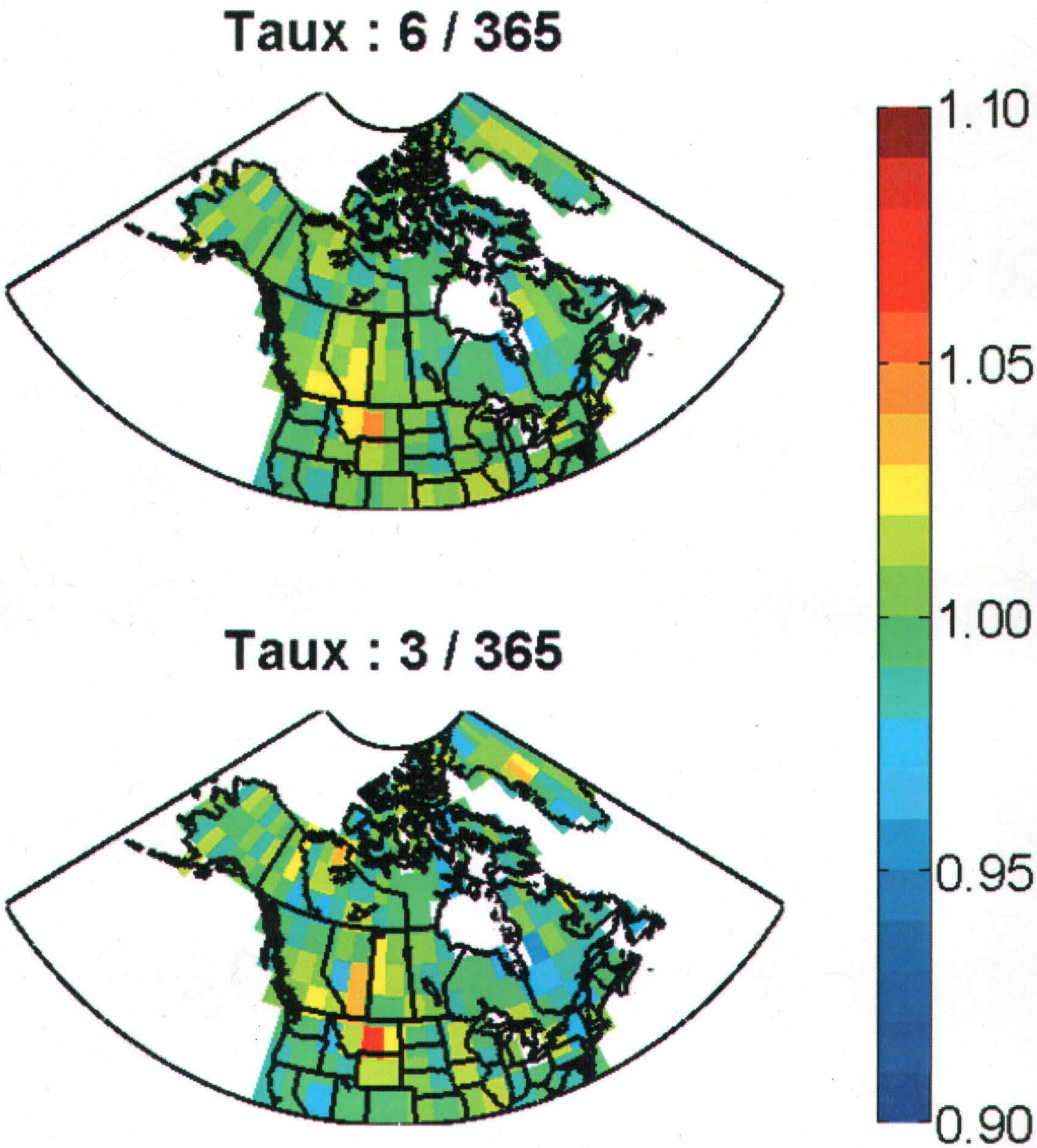

Figure 4.15 : Ratio des résultats des taux moyens $\lambda = 3$ et $\lambda = 6$ dépassements par rapport à $\lambda = 12$ des variations relatives du CJP dépassé en moyenne une fois chaque 20 ans au cours de la période 2060-2080 et de la période 1850-1950.

La Figure 4.15 montre que globalement, le taux utilisé, soit de 12 dépassements en moyenne par année, est convenable. La variation relative moyenne sur l'ensemble du territoire est respectivement de 0.999 et 0.997 pour les taux $\lambda = 6$ et $\lambda = 3$, ce qui signifie que les résultats sont approximativement les mêmes. On remarque cependant une région près des Prairies où on observe une zone cohérente spatialement contenant des résultats obtenus avec des $\lambda = 3$ et $\lambda = 6$ qui sont légèrement plus importants que pour $\lambda = 12$. Dans le cas de $\lambda = 3$, cette différence varie de 2 à 7 %. Cela pourrait signifier que le taux de 12 dépassements utilisé pour cette région n'est pas tout à fait adéquat.

Il convient également d'examiner la sensibilité des résultats saisonniers. Pour ce faire, des cartes semblables à celles de la Figure 4.12 ont été générées. On compare les augmentations relatives entre les périodes 2060-2080 et 1850-1950 du CJP dépassé en moyenne une fois en 20 ans au cours de chacune des saisons obtenues pour $\lambda = 3$ et $\lambda = 12$. La Figure 4.16 présente les résultats obtenus.

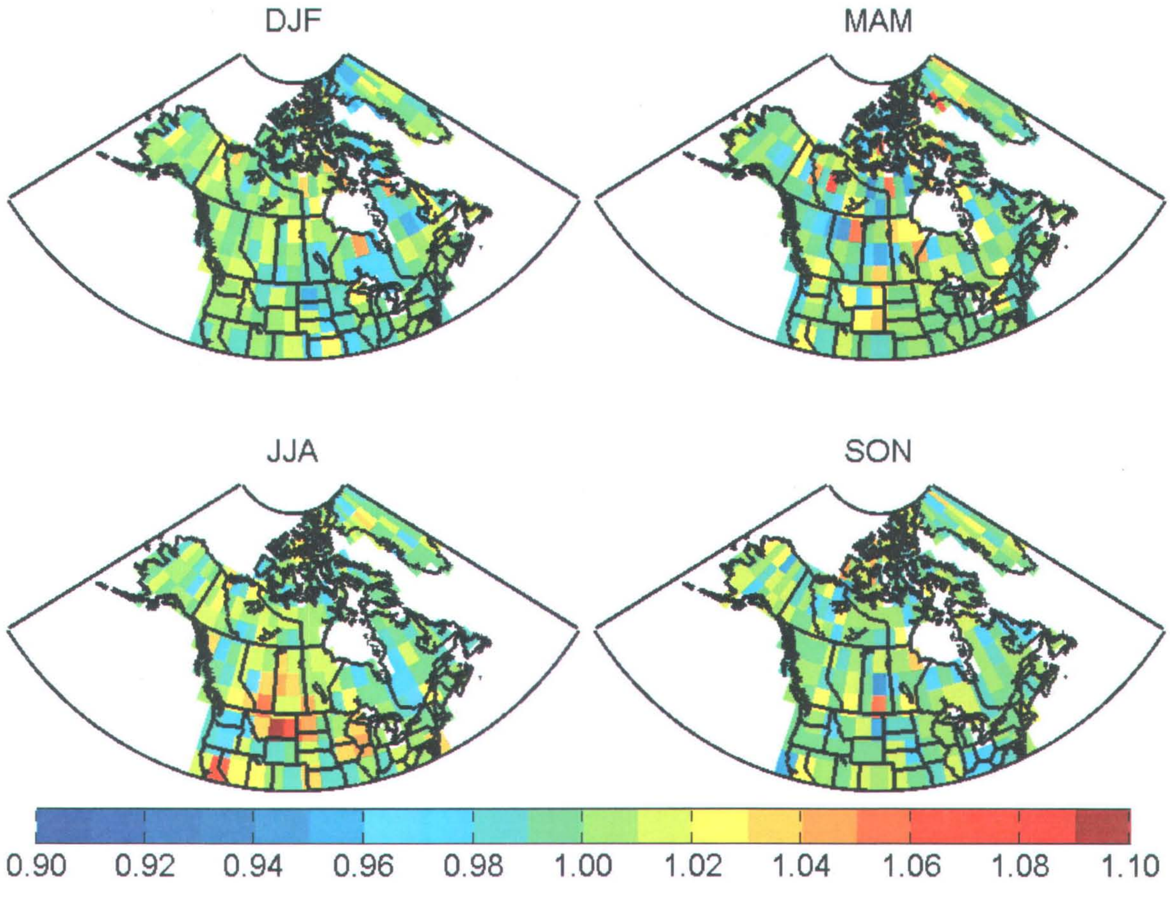

Figure 4.16 : Ratio des résultats pour $\lambda = 3$ par rapport à $\lambda = 12$ des variations relatives du CJP dépassé en moyenne une fois chaque 20 ans au cours des différentes saisons de la période 2060-2080 et de la période 1850-1950.

Pour les saisons d'automne, d'hiver et de printemps, le choix de $\lambda = 3$ ou $\lambda = 12$ ne semble pas avoir d'incidence importante sur les valeurs estimées. Pour la saison estivale, on observe cependant au niveau des Prairies canadiennes une région avec de plus grands écarts entre les valeurs estimées à partir des taux de dépassement 3 et 12. Une analyse plus approfondie de cette région (voir Figure 4.16) permet de constater que les CJP des extrêmes estivaux sont beaucoup plus importants que ceux des autres saisons. Cette variabilité intra-annuelle prononcée explique pourquoi les pluies de grande importance se produisent essentiellement durant l'été. Il est donc probable que la différence présente à la Figure 4.15 soit attribuable à la disparité observée pour la saison estivale. Pour cette région, $\lambda = 12$ semble trop élevé puisqu'on observe systématiquement un écart positif entre les résultats des taux 3 et 12. Pour pallier ce problème, on pourrait choisir d'utiliser un taux unique pour l'ensemble du territoire inférieur au taux $\lambda = 12$ initialement considéré mais, dans ce cas, on perdrait beaucoup d'information puisque le taux initial semble convenir à la majorité du territoire et des saisons.

Pour expliquer l'inadéquation du taux pour cette région, une hypothèse plausible, qui resterait à vérifier, est que l'utilisation d'un taux moyen annuel donné pourrait ne pas convenir à des régions qui présentent une variabilité intra-annuelle importante au niveau des extrêmes. Pour certains mois, le seuil correspondant au taux de dépassement fixé pourrait ne pas satisfaire aux critères asymptotiques de la GPD en retenant des valeurs pas assez extrêmes ou trop inhomogènes, par exemple. L'adéquation du taux pourrait donc dépendre de la climatologie de la région et une solution à ce problème serait l'utilisation d'un taux régional.

La sélection du taux de dépassement restera sans aucun doute toujours le talon d'Achille de la méthode SDP. Paradoxalement, ce taux est à la base de toute la pertinence de la méthode. Quoi qu'il en soit, les vérifications sur la sensibilité tendent à démontrer que, de manière générale, les résultats ne sont pas très sensibles au choix du taux dans la mesure où celui-ci se situe dans des plages raisonnables.

4.8 Comparaison avec la littérature

Pour comparer les résultats avec la littérature, le territoire est divisé en deux sous-régions qui correspondent approximativement à celles utilisées par Kharin *et al.* (2007) (Tableau 4.2). Les moyennes pour les tuiles terrestres pour chacune de ces sous-régions ont été estimées. Deux nouvelles périodes ont également été considérées en climat futur, soit 2045-2065 et 2080-2100, ainsi qu'une nouvelle période de référence, 1980-2000. Ces périodes sont celles utilisées par Kharin *et al.* (2007) dans leur étude qui portait sur un ensemble de 16 modèles globaux, incluant le MCCG3, utilisés par le GIEC. Les résultats présentés sont obtenus des 16 modèles globaux pour le scénario d'émission A2. Le Tableau 4.2 présente les écarts relatifs entre ces périodes et la période de référence pour chaque sous-région.

Tableau 4.2 : Écarts relatifs par rapport à la période 1980-2000 du CJP dépassé en moyenne une fois en 20 ans (P_{20}).

Région	Acronyme	Latitude	$\%P_{20}$			
			Résultats		Kharin *et al.* (2007)	
			2045-2065	2080-2100	2045-2065	2080-2100
Amérique du Nord	NAM	37.1 - 66.8 °N	11.8	22.1	6-16	11-23
Arctique	ARC	66.8 - 81.6 °N	16.5	32.5	8-20	18-38

Les valeurs obtenues en appliquant la méthode développée sont comparables à celles de Kharin *et al.* (2007). On notera par contre que pour la période 2080-2100, les écarts relatifs se situent dans la portion supérieure de la fourchette des valeurs obtenues par Kharin *et al.* (2007).

Selon Allen et Ingram (2002), l'augmentation relative des extrêmes de précipitation par degré Kelvin devrait être près de la valeur théorique de 6 – 7 % $°K^{-1}$ estimée à partir de l'équation de Clausius-Clapeyron (Boer, 1993). Afin de vérifier si les résultats obtenus dans le cadre de la présente analyse sont conformes à cette hypothèse, les anomalies de la température annuelle moyenne pour chacune des deux régions sur les deux fenêtres futures ont été calculées par rapport à la température annuelle moyenne pour la période 1980-2000 (Tableau 4.3).

En divisant l'augmentation relative des CJP par cette anomalie, on obtient l'augmentation relative par degré Kelvin. Les CJP dépassés en moyenne une fois en 5 ans (P_5), 10 ans (P_{10}) et 20 ans (P_{20}) sont considérés (Tableau 4.3).

Tableau 4.3 : Anomalies de température et taux d'augmentation par degré d'anomalie des CJP par rapport à la période 1980-2000.

Région	2045-2065				2080-2100			
	T [°K]	$\%P_5°K^{-1}$	$\%P_{10}°K^{-1}$	$\%P_{20}°K^{-1}$	T [°K]	$\%P_5°K^{-1}$	$\%P_{10}°K^{-1}$	$\%P_{20}°K^{-1}$
NAM	2.1	5.7	5.7	5.6	3.9	5.5	5.5	5.6
ARC	2.9	5.5	5.6	5.7	6.0	5.0	5.2	5.4

Les taux d'augmentation obtenus oscillent entre 5.0 et 5.7 $\%°K^{-1}$ et semblent indifférents au temps de retour considéré. Dans l'ensemble, les valeurs obtenues se comparent assez bien avec la valeur théorique de 6-7 % déterminée à partir de l'équation de Clausius-Clapeyron. Cela semble supporter l'hypothèse selon laquelle l'intensité des précipitations extrêmes est déterminée dans une large mesure par la capacité de l'atmosphère à emmagasiner de l'humidité.

Enfin, pour avoir une idée de l'augmentation en fréquence du CJP P_{20}, on a calculé le nombre moyen de dépassements de P_{20}, noté N, pour la période 1980-2000 (Tableau 4.4).

Tableau 4.4 : Nombre moyen de dépassements de P_{20} (période 1980-2000) durant les diverses périodes considérées

Région	N		
	1980-2000	2045-2065	2080-2100
NAM	1.0	2.1	3.6
ARC	1.0	2.0	3.4

Le Tableau 4.4 indique que le CJP dont l'espérance du nombre de dépassements est de 1 sur la période 1980-2000 sera dépassé en moyenne 2 fois pendant la période 2045-2065 et plus de 3 fois durant la période 2080-2100. Le résultat est sensiblement le même pour les régions NAM et ARC avec des valeurs respectives de 2.1 et 2.0 (pour 2045-2065) et de 3.6 et 3.4 (pour 2080-2100). Ainsi, la valeur de référence historiquement dépassée en moyenne une fois tous les 20 ans en 1980-2000 sera dépassée en moyenne tous les 10 ans vers le milieu du 21^{e} siècle et tous les 6 ans environ à la fin de ce siècle. Kharin *et al.* (2007) ont obtenu une réduction des temps de récurrence pour les régions des tropiques et de haute et moyenne latitude d'un facteur 2 pour 2045-2065 et d'un facteur 3 pour 2080-2100. À partir du modèle canadien de climat global antérieur MCCG2 et du scénario de forçage SRES A2, Kharin et

Zwiers (2005) ont pour leur part trouvé que le temps de retour de l'occurrence P_{20} (calculé en 2000) serait environ de 7.5 ans en 2090 pour NAM et ARC. Les résultats obtenus avec la présente méthode se situent dans les mêmes plages que celles mentionnées dans la littérature.

5. APPLICATION AUX SÉRIES OBSERVÉES

Les hauteurs de précipitations extrêmes et leur récurrence constituent des éléments importants dans la conception et le dimensionnement d'ouvrages hydrauliques dans une optique de gestion du risque. En climat actuel, ces occurrences extrêmes sont déterminées à partir des séries observées. À cet effet, la méthode des SMA est généralement employée. Toutefois, cette approche ne considère qu'une seule valeur par année et conduit généralement à des échantillons de petite taille. L'application de la méthode SDP aux séries observées constitue une alternative très intéressante puisqu'elle permet de pallier les faiblesses de l'approche SMA en considérant des échantillons plus grands.

Le présent chapitre se différencie du précédent et présente une application de la méthode SDP aux séries observées aux stations pluviométriques. Contrairement aux séries issues de modèles climatiques, les séries observées sont généralement moins longues. Au niveau de la caractérisation des extrêmes, cela constitue une difficulté additionnelle. Dans un premier temps, la section 5.1 donne une description des données utilisées. La section 5.2 porte sur l'estimation du seuil. La section 5.3 traite de la modélisation de la variation temporelle des paramètres par le biais de deux approches. La section 5.4 introduit le concept de régionalisation alors que la section 5.5 présente le critère AIC_c employé afin de classer les différents modèles. La section 5.6 compare différents modèles dont les paramètres peuvent être régionaux ou locaux. Finalement, la section 5.7 compare différents modèles régionaux qui intègrent diverses covariables spatiales.

5.1 Données utilisées

Les stations pluviométriques analysées sont localisées dans la partie sud du Québec et sont opérées par le Ministère du Développement Durable, de l'Environnement et des Parcs (MDDEP) et Environnement Canada. Pour ces stations, les hauteurs de précipitations maximales tombées pendant différents intervalles de temps sont disponibles quotidiennement, soient les cumuls maximaux sur 5, 10, 15 et 30 minutes ainsi que 1, 2, 6 et 12 heures. Pour les durées comprises entre cinq et 30 minutes inclusivement, les séries de 107 stations sont disponibles, alors que pour les durées 1, 2, 6 et 12 heures, on dispose des séries de 109 stations. La période couverte s'étend de mai à octobre (MAO) puisque les stations ne sont pas en opération pendant la saison hivernale. Seules les stations situées au sud du 49^{e} parallèle, dont l'historique est supérieur à 10 ans et qui comportent un minimum de 300 mesures pour chacun des six mois ont été retenues. Les historiques disponibles varient entre 10 et 51 ans. Les premières mesures datent de 1943 et les plus récentes de 2009. La Figure 5.1 résume sous forme graphique les années couvertes par chacune des 109 stations. La moyenne et l'écart-type du nombre d'années disponibles sont respectivement de 32 et 9.7 ans. La Figure 5.2 présente, quant à elle, la carte localisant les différentes stations.

Il faut porter attention au fait que les maxima dans les séries quotidiennes correspondent au maximum survenu pendant une journée, soit durant une période de 24 heures débutant à 8h du matin. Pour les plus longues durées (par exemple 6 heures et 12 heures), la probabilité que l'événement de pluie générateur du maximum chevauche deux journées est importante. Lorsqu'il y a chevauchement, la valeur reportée dans la série sous-estime la valeur réelle de l'intensité maximale de l'événement sur le pas de temps considéré (Mailhot et Talbot, 2011).

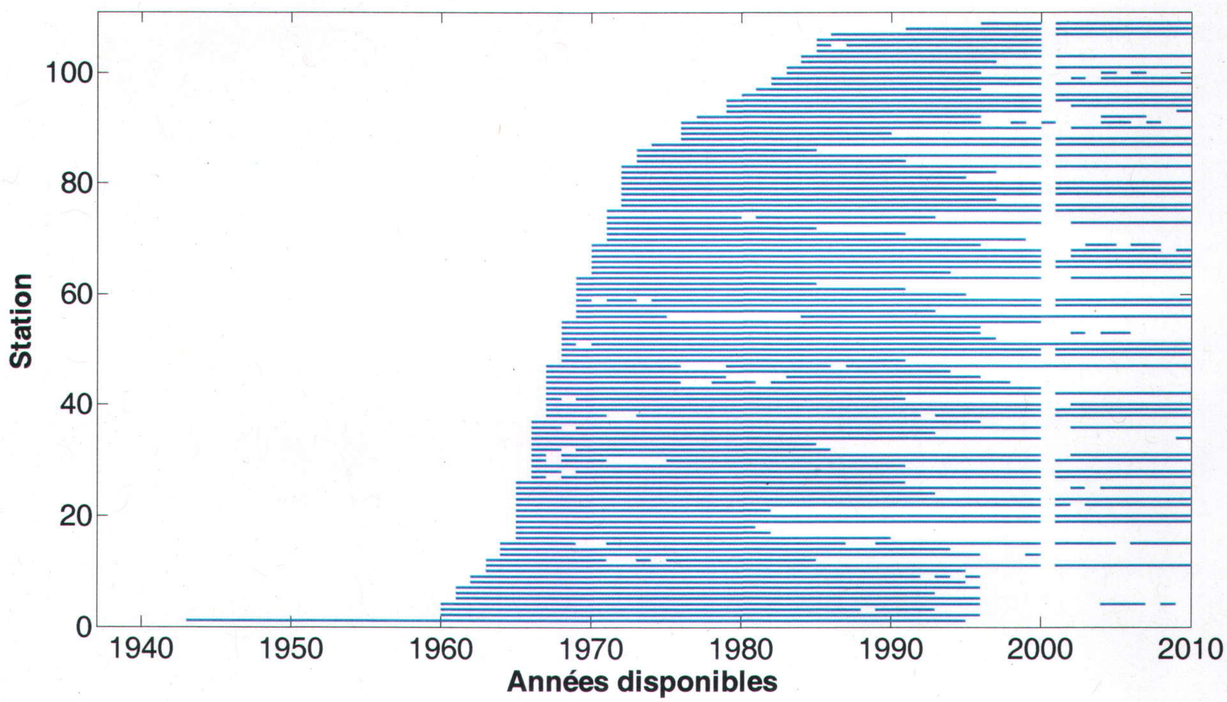

Figure 5.1 : Années pour lesquelles des mesures de mai à octobre sont disponibles pour les 109 stations. Les lignes bleues indiquent la présence de mesures.

L'absence de mesures pour l'année 2000 est le résultat d'un problème d'acquisition. Toutes les données ont été perdues (communication personnelle de Mme Catherine Savard, MDDEP).

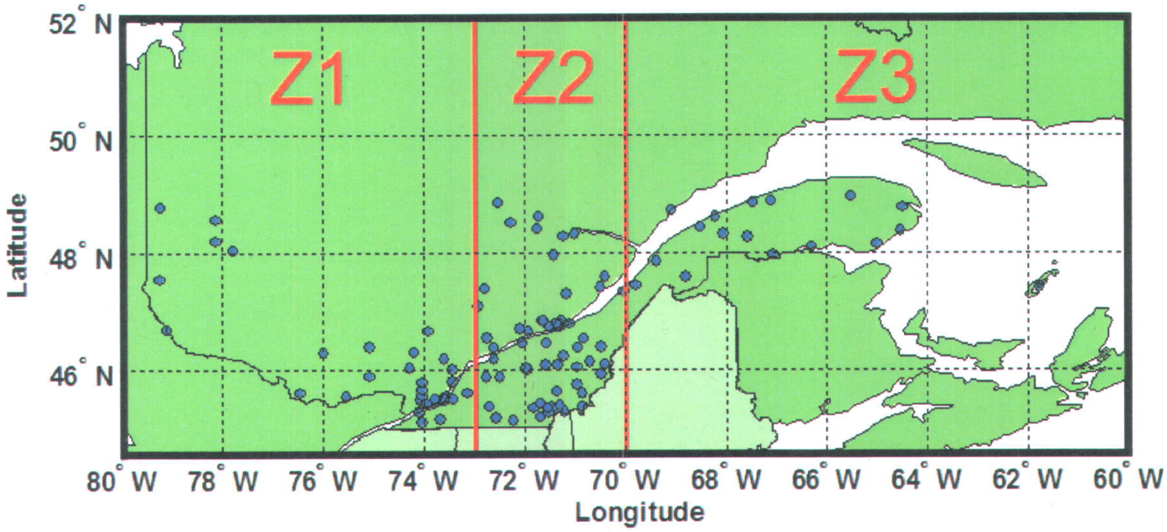

Figure 5.2 : Cartes des stations météorologiques retenues pour analyse. Les points bleus indiquent l'emplacement des stations. Les lignes rouges délimitent les trois zones présentées au Tableau 5.1.

5.2 Estimation du seuil

La méthodologie développée afin d'obtenir le seuil qui a servi à construire les SDP est globalement similaire à celle présentée à la section 4.2. De légères modifications ont cependant été apportées. L'annexe E présente les détails de la méthode. Le seuil est déterminé à partir d'un taux de dépassement unique de six dépassements par année MAO, soit en moyenne un dépassement par mois. Dans le cas des séries aux stations, tel qu'exposé à la section 5.1, les historiques ne couvrent pas plus d'une cinquantaine d'années. Le signal de changements climatiques sur les extrêmes ne sera probablement pas perceptible pour de si courtes séries. La non stationnarité interannuelle est pour cette raison ignorée. On vérifiera par ailleurs à la section 5.3 que les SDP sont bel et bien stationnaires à partir du test de Mann-Kendall (Tableau 5.2). Puisque la non stationnarité n'est pas considérée, les séries, contrairement à celles de la section 4.2, seront uniquement fragmentées en blocs mensuels (voir annexe E). Pour chacun de ces blocs et pour chacune des stations, la valeur du seuil correspondant au taux de dépassement considéré est estimée en prenant bien soin d'effectuer le *declustering* des séries. On obtient, à l'instar de la courbe bleue de la Figure 4.4, une succession de plateaux qui correspondent à la valeur de seuil déterminée pour chaque mois. Il n'existe cependant pas de variation interannuelle du seuil et on n'obtient que six plateaux par année, car seuls les mois de mai à octobre sont considérés. Un lissage est effectué afin d'obtenir un seuil continu intra-annuellement. Le seuil ainsi obtenu est propre à chaque station et est utilisé pour obtenir les SDP. Afin de visualiser l'allure générale du seuil au sein du territoire à l'étude, on a arbitrairement divisé celui-ci en trois zones distinctes selon la longitude (voir Tableau 5.1 et Figure 5.1). Les longitudes choisies assurent la présence d'un certain nombre de stations au sein de chacune des zones. Ces divisions visent essentiellement à simplifier la présentation des résultats.

Tableau 5.1 : Subdivision du territoire en trois sous-régions

Région	Nom	Longitude	Nombre de stations
Ouest	Z1	> 73 °W	33
Centre	Z2	70-73 °W	59
Est	Z3	< 70 °W	17

Pour chacune des zones, le seuil moyen pour chaque jour de l'année MAO est calculé à partir des différentes stations. Le but de cet exercice est de constater si, de manière générale, le seuil varie au sein du territoire considéré. Les résultats sont présentés à la Figure 5.3.

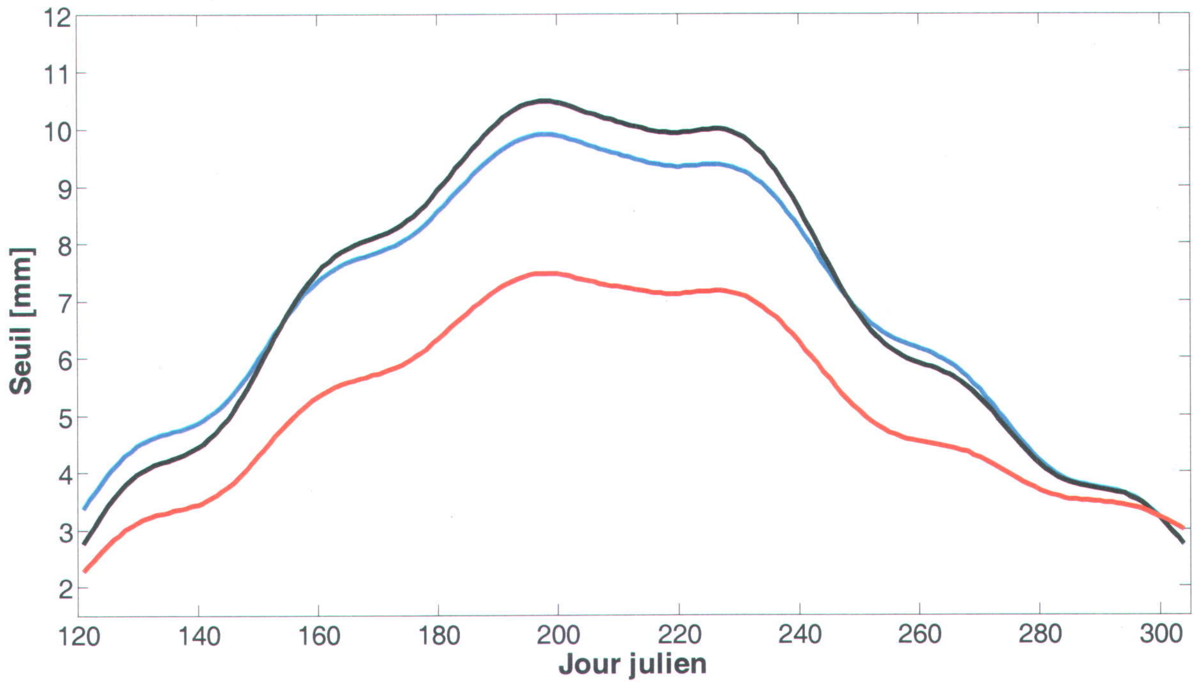

Figure 5.3 : Moyenne des seuils pour la durée 30 minutes pour un taux moyen de 6 dépassements MAO en fonction du jour julien pour les stations des trois zones. Le trait bleu représente la zone Z1, le trait noir la zone Z2 et le trait rouge la zone Z3 (voir Tableau 5.1 et Figure 5.2).

La Figure 5.3 suggère une différence substantielle entre le seuil moyen calculé pour les zones Z1 et Z2 et celui de la zone Z3. Le seuil moyen de la zone Z3, qui correspond à la péninsule gaspésienne, est plus faible. La même remarque peut-être faite pour les sept autres durées. Cette différence n'est pas constante dans le temps, elle atteint un maximum entre les 200 et 210^{e} jour julien, soit pendant la deuxième moitié de juillet. La différence marquée pour le seuil de la zone Z3 constitue un indice de l'inhomogénéité spatiale des extrêmes. On remarque également que cette inhomogénéité semble *a priori* plus ou moins prononcée selon la période de l'année. Pour la suite des choses, il sera intéressant d'analyser si cette inhomogénéité est également présente au niveau des paramètres de forme et d'échelle.

5.3 Modélisation de la variation temporelle des paramètres

Pour vérifier l'hypothèse de stationnarité émise à la section 5.2, le test de Mann-Kendall (Hamed et Rao, 1998) a été appliqué aux SDP mensuelles des six mois à l'étude. Pour chacun de ces mois et pour chacune des huit durées de précipitations étudiées, la fraction de séries considérées stationnaires par le test avec un seuil α=0.05 est estimée. Le Tableau 5.2 présente ces résultats.

Tableau 5.2 : Fraction des séries SDP considérées stationnaires par le test de Mann-Kendall avec un seuil α =0.05.

Mois	Durée							
	5 min.	10 min.	15 min.	30 min.	1h	2h	6h	12h
Mai	0.95	0.96	0.97	0.94	0.99	0.99	0.96	0.93
Juin	0.93	0.94	0.97	0.98	0.95	0.96	0.94	0.95
Juil.	0.94	0.98	0.96	0.94	0.95	0.94	0.93	0.96
Août	0.96	0.94	0.93	0.99	0.97	0.96	0.95	0.94
Sep.	0.93	0.95	0.96	0.95	0.94	0.97	0.92	0.95
Oct.	0.95	0.95	0.94	0.93	0.95	0.93	0.94	0.97

Les fractions présentées au Tableau 5.2 montrent clairement que les SDP mensuelles sont stationnaires. Ces fractions se rapprochent toutes de 0.95, soit la valeur attendue pour un test avec un seuil α =0.05. Le même constat est obtenu à partir d'un test différent (non montré) visant à vérifier si le paramètre de pente est statistiquement significatif pour une régression linéaire des SDP avec le temps comme variable explicative. Pour ces raisons, seule la modélisation de la variation intra-annuelle des paramètres de la GPD sera considérée. Deux méthodes qui permettent de modéliser cette variation seront comparées afin de tenter de déterminer la manière optimale de procéder. Elles sont décrites dans les deux sous-sections suivantes.

5.3.1 Représentation en séries de Fourier

La modélisation non paramétrique de la variation intra-annuelle pose un problème de continuité. Avec une approche similaire à celle de la section 4.3.1, on observe des discontinuités importantes entre les valeurs des différents blocs mensuels qui servent à diviser l'année. Une alternative au lissage, qui vise à pallier ce problème, est d'opter pour une modélisation paramétrique de la variabilité intra-annuelle. Considérant le caractère périodique des saisons, une représentation par séries de Fourier semble appropriée. Mendez *et al.* (2008) ont utilisé de telles séries afin de modéliser la variation intra-annuelle des valeurs de hauteurs de vagues extrêmes. Katz *et al.* (2002) ont également eu recours à des séries sinusoïdales afin de modéliser la variation intra-annuelle de la distribution des CJP supérieurs à un seuil donné. Ces deux articles ont permis de montrer que la prise en compte de la saisonnalité améliorait sensiblement la qualité des ajustements.

Les paramètres de forme et d'échelle sont donc modélisés à l'aide de séries de Fourier à m_k et m_α modes respectivement, comme suit :

$$k(n)=k_0+\sum_{j=1}^{m_k}\left\{a_j\cos\left(2jn\frac{\pi}{T}\right)+b_j\sin\left(2jn\frac{\pi}{T}\right)\right\} \tag{5.1}$$

$$\alpha(n)=\alpha_0+\sum_{j=1}^{m_\alpha}\left\{c_j\cos\left(2jn\frac{\pi}{T}\right)+d_j\sin\left(2jn\frac{\pi}{T}\right)\right\} \tag{5.2}$$

où n désigne le jour julien et T la période (365.25 jours). Le logarithme de la vraisemblance de la série est donné par :

$$\ln\left\{\prod_{i=1}^{N} f(x_i,n_i)\right\}=\sum_{i=1}^{N}\left\{-\ln(\alpha(n_i))+\left(\frac{1}{k(n_i)}-1\right)\ln\left(1-\frac{k(n_i)(x_i-\xi_i)}{\alpha(n_i)}\right)\right\} \tag{5.3}$$

Les ensembles de coefficients $\left\{a_j,b_j\right\}, j=1,\ldots,m_k$ et $\left\{c_j,d_j\right\}, j=1,\ldots,m_\alpha$ sont estimés par maximisation de la vraisemblance. Le nombre total de paramètres à déterminer est donné par $2\,(m_k+m_\alpha+1)$.

5.3.2 Séries mensuelles

L'utilisation d'une méthode paramétrique comme les séries de Fourier exige de préciser des formes fonctionnelles qui décrivent l'évolution temporelle des paramètres. Il est alors possible que certaines variations intra-annuelles ne soient pas adéquatement représentées. Afin d'éviter une contrainte sur la forme des variations et également dans le but de valider ou d'invalider l'utilisation des séries de Fourier, on considère une représentation non paramétrique des variations intra-annuelles. Il s'agit dans ce cas d'utiliser une approche similaire à celle précédemment exposée à la section 4.3.1 pour le MCCG3, à savoir une fragmentation de l'année en séries mensuelles. Cependant, on s'abstiendra cette fois d'interpoler les différents plateaux mensuels afin d'optimiser la valeur de ceux-ci depuis la méthode du maximum de vraisemblance. Évidemment, on fait face en procédant ainsi aux problèmes que l'on tentait de contourner en utilisant les séries de Fourier. Les paramètres de chaque mois sont traités indépendamment et les variations sont discontinues d'un mois à un autre. Ultimement, on espère pouvoir se prononcer quant à la méthode qui permet de mieux représenter la variation intra-annuelle en comparant le nombre de paramètres et la vraisemblance des différentes approches.

5.4 Analyse locale versus régionale

Lorsque l'on travaille avec des sorties de modèles, on dispose bien souvent d'une abondance de données, ce qui permet une meilleure inférence statistique des paramètres. Cependant, lorsque les données à analyser proviennent de stations, les séries sont généralement beaucoup moins longues, présentent souvent des données manquantes et ne couvrent que quelques mois par année. Il n'est pas rare d'obtenir, pour les stations d'une même région, des différences significatives sur les intensités pour un même temps de retour. Ces différences s'expliquent, entre autres, par la faible taille de l'échantillonnage (source de variabilité) et par le caractère ponctuel de certains évènements de pluies intenses. Ainsi, en raison de la nature des événements extrêmes, une station à moins de 10 km d'une autre peut capter un événement très intense alors que l'autre pourra enregistrer aucune précipitation. Une analyse locale, station par station, conduirait alors à des résultats forts différents lorsque les séries sont relativement courtes.

La Figure 5.4 présente le résultat d'une analyse station par station où l'on a calculé le cumul de précipitations sur 30 minutes associé au temps de retour 10 ans à partir de l'approche SDP. On a ensuite assigné la valeur calculée au plus proche voisinage de la station selon la méthode des polygones de Thiessen.

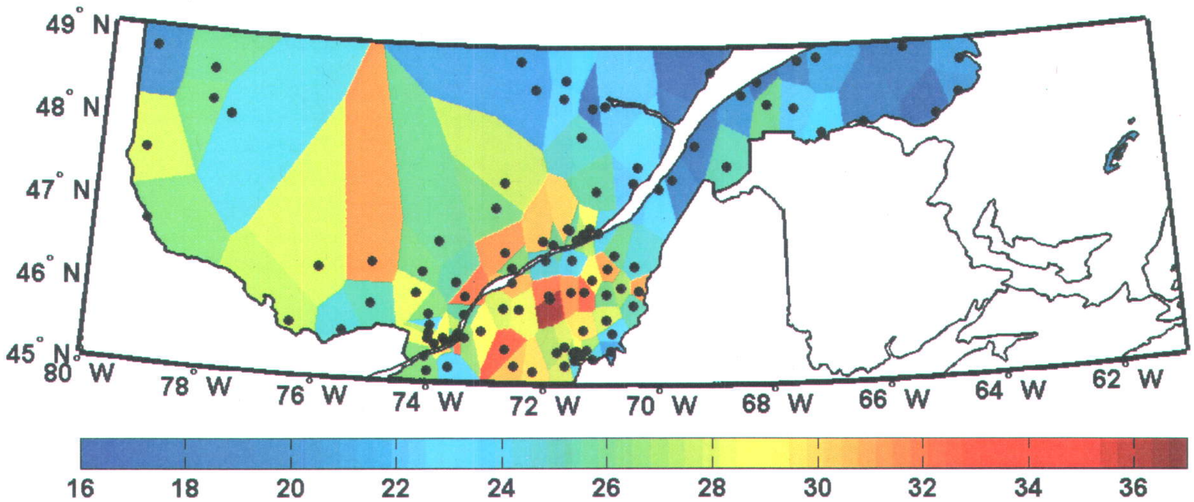

Figure 5.4 : Cumul (mm) sur 30 minutes de temps de retour 10 ans estimé selon le modèle 2L0L (voir Tableau 5.4). L'interpolation est effectuée par la méthode des polygones de Thiessen.

La Figure 5.4 permet de constater à quel point la variabilité inter-station peut être grande même pour de petits territoires. Sur la base de cette constatation, il semble hasardeux de se fier aux résultats d'une seule station si l'on désire obtenir une information qui présente une certaine cohérence spatiale. La variabilité considérable observée à la Figure 5.4 s'explique en grande partie par les courtes séries disponibles en chaque station et par la nature spatiale très hétérogène des événements étudiés.

5.4.1 Régionalisation

Pour pallier le problème d'échantillonnage, on peut avoir recours à la régionalisation. La régionalisation repose sur l'hypothèse selon laquelle, pour une région donnée, un ou plusieurs des paramètres de la distribution de la variable d'intérêt sont spatialement homogènes, ce qui permet de diminuer le nombre de paramètres à estimer. Cependant, il faut préciser dans ce cas les méthodes qui permettent d'estimer régionalement les paramètres et les critères qui rendent possible l'identification d'une région homogène. Plusieurs auteurs se sont intéressés aux différents aspects de la régionalisation (voir par exemple Madsen *et al.*, 2007b; Hosking et Wallis, 1997; Hosking et Wallis, 1993). Afin de vérifier l'homogénéité de la région, la statistique H_1 de Hosking et Wallis (1997) a été utilisée. Ce test, basé sur la théorie des *L-moments*, détermine le niveau d'homogénéité d'une région en comparant la dispersion régionale des ratios de L-moments d'un groupe de stations avec la dispersion d'un groupe homogène influencé seulement par la variabilité d'échantillonnage. Selon ce test, la région est considérée acceptablement homogène pour l'ensemble des durées.

Chaque paramètre de la distribution peut être considéré comme étant soit local ou régional. Un paramètre local signifie que sa valeur est propre à un site donné alors que pour un paramètre régional, la valeur est commune à tous les sites inclus dans la région d'intérêt. Pour l'estimation des paramètres régionaux, certaines méthodes existent, notamment la méthode des *L-moments* (Hosking et Wallis, 1997). Pour le présent travail une version régionalisée de la méthode du maximum de vraisemblance qui consiste à maximiser la somme des log-vraisemblances aux stations a été retenue. Buishand (1991) a utilisé cette procédure afin d'inférer les paramètres d'une GEV régionale. Cette approche permet, entre autres, de travailler dans un cadre bayésien et d'inclure des covariables dans l'estimation des paramètres. L'utilisation de cette méthode sous l'hypothèse régionale est également conceptuellement simple et intuitive. Il importe de préciser que, pour certaines stations, il peut exister une certaine corrélation entre les occurrences étant donné les faibles distances inter-stations et le chevauchement des historiques. Cependant, Buishand (1991) indique que la corrélation spatiale n'introduit pas de biais significatif dans l'estimation des paramètres. Par contre, cette corrélation aura un impact dans l'évaluation des intervalles de confiance.

Dans un premier temps, il faut statuer quant au caractère régional ou local des différents paramètres. On propose à la section 5.5.1 plusieurs modèles afin de tester les différentes combinaisons possibles. Dans la littérature, on suggère souvent de considérer un paramètre de forme régional (ex. : Smith 1989). Il n'est cependant pas exclu que le paramètre d'échelle le soit également.

La Figure 5.5 schématise les étapes de calcul des paramètres des modèles régionaux. Pour des valeurs de paramètres régionaux données, les valeurs des paramètres locaux à chaque station sont obtenues en maximisant la vraisemblance. Le logarithme de la vraisemblance de chaque site est ensuite additionné pour obtenir une valeur régionale du logarithme de la vraisemblance. Les valeurs du ou des paramètres régionaux sont alors optimisées afin de maximiser la valeur du logarithme de la vraisemblance régionale.

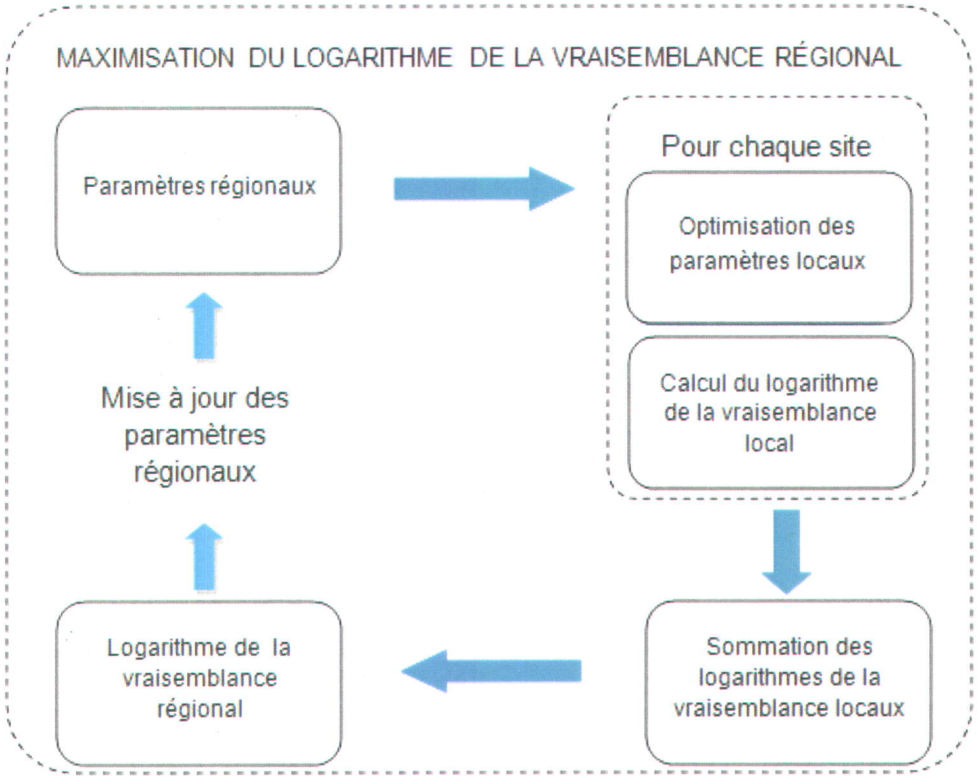

Figure 5.5 : Schéma du processus d'optimisation des paramètres pour l'analyse régionale.

Le schéma de la Figure 5.5 se traduit par l'équation du logarithme de la vraisemblance régionale suivante :

$$\mathcal{L}_{\max} = \max_{\theta} \mathcal{L}(\theta) = \max_{\theta^{(R)}} \sum_{p=1}^{S} \left[\max_{\theta^{(L)}} \sum_{i=1}^{N_p} \ln \left\{ f(x_{i,p} \,|\, \theta_p^{(L)} \, \theta^{(R)}) \right\} \right] \quad (5.4)$$

avec :

$f(.)$ la fonction de densité de probabilité de la GPD

$\theta = \left[\theta^{(L)}, \theta^{(R)} \right]$ le vecteur des paramètres de la GPD

$\theta_p^{(L)}$ le vecteur des paramètres locaux de la station p

$\theta^{(R)}$ le vecteur des paramètres régionaux

S le nombre de stations

N_p le nombre de valeurs de la SDP pour la station p

$x_{i,p}$ valeurs $i = 1 \ldots N_p$ de la SDP à la station p

5.5 Critère d'Akaike (Akaike's Information Criterion)

Plusieurs modèles pourront être proposés en considérant : 1) le caractère local ou régional de chacun des paramètres; et 2) le nombre de modes optimal pour la modélisation de type séries de Fourier de la variation intra-annuelle du paramètre. Au final, il importe de départager lesquels parmi ces modèles sont les plus aptes à représenter les séries d'observations disponibles. La plupart de ces modèles sont imbriqués (c'est-à-dire qu'un modèle plus complexe peut être développée depuis un modèle plus simple en ajoutant un ou plusieurs paramètres de valeurs définies) et donc leur vraisemblance augmentera ou stagnera avec une augmentation du nombre de paramètres. Cependant, le gain en vraisemblance ne justifie pas toujours l'accroissement de la complexité. Le critère AIC, acronyme pour *Akaike's Information Criterion*, (Akaike, 1973) permet de considérer à la fois la vraisemblance d'un modèle et sa complexité, définie à travers le nombre de paramètres.

Depuis son développement en 1974, le critère AIC a été utilisé dans bon nombre de disciplines et de champs d'activités. Strupczweski *et al.* (2001) ont eu recours à ce critère afin d'identifier le modèle optimal qui décrit la fréquence des crues parmi 56 modèles considérés. Mendez *et al.* (2008) l'ont utilisé dans une optique similaire, s'intéressant cette fois aux hauteurs de vagues extrêmes. Récemment, Burnham et Anderson (2004) ont publié un article où ils exposent les limites du critère AIC et certaines erreurs ou mésinterprétations couramment observées lors de son utilisation. Burnham et Anderson (2004) recommandent, entre autres choses, l'utilisation du critère corrigé AIC_c qui apporte une correction de deuxième ordre permettant de tenir compte du rapport entre la taille de l'échantillon et le nombre de paramètres du modèle. Les deux tests convergent vers le même résultat dans le cas où la taille de l'échantillon tend vers l'infini. Le score AIC_c se calcule à partir de l'équation suivante :

$$AIC_c = -2\log\left\{\mathcal{L}_{\max}\left(\hat{\theta}\right)\right\} + 2K + \frac{2K(K+1)}{n-K-1} \qquad (5.5)$$

$\mathcal{L}_{\max}\left(\hat{\theta}\right)$ correspond au maximum de vraisemblance, K au nombre de paramètre et n à la taille de l'échantillon. Le modèle optimal est celui pour lequel la valeur AIC_c est la plus faible. On remarque que le score diminue avec un accroissement du logarithme de la vraisemblance, donc la qualité de l'ajustement, et augmente avec le nombre de paramètres.

La valeur absolue du score AIC_c ne possède aucune signification particulière, ce qui importe étant l'écart entre les scores des différents modèles. Afin de faciliter la comparaison de ces derniers, on a recours au *delta* AIC_c calculé à l'aide de l'équation suivante :

$$\Delta_i = AIC_{c_i} - \min_i\left[AIC_{c_i}\right] \qquad (5.6)$$

Le terme $\min_i[AIC_{c_i}]$ renvoie la valeur minimale de AIC_c parmis les modèles candidats. Le *delta* AIC_c calculé à partir de l'équation (5.6) donne, pour chacun des modèles, l'écart entre son score AIC_c et celui du meilleur candidat parmi le groupe de M candidats. Les valeurs de Δ_i permettent une comparaison rapide des différents modèles. Plus le Δ_i est élevé, moins le modèle est susceptible d'être retenu.

La vraisemblance d'un modèle peut être évaluée à partir des valeurs de Δ_i. Il convient cependant de pondérer la vraisemblance de ces modèles. Les poids *Akaike* ω_i sont calculés à partir de l'équation suivante :

$$\omega_i = \frac{\exp(-\Delta_i / 2)}{\sum_{m=1}^{M} \exp(-\Delta_m / 2)} \tag{5.7}$$

Les poids *Aikake* sont intéressants car ils correspondent à la probabilité relative que chacun des modèles soit le modèle offrant le meilleur ajustement parmi les modèles proposés.

Le critère AIC n'est pas le seul critère ou test permettant de guider la sélection d'un modèle. Semblable au critère AIC, le critère BIC peut également être utilisé. Comparativement au test AIC, le test BIC pénalise davantage le nombre paramètres avec une pénalité de $k\ln(N)$ comparativement au terme de pénalité $2k$ du critère AIC avec k le nombre de paramètres et N la taille de l'échantillon (Burnham et Anderson, 2004). Pour une taille d'échantillon moyenne $N = 16800$, considérant les huit durées, le test BIC revient à pénaliser en moyenne 4.7 fois plus l'ajout de paramètres que le critère AIC. Burnham et Anderson (2002) argumente cependant en faveur du critère AIC et affirment que celui-ci possède plusieurs avantages sur son concurrent. Premièrement ce critère se fonde sur la théorie de l'information. Le terme de pénalité dans le critère AIC n'est pas empirique, celui-ci origine de l'estimation de l'information de Kullback-Leibler, élément fondamental de la théorie de l'information. Deuxièmement, les auteurs sont critiques face au choix du *prior* dans le développement du critère BIC. Celui-ci est égal à $1/R$ où R est le nombre de modèles considérés. Ce *prior* ne considère ni le nombre de paramètres du modèle ni la taille de l'échantillon ce qui leur apparait déraisonnable. Finalement, les auteurs montrent qu'à l'instar du critère BIC, le critère AIC peut être développé dans un cadre bayésien en utilisant simplement un *prior* différent et peut par conséquent être considéré comme un résultat bayésien.

Outre les critères AIC et BIC, le test du rapport de vraisemblance est, dans le cas de modèles imbriqués, également applicable. Ce test consiste à vérifier si l'ajout de un ou plusieurs paramètres est statistiquement significatif via un test d'hypothèse. Bien qu'élégant ce test est un peu moins général que le critère AIC puisque son application est restreintes aux modèles imbriqués c'est-à-dire lorsque le modèle « nul » est un cas spécial du modèle alternatif. Par exemple, il ne pourrait pas être utilisé afin de comparer la représentation de la variabilité intra-annuelle par bloc mensuel à la représentation par série de Fourier.

Le but du présent mémoire n'étant pas de comparer ces différents tests et critère, seul le critère AIC a été retenu pour la suite des choses. L'attention est davantage portée sur le cadre méthodologique et les résultats généraux. Il ne serait cependant pas difficile de substituer un critère au profit d'un autre et cela constituerait certainement un exercice intéressant qui permettrait de constater l'impact sur les résultats.

5.6 Modèles locaux et régionaux

5.6.1 Proposition de modèles

En considérant conjointement la description en séries de Fourier, la représentation en blocs mensuels et le caractère local ou régional de chacun des paramètres, on est confronté à un nombre important de combinaisons de modèles possibles. La Figure 5.6 permet de visualiser les combinaisons possibles pour chacun des paramètres de la GPD. De plus la description en séries de Fourier exige de préciser le nombre de modes à considérer pour décrire la variabilité intra-annuelle des paramètres. Rappelons que le seuil (paramètre de position) est quant à lui toujours considéré comme local.

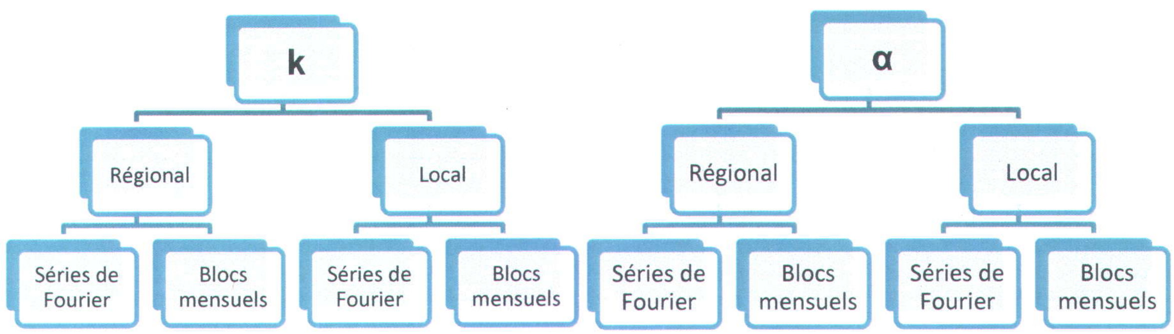

Figure 5.6 : Arborescence des modèles possibles.

Dans le cas de l'approche par séries de Fourier, les modes $m = 0,1,2$ seront considérés. Le choix d'examiner un maximum de deux modes s'appui sur la présomption que ce nombre est suffisant pour représenter la variabilité intra-annuelle des extrêmes pour les mois de mai à octobre. Ce choix permet également d'éviter d'avoir à optimiser des fonctions inutilement complexes. Les combinaisons alliant une représentation d'un paramètre par série de Fourier et de l'autre par blocs mensuels ne sont pas considérées.

Les Tableaux 5.3, 5.4 et 5.5 résument la liste des modèles retenus à partir des différentes combinaisons possibles. Les noms de modèles sont composés de quatre caractères. Le premier réfère au nombre de modes pour le paramètre d'échelle (α); le deuxième indique si le paramètre d'échelle est local (L) ou régional (R); le troisième caractère réfère au nombre de modes pour le paramètre de forme (k); et le quatrième indique si ce paramètre de forme est local (L) ou régional (R). Pour les modèles qui utilisent une représentation de la variation intra-annuelle à l'aide de blocs mensuels stationnaires au lieu des séries de Fourier, le caractère qui identifie le nombre de modes dans la nomenclature est remplacé par la lettre «b». La Figure 5.7 donne un exemple de la nomenclature utilisée.

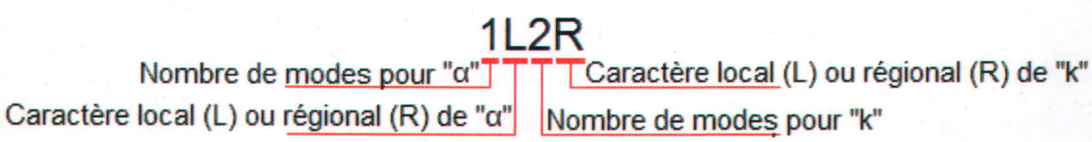

Figure 5.7 : Exemple pour le modèle 1L2R de la nomenclature utilisée.

Le chiffre entre parenthèses à côté du nom des modèles correspond à titre indicatif au nombre de paramètres associés dans le cas où l'on considère 107 stations (durées 5, 10, 15, 30 minutes). Pour 109 stations, ce nombre sera légèrement plus élevé s'il comporte au moins un paramètre local.

La minimisation d'une fonction dans un espace multidimensionnel n'est certes pas chose facile. L'algorithme d'optimisation utilisé, la fonction *fminsearch* de Matlab ® basée sur la méthode de Nelder-Mead (Lagarias *et al.*, 1998), a cependant donné des résultats cohérents. Pour les modèles les plus simples, la convergence de l'optimisation est relativement rapide. La convergence a été vérifiée en prenant différentes valeurs de départ pour l'optimisation. Les modèles les plus simples sont d'abord optimisés. Les paramètres obtenus sont par la suite utilisés comme point de départ pour l'optimisation de modèles plus complexes. Par exemple, le résultat de l'optimisation du modèle 0R0R servira de point de départ à l'optimisation 1R0R avec une valeur nulle pour les nouveaux paramètres. Le résultat de l'optimisation de 1R0R servira ensuite de point de départ à l'optimisation de 2R0R, etc. Parfois, deux jeux de paramètres peuvent servir de point de départ. Par exemple, pour optimiser le modèle 1R1R, il est possible d'utiliser le résultat de l'optimisation du modèle 1R0R ou encore du modèle 0R1R. Dans de tels cas, l'optimisation est effectuée à partir des deux points et; advenant qu'il existe une différence entre les résultats, l'optimisation avec la plus haute vraisemblance est retenue. Cependant, force est d'admettre que cela ne signifie pas pour autant que les valeurs trouvées sont des minima globaux de la fonction de vraisemblance. Pour en être certain, il faudrait vérifier la valeur de la vraisemblance pour tout l'espace paramétrique et s'assurer que le résultat de l'optimisation corresponde bel et bien au minimum global. Cet exercice est long car il nécessite un nombre d'itérations très élevé. Étant donné qu'une seule optimisation des 40 modèles nécessite approximativement 5 jours de temps de calcul avec un processeur de 2.4 GHz, l'exercice n'a pas été réalisé. Les résultats obtenus sont toutefois cohérents d'un modèle à l'autre.

Tableau 5.3 : Nomenclature des modèles présentant un paramètre de forme régional. Les chiffres entre parenthèses indiquent le nombre de paramètres des modèles lorsque 107 stations sont considérées.

			Forme (k) régional		
		Mode(s)	0	1	2
Échelle (α)	**Régional**	0	0R0R (2)	0R1R (4)	0R2R (6)
		1	1R0R (4)	1R1R (6)	1R2R (8)
		2	2R0R (6)	2R1R (8)	2R2R (10)
	Local	0	0L0R (108)	0L1R (110)	0L2R (112)
		1	1L0R (322)	1L1R (324)	1L2R (326)
		2	2L0R (536)	2L1R (538)	2L2R (540)

Tableau 5.4 : Nomenclature des modèles présentant un paramètre de forme local. Les chiffres entre parenthèse indiquent le nombre de paramètres des modèles lorsque 107 stations sont considérées.

			Forme (k) local		
		Mode(s)	0	1	2
Échelle (α)	Régional	0	0R0L (108)	0R1L (322)	0R2L (536)
		1	1R0L (110)	1R1L (324)	1R2L (538)
		2	2R0L (112)	2R1L (326)	2R2L (540)
	Local	0	0L0L (214)	0L1L (428)	0L2L (642)
		1	1L0L (428)	1L1L (642)	1L2L (856)
		2	2L0L (642)	2L1L (856)	2L2L (1070)

Tableau 5.5 : Nomenclature des modèles construits en blocs mensuels. Les chiffres entre parenthèses indiquent le nombre de paramètres des modèles lorsque 107 stations sont considérées.

		Forme (k)	
		Régional	Local
Échelle (α)	Régional	bRbR (12)	bRbL (648)
	Local	bLbR (648)	bLbL (1284)

5.6.2 Classement AIC_c des modèles

Plusieurs questions sont *a priori* intéressantes dans la comparaison de ces modèles : Quelle modèle permet de mieux reproduire la variabilité intra-annuelle des paramètres ? La méthode paramétrique par séries de Fourier ou la méthode non paramétrique par blocs mensuels ? Dans le cas où la méthode paramétrique est retenue, existe-t-il un nombre de modes optimal ? Et finalement, de quelle manière vaut-il mieux considérer les paramètres d'échelle et de forme, localement ou régionalement ? Pour tenter de répondre à ces questions, le score AIC_c a été calculé pour chacun des modèles, ensuite les modèles ont été classés en fonction de leur Δ_i respectif. La Figure 5.8 présente, pour chacune des durées, les cinq meilleurs modèles du point de vue AIC_c sur les 40 modèles proposés.

La Figure 5.8 montre que : 1) tous les modèles retenus intègrent une variabilité intra-annuelle des paramètres; 2) le modèle régional (paramètre de forme et d'échelle régionaux) est supérieur aux autres modèles. À seulement deux reprises un modèle autre que le modèle régional se retrouve dans le palmarès des cinq meilleurs modèles. Pour les six durées les plus courtes, de 5 minutes à 2 heures, le modèle 2R2R est identifié comme le meilleur modèle. Pour les durées de 6 heures et 12 heures ce sont respectivement les modèles 1R2R et 2R1R qui sont retenus au premier rang. On remarque cependant que pour ces deux durées, les cinq valeurs de Δ_i sont très proches les unes des autres, indice d'une vraisemblance similaire des modèles. Il faut attendre le 7^e rang (non montré sur la figure) avant d'observer une différence substantielle dans les valeurs de Δ_i. En ce qui a trait au premier rang de chaque durée, dans sept des huit cas, deux modes sont retenus pour les paramètres de forme et d'échelle.

La Figure 5.8 présente une synthèse des résultats obtenus en ne s'intéressant qu'aux cinq meilleurs modèles pour chacune des durées. Pour avoir un meilleur aperçu de l'allure générale des résultats, on peut se référer à la Figure 5.9 qui présente le classement des 40 modèles pour la durée 30 minutes. Il est intéressant de remarquer la position du modèle 0L0L au classement général puisqu'il s'agit du modèle « naïf » utilisé si on choisit d'ignorer la variabilité intra-annuelle des paramètres et si tous les paramètres sont considérés locaux. Pour la durée 30 minutes, il se classe au 37^e rang sur 40 modèles et pour toutes les durées considérées, il se retrouve au mieux au 30^e rang sur 40 modèles. Il semble également que les modèles qui ne considèrent aucun mode sur le paramètre d'échelle ($m_\alpha = 0$) performent très mal. Sur 12 modèles possibles, 11 se retrouvent en queue de peloton. Le nombre de modes sur le paramètre de forme semble quant à lui avoir une incidence beaucoup moins importante sur le classement. Finalement, les Figures 5.8 et 5.9 montrent que la prise en compte de la variabilité intra-annuelle à l'aide de séries de Fourier semble plus intéressante que la représentation par blocs mensuels (modèles bRbR, bLbR, bRbL et bLbL).

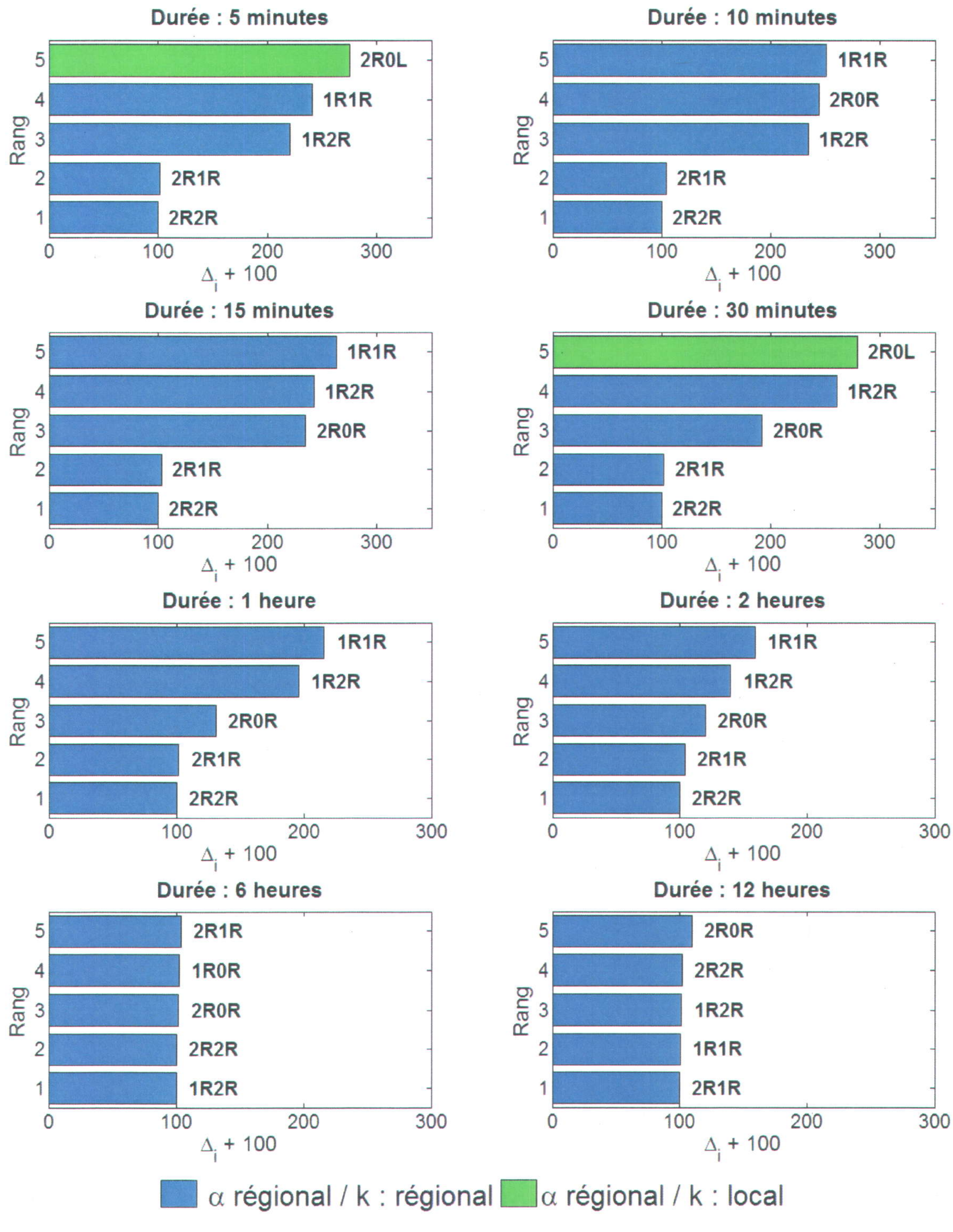

Figure 5.8 : Classement des cinq meilleurs modèles sur les 40 modèles proposés. Le nom du modèle apparaît à droite de chaque barre.

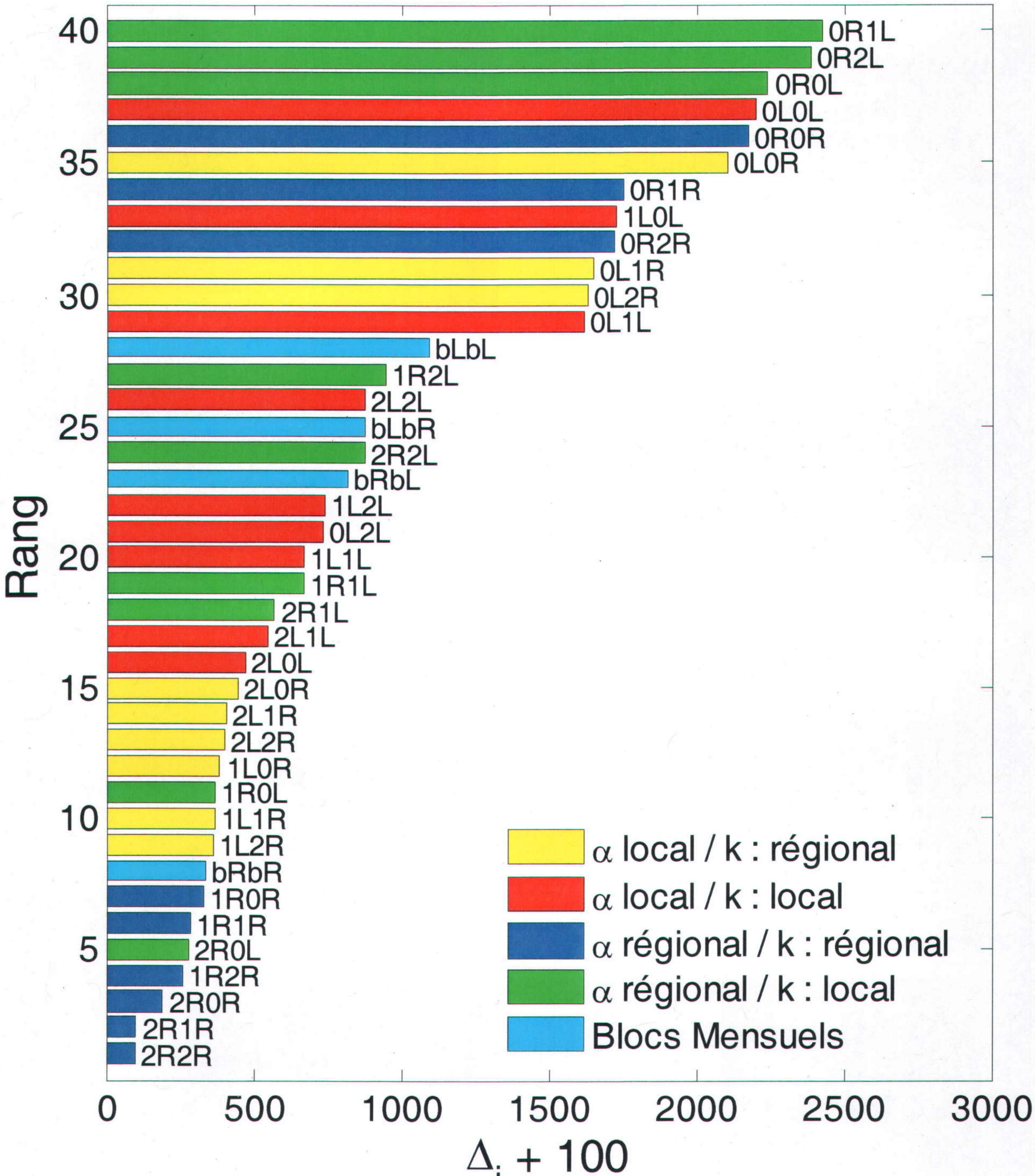

Figure 5.9 : Classement AIC$_c$ des 40 modèles proposées pour une durée 30 minutes. Le nom du modèle apparaît à droite de chaque barre.

5.7 Modèles régionaux avec covariables spatiales

Les résultats obtenus sont, dans une certaine mesure, étonnants puisque par rapport au territoire couvert, les modèles régionaux laissent peu de place au caractère local étant donné que seul le seuil diffère pour chacune des stations. Cela sous-entend que les cumuls de précipitations extrêmes sont à peu près les mêmes pour l'ensemble du territoire. L'homogénéité apparente de la région peut cependant masquer une structure spatiale des paramètres qu'il n'est possible de révéler qu'avec l'ajout d'information. Dans cette optique, de nouveaux modèles dérivés à partir des modèles régionaux ont été proposés. L'idée est d'inclure une information supplémentaire susceptible d'améliorer l'estimation des paramètres en tenant compte d'une possible dépendance spatiale. La latitude (L), la longitude (G) et l'élévation (E) sont les covariables (variables explicatives) considérées. Plutôt que d'être homogène sur l'ensemble du territoire, les paramètres de forme et d'échelle de la GPD sont considérés comme des fonctions des covariables spatiales. Les modèles obtenus représentent une complexité intermédiaire entre les modèles régional/régional et régional/local.

5.7.1 Modélisation de la dépendance spatiale des paramètres

Pour chacune des stations, on dispose de la latitude, de la longitude ainsi que de l'élévation. Ces covariables ont été introduites au niveau des paramètres des séries de Fourier à partir d'un modèle de covariance linéaire. Pour l'équation 5.1 du paramètre de forme (k), par exemple, les paramètres k_0, a_j et b_j de la série de Fourier sont maintenant fonction des covariables spatiales. On obtient :

$$k_0 = k_{0,0} + k_{0,L}L + k_{0,G}G + k_{0,E}E \tag{5.8}$$

$$a_j = a_{j,0} + a_{j,L}L + a_{j,G}G + a_{j,E}E \tag{5.9}$$

$$b_j = b_{j,0} + b_{j,L}L + b_{j,G}G + b_{j,E}E \tag{5.10}$$

avec :

k_0, a_j et b_j pour $j = 1 \ldots, m_k$ les paramètres de la série de Fourier à m_k modes du paramètre de forme.

$[k_{0,0}\ k_{0,L}\ k_{0,G}\ k_{0,E}]$ le vecteur de paramètres du modèle linéaire de k_0 pour chaque covariable.

$[a_{j,0}\ a_{j,L}\ a_{j,G}\ a_{j,E}]$ le vecteur de paramètres du modèle linéaire de a_j pour chaque covariable.

$[b_{j,0}\ b_{j,L}\ b_{j,G}\ b_{j,E}]$ le vecteur de paramètres du modèle linéaire de b_j pour chaque covariable.

L la latitude, G la longitude et E l'élévation

Le paramètre d'échelle (α) est modélisé de la même manière. L'approche résultant permet une variation linéaire de la moyenne, de l'amplitude et de la phase des fonctions sinusoïdales qui représentent la variabilité intra-annuelle des paramètres. Le modèle linéaire est retenu pour sa simplicité. Cependant, différents modèles de dépendance pourraient être utilisés. La dépendance aux covariables est appliquée à tous les modes de tous les paramètres de la distribution.

5.7.2 Proposition de modèles régionaux avec covariables spatiales

Afin de réduire le nombre de modèles à analyser, on ne considèrera que les modèles les plus pertinents. La Figure 5.8 et la Figure 5.9 permettent à cet effet de constater que les approches qui ne présentent pas de variation intra-annuelle du paramètre d'échelle ne sont pas très intéressantes et se classent en queue de peloton. Les modèles 0R0R 0R1R et 0R2R ont pour cette raison été écartés. Les six modèles régionaux restants (1R0R, 1R1R, 1R2R, 2R0R, 2R1R et 2R2R) ont été divisés en cinq nouvelles classes. Chacune de ces classes correspond à un usage particulier des covariables spatiales (Tableau 5.6).

Tableau 5.6 : Classe des modèles qui présentent au moins une covariable spatiale.

Classe	Covariable(s)
1	Latitude (L)
2	Longitude (G)
3	Élévation (E)
4	Latitude + Longitude (LG)
5	Latitude + Longitude + Élévation (LGE)

La nomenclature employée pour les modèles avec covariables spatiales est détaillée à la Figure 5.10. Les deux premiers chiffres correspondent respectivement au nombre de modes sur les paramètres d'échelle et de forme. La séquence de une à trois lettres indique les covariables introduites dans le modèle.

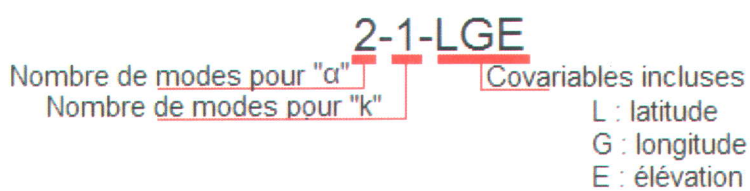

Figure 5.10 : Exemple de nomenclature pour le modèle 2-1-LGE.

Tableau 5.7 : Nomenclature des modèles régionaux avec covariables spatiales proposés. Les chiffres entre parenthèses indiquent le nombre de paramètres des modèles.

			Forme – k		
	Covariable(s)	Nombre de modes	0	1	2
Échelle – α	Aucune	1	1R0R (4)	1R1R (6)	1R2R (8)
		2	2R0R (6)	2R1R (8)	2R2R (9)
	Latitude (L)	1	1-0-L (8)	1-1-L (12)	1-2-L (16)
		2	2-0-L (12)	2-1-L (16)	2-2-L (20)
	Longitude (G)	1	1-0-G (8)	1-1-G (12)	1-2-G (16)
		2	2-0-G (12)	2-1-G (16)	2-2-G (20)
	Élévation (E)	1	1-0-E (8)	1-1-E (12)	1-2-E (16)
		2	2-0-E (12)	2-0-E (16)	2-2-E (20)
	Latitude + Longitude (LG)	1	1-0-LG (12)	1-1-LG (18)	1-2-LG (24)
		2	2-0-LG (18)	2-1-LG (24)	2-2-LG (30)
	Latitude + Longitude + Élévation (LGE)	1	1-0-LGE (16)	1-1-LGE (24)	1-2-LGE (32)
		2	2-0-LGE (24)	2-1-LGE (32)	2-2-LGE (40)

5.7.3 Classement AIC_c des modèles régionaux avec dépendance spatiale

Une fois de plus, à partir du test AIC_c, les performances relatives des différents modèles ont été comparées (Figure 5.11).

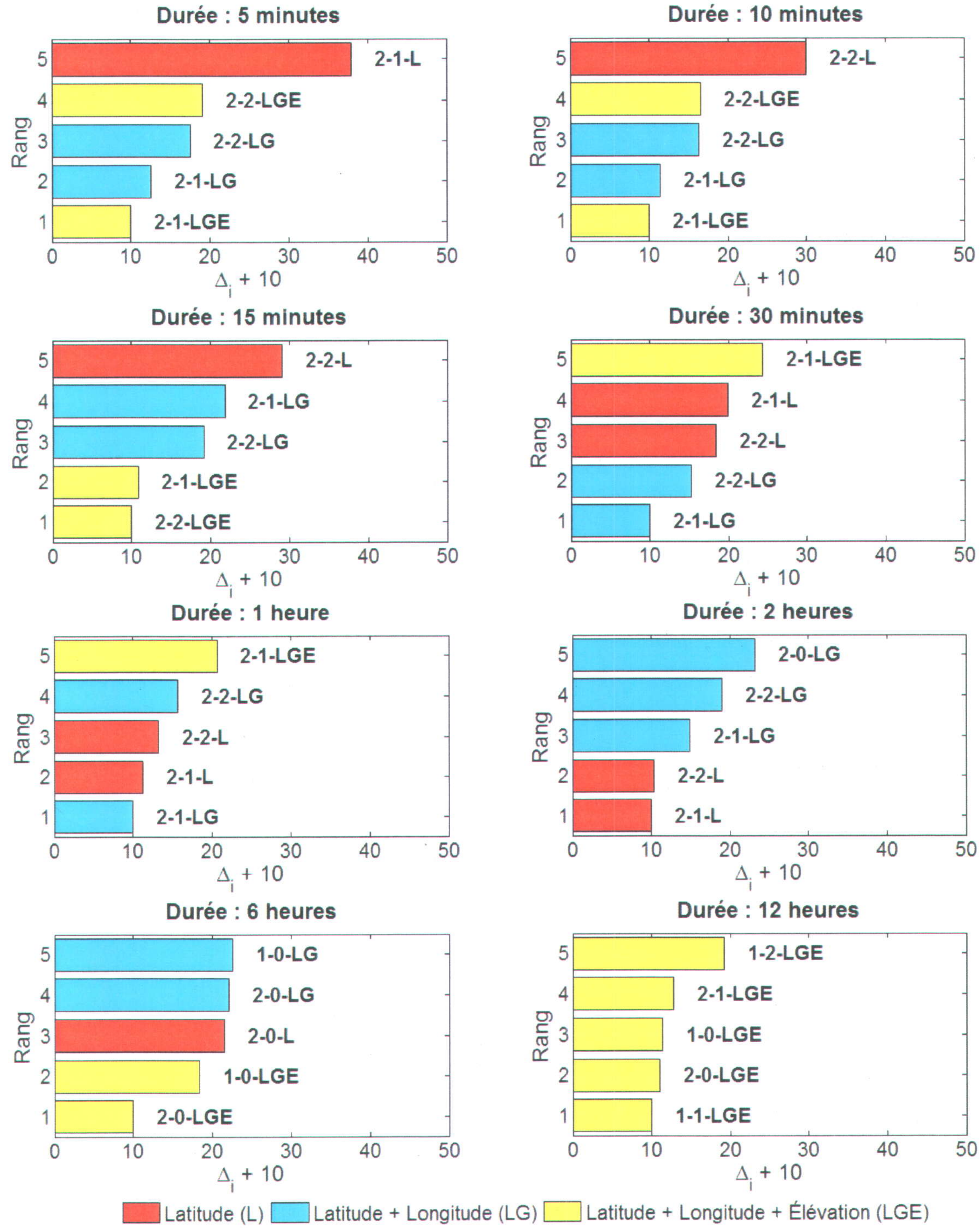

Figure 5.11 : Classements pour les huit durées des cinq meilleurs modèles sur les 36 modèles proposés.

Cette fois, on remarque que, dépendamment des durées, le classement des modèles varie beaucoup plus qu'à la Figure 5.9. Le critère conclut essentiellement que pour toutes les durées, il convient d'inclure une ou plusieurs covariables spatiales dans l'estimation des paramètres. Parmi les cinq meilleurs modèles retenus, et pour toutes les durées, tous considèrent au moins une covariable spatiale. Pour toutes les durées considérées, les meilleurs modèles sans covariables n'apparaissent jamais avant le 12^{e} rang (durée 30 minutes, Figure 5.12) et les valeurs de Δ_i pour ces modèles sont généralement de l'ordre de 50 à 150. Comme alternative au critère AIC_c, le test du rapport de vraisemblance a été appliqué (non-montré) entre le modèle au 1er rang et les 35 autres modèles considérés. Le modèle optimal retenu par ce test est le même que pour le critère AIC_c.

Les résultats de la Figure 5.11 tendent également à indiquer une préférence pour les modèles qui considèrent deux modes pour le paramètre d'échelle (α). Pour la durée 30 minutes, la Figure 5.12 montre très clairement ce résultat. Seule pour la durée 12 heures, le premier rang est occupé par un modèle qui ne considère qu'un seul mode sur ce paramètre. Pour les durées comprises entre 5 minutes et 2 heures, le nombre de modes sur le paramètre de forme est de un ou deux avec des écarts relativement faibles entre l'une ou l'autre des options. Pour la durée 6 heures cependant, les cinq meilleurs modèles ne considèrent aucun mode pour le paramètre de forme. Dans le cas de la durée 12 heures, les écarts sont très faibles entre les quatre premiers modèles qui considèrent soit zéro, soit un mode sur ce paramètre. Pour les durées plus longues, la présence de modes sur le paramètre de forme est moins évidente.

En ce qui a trait à la variable qu'il convient de considérer, il ne semble pas y avoir de réponse unique. En fait, on peut s'attendre, étant donné la nature du critère AIC, à ce que le modèle retenu change avec la série de données analysée. Le choix du modèle à partir du score AIC est spécifique à un échantillon et une taille donnée (Burnham et Anderson, 2004). Par exemple, en modifiant le taux de dépassements moyen servant à obtenir les SDP, on obtiendra vraisemblablement des résultats différents. Pour des échantillons de taille moindre, il n'est peut-être pas aussi pertinent d'inclure certaines covariables puisque l'on ne dispose plus d'assez de données pour estimer leur influence correctement. Cependant, bien que le « meilleur » modèle puisse changer en fonction de la taille de l'échantillon, on s'attend à ce que le résultat final, soit l'estimation d'un quantile donné, n'y soit pas très sensible. Pour obtenir un portrait plus général du classement obtenu, on trace à la Figure 5.12 le classement de tous les modèles proposés pour la durée 30 minutes. Ce résultat est représentatif du comportement des durées comprises entre 5 minutes et 6 heures.

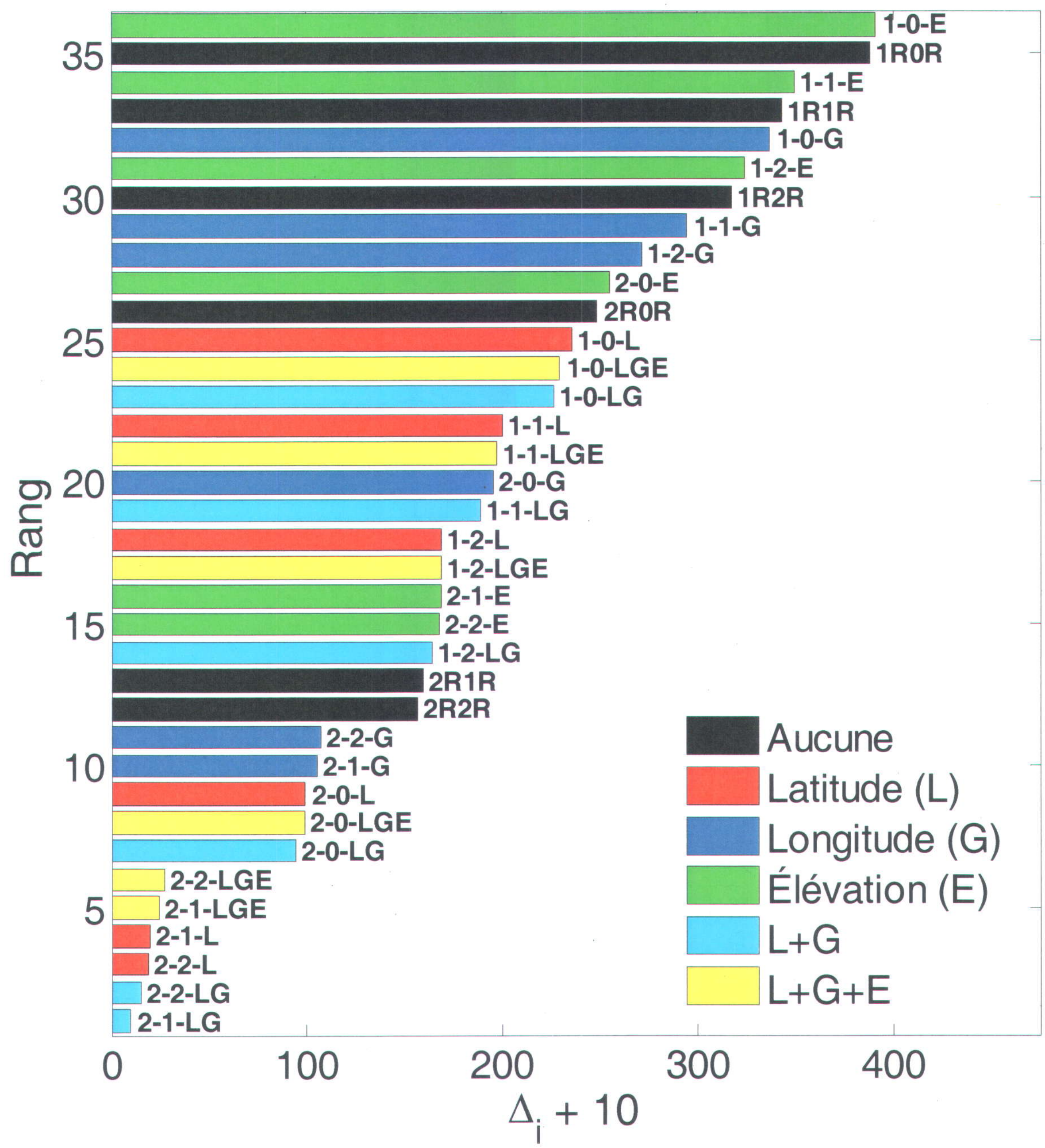

Figure 5.12 : Classement AIC_c des 40 modèles proposés pour la durée 30 minutes.

En effectuant les correspondances appropriées, on remarque que le modèle 2R2R au 12^{e} rang à la Figure 5.12 se retrouve au premier rang à la Figure 5.9. Ce dernier était alors, aux yeux du critère, le meilleur modèle parmi les 40 initialement proposés. La Figure 5.12 montre clairement que l'ajout de covariables spatiales permet d'améliorer substantiellement la qualité des modèles. Dans le cas de la durée 30 minutes, on observe un gain très important de plus de 100 points. La bonne performance des modèles qui considèrent deux modes sur le paramètre d'échelle est nettement mise en évidence par la Figure 5.12.

Cette figure montre également que la variable qui apporte le plus d'information est la latitude. Il ressort également qu'il est intéressant de considérer la longitude. Le rôle de l'élévation est moins évident et, dans le cas de la durée 30 minutes, cette covariable ne semble pas apporter d'information valable. La durée 12 heures présente quant à elle un comportement différent. Pour cette durée, l'élévation devance la latitude et la longitude en termes de score AIC_c. Ce résultat se détache des précédents, mais pourrait s'expliquer par la nature différente des événements générateur d'extrêmes sur de courtes et de longues durées. Il serait intéressant d'observer le rôle des différentes covariables pour d'autres durées, par exemple 24 et 48 heures, afin de constater si des résultats similaires sont obtenus.

Les modèles proposés jusqu'à maintenant pourraient être raffinés davantage. Dans notre cas une dépendance linéaire est considérée sur chacun des paramètres de la série de Fourier (moyenne, amplitude et phase) pour les deux paramètres de la GPD et pour chacun des modes. Or, il serait possible de diminuer la complexité des modèles en n'incluant par exemple une dépendance que sur la moyenne des séries de Fourier ou encore en n'introduisant une dépendance que sur un seul des paramètres de la GPD. D'autres modèles de dépendance pourraient également être testés pour les différentes covariables spatiales.

5.7.4 Variation spatiale et intra-annuelle des paramètres

La Figure 5.13 montre la variation intra-annuelle des paramètres en différents endroits du territoire pour le modèle 2-1-LG (meilleur modèle selon le critère considéré) pour la durée 30 minutes. Ce modèle identifie une structure au niveau des paramètres en fonction de la latitude et de la longitude. Trois courbes situées le long de l'axe sud-ouest / nord-est sont présentées, ce qui permet de visualiser les valeurs des paramètres aux extrémités du territoire considéré ainsi qu'en son centre. On remarque que l'écart entre les paramètres n'est pas constant et qu'il atteint un maximum dans les environs du 210^{e} jour julien, à l'instar du seuil (voir Figure 5.2), soit la deuxième moitié de juillet. La Figure 5.13 montre que la variabilité intra-annuelle des paramètres est plus importante au sud-est qu'au nord-ouest et qu'elle semble progresser selon cet axe.

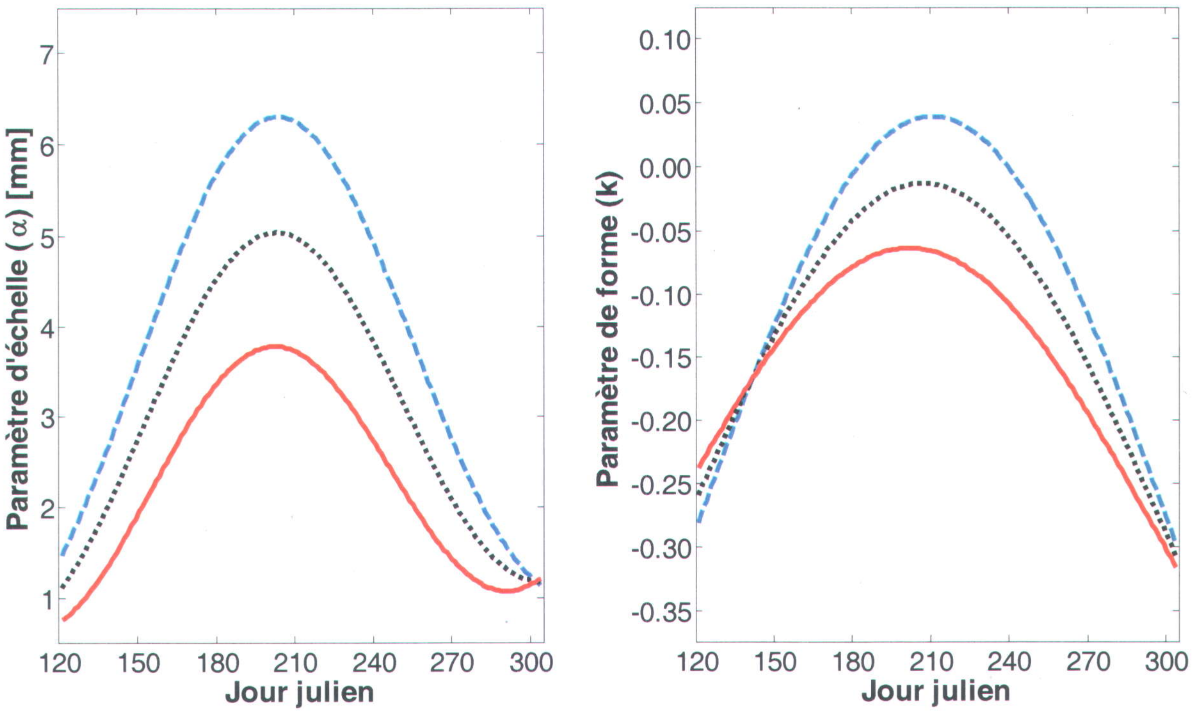

Figure 5.13 : Valeur du paramètre d'échelle (à gauche) et du paramètre de forme (à droite) en fonction du jour julien en différents points du territoire. Le trait bleu en tirets correspond au coin sud-ouest (45°N 77°W) près de Gatineau, le trait pointillé noir au centre du territoire (47°N 71°W) près de Québec et le trait continu rouge au coin nord-est (49°N 65°W) près de Gaspé. Les résultats présentés sont ceux du modèle 2-1-LG pour la durée 30 minutes.

5.7.5 Estimations des cumuls de précipitations extrêmes

À la suite du classement AIC_c, on obtient, pour une série d'observations, quel modèle il convient le mieux d'utiliser. Bien que les modèles permettent de déterminer la valeur des paramètres d'échelle et de forme en tout point du territoire, il demeure que le seuil est déterminé localement et est spécifique à chaque station. Il faut donc recourir au krigeage afin d'interpoler l'information disponible aux stations sur l'ensemble du territoire. Plutôt que d'interpoler le seuil, il a été choisi d'interpoler directement le champ physique observable, à savoir les cumuls de précipitation calculés aux stations. La procédure d'interpolation employée est le krigeage ordinaire sur résidu (Zhang et Srinivasan, 2009; Goovaerts, 1997).

Pour obtenir la valeur x du cumul de précipitation dépassé en moyenne une fois chaque T ans, on résout numériquement l'équation suivante :

$$\frac{1}{T}=\sum_{j=j_1}^{j_2}\frac{\lambda}{N}\left\{1-F\left(x\mid k_j,\alpha_j,\xi_j\right)\right\} \tag{5.11}$$

Ici, λ est égal à 6 et correspond au nombre moyen de dépassements par année MAO et N = 184 correspond au nombre de jours de cette période. j_1 est le premier jour de l'année MAO (1er mai) et j_2 le dernier (31 octobre). $F\left(x\mid k_j,\alpha_j,\xi_j\right)$ est la fonction de répartition de la GPD (équation 2.2).

Une fois les valeurs calculées à chacune des stations, elles sont interpolées spatialement sur l'ensemble du territoire. Les résultats obtenus pour la durée 30 minutes et T = 10 ans sont présentés à la Figure 5.14. Le modèle 2-1-LG a été retenu pour cette durée, tel qu'indiqué aux Figures 5.11 et 5.12. Ce modèle retient la latitude et la longitude comme covariables spatiales.

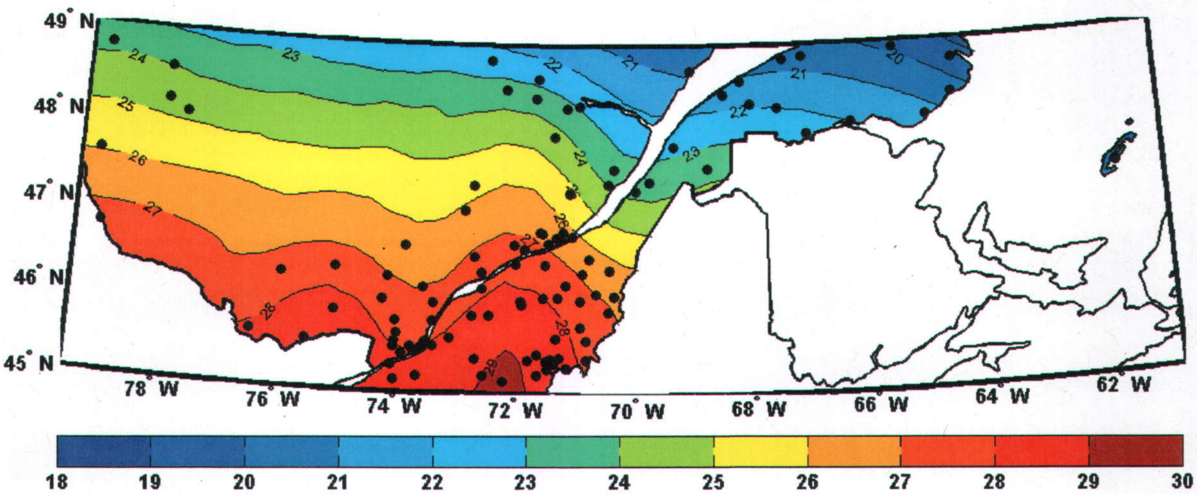

Figure 5.14 : Cumul de précipitation sur 30 minutes (mm) de temps de retour 10 ans selon le modèle 2-1-LG.

Les résultats de la Figure 5.14 mettent en lumière un gradient nord-sud assez évident. La valeur varie entre ~19 mm pour la pointe nord de la Gaspésie et ~28 mm à l'extrême sud de la province. Pour un regroupement des stations situées sous le 46°N, une estimation empirique du cumul de précipitations de temps de retour 10 ans conduit à une valeur de 27.7 mm. Les résultats présentés à la Figure 5.14 sont en cohérence avec cette valeur.

L'annexe F présente les cartes obtenues pour les autres durées considérées. Généralement, pour les durées de six heures et moins, on observe un gradient et un patron similaires à ceux de la Figure 5.14. Pour la durée12 heures, le comportement est légèrement différent, le gradient nord-sud est moins marqué et il semble y avoir une certaine influence de l'élévation.

5.7.6 Sensibilité des résultats au modèle retenu

Tel que discuté précédemment, il existe une incertitude associée au choix d'un modèle. Les poids AIC permettent de considérer cette dernière en pondérant chacun des modèles en fonction de leur performance respective. Pour illustrer la sensibilité des résultats au choix du modèle, on peut aisément comparer les résultats issus des différents modèles. Par contre, plutôt que de comparer des cartes pour lesquelles les principales différences sont attribuables à l'étape de krigeage, il est préférable de comparer les valeurs calculées directement aux stations.

Dans l'exemple utilisé, le modèle optimal 2-1-LG termine en tête de liste avec un poids de 91 % alors que les modèles 2-2-LG et 2-2-L suivent avec des poids respectifs de 7 % et 1 %. Le 1 % restant est réparti parmi les autres modèles. On s'attend à ce que les résultats obtenus à partir des modèles dont les poids sont non négligeables soient relativement similaires. Afin de vérifier que c'est bien le cas, la Figure 5.15 présente l'écart relatif des valeurs prédites aux stations entre les modèles 2-2-LG et 2-2-L et le modèle 2-1-LG pour un cumul de précipitation sur 30 minutes de temps de retour 10 ans. L'axe des abscisses présente la longitude des stations. Cela permet de mieux visualiser l'impact de considérer des modèles avec la latitude et la longitude (2-1-LG et 2-2-LG) plutôt qu'un modèle basé uniquement sur la latitude (2-2-L),

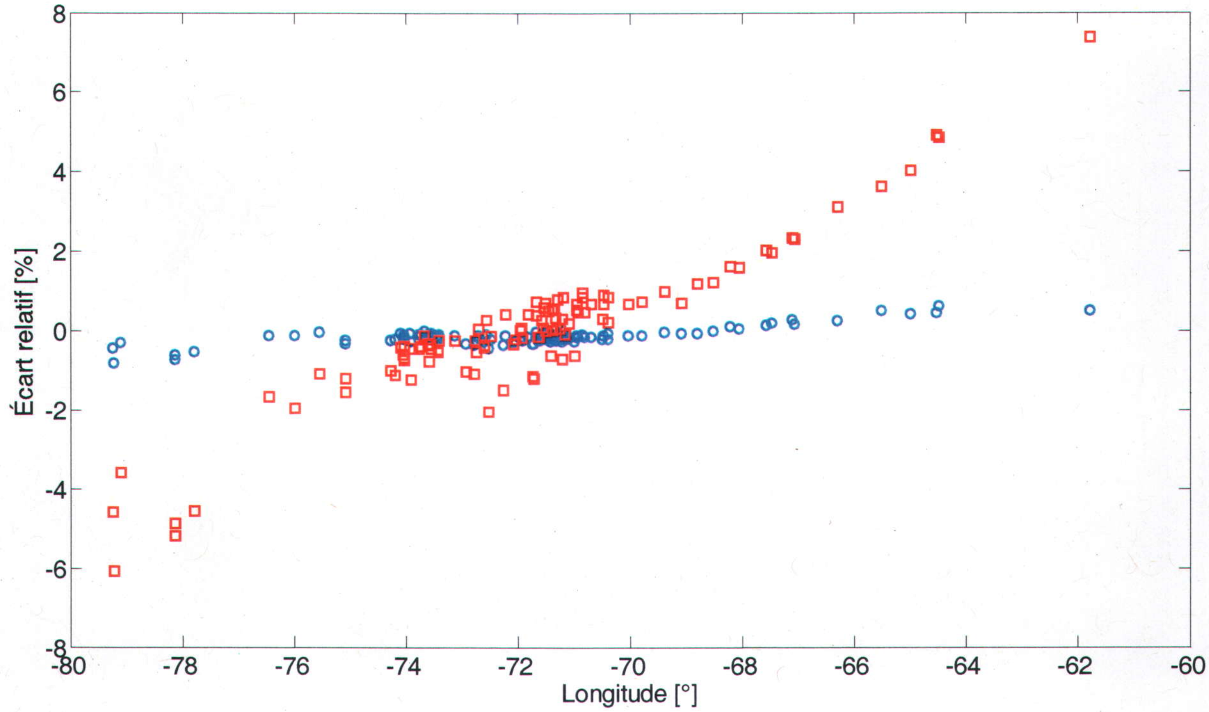

Figure 5.15 : Écart relatif, par rapport au modèle 2-1-LG, des valeurs de cumul de précipitation sur 30 minutes de temps de retour 10 ans prédites par le modèle 2-2-LG (cercles bleus) et le modèle 2-2-L (carrés rouges) pour chacune des 107 stations.

Les résultats montrent que les modèles 2-1-LG et 2-2-LG donnent des résultats similaires puisque l'écart relatif pour l'ensemble des stations est inférieur à 1%. Ces deux modèles considèrent la latitude et la longitude et ne diffèrent que dans le nombre de modes qui décrivent le paramètre de forme. Pour le modèle 2-2-L, qui ne considère pas les mêmes covariables (latitude seulement), l'écart relatif maximal est plus important et atteint 7.4%; il faut toutefois souligner que le poids de ce modèle est faible (1 %). On remarque de plus que l'absence de la longitude comme covariable se traduit par des écarts négatifs pour les stations à l'ouest et des écarts positifs pour les stations à l'est. Les modèles 2-1-LG et 2-2-LG permettent de tenir compte d'un certain gradient selon la longitude, ce qui n'est pas le cas du modèle 2-2-L.

5.7.7 Sensibilité au taux annuel moyen retenu pour définir le seuil

On peut encore une fois s'interroger de l'impact sur les résultats du taux annuel moyen de dépassement servant à obtenir les SDP. La question est d'autant plus pertinente, qu'étant donné la nature du critère AIC, la taille de l'échantillon possède une incidence certaine sur le choix et la pondération des différents modèles. Afin de vérifier cette sensibilité, trois autres taux ont été également considérés, soit 1, 1.5 et 3 dépassements en moyenne pour l'année MAO. L'analyse décrite jusqu'à présent a été reproduite pour chacun des taux et chacune des durées. Pour des raisons de concision, la Figure 5.16 présente les résultats pour la durée 30 minutes uniquement. Ces résultats illustrent cependant le comportement général de l'effet de la diminution du taux sur les modèles sélectionnés pour l'ensemble des durées.

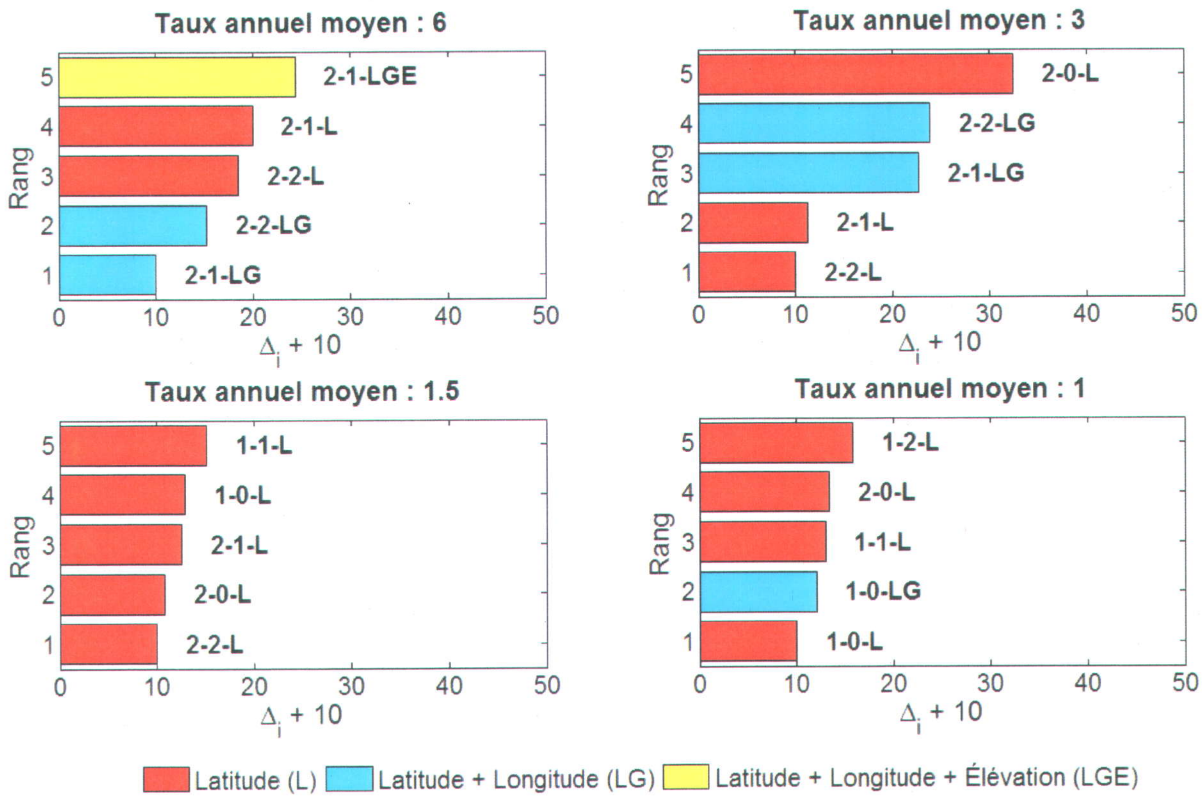

Figure 5.16 : Classement des cinq meilleurs modèles pour la durée 30 minutes en fonction des différents taux annuels moyens de dépassement.

Un des principaux constats de la Figure 5.16 est que, de manière générale, il semble que la complexité des modèles retenus décroisse avec une diminution du taux annuel moyen. Cette tendance est observable pour chacune des durées. Plus le taux de dépassement choisi afin de construire les SDP est bas, moins les modèles tendent à être complexes. Cette baisse de complexité se traduit notamment par une diminution du nombre de modes qui servent à décrire la variabilité intra-annuelle des paramètres. On observe également que de façon générale, les modèles retenus pour les taux plus faibles considèrent moins de covariables spatiales, généralement une seule, soit celle qui fournit le plus d'information. Cependant, pour le plus faible taux considéré et pour toutes les durées, les modèles avec covariables spatiales ont tout de même été jugés préférables aux modèles sans covariable. D'une part, cette observation s'explique par le fait que les SDP comportent moins de données lorsque le seuil augmente (ou de façon équivalente lorsque le taux de dépassement diminue). Il est alors plus difficile de justifier l'ajustement de modèles plus complexes car on dispose de moins d'information. Par exemple, la série devient trop courte pour permettre de déceler adéquatement la structure spatiale ou temporelle d'un paramètre. À complexité égale, le rapport entre le nombre de paramètres à estimer et le nombre d'observations augmente. D'autre part, la nature des données retenues peut également jouer un rôle. Avec des taux plus élevés, on retiendra davantage de valeurs. Conséquemment, des valeurs un peu plus fréquentes seront considérées. Il est possible que pour ces quantiles moins élevés, les variations intra-annuelles soient plus importantes, donc plus facilement détectables, ou encore que le rôle de certaines covariables soit plus évident.

Au niveau des estimations, on s'attend à ce que la diminution du taux annuel moyen qui sert à construire les SDP se traduise au niveau des résultats par une augmentation de la variabilité. Toutefois, on suppose également que cela n'introduit pas de biais. Afin d'obtenir un portrait global de la situation, la Figure 5.17 présente l'écart relatif entre les valeurs du cumul de précipitation sur 30 minutes de temps de retour 10 ans des modèles retenus pour les taux de dépassement annuel moyen de 1, 1.5 et 3 et le modèle optimal pour le taux 6.

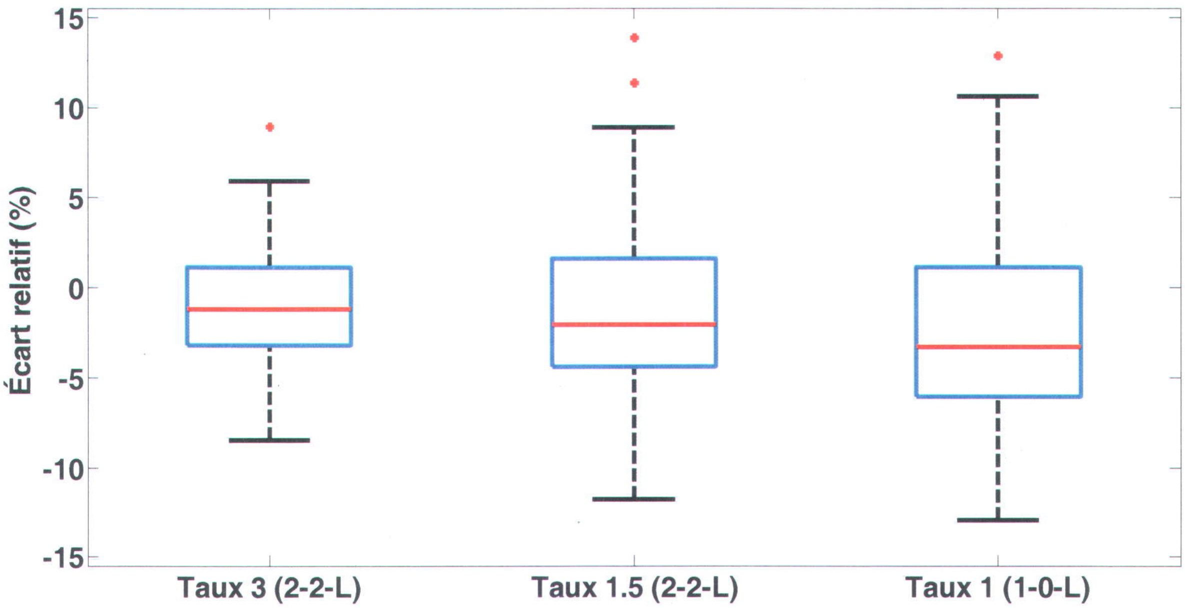

Figure 5.17 : Écarts relatifs (%) des prédictions entre le taux considéré et le taux de 6 dépassements pour chaque station du cumul de précipitation sur 30 minutes de temps de retour 10 ans. Le libellé en abscisse indique le taux considéré et le modèle retenu est indiqué entre parenthèses.

La Figure 5.17 montre que la diminution du taux annuel moyen conduit à une hausse de la variabilité. On obtient respectivement des écarts-types de 3.2, 4.6 et 5.2 % en comparant les résultats des taux 1, 1.5 et 3 avec ceux du taux 6. En ce qui a trait aux écarts moyens, leur valeur absolue augmente légèrement avec une diminution du taux. On obtient respectivement des valeurs de -0.9, -1.3 et -2.4 %. Pour de telles valeurs, l'hypothèse de moyenne nulle ($\mu = 0$) est rejetée pour un seuil de signification de $\alpha = 5\%$. Il semble donc exister un léger écart entre les résultats des modèles sélectionnés pour différents taux. Cependant, il n'est pas possible de départager si cette différence est induite par le taux utilisé ou par la complexité du ou des modèles retenus qui varie parallèlement au taux. Il faut également garder à l'esprit qu'ultimement, on recherche un certain compromis entre la variance et le biais d'un modèle. Il faudrait également comparer les valeurs pour différents temps de retour. En effet il est possible qu'un modèle/taux soit meilleur pour estimer certaines récurrences.

6. CONCLUSION

Le principal objectif de ce travail consistait à développer une méthodologie qui permette d'utiliser l'approche par série de durées partielles en régime non stationnaire. La méthodologie proposée permet de considérer le caractère non homogène des séries attribuables à la fois aux changements climatiques et aux cycles saisonniers. La sélection d'un seuil adéquat est toutefois primordiale pour la mise en œuvre de l'approche. Un taux annuel moyen de dépassement a été fixé afin de construire les SDP. Cette approche a été retenue pour deux raisons principales. La première tient à sa simplicité. En effet, si certaines méthodes proposées permettent l'identification d'un seuil, toutes reposent sur un certain degré de subjectivité, rendant par conséquent la procédure difficile à automatiser. La seconde raison s'appuie sur des considérations physiques. L'utilisation d'un taux fixe permet de s'intéresser *grosso modo* à la même « fraction », aux mêmes quantiles de la distribution. On obtient une information quant aux changements dans la distribution de ces mêmes quantiles. Une valeur dite « extrême » correspond alors au dépassement d'un certain quantile de la distribution.

Dans un premier temps, l'approche a été appliquée à l'analyse des extrêmes de précipitations simulés par le MCCG3. Le modèle simule notamment les CJP pour la période 1850-2100. Le maillage du modèle est de 3.75° en latitude et en longitude, ce qui correspond approximativement à des tuiles de 390 x 260 km à une latitude de 47°. Le territoire couvert est composé de 420 tuiles qui représentent l'ensemble du Canada et la partie nord des États-Unis. Cinq membres issus de simulations considérant des conditions initiales différentes sont disponibles. Ces cinq membres ont été combinés en supposant que l'hypothèse d'ergodicité (voir section 4.1) était respectée. Les simulations considérées emploient le scénario d'émission de gaz à effet de serre et aérosols SRES A2. Un taux moyen de 12 dépassements par année a été retenu afin de fixer le seuil. En plus d'être stabilisé inter-annuellement, le taux de dépassement est constant intra-annuellement. La probabilité de générer une valeur de CJP au-dessus du seuil est donc la même pour tous les jours de l'année.

La variation intra-annuelle des extrêmes a été examinée afin de vérifier si les changements climatiques auront une influence particulière sur le patron intra-annuel d'occurrence des extrêmes. La prise en compte de ce type de variabilité se justifie également par le fait que les paramètres de la GPD varient sensiblement d'un mois à l'autre. Une approche non paramétrique, qui consiste à subdiviser en périodes pseudo-stationnaires la plage 1850-2100, a été appliquée afin de prendre en compte la non-stationnarité. La méthode des moments a été utilisée pour inférer les paramètres sur chacune de ces périodes. Les valeurs de paramètres obtenues ont ensuite été lissées afin de décrire la variation des paramètres de manière continue.

La période 1850-1950, considérée stationnaire (voir Tableau 4.1), a été utilisée comme période de référence afin d'évaluer la variation relative des CJP extrêmes en climat futur. La valeur du CJP dépassé en moyenne une fois chaque 20 ans sur la période 1850-1950 a été calculée de même que celles pour les périodes 2020-2040 et 2060-2080. Les variations relatives par rapport à la période de référence varient, selon les tuiles du modèle, entre 0 et 20 % pour la période 2020-2040 et entre 10 et 40 % pour la période 2060-2080. Les augmentations relatives sont plus importantes pour la région arctique (66.8 - 81.6 °N). Il faut cependant garder à l'esprit que les valeurs de CJP extrêmes dans ces régions sont, en termes absolus, moins importantes que pour les régions plus au sud.

Les valeurs du CJP dépassées en moyenne une fois tous les 20 ans au cours des différentes saisons de la période 2060-2080 et de la période 1850-1950 ont été comparées. L'impact des changements climatiques sur la répartition intra-annuelle des extrêmes peut ainsi être mis en évidence. Pour l'ensemble du territoire, les résultats montrent que les extrêmes de la saison hivernale (DJF) connaitront les variations les plus importantes, soit une augmentation de l'ordre de 36 % en moyenne pour la période 2060-2080, comparativement à 1850-1950. Les changements appréhendés pour le printemps (MAM) seront sensiblement moins importants avec une augmentation moyenne avoisinant les 27 %.

La saison estivale (JJA) est caractérisée par une diminution moyenne de 10 % des valeurs de CJP dépassées en moyenne une fois chaque 20 ans pour les tuiles au sud du 50°N. Pour les tuiles au nord, on anticipe une augmentation moyenne de 24 %. La saison automnale (SON) se caractérise, quant à elle, par un léger gradient nord-sud avec des augmentations moyennes allant de 15 % pour les tuiles méridionales jusqu'à 35 % pour les tuiles septentrionales.

Les saisons ou les mois les plus propices à générer des CJP de fortes valeurs ne sont pas les mêmes pour toutes les tuiles du territoire. La variabilité mensuelle des extrêmes a été présentée à la Figure 4.14 pour six tuiles qui présentent une typologie des différents patrons mensuels observés. La valeur du CJP dépassée en moyenne une fois tous les 20 ans au cours de chacun des mois a été estimée pour les périodes 1850-1950 et 2060-2080. Pour certaines tuiles, celle qui couvre le Montana et le nord du Québec par exemple, l'été est la saison la plus propice pour générer des extrêmes tandis que pour d'autres, comme celle qui couvre la Côte Ouest, la probabilité d'observer une valeur élevée de CJP en été est pratiquement nulle.

L'examen de la Figure 4.14 révèle également que les CC auront, dans certaines régions, l'effet de déplacer la probabilité d'occurrence d'extrêmes vers d'autres mois ou saison. Par exemple, pour la tuile qui couvre l'État de New-York pour la période 1850-1950, le mois d'octobre se voit associé les valeurs de CJP les plus élevées, alors que pour la période 2060-2080, ce sera plutôt au cours des mois de décembre à mars. Une redistribution de la probabilité intra-annuelle d'occurrence des extrêmes est une conséquence très probable des CC.

Les résultats obtenus sont, dans leur ensemble, cohérents avec ceux reportés dans la littérature, notamment ceux de Kharin *et al.* (2007). L'analyse du taux d'augmentation relatif des extrêmes par degré Kelvin suggère qu'il suit approximativement celui établi sur la base de l'équation de Clausius-Clapeyron. Le taux obtenu varie entre 5.0 et 5.7 %/°K selon la récurrence et la zone considérée, alors que le taux théorique de Clausius-Clapeyron est de 6-7 %/°K. Pour l'ensemble du territoire, il semble que la valeur de CJP dépassée en moyenne une fois sur la période 1980-2000 sera dépassée environ deux fois pour la période 2045-2065 et 3.5 fois pour la période 2080-2100, résultat cohérent avec ceux reportés dans la littérature.

Afin d'investiguer l'impact du choix, somme toute arbitraire, de retenir en moyenne 12 dépassements par année, les analyses ont été répétées en utilisant des taux moyens de 3 et 6 dépassements par année. Les résultats obtenus à partir de ces taux ont été comparés au taux moyen de 12 dépassements par année. Il semble que, pour une zone de quelques tuiles qui correspondent grossièrement à la région des Prairies, le taux de 12 dépassements soit inadéquat puisque les valeurs obtenues en considérant les autres taux étaient systématiquement supérieures. Une solution serait d'employer un taux régional ajusté en fonction de la climatologie de la région. L'écart pour ces régions n'était toutefois pas important (5-10 %) et ne change en rien les conclusions générales.

La prise en compte de la variabilité intra-annuelle des extrêmes constitue un avantage indéniable de l'approche. En plus de l'étude des évènements extrêmes à l'échelle annuelle, cela rend possible l'analyse des extrêmes pour des périodes spécifiques de l'année, par exemple un mois ou une saison. Sachant que les évènements extrêmes sont générés par différents processus météorologiques selon leur moment d'occurrence dans l'année, il va de soi que cet aspect peut être mis à profit pour une meilleure compréhension et anticipation de l'impact des CC. La méthodologie développée peut facilement être appliquée afin d'analyser les séries issues d'autres modèles. Une application aux données du projet NARCCAP (North American Regional Climate Change Assessment Program) serait très intéressante et permettrait de comparer les résultats de différents modèles régionaux et pilotes. L'utilisation de différents scénarios d'émission de GES permettrait, quant à elle, de dresser un éventail complet des impacts possibles des CC en considérant l'incertitude liée au scénario d'émission. Le choix du scénario aurait, à l'horizon 2100, le plus d'impact sur les résultats (Hawkins et Sutton, 2009). Il faut donc rester critique face aux résultats présentés jusqu'à présent puisque d'une part, l'incertitude n'a pas été évaluée, et d'autre part, Min *et al.* (2010) montrent que les GCM semblent sous-estimer l'augmentation des précipitations extrêmes due au réchauffement climatique. Il faut également garder à l'esprit qu'à l'échelle de l'hydrologie opérationnelle, on travaille à des résolutions spatiales beaucoup plus fines que celles des mailles de calcul des GCM. L'application de la méthode à des séries issues de RCM de résolution plus fine semble donc tout à fait indiquée.

La seconde partie du mémoire porte sur l'analyse de données de stations météorologiques réparties au sud de la province de Québec. Une version régionalisée de l'approche SDP a été développée afin d'améliorer la qualité des estimateurs. Les séries observées sont des cumuls maxima correspondant à différentes durées au cours d'une journée. Pour les durées de 5, 10, 15 et 30 minutes, 107 stations de mesure sont disponibles, alors que les séries en 109 stations sont renseignées pour les durées 1, 2, 6 et 12 heures. Les mesures couvrent la période du 1er mai au 31 octobre. Les historiques des stations considérées varient entre 10 et 51 ans. Les données proviennent du MDDEP et d'Environnement Canada. Les SDP sont stationnaires selon le test de Mann-Kendall, il n'existe donc pas de variation interannuelle des paramètres. Cependant, la variabilité intra-annuelle doit, quant à elle, être considérée.

Différentes hypothèses sont émises quant au caractère régional ou local des différents paramètres et à ce titre plusieurs modèles ont été proposés. Parallèlement, la variation intra-annuelle des paramètres est décrite à partir d'une série de Fourier à 0, 1 ou 2 modes. Une alternative non paramétrique qui s'appuie sur la description de la variabilité intra-annuelle à l'aide de blocs mensuels est également proposée.

Les performances des différents modèles ont été comparées à partir du critère d'Akaike corrigé (AIC_c). Pour toutes les durées considérées, les modèles qui comportent une variabilité intra-annuelle des valeurs régionales des paramètres d'échelle et de forme sont de loin les meilleurs candidats. La variabilité du paramètre d'échelle semble particulièrement importante. Le critère AIC_c conclut également que la modélisation de la variabilité intra-annuelle à partir des séries de Fourier est supérieure à la représentation offerte par la méthode non paramétrique.

Les modèles régionaux ont été développés afin d'introduire une covariance spatiale des paramètres. L'introduction de ces covariables spatiales, à savoir la latitude, la longitude et l'élévation, a permis de révéler une structure spatiale au niveau des paramètres. Aux yeux du critère AIC_c, la prise en compte de ces covariables spatiales améliore considérablement la vraisemblance des modèles. Pour le territoire considéré, à l'exception de la durée 12 heures, la latitude représente la covariable spatiale qui s'est avérée la plus importante. Pour la durée 12 heures, c'est l'élévation qui améliorait le plus l'inférence des paramètres. Étendre l'analyse aux durées de 24 et 48 heures permettrait d'évaluer le rôle potentiel de la topographie pour les cumuls sur de plus longues durées. La prise en compte de la durée 24 heures serait d'autant plus intéressante que c'est un pas de temps très usuel dans l'analyse de la variabilité hydro-climatique.

Suivant la méthodologie développée, le modèle optimal parmi tous ceux proposé a été identifié pour chacune des durées. Les cartes du cumul de précipitation dépassé en moyenne une fois chaque 10 ans ont été tracées. Pour tous les pas de temps, à l'exception de 12 heures, un gradient nord-sud avec des valeurs plus élevées au sud est clairement mis en évidence.

Afin de visualiser l'impact du choix du modèle sur les résultats, les valeurs obtenues à partir du modèle optimal ont été comparées aux valeurs obtenues à partir d'autres modèles dont les poids AIC_c étaient significatifs. L'écart relatif maximal observé aux stations entre le modèle classé au premier et deuxième rang était inférieur à 1% et d'environ 7.4% entre le modèle au premier et troisième rang pour la durée 30 minutes. Les résultats se sont avérés davantage sensibles au choix du taux annuel moyen de dépassement servant à obtenir les SDP. Pour parvenir à ce constat, on a répété la procédure de sélection du modèle sur les SDP obtenues à partir de différents taux annuels moyens de dépassement, soit λ = 1, 1.5, 3 et 6. La complexité des modèles sélectionnés tend à décroître lorsque le taux décroît. Une hypothèse pour expliquer ceci est qu'avec un taux plus bas, les SDP sont plus courtes et donc, à complexité égale, le rapport entre le nombre de paramètres et le nombre de données augmente. Il y a donc moins d'informations qui permettent de détecter la structure temporelle et spatiale des paramètres.

Pour la durée 30 minutes, les écarts relatifs des résultats obtenus à partir de différents taux inférieurs et du taux de 6 dépassements ont été calculés. On constate que l'écart-type sur les écarts relatifs augmente avec une diminution du taux. La valeur absolue de l'erreur moyenne augmente également. Alors qu'une augmentation de la variance des écarts peut s'expliquer par une diminution de la taille des échantillons, une augmentation de l'écart relatif moyen pourrait être un indice de l'existence d'un biais. Par exemple, le taux utilisé (6) pourrait être trop élevé. Cependant, ce biais pourrait également provenir du modèle retenu, qui change avec la taille de l'échantillon. Pour les valeurs aux stations, le plus grand écart relatif observé est inférieur à 15 %. L'écart moyen entre les différents taux est quant à lui compris entre -0.9 et -2.4 %.

Les résultats montrent dans leur ensemble l'importance de considérer la variabilité intra-annuelle des paramètres. Cette variabilité intra-annuelle est devenue particulièrement évidente en considérant le caractère régional des paramètres. L'introduction de covariables spatiales dans l'estimation des paramètres a permis de révéler une structure spatiale des paramètres très intéressante du point de vue du critère AIC et ce, même si la région était considérée comme raisonnablement homogène selon la statistique H_1 de Hosking et Wallis (1997). La vraisemblance des modèles est grandement améliorée par l'introduction de ces covariables et justifie amplement l'augmentation en complexité. Pour le territoire étudié, la latitude semble très importante au niveau de la structure des paramètres pour les sept plus courtes durées considérées. Il pourrait être intéressant de comparer davantage de modèles, par exemple en considérant des modèles qui ne présentent une dépendance spatiale que sur l'un ou l'autre des paramètres de la GPD. Il serait également intéressant de vérifier si un modèle de dépendance autre que le modèle linéaire permettrait d'améliorer l'inférence.

D'autres éléments pourraient encore être approfondis. Par exemple, il serait intéressant d'évaluer la sensibilité des performances du modèle vis-à-vis la densité et l'organisation spatiale du réseau de stations. La capacité prédictive du modèle pourrait aussi être évaluée en des points non-jaugés soit par une approche de type validation croisée ou par l'ajout de nouvelles stations. Finalement en ce qui à trait à l'interpolation, il serait intéressant de tenter d'obtenir une carte des paramètres de la distribution en interpolant le paramètre de seuil. On pourrait ensuite déduire directement les cumuls de précipitations de différents temps de retour sur l'ensemble du territoire et comparer ces résultats à ceux obtenus en krigeant les valeurs aux stations.

BIBLIOGRAPHIE

Akaike H (1973) {Information Theory and an Extension of the Maximum Likelihood Principle}. *International Symposium on Information Theory, 2 nd, Tsahkadsor, Armenian SSR.* p 267-281.

Allen MR & Ingram WJ (2002) Constraints on future changes in climate and the hydrologic cycle. *Nature* 419(6903):224-232. Doi: 10.1038/nature01092

Bermudez PD & Kotz S (2009) Parameter estimation of the generalized Pareto distribution-Part I. *J. Stat. Plan. Infer.* 140(6):1353-1373. Doi: 10.1016/j.jspi.2008.11.019

Boer GJ (1993) Climate change and the regulation of the surface moisture and energy budgets. *Climate Dynamics* 8(5):225-239. Doi: 10.1029/2005JD006548

Buishand TA (1991) Extreme rainfall estimation by combining data from several sites. *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques* 36(4):345-365. Doi: 10.1080/02626669109492519

Burnham KP & Anderson DR (2002) *Model selection and multimodel inference : a practical information-theoretic approach.* Springer, New York, USA. 488 p

Burnham KP & Anderson DR (2004) Multimodel inference - understanding AIC and BIC in model selection. *Sociol. Methods. Res.* 33(2):261-304. Doi: 10.1177/0049124104268644

Cleveland WS & Devlin SJ (1988) Locally weighted regression: An approach to regression analysis by local fitting. *Journal of the American Statistical Association* 83(403):596-610. Doi: 10.2307/2289282

Coelho CAS, Ferro CAT, Stephenson DB & Steinskog DJ (2008) Methods for exploring spatial and temporal variability of extreme events in climate data. *J. Clim.* 21(10):2072-2092. Doi: 10.1175/2007jcli1781.1

Coles S (2001) *An Introduction to Statistical Modeling of Extreme Values.* Springer, London, UK. 228 p

Coles S, Pericchi LR & Sisson S (2003) A fully probabilistic approach to extreme rainfall modeling. *Journal of Hydrology* 273(1-4):35-50. Doi: 10.1016/S0022-1694(02)00353-0

Cunnane C (1978) Unbiased plotting positions - A review. *Journal of Hydrology* 37(3-4):205-222. Doi: 10.1016/0022-1694(78)90017-3

Dai A, Trenberth KE & Karl TR (1998) Global variations in droughts and wet spells: 1900-1995. *Geophys. Res. Lett.* 25(17):3367-3370. Doi: 10.1029/98GL52511

Davison AC & Smith RL (1990) Models for exceedances over high thresholds. *Journal of the Royal Statistical Society. Series B (Methodological)* 52(3):393-442. Doi: 10.2307/2345667

de Elia R, Caya D, Cote H, Frigon A, Biner S, Giguere M, Paquin D, Harvey R & Plummer D (2008) Evaluation of uncertainties in the CRCM-simulated North American climate. *Climate Dynamics* 30(2-3):113-132. Doi: 10.1007/s00382-007-0288-z

Dierckx G & Teugels JL (2010) Change point analysis of extreme values. *Environmetrics* 21(7-8):661-686. Doi: 10.1002/env.1041

El Adlouni S, Ouarda T, Zhang X, Roy R & Bobee B (2007) Generalized maximum likelihood estimators for the nonstationary generalized extreme value model. *Water Resour. Res.* 43(3):13. Doi: 10.1029/2005wr004545

Emori S & Brown SJ (2005) Dynamic and thermodynamic changes in mean and extreme precipitation under changed climate. *Geophys. Res. Lett.* 32(17):5. Doi: 10.1029/2005gl023272

Flato GM & Boer GJ (2001) Warming asymmetry in climate change simulations. *Geophys. Res. Lett.* 28(1):195-198. Doi: 10.1029/2000gl012121

GIEC (2007) Climate change 2007 : Synthesis Report. (IPCC, Geneva, Switzerland), p 104.

Goovaerts P (1997) *Geostatistics for Natural Resources Evaluation.* Oxford University Press, New York. 483 p

Gregory JM & Oerlemans J (1998) Simulated future sea-level rise due to glacier melt based on regionally and seasonally resolved temperature changes. *Nature* 391(6666):474-476. Doi: 10.1038/35119

Hamed KH & Ramachandra Rao A (1998) A modified Mann-Kendall trend test for autocorrelated data. *Journal of Hydrology* 204(1-4):182-196. Doi: 10.1016/s0022-1694(97)00125-x

Hawkins E & Sutton R (2009) The potential to narrow uncertainty in regional climate predictions. *Bull. Amer. Meteorol. Soc.* 90(8):1095-+. Doi: 10.1175/2009bams2607.1

Hines W & Montgomery D (1990) *Probability and Statistics in Engineering and Management Science.* Wiley, New York, USA. 732 p

Hosking JRM & Wallis JR (1987) Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics* 29(3):339-349. Doi: 10.2307/1269343

Hosking JRM & Wallis JR (1993) Some statistics useful in regional frequency analysis. *Water Resour. Res.* 29(2):271-281. Doi: 10.1029/92WR01980

Hosking JRM & Wallis JR (1997) *Regional frequency analysis: an approach based on L-moments.* Cambridge University Press, Cambridge, UK. 224 p

Katz RW, Parlange MB & Naveau P (Statistics of extremes in hydrology. *Adv. Water Resour.* 25(8-12):1287-1304. Doi: 10.1016/s0309-1708(02)00056-8

Khaliq MN, Ouarda T, Ondo JC, Gachon P & Bobee B (2006) Frequency analysis of a sequence of dependent and/or non-stationary hydro-meteorological observations: A review. *Journal of Hydrology* 329(3-4):534-552. Doi: 10.1016/j.jhydrol.2006.03.004

Kharin VV & Zwiers FW (2005) Estimating extremes in transient climate change simulations. *J. Clim.* 18(8):1156-1173. Doi: 10.1029/2003GL017324

Kharin VV, Zwiers FW, Zhang XB & Hegerl GC (2007) Changes in temperature and precipitation extremes in the IPCC ensemble of global coupled model simulations. *J. Clim.* 20(8):1419-1444. Doi: 10.1175/jcli4066.1

Kysely J, Picek J & Beranova R (2010) Estimating extremes in climate change simulations using the peaks-over-threshold method with a non-stationary threshold. *Glob. Planet. Change* 72(1-2):55-68. Doi: 10.1016/j.gloplacha.2010.03.006

Lagarias JC, Reeds JA, Wright MH & Wright PE (1998) Convergence properties of the Nelder—Mead simplex method in low dimensions. *SIAM Journal of Optimization* 9:112-147. Doi: 10.1.1.120.6062

Lang M, Ouarda T & Bobee B (1999) Towards operational guidelines for over-threshold modeling. *Journal of Hydrology* 225(3-4):103-117. Doi: 10.1016/S0022-1694(99)00167-5

Madsen H, Pearson CP & Rosbjerg D (1997a) Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events: 2. Regional modeling. *Water Resour. Res.* 33(4):759-769. Doi: 10.1029/96wr03849

Madsen H, Rasmussen PF & Rosbjerg D (1997b) Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events .1. At-site modeling. *Water Resour. Res.* 33(4):747-757. Doi: 10.1029/96WR03848

Mailhot A, Duchesne S, Caya D & Talbot G (2007) Assessment of future change in intensity-duration-frequency (IDF) curves for southern Quebec using the Canadian regional climate model (CRCM). *Journal of Hydrology* 347(1-2):197-210. Doi: 10.1016/j.jhydrot.2007.09.019

Mailhot A, Kingumbi A, Talbot G & Poulin A (2010) Future changes in intensity and seasonal pattern of occurrence of daily and multi-day annual maximum precipitation over Canada. *Journal of Hydrology* 388(3-4):173-185. Doi: 10.1016/j.jhydrol.2010.04.038

Mailhot A & Talbot G (2011) Mise à jour des estimateurs intensité-durée-fréquence (IDF) pour le sud-Québec : Tome I - Données et méthodes. (INRS-ETE, Québec, Canada), p 60.

McFarlane NA, Boer GJ, Blanchet JP & Lazare M (1992) The canadian climate center 2nd-generation general circulation model and its equilibrium climate. *J. Clim.* 5(10):1013-1044. Doi: 10.1175/1520-0442

McFarlane NA, J. F. Scinocca, M. Lazare, R. Harvey, D. Verseghy, and J. Li (2005) The CCCma third generation atmospheric general circulation model. (CCCma), p 25.

Mekis E & Hogg WD (1999) Rehabilitation and analysis of Canadian daily precipitation time series. *Atmos.-Ocean* 37(1):53-85. Doi: 10.1080/07055900

Mendez FJ, Menendez M, Luceno A, Medina R & Graham NE (2008) Seasonality and duration in extreme value distributions of significant wave height. *Ocean Eng.* 35(1):131-138. Doi: 10.1016/j.oceaneng.2007.07.012

Meylan P, Favre A-C & Musy A (2008) *Hydrologie fréquentielle : une science prédictive*. Presses polytechniques et universitaires romandes, Lausanne, Suisse. 173 p

Min S-K, Zhang X, Zwiers FW & Hegerl GC (2011) Human contribution to more-intense precipitation extremes. *Nature* 470(7334):378-381. Doi: 10.1038/nature09763

Nakicenovic N & Swart R (2000) Special report on emissions scenarios: a special report of Working Group III of the Intergovernmental Panel on Climate Change. (IPCC Geneva, Switzerland), p 21.

Olsen JR, Lambert JH & Haimes YY (1998) Risk of extreme events under nonstationary conditions. *Risk Anal.* 18(4):497-510. Doi: 10.1111/j.1539-6924

Pickands J (1975) Statistical inference using extreme order statistics. *The Annals of Statistics* 3(1):119-131. Doi: 10.1214/aos/1176343003

Quessy J-F, Favre A-C, Saïd M & Champagne M (2011) Statistical inference in Lombard's smooth-change model. *Environmetrics* 22(7):882-893. Doi: 10.1002/env.1108

Smith JA (1989) Regional flood frequency analysis using extreme order statistics of the annual peak record. *Water Resour. Res.* 25(2):311-317. Doi: 10.1029/WR025i002p00311

Strupczewski WG, Singh VP & Feluch W (2001) Non-stationary approach to at-site flood frequency modelling I. Maximum likelihood estimation. *Journal of Hydrology* 248(1-4):123-142. Doi: 10.1016/S0022-1694(01)00397-3

Sugahara S, da Rocha RP & Silveira R (2009) Non-stationary frequency analysis of extreme daily rainfall in Sao Paulo, Brazil. *International Journal of Climatology* 29(9):1339-1349. Doi: 10.1002/joc.1760

Titus JG & Narayanan V (1996) The risk of sea level rise : A delphic Monte Carlo analysis in which twenty researchers specify subjective probability distributions for model coefficients within their respective areas of expertise. *Climatic Change* 33(2):151-212. Doi: 10.1.1.169.3026

Zhang XB, Hogg WD & Mekis E (2001) Spatial and temporal characteristics of heavy precipitation events over Canada. *J. Clim.* 14(9):1923-1936. Doi: 10.1175/1520-0442

Zhang XB, Zwiers FW, Hegerl GC, Lambert FH, Gillett NP, Solomon S, Stott PA & Nozawa T (2007) Detection of human influence on twentieth-century precipitation trends. *Nature* 448(7152):461-U464. Doi: 10.1038/nature06025

Zhang XS & Srinivasan R (2009) GIS-Based Spatial Precipitation Estimation: A Comparison of Geostatistical Approaches(1). *J. Am. Water Resour. Assoc.* 45(4):894-906. Doi: 10.1111/j.1752-1688.2009.00335.x

ANNEXE A : DÉTAILS SUR L'INTERPOLATION DES PALIERS DES VALEURS MENSUELLES DES PARAMÈTRES DE LA GPD.

Pour chacun des mois de l'année, les valeurs des paramètres de la GPD pour différentes périodes doivent être interpolées avant d'obtenir une courbe continue. Différents nœuds, indiqués par les carrés rouges à la Figure A-1, sont d'abord ajoutés aux différents paliers identifiés lors de l'étape de sélection du seuil.

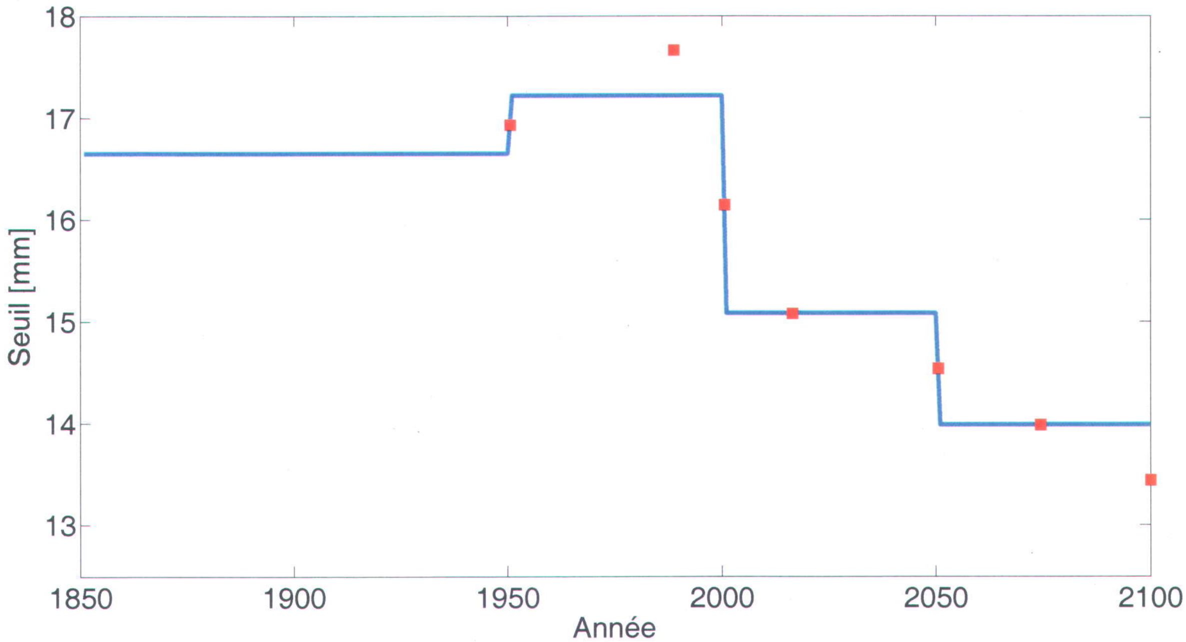

Figure A-1 : Valeur du seuil pour le mois d'août de la tuile en encadré rouge de la Figure 4.1. Les carrés rouges indiquent l'emplacement des différents nœuds utilisés par la méthode.

Les nœuds situés aux transitions entre deux plateaux sont placés à mi-hauteur. Pour le nœud en 2100, on supposera $h_1 = h_2$ (voir Figure A-2). La position en abscisse du nœud situé à l'intérieur de chaque palier est définie en fonction des différences entres les valeurs des paliers en périphérie (équation E-1). La Figure A-2 présente un schéma du positionnement des nœuds.

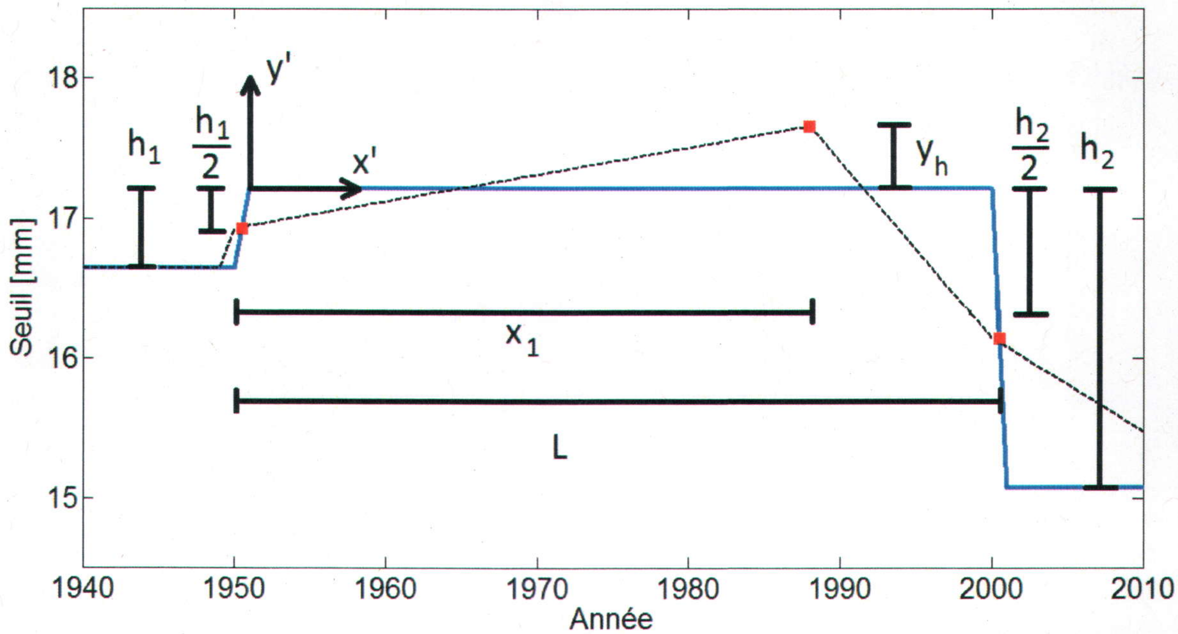

Figure A-2 : Positionnements des nœuds pour la période 1950-2000 de la Figure A-1. Les carrés rouges indiquent l'emplacement des nœuds. Les courbes en tirets noirs sont des droites reliant les différents nœuds.

La distance x_1 est définie par :

$$x_1 = \frac{L}{1+\left|\frac{h_1}{h_2}\right|} \qquad \text{si } h_1 \text{ et } h_2 \neq 0$$

$$x_1 = \frac{L}{2} \qquad \text{si } h_1 \text{ ou } h_2 = 0 \qquad \text{(A-1)}$$

L'ordonnée y_h est définie de sorte qu'en reliant les nœuds par des droites (lignes pointillées noires des Figures A-2 et A-3), les aires au-dessus et au-dessous du palier sont égales. Cette valeur peut être trouvée analytiquement. Dans le référentiel prime les équations de ces droites sont données par :

$$f_1(x') = \frac{(y_h - 0.5h_1)}{x_1} x' + 0.5h_1 \quad \text{(A-2)}$$

$$f_2(x'') = \left[\frac{0.5h_2 - y_h}{L - x_1} \right] x'' + y_h \quad \text{(A-3)}$$

avec $x'' = x' - x_1$.

On cherche :

$$\int_0^{x_1} f_1(x')dx' + \int_0^{L-x_1} f_2(x'')dx'' = 0 \quad \text{(A-4)}$$

$$\frac{(y_h - 0.5h_1)}{x_1}\frac{x_1^2}{2} + 0.5h_1x_1 + \left[\frac{0.5h_2 - y_h}{L - x_1}\right]\frac{(L-x_1)^2}{2} + y_h(L - x_1) = 0 \quad \text{(A-5)}$$

La solution de l'équation A-5 est :

$$y_h = \frac{-1}{2L}\left[h_1x_1 + h_2(L - x_1)\right] \quad \text{(A-6)}$$

Une fois leur position déterminée, les différents nœuds sont reliés par des droites. Les lignes en tirets noirs de la Figure A-3 présentent le résultat obtenu. Ces droites sont ensuite lissées à l'aide de la méthode LOESS avec une portée de 75 ans. Le résultat de ce lissage est utilisé comme valeur de paramètre et correspond à la courbe rouge de la Figure A-3.

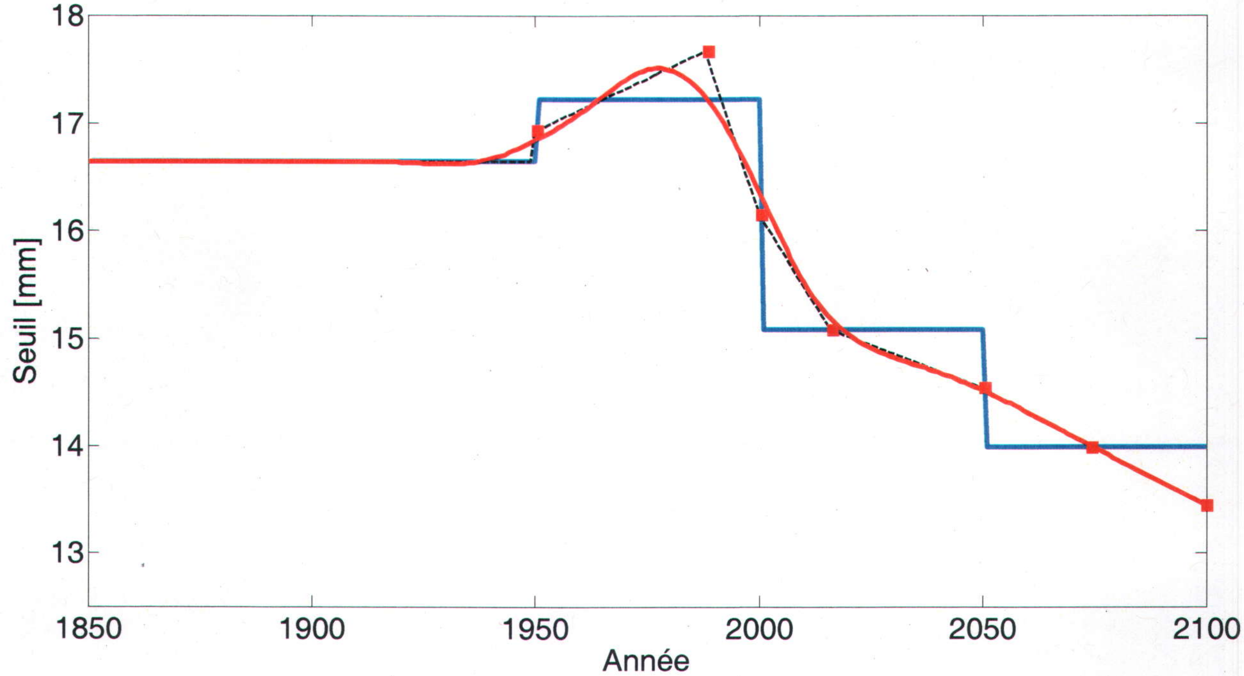

Figure A-3 : Courbe obtenue à la suite du lissage des droites en tirets noirs reliant les différents nœuds par la méthode LOESS avec une portée de 75 ans pour le seuil du mois d'août de la tuile en encadré rouge de la Figure 4.1.

ANNEXE B : VALEUR DE CJP DÉPASSÉE EN MOYENNE N FOIS EN 20 ANS POUR DIFFÉRENTES PÉRIODES.

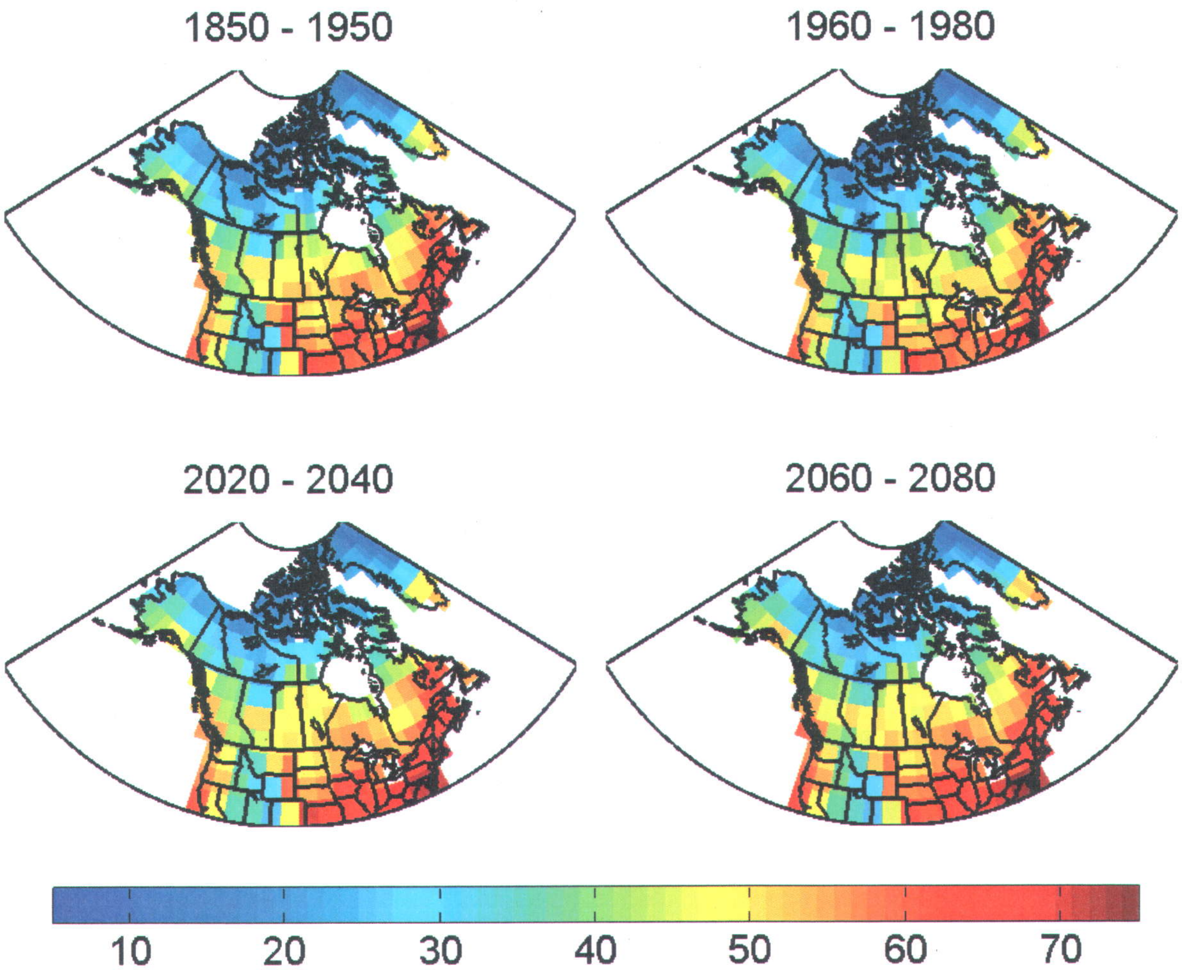

Figure B-1 : CJP (mm) dépassé en moyenne deux fois en 20 ans sur chaque période considérée.

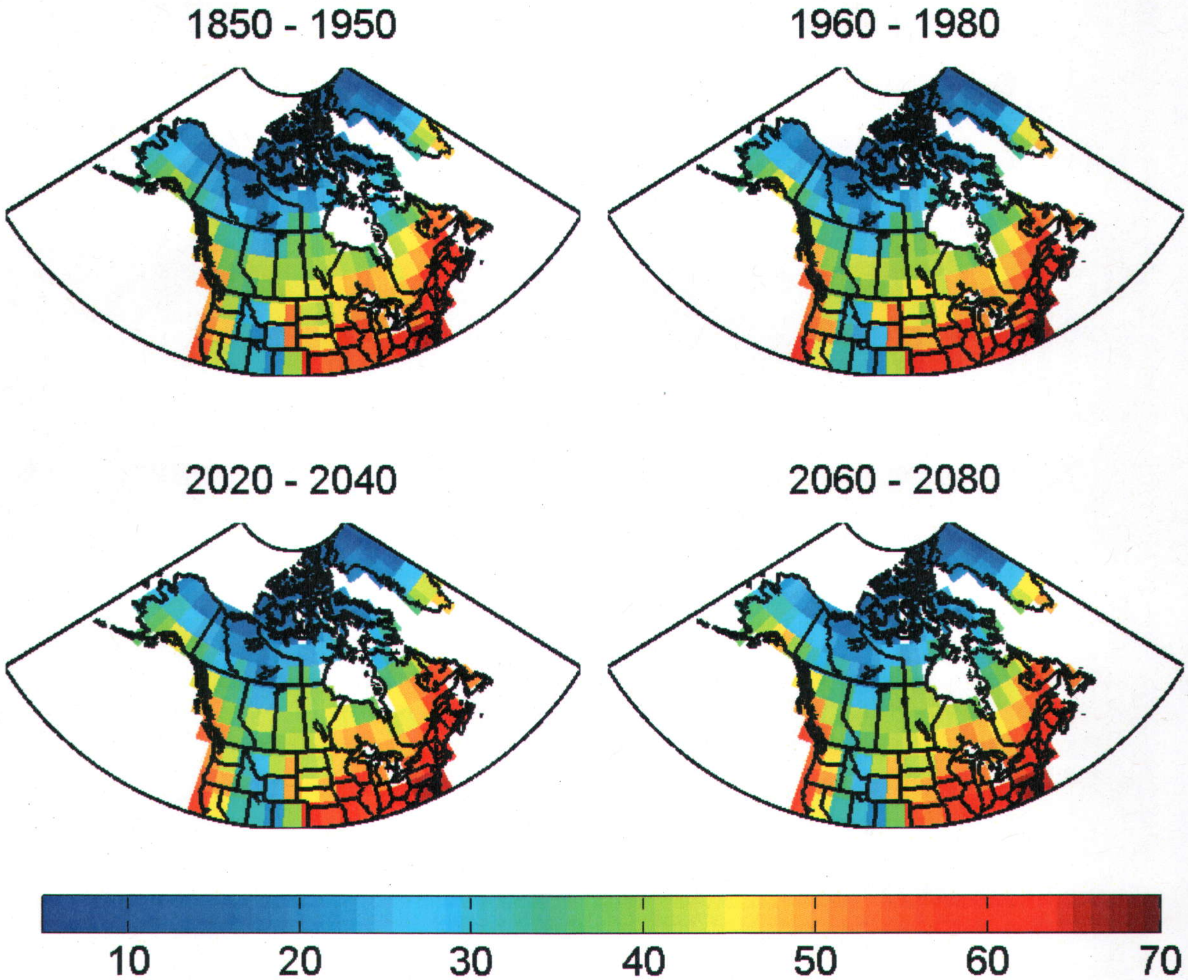

Figure B-2 : CJP (mm) dépassé en moyenne quatre fois en 20 ans sur chaque période considérée.

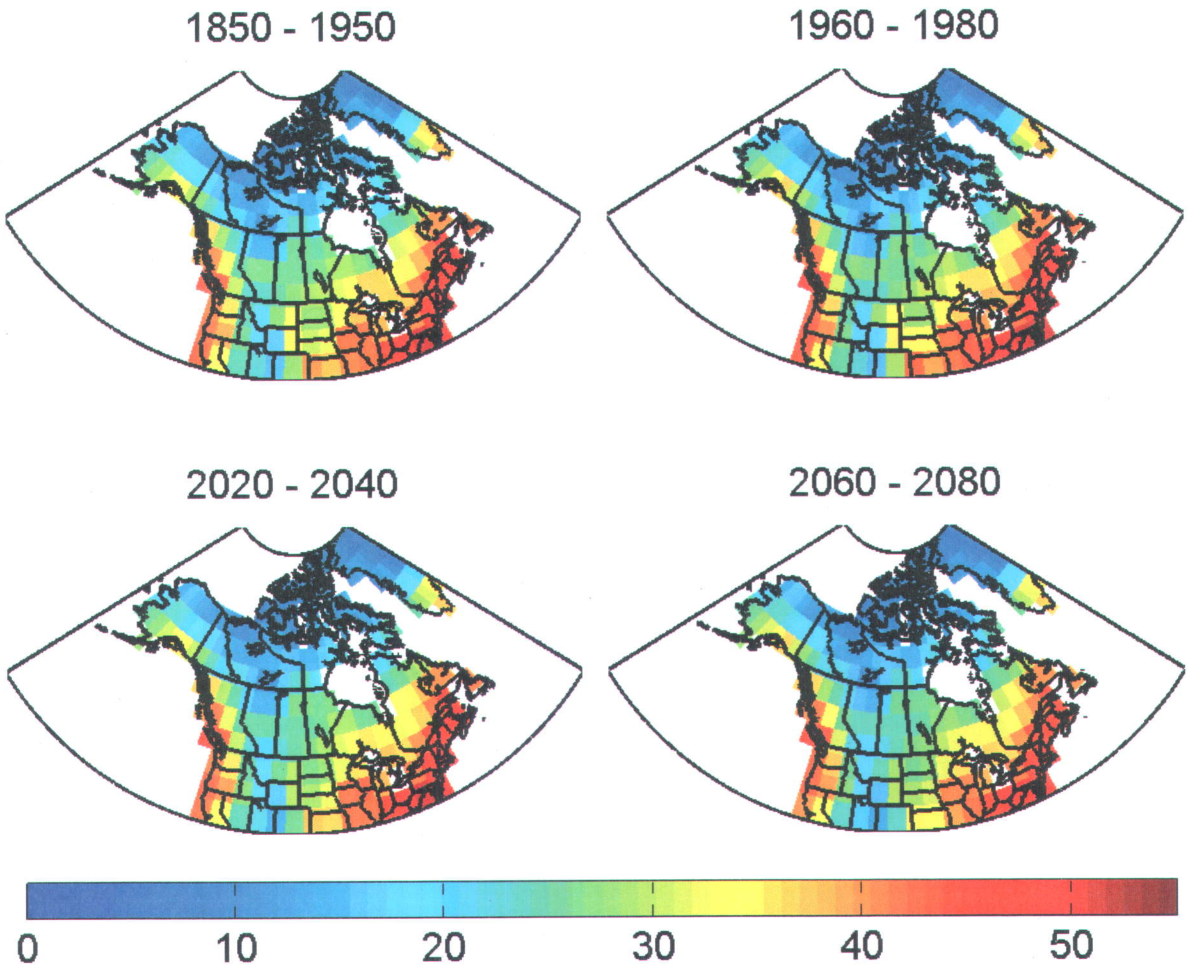

Figure B-3 : CJP (mm) dépassé en moyenne 20 fois en 20 ans sur chaque période considérée.

ANNEXE C : AUGMENTATION RELATIVE DES VALEURS DE CJP DÉPASSÉES EN MOYENNE N FOIS EN 20 ANS EN CLIMAT FUTUR COMPARATIVEMENT À LA PÉRIODE 1850-1950.

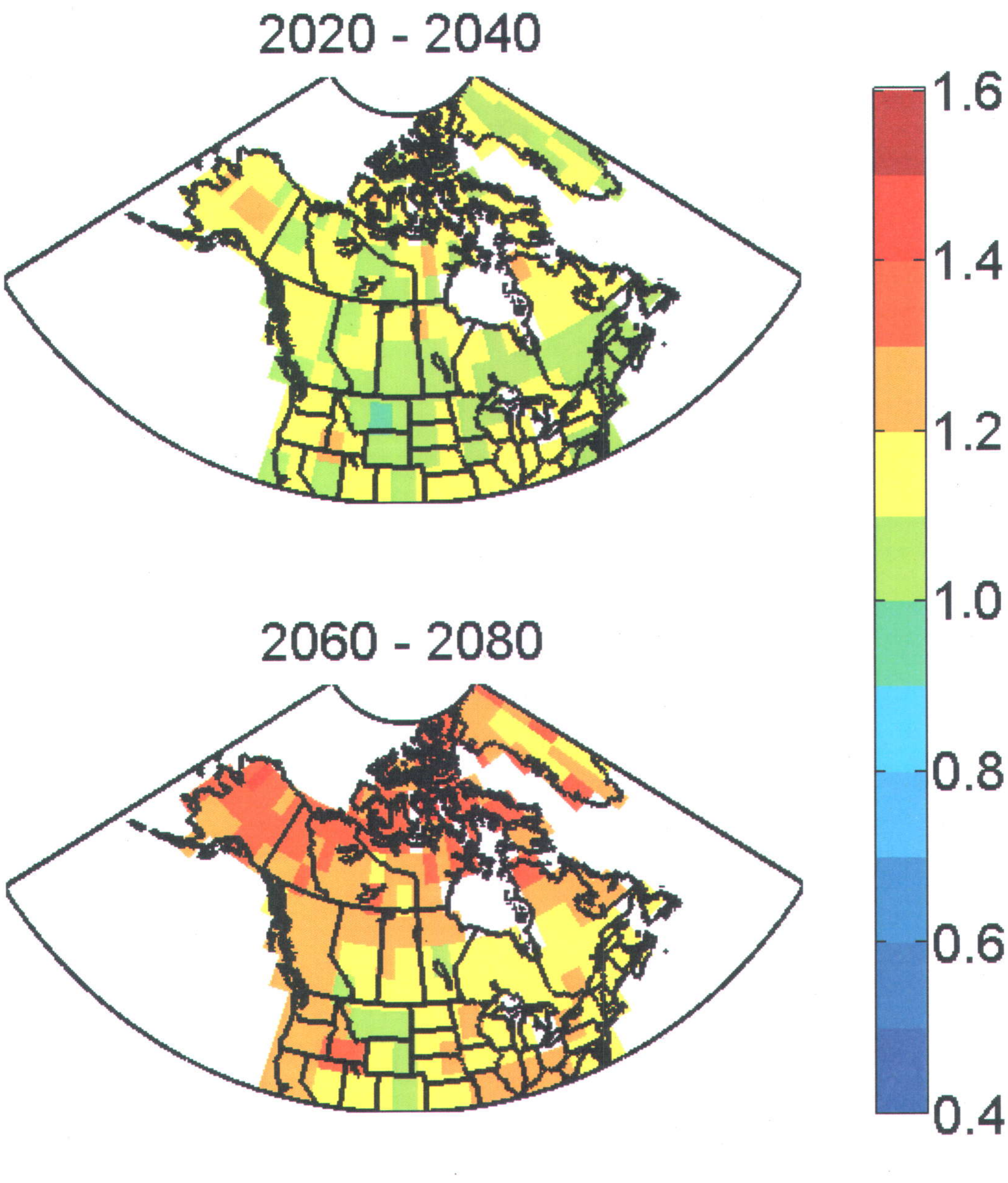

Figure C-1 : Rapports des CJP dépassés en moyenne deux fois en 20 ans pour les périodes 2020-2040 et 2060-2080 comparativement à la période de référence 1850-1950.

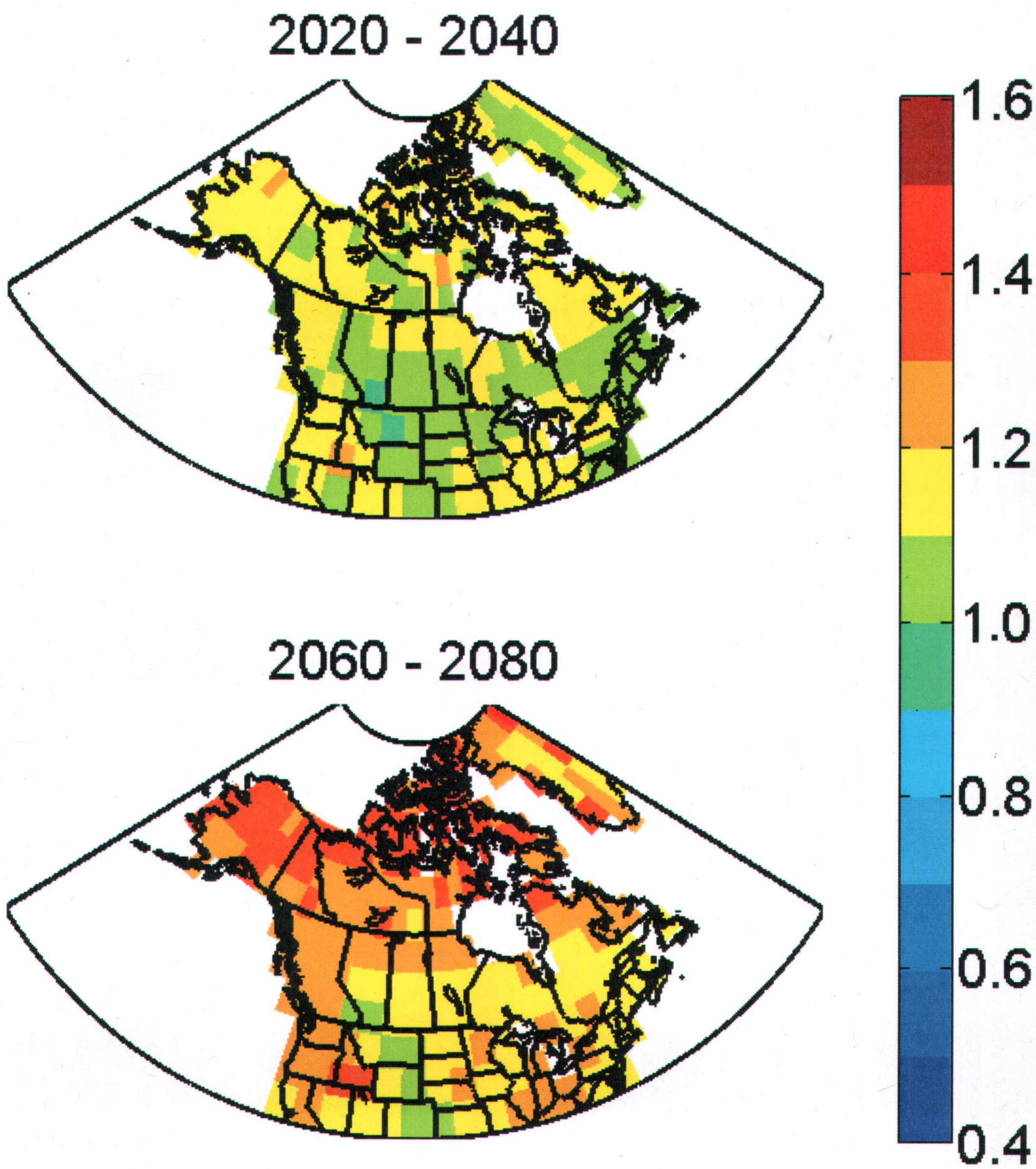

Figure C-2 : Rapports des CJP dépassés en moyenne quatre fois en 20 ans pour les périodes 2020-2040 et 2060-2080 comparativement à la période de référence 1850-1950.

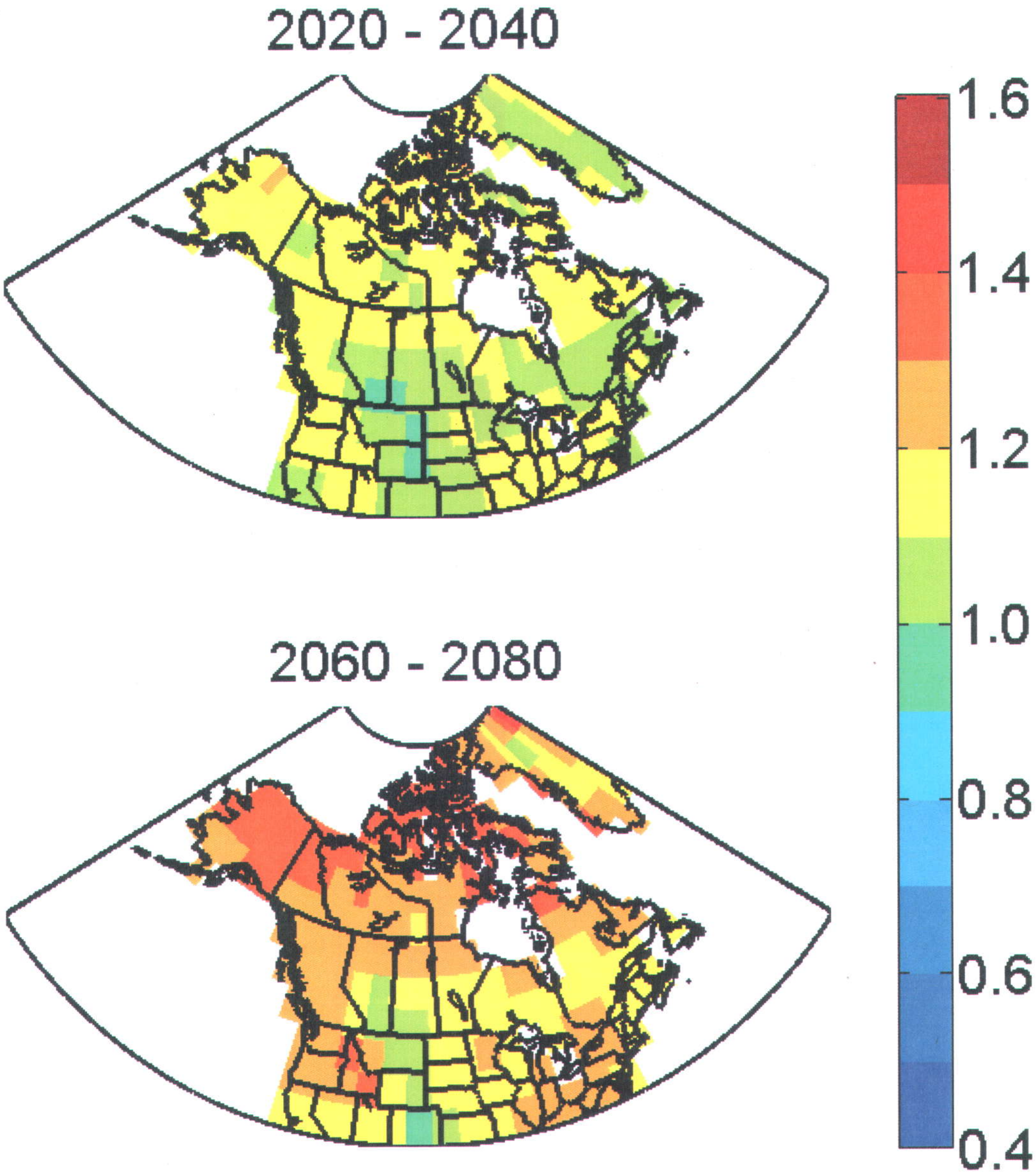

Figure C-3 : Rapports des CJP dépassés en moyenne 20 fois en 20 ans pour les périodes 2020-2040 et 2060-2080 comparativement à la période de référence 1850-1950.

ANNEXE D : MOYENNE SUR LES TUILES D'UNE LATITUDE DONNÉE DE LA VARIATION RELATIVE DU CJP DÉPASSÉ EN MOYENNE N FOIS EN 20 ANS AU COURS DE DIFFÉRENTES SAISONS.

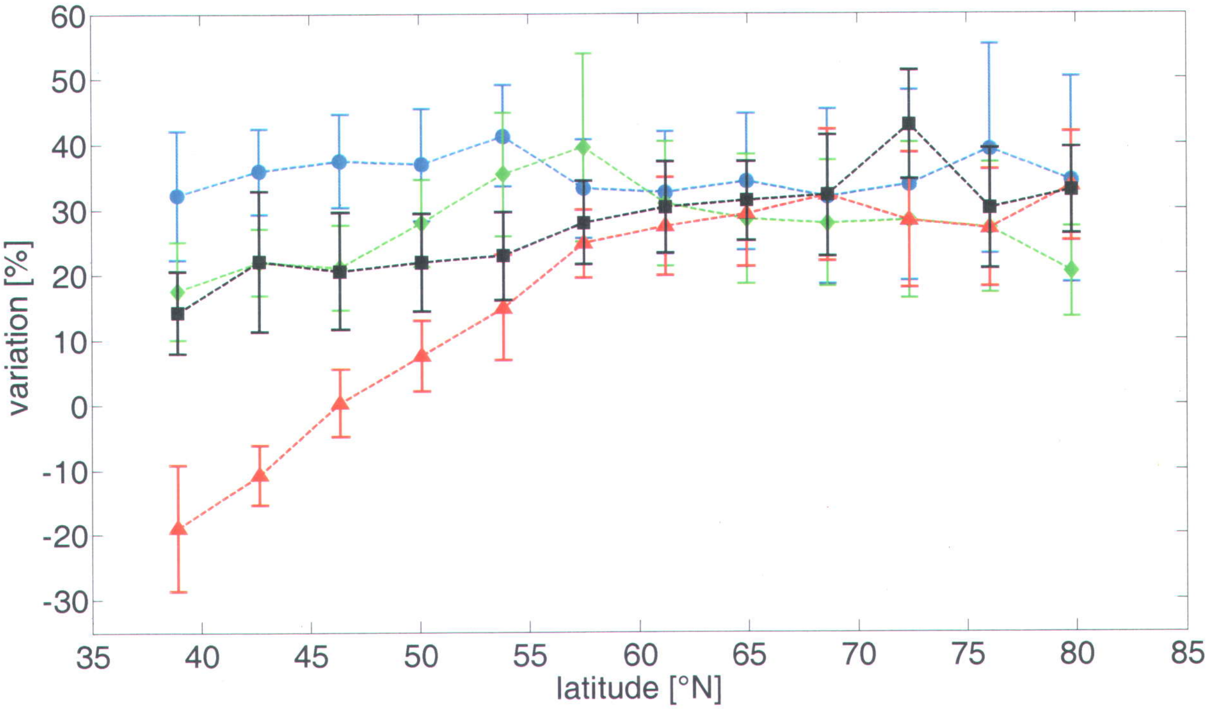

Figure D-1 : Moyenne sur les tuiles d'une latitude donnée de la variation relative du CJP dépassé en moyenne 0.5 fois en 20 ans ($N = 0.5$) au cours des différentes saisons de la période 2060-2080 et leurs homologues de la période 1850-1950. Les cercles bleus correspondent aux mois hivernaux (DJF), les losanges verts aux mois printaniers (MAM), les triangles rouges aux mois estivaux (JJA) et les carrés noirs aux mois automnaux (SON). Les barres d'erreur représentent l'écart-type sur les valeurs des tuiles pour une latitude donnée.

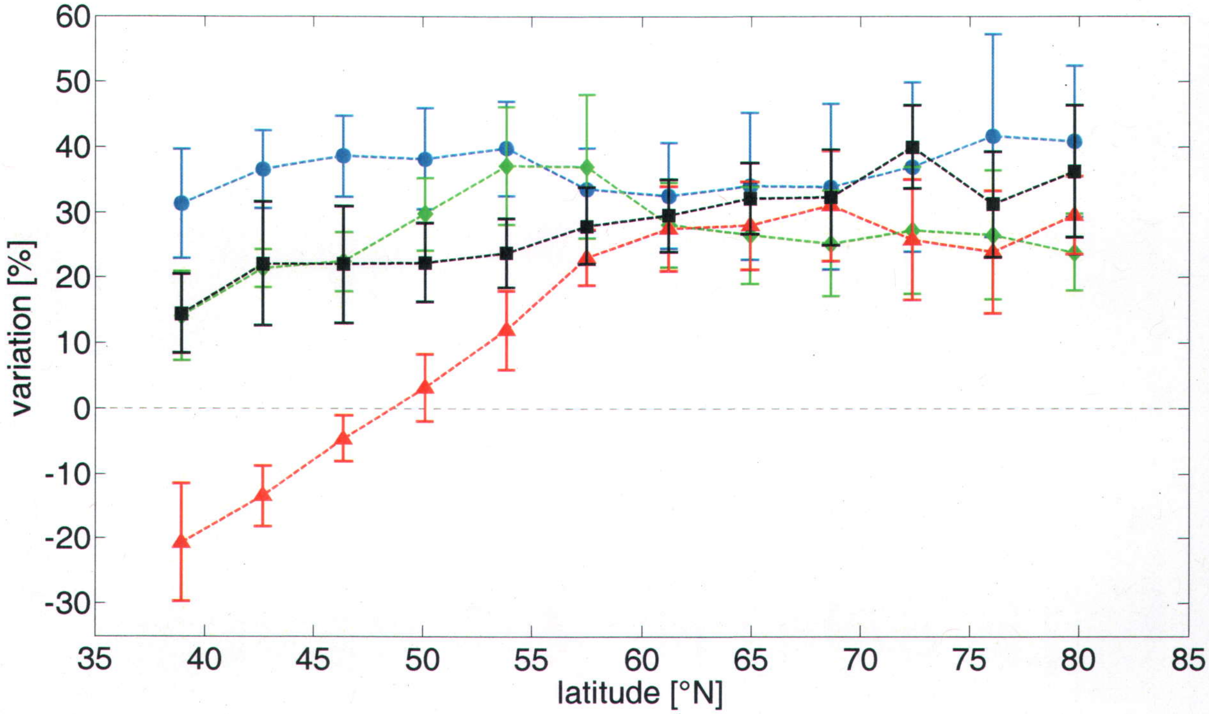

Figure D-2 : Moyenne sur les tuiles d'une latitude donnée de la variation relative du CJP dépassé en moyenne deux fois en 20 ans ($N = 2$) au cours des différentes saisons de la période 2060-2080 et leurs homologues de la période 1850-1950. Les cercles bleus correspondent aux mois hivernaux (DJF), les losanges verts aux mois printaniers (MAM), les triangles rouges aux mois estivaux (JJA) et les carrés noirs aux mois automnaux (SON). Les barres d'erreur représentent l'écart-type sur les valeurs des tuiles pour une latitude donnée.

ANNEXE E : DÉTAILS DE LA MÉTHODE D'OBTENTION DU SEUIL POUR LES SÉRIES DE STATIONS MÉTÉOROLOGIQUES

Dans le cas des séries observées, de légères modifications doivent être apportées à la méthode présentée à la section 4.2. D'abord, seule la variabilité intra-annuelle est considérée. Ensuite, une opération supplémentaire doit être effectuée en raison du fait que, dans le cas des stations pluviométriques, seuls 6 des 12 mois de l'année sont disponibles. La première étape consiste à choisir un taux de dépassement. Pour le présent travail, 6 dépassements par année MAO furent considérés, soit en moyenne un dépassement par mois. Les données disponibles sont regroupées en six séries mensuelles, une pour chaque mois.

Pour chacune des séries mensuelles, un algorithme détermine la valeur de seuil qui permet d'atteindre le taux retenu. Afin d'assurer l'indépendance des occurrences retenues, les séries sont soumises à l'exercice de *declustering* discuté à la section 2.3. À la suite de cette étape, les valeurs de seuil obtenues se présentent comme une succession de plateaux mensuels représentés par le trait bleu uni à la Figure E-1. À l'instar de la méthode de la section 4.2, ces plateaux sont ensuite lissés par le biais de la méthode LOESS avec une portée de 60 jours. Cependant, pour que cette procédure fonctionne efficacement pour les mois de mai et octobre, il faut considérer deux valeurs de plateaux additionnelles, l'une pour la période pré-mai et l'une pour la période post-octobre. Les valeurs de ces plateaux ont une incidence sur le lissage et, par conséquent, sur les valeurs de seuil des mois de mai et d'octobre. Un algorithme optimise les valeurs de ces plateaux afin que la courbe de seuil lissée permette d'atteindre le taux de dépassement voulu. Ces plateaux sont représentés par les traits bleus en tirets. Pour la période pré-mai, la valeur initiale fournie à l'algorithme pour optimisation résulte de la tendance linéaire entre le mois de mai et de juin; pour la période post-octobre, la valeur est le résultat de la tendance linéaire entre le mois de septembre et d'octobre.

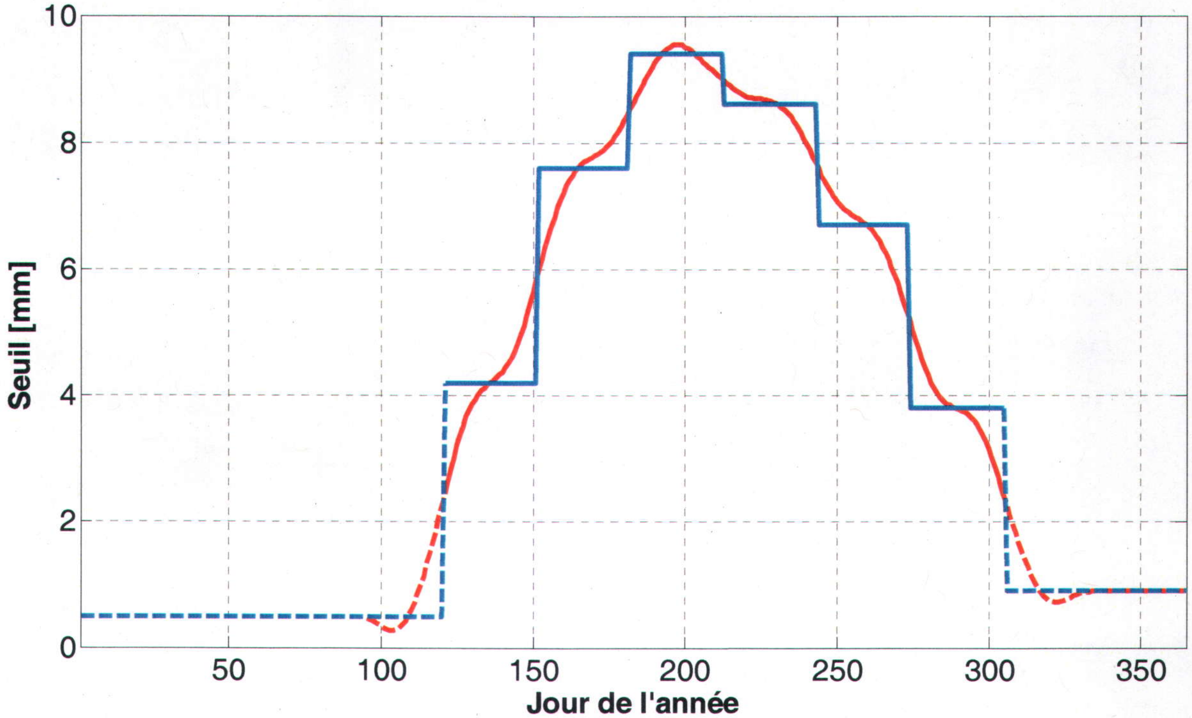

Figure E-1 : Exemple de la sélection du seuil pour la station de l'aéroport Pierre Elliott Trudeau pour la durée 30 minutes. Le trait bleu continu correspond aux valeurs mensuelles de seuil. Le trait rouge continu est le résultat du lissage à partir de la méthode LOESS avec une portée de 60 jours. Pour les deux couleurs, le segment en trait continu délimite la période MAO. Le trait bleu en tirets correspond aux valeurs de plateaux identifiés par l'algorithme.

ANNEXE F : CARTES POUR LE SUD DU QUÉBEC DU CUMUL DE PRÉCIPITATION DE TEMPS DE RETOUR 10 ANS POUR DIFFÉRENTES DURÉES.

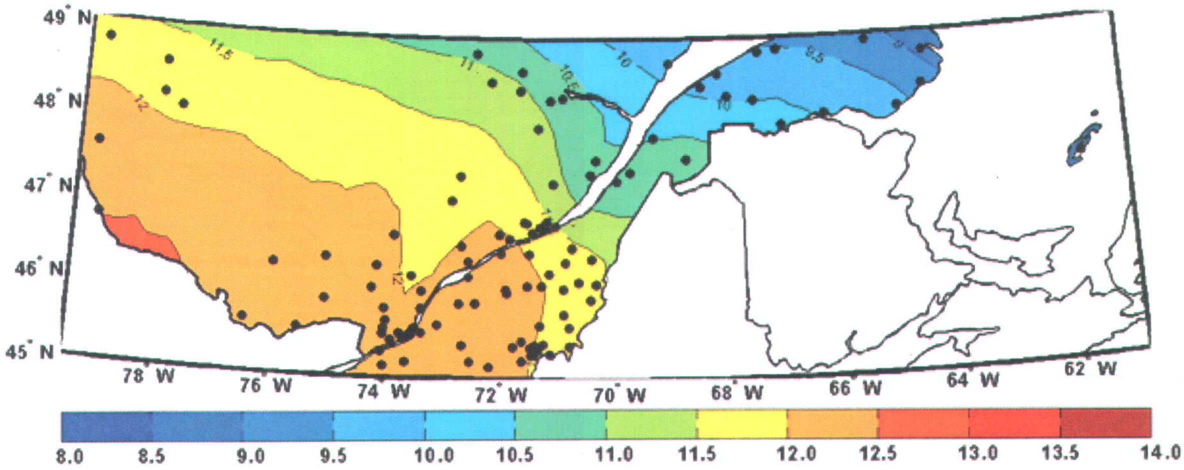

Figure F-1 : Carte du cumul de précipitation (mm) sur 5 minutes de temps de retour 10 ans selon le modèle 2-1-LGE.

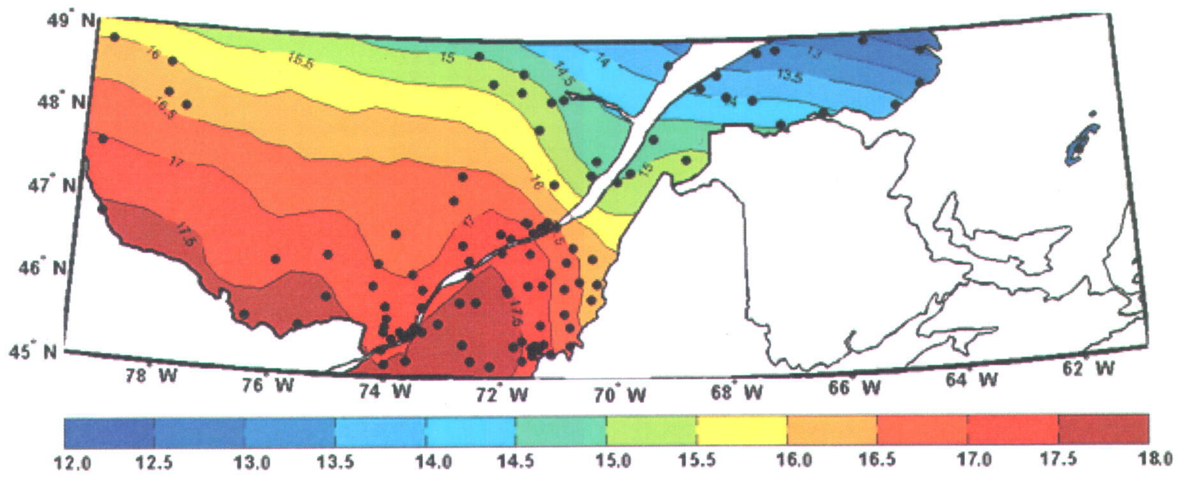

Figure F-2 : Carte du cumul de précipitation (mm) sur 10 minutes de temps de retour 10 ans selon le modèle 2-1-LGE.

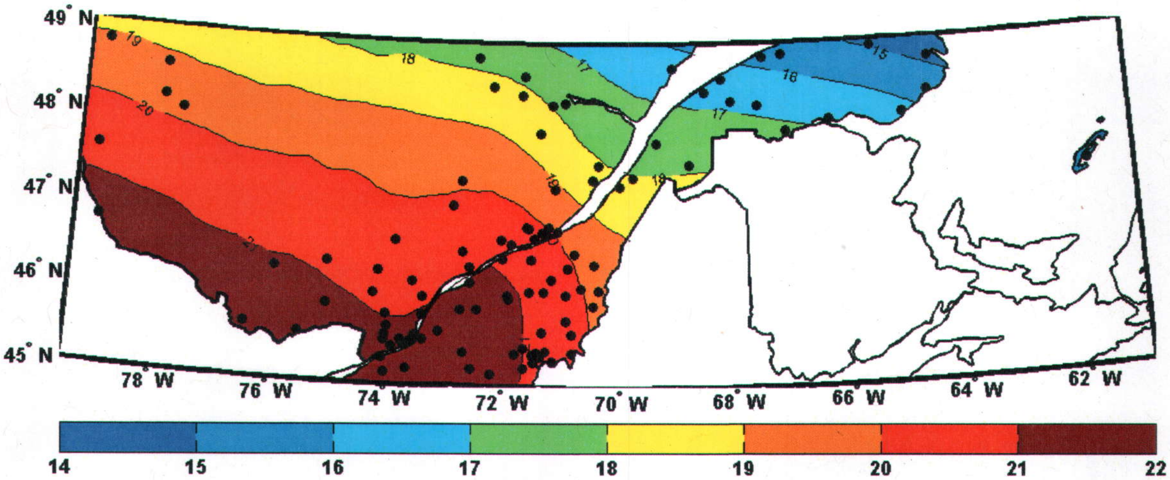

Figure F-3 : Carte du cumul de précipitation (mm) sur 15 minutes de temps de retour 10 ans selon le modèle 2-2-LGE.

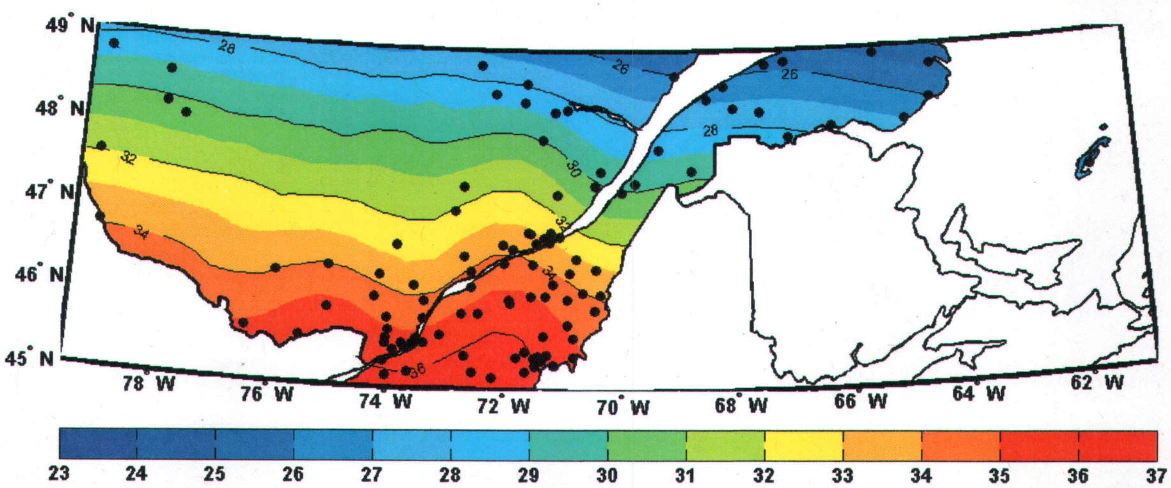

Figure F-4 : Carte du cumul de précipitation sur 1 heure (mm) de temps de retour 10 ans selon le modèle 2-1-LG.

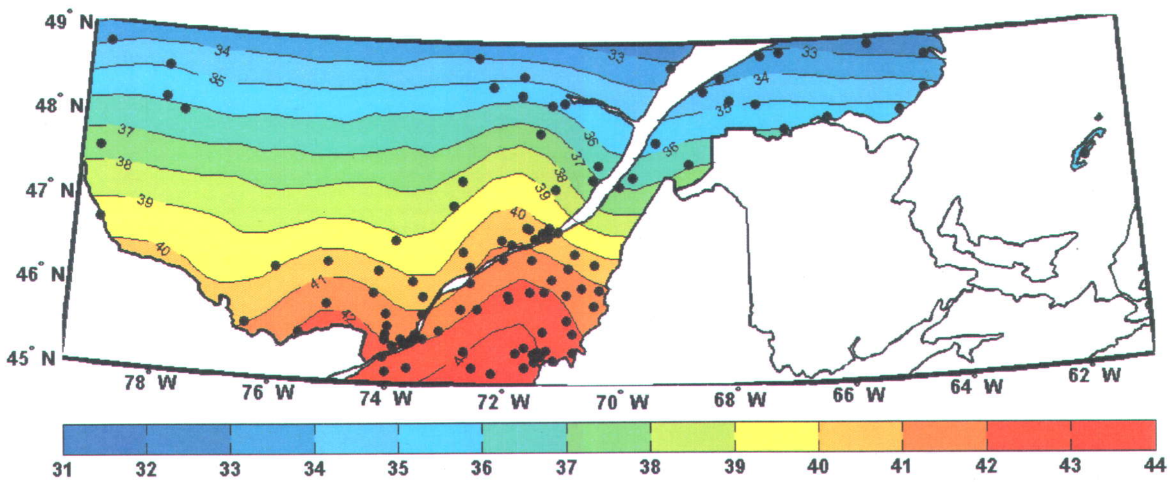

Figure F-5 : Carte du cumul de précipitation (mm) sur 2 heures de temps de retour 10 ans selon le modèle 2-1-L.

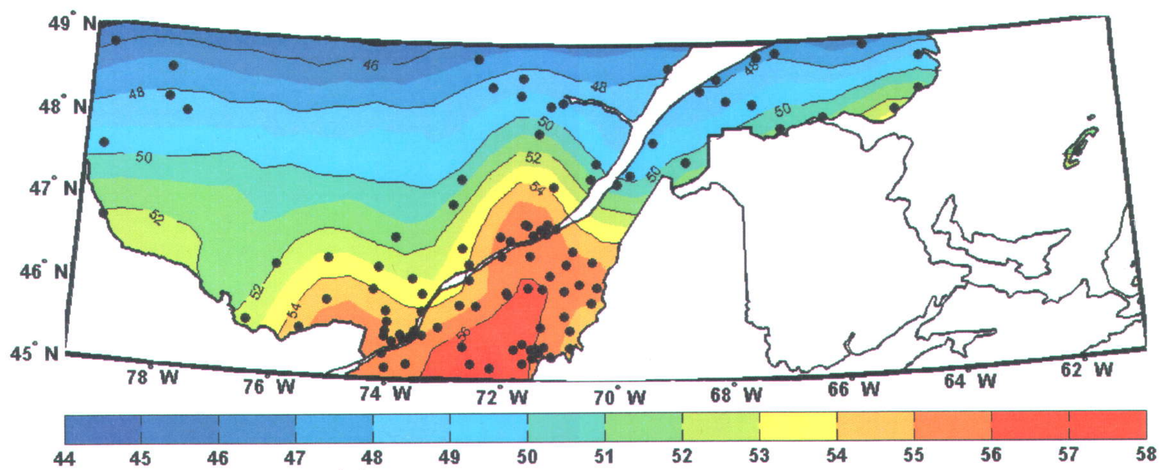

Figure F-6 : Carte du cumul de précipitation (mm) sur 6 heures de temps de retour 10 ans pour le modèle 2-0-LGE.

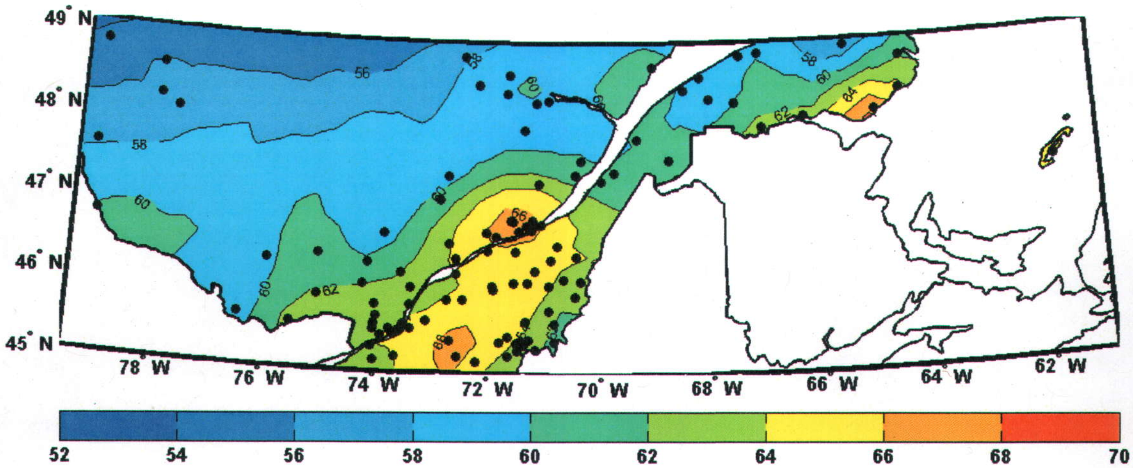

Figure F-7 : Carte du cumul de précipitation (mm) sur 12 heures de temps de retour 10 ans selon le modèle 1-1-LGE.