

Université du Québec
Institut National de la Recherche Scientifique
Centre Eau-Terre-Environnement

**DEVELOPPEMENT DE METHODES D'ANALYSE FREQUENTIELLE
NON-STATIONNAIRE AVEC L'APPROCHE DES DÉPASSEMENTS DE
SEUIL ET APPLICATION AVEC LA PRÉCIPITATION TOTALE
JOURNALIÈRE DANS LE SUD-EST DU CANADA**

Par
Alida Nadège Thiombiano

Thèse présentée pour l'obtention du grade de
Philosophiae Doctor (Ph.D) en sciences de l'eau

Jury d'évaluation

Examinateur externe	Naveed Khaliq Conseil National de la Recherche Canada
Examinateur externe	François Brissette École de Technologie Supérieure
Examinateur interne et Président du Jury	Fateh Chebana Institut National de la Recherche Scientifique
Codirecteur de recherche	Taha BMJ Ouarda Institut National de la Recherche Scientifique
Codirecteur de recherche	Salah-Eddine El Adlouni Université de Moncton
Directeur de recherche	André St-Hilaire Institut National de la Recherche Scientifique

DÉDICACE

À

Mon père Vincent Thiombiano et à ma mère Adiaratou Sophie Toguyeni/Thiombiano, tous deux décédés il y a de cela 20 ans. Merci de m'avoir transmis l'amour pour le travail bien fait dès mon jeune âge.

Mon époux Bonaventure Ouédraogo et mon fils Nathan Ouédraogo. Merci pour votre patience et votre soutien dans l'accomplissement de ma formation doctorale.

Mon beau-frère et tuteur Zacharie Zida, son épouse Alice Thiombiano/Zida (ma sœur) et leurs enfants Rosine, Emmanuel et Aimé Zida. Merci de m'avoir accueillie dans votre famille devenue la mienne.

Mes frères Achille, Anatole, Fadel, Hubert, Adolphe, Wilfried, Aristide et Auguste. Merci pour l'amour et le soutien inconditionnel reçus de chacun de vous.

Je suis arrivée jusque-là grâce à chacun de vous !
Puisse Dieu nous garder toujours unis dans Son Amour, amen !

REMERCIEMENTS

Je suis très reconnaissante à mes superviseurs pour toutes les connaissances que j'ai acquises sous leurs encadrements respectifs durant ma formation au programme de Doctorat en Sciences de l'Eau. Au professeur André St-Hilaire : je t'exprime ma gratitude pour ta patience et les accommodements que tu m'as permis de faire. Merci aussi pour ta disponibilité, ton leadership, ta rigueur et ta grande contribution dans l'amélioration de mon esprit d'analyse et celle de ma communication orale et écrite. Au professeur Salah-Eddine El Adlouni : j'ai tout appris en statistiques et en programmation avec toi. Je te remercie pour tous ces acquis techniques qui viennent enrichir mes compétences en recherche appliquée. Au Professeur Taha Ouarda : merci pour la proposition du sujet de recherche et pour ton esprit critique qui a toujours aidé à orienter les travaux de recherche et d'analyses. Je tiens également à remercier le professeur Nassir El-Jabi de l'Université de Moncton, qui m'a beaucoup encouragée à entreprendre ces études doctorales. Nassir, reçois en ces modestes lignes, ma profonde gratitude pour cet accomplissement. Merci à vous quatre pour le soutien financier reçu durant ma formation.

Mes remerciements s'adressent également à mes collègues Christian Saad, Iris Klein, Bouchra Nasri, Audrey Maheu, Sébastien Ouellet-Proulx, pour le support reçu en programmation, en statistiques et en cartographie.

Je voudrais à présent témoigner ma profonde gratitude à ma cousine Charlotte Thiombiano/Girard, son mari Mario Girard et leur fille Yasmine Girard pour leurs soutiens multiformes. Merci aussi à mon cousin Charles Thiombiano. Vous êtes ma famille au Canada et je suis chanceuse de vous avoir toujours à mes côtés.

Mes sincères remerciements à mes amis Issouf, Sarata, Agathe, Bassirou, Bertrand et Dorma, ainsi qu'aux familles Sinon Moussa, Ben Ammar Mehdi, Bonou Rémi, Sandwidi Moïse et Traoré Ben Salif, pour les encouragements multiformes.

Je ne peux terminer sans adresser mes remerciements aux membres du Jury qui ont bien voulu évaluer cette thèse de doctorat.

RESUMÉ

L'analyse fréquentielle (AF) est un outil statistique très utilisé en hydrologie pour la prédiction des quantiles d'évènements extrêmes. L'objectif général de la présente thèse était de proposer de nouvelles approches d'AF des extrêmes basées sur la méthode des dépassemens de seuil ou *peaks-over-threshold* (POT) en anglais.

En effet, avec la problématique des changements climatiques (CC), les indices climatiques sont fréquemment utilisés comme covariables dans le développement de modèles d'AF non-stationnaire du fait de la présence de tendances, dépendances, cycles ou ruptures dans les séries d'observations hydroclimatiques. Le premier objectif spécifique de cette thèse visait à analyser les téléconnexions entre les oscillations climatiques à grande échelle et les processus hydrologiques locaux/régionaux, afin de développer des modèles POT avec covariables. Les précipitations étant par essence très variables dans le temps et dans l'espace, la relation de dépendance entre plusieurs indices climatiques et des séries temporelles décrivant l'intensité et la fréquence des évènements extrêmes de la précipitation totale journalière, a été examinée à l'échelle du Sud-Est du Canada. Deux méthodes d'évaluation complémentaires, à savoir l'analyse de corrélation des rangs par le calcul du tau de Kendall ainsi que l'analyse par ondelettes, ont été employées pour confirmer statistiquement les interactions existantes. Cette analyse préliminaire a permis d'identifier deux indices climatiques ayant une influence significative sur l'intensité et la fréquence des précipitations extrêmes à l'échelle de la zone d'étude choisie : l'indice de l'oscillation arctique (AO pour *Arctic Oscillation*) et l'indice du Pacifique-Amérique du Nord (PNA pour *Pacific North American*). Des modèles de Pareto Généralisée (GPD pour *Generalized Pareto Distribution*) non-stationnaires ont été alors développés en faisant varier le paramètre d'échelle de la GPD en fonction de l'indice AO ou PNA au moyen de l'utilisation de fonctions semi-paramétriques, les fonctions B-splines. Celles-ci permettent de capturer aussi bien les relations simples (c'est à dire linéaires) que complexes (c'est à dire non-linéaires).

Les quantiles de la précipitation extrême ont été estimés à partir du modèle de la GPD stationnaire (où tous les paramètres sont constants) et de différents modèles GPD-B-splines qui ont été développés en faisant varier deux paramètres déterminants de la fonction B-spline: le degré (d) du polynôme et le nombre de nœuds internes (k). Les modèles GPD non-stationnaires ont montré une meilleure performance que le modèle GPD stationnaire à la lumière des valeurs minimales du critère d'information d'Akaïke observées surtout avec le modèle GPD-B-spline ($k=2$ et $d=1$).

La structure physique des indices climatiques comportant une phase négative et une phase positive, les modèles GPD-B-splines se sont montrés flexibles. En effet, des réponses linéaires et non-linéaires des précipitations extrêmes aux indices AO et PNA ont pu être détectées au Sud-Est du Canada. D'où des possibilités de régionalisation de ces modèles locaux pour fournir un outil robuste de prédiction des quantiles.

Toutefois, l'interprétation pratique des quantiles est généralement associée à la notion de période de retour. Pour cette raison, il y a un intérêt à incorporer l'information sur le taux moyen annuel des dépassements de seuil dans la modélisation des dépassements de seuil. Pour ce faire, l'étude de la surdispersion dans l'occurrence annuelle de ces événements est suggérée au préalable pour valider ou non l'hypothèse que la fréquence des dépassements de seuil est distribuée suivant un processus de Poisson. En effet, dans le cas échéant, les distributions Binomiale (B) et Binomiale Négative (BN) sont des alternatives à considérer. A l'échelle de la région d'étude, la distribution de Poisson (PD) a été validée à 90% des stations à l'aide du calcul de l'indice de dispersion et du test statistique du Khi-carré qui lui est associé. D'où la formulation d'un modèle couplé PD-GPD. L'incorporation additionnelle de la non-stationnarité au niveau du paramètre d'intensité de la PD a ainsi permis d'analyser son effet sur les quantiles conditionnels.

Les résultats de la modélisation effectuée pour répondre au premier objectif spécifique sont basés seulement sur l'échantillon des dépassements de seuil. Ceci implique que le seuil est fixé au préalable et qu'il n'y a pas de quantification de l'incertitude liée à sa sélection durant l'étape subséquente qu'est l'inférence statistique.

D'où la formulation du second objectif spécifique qui avait pour finalité d'utiliser l'ensemble des données disponibles pour obtenir une estimation du seuil comme un paramètre à part entière. L'approche des modèles dits « mixtes » a été donc utilisée pour la réalisation de cet objectif, en développant des modèles hybrides paramétriques à deux composantes, la première décrivant les observations en-dessous du seuil, et la seconde, celles au-dessus du seuil. Les distributions Gamma et Normale ont été évaluées séparément pour l'ajustement de la première composante, tandis que la GPD a été naturellement choisie pour l'ajustement de la deuxième composante. La location du seuil est au point de jonction des deux composantes. A cet effet, les modèles hybrides Gamma-GPD sans contrainte de continuité (modèle M1) et avec contrainte de continuité (modèle M2) au point du seuil de la fonction de densité de probabilité résultante, ainsi que le modèle Normale-GPD (M3) avec la même contrainte de continuité que M2, ont été développés dans un cadre Bayésien.

La comparaison de ces modèles entre eux et avec la GPD classique (modèle M0) a été faite au moyen de simulations en comparant le biais relatif et la racine de l'erreur quadratique moyenne relative des estimations. Les modèles hybrides avec une distribution Gamma (M1 et M2) permettent d'obtenir de meilleures estimations (faibles biais) que les modèles M3 et M0. L'un des avantages avec les modèles M1, M2 et M3 est l'automatisation de la détermination du seuil comme il fait partie du vecteur des paramètres du modèle hybride à estimer. L'application des modèles M1, M2 et M3 à des données de la précipitation totale journalière à une station spécifique a permis de constater leur fonctionnement pratique. De plus, les quantiles estimés à partir des modèles M0, M1, M2 et M3 sont comparables.

Les différents développements réalisés dans le cadre de cette thèse avec l'approche des dépassements de seuil (modèles hybrides et modèles POT), ont permis de proposer des modèles qui peuvent prendre en compte le maximum de données et d'informations disponibles pour améliorer la qualité des estimations.

Mots clés : Modèle POT ; Fonctions B-splines ; Covariables climatiques ; Modèle hybride ; Quantiles ; Précipitation ; Canada.

ABSTRACT

Frequency analysis (FA) is widely used in statistical hydrology for the analysis of extreme events. The overall objective of this thesis was to propose a number of new approaches for modeling extreme events based on the peaks-over-threshold (POT) sampling strategy.

Large scale climate indices are frequently used as covariates for the development of nonstationary FA models due to the presence of trends, dependencies, cycles or jumps in the time series of hydroclimatic variables. Hence, the first specific objective of this thesis was to analyze the teleconnection between large scale climate patterns and local/regional level hydrological processes, to develop POT models with covariates. Since precipitation is inherently characterized by strong spatial and temporal variability, the dependence between several large scale climate patterns and the time series of extreme daily precipitation amounts was examined for Southeastern Canada. Two complementary assessment methods, the Rank-based correlation analysis using the Kendall's tau and the wavelet analysis were used to statistically test the relationships between large scale climate indices and precipitation extremes. This preliminary analysis allowed identification of two climate indices that significantly influence the intensity and frequency of extreme precipitation over the study area. These indices are the Arctic Oscillation index (AO) and the Pacific North American index (PNA). Nonstationary FA approach based on the Generalized Pareto Distribution (GPD) was developed by allowing the GPD scale parameter to vary as a function of the AO or PNA index. The B-spline functions were used as semi-parametric link functions to capture both the simple (i.e. linear) and complex (i.e. nonlinear) relationships.

Estimation of extreme precipitation quantiles was carried out using the stationary GPD (by keeping all parameters constant) and different GPD-B-spline models. These models were developed by varying two key parameters of the B-spline functions, i.e. the degree (d) of the polynomial and the number of internal nodes (k).

The nonstationary GPD models showed better performance than the stationary GPD model in terms of minimum values of the Akaike Information Criterion. This feature was observed mainly for the GPD-B-spline model with $k=2$ and $d=1$. The GPD-B-spline models were found flexible regardless of the negative or positive phase of the large scale climate patterns. Indeed, both linear and nonlinear responses of extreme precipitation to the AO and PNA indices were detected in Southeastern Canada. This suggests the possibility of exploring further the developed models for estimating various quantiles at the regional scale.

The quantile estimates are generally interpreted in association with the return period concept in practical engineering. The annual average number of POT events must be considered at this end to obtain the GPD based quantiles with a T-years return period. To do so, overdispersion in the frequency of peaks needs to be assessed in order to validate the Poisson process hypothesis. If this prior assumption fails, then the Binomial (B) and Negative Binomial (BN) distributions can be considered instead. For the selected study area, the adequacy of the Poisson distribution (PD) was validated at 90% of the stations by computing the dispersion index and using the Chi-square statistical test. Hence formulation of a combined PD-GPD model and assessment of the effects of non-stationarity in the PD rate parameter on the conditional quantiles were studied.

Only the sample of peaks above a fixed threshold was used to carry out the modeling in order to achieve the first specific objective of the study. This implies that the threshold is fixed a priori with no possibility to further quantify the uncertainty associated with its subjective selection. The second specific purpose of this thesis was to use all available data for an objective estimation of the threshold as an explicit parameter. The so-called "mixture" models were used to achieve this objective, by developing two-component parametric hybrid models. The Gamma and Normal distributions were evaluated separately for fitting the first component, while the GPD was selected to fit the second component on the basis of theoretical grounds. These two components describe respectively the observations below and above the threshold to be estimated, whose location is assumed at the junction point of the two components.

The hybrid models Gamma-GPD without continuity constraint (referred to model M1) and with continuity constraint (referred to model M2) of the resulting probability density function at the threshold location, as well as the Normal-GPD model (referred to model M3) with enforced continuity as with model M2, were developed in a Bayesian framework. Comparison of these models with each other and with the classical GPD (i.e. model M0) was performed through a simulation study and by computing the relative bias and relative root mean square error of the quantile estimates. The hybrid GPD models with a Gamma distribution (i.e. models M1 and M2) provided best performance for quantile estimates (low biases) than models M3 and M0.

One of the advantages of using models M1, M2 and M3 is the explicit estimation of the threshold which is included in the hybrid model where all parameters need to be estimated. A case study of models M1, M2 and M3 with a sample of observed daily precipitation amounts, revealed their practical usefulness. Moreover, the quantiles estimated from models M0, M1, M2 and M3 were found comparable.

The mixture modeling approach presented in this thesis with the POT approach allows models which can use the full range of available data to improve the quality of estimations.

Key words: POT model; B-spline functions; Climatic covariates; Hybrid model; Quantiles; Precipitation; Canada.

AVANT-PROPOS

Les travaux de recherche effectués dans le cadre de cette thèse s'inscrivent dans le vaste programme de recherche « Faire-face Aux Changements Ensemble (FACE) : Mieux s'adapter aux Changements Climatiques au Canada et en Afrique de l'Ouest dans le domaine des ressources en eau ». (<http://face.ete.inrs.ca/fr>)

J'adresse toute ma reconnaissance au Centre de Recherches pour le Développement International (CRDI) ainsi qu'au Conseil de Recherches en Sciences Naturelles et Génie (CRSNG), pour le financement de ma thèse.

Un des objectifs clés du programme FACE était d'évaluer la viabilité et la résilience des systèmes naturels et humains et ce, en développant entre autres, des approches d'analyse et de prédiction qui tiennent compte des changements climatiques dans la conception et la gestion des ouvrages hydrauliques. Ces approches devront ensuite permettre d'optimiser la prédiction de la valeur des quantiles des variables hydro-climatologiques.

La structure du présent document est celle d'une thèse par articles comportant deux grandes parties : une première section destinée à la synthèse générale de la recherche effectuée, suivie de la partie II où sont présentés les trois articles réalisés.

ARTICLES ET CONTRIBUTION DES AUTEURS

Article 1: **Thiombiano, A.N.**, El Adlouni, S., St-Hilaire, A., Ouarda, T.B.M.J., El-Jabi, N. (2016) Nonstationary frequency analysis of extreme daily precipitation amounts in Southeastern Canada using a peaks-over-threshold approach. [Publié] *Theoretical and Applied Climatology* 124(1-2), doi: 10.1007/s00704-016-1789-7

Article 2: **Thiombiano, A.N.**, St-Hilaire, A., El Adlouni, S., Ouarda, T.B.M.J. (2017). Nonlinear response of precipitation to climate indices using a nonstationary Poisson-Generalized Pareto model: A case study of Southeastern Canada. [Soumis]

Article 3: **Thiombiano, A.N.**, El Adlouni, S., St-Hilaire, A., Ouarda, T.B.M.J. (2017) Bayesian inference with hybrid Generalized Pareto model. [À soumettre].

Le choix de l'approche des dépassements de seuil comme sujet de recherche général a été proposé par mon co-directeur, le professeur Taha Ouarda. Celui-ci a également suggéré une analyse complémentaire, celle par ondelettes, pour le 1^{er} article, la structuration du 2^{ème} article, et des pistes de discussion des résultats pour le 3^{ème} article. Il a enfin contribué à la révision de tous les trois articles réalisés.

L'option de la précipitation comme variable d'étude a été suggérée par le professeur Nassir El-Jabi de l'Université de Moncton, qui a apporté une contribution dans la définition de ma problématique générale de recherche durant les deux premières années de ma thèse.

Les développements méthodologiques proposés dans le cadre de cette thèse ont été élaborés en étroite collaboration avec mon co-directeur de recherche, le professeur Salah-Eddine El Adlouni, qui a d'ailleurs suggéré les idées originales à la base de leurs conceptions. Il a également révisé les trois articles réalisés.

Mon directeur de recherche, le professeur André St-Hilaire, a apporté de façon régulière et rigoureuse des orientations le long du processus de définition des objectifs de recherche, de traitement des données, d'analyse et de discussion des résultats. Pour le 1^{er} article, il a suggéré l'interprétation des résultats en termes de gradient Est-Ouest et a proposé une approche de sélection spatialement cohérente des indices climatiques pour les développements et applications présentés dans les articles 1 et 2. Il a particulièrement révisé tous les rapports et articles que j'ai produits durant ma formation doctorale.

L'étudiante que je suis a conduit toute la revue de la littérature en lien avec les objectifs de recherche fixés. J'ai aussi obtenu, traité et analysé les données de précipitations du Sud-Est du Canada ainsi que les données des différents indices climatiques utilisés dans le présent travail. Pour chacun des articles ci-dessus énumérés, j'ai développé les modèles proposés, effectué leurs applications et évalué leur robustesses, puis synthétisé et interprété tous les résultats, avant de procéder à la rédaction complète de chacun des trois articles.

CONTRIBUTION ORIGINALE DE LA THÈSE

En analyse fréquentielle, l'approche des dépassemens de seuil demeure très peu populaire en comparaison avec celle des blocs de maxima. Les travaux réalisés dans le cadre de cette thèse avaient une finalité double : (1) développer des modèles d'analyse fréquentielle non-stationnaire pour l'estimation conditionnelle de quantiles d'extrêmes hydro-climatologiques avec l'approche POT classique et (2) développer des modèles hybrides permettant une estimation plus objective du seuil au-dessus duquel les extrêmes sont considérés. Pour ce faire, les fonctions B-splines ont été utilisées comme alternative aux modèles linéaires ou quadratiques pour modéliser le lien entre les paramètres de la distribution de Pareto généralisé (GPD) et les covariables décrivant la non-stationnarité. Ceci représente le premier développement méthodologique du genre avec la méthode POT. Le modèle GPD non-stationnaire ainsi développé a été étendu pour inclure la non-stationnarité dans la fréquence des dépassemens de seuil en modélisant le paramètre d'intensité de la loi de Poisson avec des covariables. L'évaluation de l'approche méthodologique proposée pour le développement des modèles POT avec covariables a permis de proposer la première analyse de la précipitation totale journalière à l'échelle du Sud-Est du Canada en lien avec différentes téléconnexions climatiques. La complexité dans le choix du seuil étant une des principales limites à l'utilisation de la méthode POT, un modèle plus complet a été proposé. Il consiste à ajuster une distribution différente sous le seuil que celle utilisée au-dessus du seuil de sorte à ce que tous les paramètres de ce modèle, le seuil y compris, soient automatiquement estimés. Ces modèles hybrides existent déjà dans la littérature mais demeurent très peu utilisés en raison de leur faible performance au niveau des applications. La contribution apportée à ce niveau a été de partir des fondements théoriques d'une étude en économétrie pour proposer une adaptation de deux modèles hybrides paramétriques : les combinaisons Gamma-GPD et Normale-GPD.

TABLE DES MATIÈRES

DÉDICACE	III
REMERCIEMENTS	V
RESUMÉ	VII
ABSTRACT	XI
AVANT-PROPOS	XV
ARTICLES ET CONTRIBUTION DES AUTEURS	XVII
CONTRIBUTION ORIGINALE DE LA THESE	XIX
TABLE DES MATIÈRES.....	XXI
LISTE DES TABLEAUX.....	XXIII
LISTE DES FIGURES.....	XXV
LISTE DES ANNEXES.....	XXVII
PARTIE 1 : SYNTHESE GÉNÉRALE	1
1 INTRODUCTION GENERALE.....	3
1.1 MISE EN CONTEXTE	3
1.2 HYPOTHESES ET OBJECTIFS DE LA RECHERCHE.....	6
1.2.1 <i>Hypothèses</i>	6
1.2.2 <i>Objectifs</i>	7
2 REVUE DE LITTERATURE	9
2.1 VARIABILITE DANS LES SERIES DE LA PRECIPITATION ET ASSOCIATION AVEC LES CHANGEMENTS CLIMATIQUES.....	9
2.2 DISTRIBUTIONS ASYMPTOTIQUES.....	13
2.3 DETERMINATION DU SEUIL EN APPROCHE POT	17
2.4 METHODES D'ESTIMATION DES PARAMETRES	20
2.5 ANALYSE FREQUENTIELLE NON-STATIONNAIRE DES EXTREMES.....	21
3 MATÉRIELS ET MÉTHODES PROPOSÉS	23
3.1 MATERIELS D'APPLICATION.....	23
3.1.1 <i>Variable hydrologique</i>	23
3.1.2 <i>Covariables climatiques</i>	24
3.1.3 <i>Région d'étude</i>	25

3.2	DEVELOPPEMENTS METHODOLOGIQUES.....	27
3.2.1	<i>Interactions entre les précipitations et les covariables climatiques</i>	27
3.2.2	<i>Modèle GPD non-stationnaire.....</i>	28
3.2.3	<i>Modèle couplé PD-GPD non-stationnaire</i>	31
3.2.4	<i>Modèle hybride GPD</i>	32
4	RESULTATS DES APPLICATIONS.....	39
4.1	SELECTION DES STATIONS.....	39
4.2	SELECTION DE LA COVARIABLE CLIMATIQUE REGIONALE	41
4.3	VARIABILITE ET PERIODICITE	45
4.4	ANALYSE FREQUENTIELLE NON-STATIONNAIRE AVEC L'APPROCHE POT CLASSIQUE	46
4.5	MODELES HYBRIDES GPD ET ESTIMATION BAYESIENNE DES QUANTILES	50
5	CONCLUSION GENERALE	56
5.1	SYNTHESE.....	56
5.2	RETOMBEES PRATIQUES.....	58
5.3	LIMITES ET PERSPECTIVES	59
6	BIBLIOGRAPHIE GÉNÉRALE.....	61
PARTIE 2 : ARTICLES SCIENTIFIQUES		69
7	ARTICLE 1: NONSTATIONARY FREQUENCY ANALYSIS OF EXTREME DAILY PRECIPITATION AMOUNTS IN SOUTHEASTERN CANADA USING A PEAKS-OVER-THRESHOLD APPROACH	71
8	ARTICLE 2: NONLINEAR RESPONSE OF PRECIPITATION TO CLIMATE INDICES USING A NONSTATIONARY POISSON-GENERALIZED PARETO MODEL: A CASE STUDY OF SOUTHEASTERN CANADA.....	115
9	ARTICLE 3: BAYESIAN INFERENCE WITH HYBRID GENERALIZED PARETO MODEL..	151
ANNEXES		191

LISTE DES TABLEAUX

TABLEAU 1: MODELES GPD NON-STATIONNAIRES EVALUES	29
TABLEAU 2: REPARTITION DU NOMBRE DE STATIONS PAR PROVINCE ET EN FONCTION DE HUIT CRITERES.....	41
TABLEAU 3: REPARTITION DES STATIONS DECRIVANT V1, EN FONCTION DE LEURS CORRELATIONS SIGNIFICATIVES AVEC LES INDICES CLIMATIQUES.....	42
TABLEAU 4: REPARTITION DES STATIONS DECRIVANT V2, EN FONCTION DE LEURS CORRELATIONS SIGNIFICATIVES AVEC LES INDICES CLIMATIQUES.....	42
TABLEAU 5: INFLUENCE REGIONALE DES INDICES CLIMATIQUES SUR LA VARIABILITE DANS L'INTENSITE ET LA FREQUENCE DES PRECIPITATIONS EXTREMES	43
TABLEAU 6 : COMPARAISON DES VALEURS DU RB ET DU RRMSE DES QUANTILES POUR L'ERREUR DE TYPE I	51
TABLEAU 7 : COMPARAISON DES VALEURS DU RB ET DU RRMSE DES QUANTILES POUR L'ERREUR DE TYPE II	52

LISTE DES FIGURES

FIGURE 1: CADRE GEOGRAPHIQUE DE LA ZONE D'ETUDE	25
FIGURE 2: SCHEMAS ILLUSTRATIF DU CALCUL DES PROBABILITES AU NON-DEPASSEMENT POUR L'ESTIMATION DES QUANTILES A PARTIR DES MODELES HYBRIDES M1, M2 ET M3.....	35
FIGURE 3: REGION SUD-EST DU CANADA (A) ET LOCALISATION DES 329 STATIONS COMPORTANT AU MOINS 30 ANNEES DE DONNEES JOURNALIERES COMPLETES (B)	39
FIGURE 4: EXEMPLE ILLUSTRATIF DES MODELES M1 (GAUCHE) ET M2 (DROITE) AVEC LA FONCTION DE DENSITE DE PROBABILITE EN BLEU ET LA LOCATION DU SEUIL ESTIME EN ROUGE	50
FIGURE 5: AJUSTEMENT DU MODELE M1 A LA PRECIPITATION TOTALE JOURNALIERE DE LA STATION 7025250	53
FIGURE 6: AJUSTEMENT DU MODELE M2 A LA PRECIPITATION TOTALE JOURNALIERE DE LA STATION 7025250	54

LISTE DES ANNEXES

ANNEXE 1: DISTRIBUTION DE LA PRECIPITATION TOTALE JOURNALIERE A L'AEROPORT DE MIRAMICHI (STATION 8101000) DANS LA PROVINCE DU NB, SUR LA PERIODE 1945-2005	193
ANNEXE 2: FONCTION DE L'AUTOCORRELATION ENTRE LES DEPASSEMENTS DE SEUIL A LA STATION 8101000 ...	193
ANNEXE 3: ÉVOLUTION DE LA MAGNITUDE MOYENNE DES DEPASSEMENTS DE SEUIL A LA STATION 8101000	194
ANNEXE 4: ÉVOLUTION DU PARAMETRE DE FORME DE LA GPD AVEC LES DEPASSEMENTS DE SEUIL DE LA STATION 8101000.....	194
ANNEXE 5: ÉVOLUTION DU PARAMETRE D'ECHELLE DE LA GPD AVEC LES DEPASSEMENTS DE SEUIL DE LA STATION 8101000	195
ANNEXE 6: VALIDITE DU PROCESSUS DE POISSON POUR LA MODELISATION DE LA FREQUENCE ANNUELLE DES DEPASSEMENTS DE SEUIL INDEPENDANTS A L'ECHELLE DU SUD-EST (B) DU CANADA (A)	195
ANNEXE 7: STATIONS OU V1 ET V2 ONT MONTRE UNE CORRELATION SIGNIFICATIVE AVEC L'INDICE AMO	196
ANNEXE 8: STATIONS OU V1 ET V2 ONT MONTRE UNE CORRELATION SIGNIFICATIVE AVEC L'INDICE AO	196
ANNEXE 9: STATIONS OU V1 ET V2 ONT MONTRE UNE CORRELATION SIGNIFICATIVE AVEC L'INDICE NAO	197
ANNEXE 10: STATIONS OU V1 ET V2 ONT MONTRE UNE CORRELATION SIGNIFICATIVE AVEC L'INDICE PDO	197
ANNEXE 11: STATIONS OU V1 ET V2 ONT MONTRE UNE CORRELATION SIGNIFICATIVE AVEC L'INDICE PNA	198
ANNEXE 12: STATIONS OU V1 ET V2 ONT MONTRE UNE CORRELATION SIGNIFICATIVE AVEC L'INDICE SOI.....	198
ANNEXE 13: STATIONS OU V1 ET V2 ONT MONTRE UNE CORRELATION SIGNIFICATIVE AVEC L'INDICE WHWP	199
ANNEXE 14: DESCRIPTION DES 96 STATIONS DES GROUPES AO (EN GRAS) ET PNA.....	200
ANNEXE 15: INFLUENCE DE L'INDICE AO SUR V1 ET V2.....	206
ANNEXE 16: INFLUENCE DE L'INDICE PNA SUR V1 ET V2	207
ANNEXE 17: TRANSFORMEE EN ONDELETTE CONTINUE POUR DES SERIES DE V1.....	208
ANNEXE 18: TRANSFORMEE EN ONDELETTE CONTINUE POUR DES SERIES DE V2.....	209
ANNEXE 19: TRANSFORMEE EN ONDELETTE CONTINUE DES 12 SERIES ANNUELLES (A-L) CONSTITUEES POUR L'INDICE AO	210
ANNEXE 20: TRANSFORMEE EN ONDELETTE CONTINUE DES 12 SERIES ANNUELLES (A-L) CONSTITUEES POUR L'INDICE PNA.....	210
ANNEXE 21: MODELE GPD AVEC FONCTION B-SPLINE A DEUX NŒUDS ET DE DEGRE 2 : STATION 6158255	211
ANNEXE 22: MODELE GPD AVEC FONCTION B-SPLINE A DEUX NŒUDS ET DE DEGRE 2 : STATION 7060400	211
ANNEXE 23: QUASI-EQUIVALENCE DES QUANTILES STATIONNAIRES ET NON-STATIONNAIRES A LA STATION 8300100	212
ANNEXE 24: RESULTATS DE L'AF POUR LES 46 STATIONS ASSOCIEES A L'INDICE AO	213
ANNEXE 25: RESULTATS DE L'AF POUR LES 46 STATIONS ASSOCIEES A L'INDICE PNA	216
ANNEXE 26: CHANGEMENT DANS L'INTENSITE DES PRECIPITATIONS EXTREMES DURANT LES DEUX PHASES DE LA COVARIABLE CLIMATIQUE : EXEMPLE 1 AVEC LA STATION 8300500	219

ANNEXE 27: CHANGEMENT DANS L'INTENSITE DES PRECIPITATIONS EXTREMES DURANT LES DEUX PHASES DE LA COVARIABLE CLIMATIQUE : EXEMPLE 2 AVEC LA STATION 6057592	219
ANNEXE 28: CHANGEMENT DANS L'INTENSITE DES PRECIPITATIONS EXTREMES DURANT LA PHASE POSITIVE DE LA COVARIABLE CLIMATIQUE : EXEMPLE 1 AVEC LA STATION 8403600	220
ANNEXE 29: CHANGEMENT DANS L'INTENSITE DES PRECIPITATIONS EXTREMES DURANT LA PHASE POSITIVE DE LA COVARIABLE CLIMATIQUE : EXEMPLE 2 AVEC LA STATION 7020040	220
ANNEXE 30: CHANGEMENT DANS L'INTENSITE DES PRECIPITATIONS EXTREMES DURANT LA PHASE NEGATIVE DE LA COVARIABLE CLIMATIQUE : EXEMPLE 1 AVEC LA STATION 7022375	221
ANNEXE 31: CHANGEMENT DANS L'INTENSITE DES PRECIPITATIONS EXTREMES DURANT LA PHASE NEGATIVE DE LA COVARIABLE CLIMATIQUE : EXEMPLE 2 AVEC LA STATION 8300525	221
ANNEXE 32: DEPENDANCE LINEAIRE STATISTIQUEMENT SIGNIFICATIVE : EXEMPLE AVEC LA STATION 7051120 .	222
ANNEXE 33: DEPENDANCE LINEAIRE STATISTIQUEMENT NON-SIGNIFICATIVE : EXEMPLE AVEC LA STATION 8203700	222
ANNEXE 34: QUANTILES CONDITIONNELS A AO AVEC LE MODELE P-GPD A LA STATION 8403800	223
ANNEXE 35: QUANTILES CONDITIONNELS A AO AVEC LE MODELE P-GPD A LA STATION 8203535	224

PARTIE 1 : SYNTHESE GÉNÉRALE

1 INTRODUCTION GÉNÉRALE

Cette section présente le contexte justificatif des travaux de recherche réalisés. La mise en contexte, les hypothèses et les objectifs y sont donc décrits.

1.1 Mise en contexte

L'analyse fréquentielle (AF) des valeurs extrêmes est un outil privilégié en hydrologie statistique (Meylan et al. 2008) car elle permet de caractériser les séries temporelles des variables à l'aide de lois de probabilités. L'utilisation de ces lois est fondée sur la théorie de la convergence asymptotique (*EVT-Extreme Value Theory* en anglais) de Fisher et Tippett (1928), et a pour finalité l'ajustement d'une distribution de probabilité à un échantillon d'événements extrêmes, en vue de définir les probabilités d'apparition futures de certains événements d'intérêt communément appelés quantiles (Khaliq et al. 2006). Dans son approche classique (AF stationnaire-AFs), les observations d'une variable aléatoire donnée, doivent au préalable valider les hypothèses d'indépendance, d'homogénéité et de stationnarité. Toutefois, la présence de certains types de variabilité dans les séries temporelles de différents processus hydrologiques (exemples : tendance, variation cyclique, dépendance avec d'autres variables explicatives dites covariables, etc.), peut remettre en cause l'hypothèse de la stationnarité, surtout face à la problématique des changements climatiques (CC) (Khaliq et al. 2006). Ainsi, même si l'AFs a toujours ses défenseurs comme on peut le voir par exemple avec les travaux de Serinaldi et Kilsby (2015) ainsi que Serinaldi (2015) qui sonnent l'alarme à propos de l'augmentation des risques d'incertitudes avec son extension au cadre non-stationnaire (AFns), l'AFns se présente néanmoins comme une nécessité. En effet, dans de nombreux autres travaux de recherche, l'idée de la stationnarité est perçue comme une utopie (Salas et Obeysekera 2014; Katz 2013; Chavez-Demoulin et Davison 2012).

Ainsi, la finalité des développements qui intègrent la variabilité naturelle des processus hydro-climatologiques ou celle associée aux CC, visent toujours à proposer des modèles d'AF plus réalistes afin de réduire les erreurs ou, à tout le moins, de mieux les définir de manière quantitative dans les estimations qui leurs sont associées.

Selon les données disponibles et les objectifs de la recherche, deux approches de définition de l'échantillon des évènements extrêmes sont couramment utilisées, celle des blocs de maxima (généralement l'échantillonnage des maxima annuels MA) et celle des dépassements de seuil ou approche POT (de l'expression anglaise *Peaks-Over-Threshold*). Dans le premier cas, c'est la valeur maximale par année qui est sélectionnée pour constituer la série des extrêmes pour la période d'analyse. Dans le second cas, un seuil adéquat doit être fixé à l'avance et l'échantillon des extrêmes est composé par tous les dépassements de seuil indépendants avec la possibilité d'en observer plus d'un par année. Ainsi, chacune de ces deux approches présente des avantages et des limites principalement au niveau de la taille et de la qualité de l'échantillon des extrêmes, des conditions de régularité à remplir (exemple de la validation du critère d'indépendance entre les valeurs extrêmes) mais aussi dans l'automatisation de l'analyse numérique (Bezak et al. 2014). Toutefois, en présence de données à fine échelle (exemple : pas de temps journalier), l'approche POT peut être une alternative à la méthode des MA. En effet, avec cette technique d'échantillonnage, un plus grand échantillon des évènements extrêmes peut être obtenu avec la possibilité de caractériser l'intensité et/ou la fréquence de ces évènements (Lang et al. 1999). L'intensité de l'évènement extrême désigne ici la valeur du dépassement de seuil, tandis que sa fréquence annuelle, représente le nombre de dépassements de seuil observé par année. Cependant, la complexité associée à la sélection du seuil demeure une limite à sa popularité étant donné que la validité des résultats de l'AF avec l'utilisation de cette approche dépend du seuil choisi a priori (Zoglat et al. 2014). Toutefois, des solutions existent pour aider à déterminer la valeur du seuil comme c'est le cas des techniques graphiques et statistiques (Zoglat et al. 2014; Onöz et Bayazit 2001; Lang et al. 1999; Caissie et El-Jabi 1992; Davison et Smith 1990), de l'analyse de sensibilité (Pujol et al. 2007), ou encore de l'utilisation des modèles mixtes (Scarrott et MacDonald 2012). Dans ce dernier cas spécifiquement, c'est une estimation objective du seuil en tant que paramètre à part entière, qui est proposée en comparaison avec les précédentes techniques traditionnelles subjectives.

Les outils d'AF font l'objet de développement continu et de nombreuses applications en hydrologie (AghaKouchak et al. 2013; Khalil et al. 2006; Katz et al. 2002). L'hydrologie se consacre à l'étude du cycle de l'eau dont les précipitations constituent la donnée d'entrée principale. À cet effet, la compréhension et la caractérisation d'une telle variable sont capitales dans les champs d'application de l'hydrologie comme la conception des réservoirs d'eau (Onibon et al. 2004) et la gestion du réseau urbain de drainage (Mailhot et Duchesne 2005). En effet, malgré toute la complexité qui entoure la dynamique des phénomènes hydrologiques, les gestionnaires des ressources hydriques doivent pouvoir continuellement disposer d'outils d'aide à la décision dans l'accomplissement de leurs tâches. Par exemple, la construction et la gestion d'ouvrages hydrauliques (barrages, réseaux d'assainissement, etc.) demande des informations sur des courbes d'intensité-durée-fréquence (IDF) de précipitation ainsi que des estimations des quantiles associés aux observations extrêmes du débit, qui soient les plus précises possibles (Dörte 2013; Soulis et al. 2016). Le quantile est généralement interprété en termes de probabilité au dépassement ou au non-dépassement spécifiée, ou encore pour une période de retour donnée (Rootzén et Katz 2013; Salas et Obeysekera 2014). Quant aux courbes IDF, elles informent sur la récurrence d'un évènement de précipitation d'intensité et de durée donnée (Mailhot et Duchesne 2005).

Les séries de précipitation se caractérisent déjà par une variabilité interannuelle naturelle à la fois déterministe et aléatoire en raison des facteurs physiographiques et des conditions climatiques qui varient dans l'espace et dans le temps (Anctil et al. 2005). Le groupe d'experts intergouvernemental sur l'évolution du climat (GIEC ou IPCC pour *Intergovernmental Panel on Climate Change* en anglais) indique avec un niveau de confiance élevé que les CC en cours vont amplifier cette variabilité (GIEC 2007). La probabilité que la prédiction des précipitations soit davantage complexe apparaît alors grande.

En effet, les CC ont un impact significatif sur les variables climatiques et particulièrement sur leur variabilité et leurs extrêmes dont l'augmentation de la fréquence et de l'intensité peut entraîner des pertes de vies humaines et des dommages matériels et écologiques importants (IPCC 2012). D'ailleurs, l'intensification du cycle hydrologique associée à ces CC (Trenberth et al. 2007), donne lieu à des événements de précipitation de plus en plus extrêmes avec au nombre des conséquences manifestes, des phénomènes d'inondation récurrents et historiques. Un exemple typique des problèmes d'inondations dans les prairies Canadiennes est décrit par Rannie (2016) et Wazney et Clark (2016). En effet, la durée et la répartition spatiale des précipitations dans un bassin versant donné, constituent des facteurs déterminants dans l'occurrence des inondations (Teegavarapu 2012), d'où l'importance d'une bonne caractérisation des précipitations afin de fournir des outils d'aide à la décision pour des aménagements hydrauliques durables.

1.2 Hypothèses et objectifs de la recherche

1.2.1 Hypothèses

Les travaux de recherche réalisés dans le cadre de cette thèse sont basés sur trois hypothèses principales :

(1) Les indices climatiques sont des valeurs numériques qui décrivent des oscillations climatiques de grande échelle. Ils peuvent porter en eux les signaux du CC et aider à expliquer en partie la variabilité dans les séries chronologiques des variables hydrologiques comme la précipitation.

En effet, les CC induits aussi bien par les processus naturels qu'humains, ont des répercussions sur les conditions environnementales régissant les processus hydrologiques à toutes les échelles spatio-temporelles.

(2) Les modèles asymptotiques avec covariables peuvent améliorer les estimations des quantiles pour des prédictions plus réalistes.

Les modèles des valeurs extrêmes ont des bases théoriques très solides permettant d'analyser des phénomènes par la formulation et l'évaluation d'hypothèses statistiques.

(3) Les modèles dits "mixtes" sont une solution au choix subjectif du seuil inhérent à la méthode POT classique.

Ce type de modèles utilise l'ensemble des observations disponible (extrêmes et non-extrêmes), pour développer un mélange de lois statistiques en vue d'une estimation objective du seuil.

1.2.2 Objectifs

L'objectif général de cette thèse était de proposer de nouvelles approches d'AF des extrêmes basées sur la méthode des dépassemens de seuil. Deux objectifs spécifiques s'y rattachent :

- analyser les téléconnexions entre les oscillations climatiques de grande échelle et les processus hydrologiques locaux/régionaux, afin de développer des modèles POT avec covariables.
- utiliser l'ensemble des données disponibles pour développer des modèles hybrides permettant une estimation plus objective du seuil.

2 REVUE DE LITTÉRATURE

Cette section propose une recension des écrits de travaux antérieurs associés à nos objectifs de recherche, afin de mettre en lumière la contribution originale de la thèse.

2.1 Variabilité dans les séries de la précipitation et association avec les changements climatiques

Dans cette thèse, les définitions du glossaire du GIEC (2007) sont adoptées pour les notions de variabilité et changement climatiques. Ainsi, on entend par changement climatique, « *la variation de l'état du climat, que l'on peut déceler (par exemple au moyen de tests statistiques) par des modifications de la moyenne et/ou de la variabilité de ses propriétés et qui persiste pendant une longue période, généralement pendant des décennies ou plus. Les changements climatiques peuvent être dus à des processus internes naturels, à des forçages externes ou à des changements anthropiques persistants dans la composition de l'atmosphère ou dans l'utilisation des terres* ». Quant à la variabilité climatique, elle désigne « *des variations de l'état moyen et d'autres variables statistiques (écart types, phénomènes extrêmes, etc.) du climat à toutes les échelles temporelles et spatiales au-delà de la variabilité propre à des phénomènes climatiques particuliers* ».

Les précipitations ont des distributions spatiales et temporelles très variables en raison des processus hydro-météorologiques propres à chaque région. De nombreuses analyses de séries chronologiques de la précipitation à l'échelle mondiale ont permis de conclure à la présence de tendances significatives dans les données à plusieurs endroits, surtout en ce qui concerne l'intensité et la fréquence de ses observations extrêmes (Easterling et al. 2000). D'ailleurs, comme l'on souligné Barrow et al. (2004), les systèmes écologiques et socio-économiques sont plus vulnérables à ces comportements extrêmes qu'à un changement de l'état moyen du climat.

Alexander et al. (2006) ont par exemple analysé à l'échelle mondiale, les évènements au-dessus des 95^{ème} et 99^{ème} percentiles de la précipitation liquide totale, leurs maxima annuels suivant un cumul journalier et au cinq (05) jours, ainsi que le nombre de jours avec des précipitations >10 mm puis 20 mm. Leurs analyses ont indiqué une augmentation générale significative des précipitations extrêmes sur la période 1951-2003 qui est consistante avec l'observation d'une tendance à l'humidification du climat qu'ils ont faite. Cette récurrence des précipitations extrêmes s'avèrent davantage significative en Amérique du Nord (IPCC 2012), avec une dominance des précipitations liquides sur celles solides surtout dans la partie Est de cette région (Trenberth et al. 2007). À l'échelle spécifique du Canada, les analyses de Thistle et Caissie (2013), Mekis et Vincent (2011), Vincent et Mekis (2006), Zhang et al. (2000; 2001; 2010) concluent également à une humidification et un réchauffement du climat canadien. Par exemple, Zhang et al. (2001) ont indiqué que plus de 71% des précipitations sont tombées sous forme liquide au sud du Canada sur la période 1900-1998, et que c'est ce type de précipitation qui influence principalement les précipitations annuelles extrêmes. Ils ont aussi détecté une variabilité décennale dans la fréquence et l'intensité des précipitations extrêmes. D'ailleurs, d'après Vincent et Mekis (2006), le nombre de jours avec de la précipitation totale et de la précipitation liquide ont connu une augmentation significative de 20 et 30 jours respectivement sur la période 1950-2003. Mekis et Vincent (2011) ont aussi montré que la précipitation totale annuelle ainsi que celle saisonnière ont connu une augmentation significative de 12.5% et 8.7% respectivement à l'échelle du Canada sur la période 1950-2009, et dans sa partie Sud-Est de 1900 à 2009 surtout au printemps et à l'automne.

De nombreuses études ont ainsi associé ces variations consistantes dans le temps, avec les CC en cours depuis le 20^{ème} siècle (GIEC 2007; IPCC 2012). En effet, il existe une corrélation significative entre la température et les processus qui contrôlent le cycle hydrologique (Bates et al. 2008) dont la précipitation est la principale composante. D'ailleurs, une des conséquences des CC est l'intensification de ce cycle particulièrement aux latitudes élevées et dans les régions tropicales entraînant généralement une hausse de la fréquence des précipitations extrêmes (IPCC 2012).

Ces évidences sur la réalité des CC ont ouvert la porte à l'évaluation des interactions qui pourraient exister entre des variables explicatives telles que les indices climatiques et les processus hydro-climatologiques. En effet, ces covariables sont associées aux CC puisqu'elles portent en elles les signaux de la variabilité dans la circulation atmosphérique et océanique (Teegavarapu 2012; Rossi et al. 2011; Trenberth et al. 2007). Obtenus à partir de la normalisation des différences de pression au niveau de la mer entre deux régions proches des centres d'action d'un indice donné (Beranova et Kysely 2015), les indices climatiques sont des séries de valeurs numériques qui permettent de quantifier l'évolution temporelle d'un processus climatique pour une région particulière (Lee et al. 2013). Ils sont également connus sous le nom de téléconnexions puisque ce sont des oscillations atmosphérique et océanique qui se manifestent à grande échelles (Bonsal et Shabbar 2011).

C'est dans ce contexte que de nombreuses recherches ont permis d'établir une relation entre les observations d'évènements extrêmes de la précipitation et différents indices climatiques. Les plus connus sont l'El Nino-oscillation australe (*ENSO-El Nino-Southern Oscillation*), l'oscillation de l'Atlantique Nord (*NAO-North Atlantic Oscillation*), l'oscillation de l'Arctique (*AO-Arctic Oscillation*), la téléconnexion du Pacifique-Amérique du Nord (*PNA-Pacific North America*), l'oscillation décennale du Pacifique (*PDO-Pacific Decadal Oscillation*) et l'oscillation multi-décennale de l'Atlantique (*AMO-Atlantic Multi-decadal Oscillation*). Une description détaillée de ces indices est proposée par Rossi et al. (2011) ainsi que Teegavarapu (2012) qui a schématisé une illustration approximative des centres d'actions des indices AO, NAO, ENSO et PDO (voir son Chapitre 6). Toutefois, des interactions existent également entre ces indices climatiques avec comme conséquence des influences combinées sur les processus affectés. Par exemple, Teegavarapu (2012) souligne que le SOI et l'ENSO ont le même centre de location dans les régions NINO 1-4 avec des valeurs négatives de SOI durant les épisodes El Nino (phase chaude ou positive de ENSO) et des SOI positifs durant les épisodes La Nina (phase froide ou négative de ENSO). Ces derniers s'avèrent également, d'après le même auteur, plus intenses durant les phases négatives de PDO.

Rossi et al. (2011) ont également indiqué qu'il y a une association entre la phase positive (négative) de l'indice PNA et l'occurrence des événements La Nina (El Nino). De plus, leurs analyses ont permis de conclure que la composition spectrale de l'indice PNA est similaire à celle de PDO et plus encore, que l'indice PNA résume les caractéristiques climatologiques des indices SOI et PDO, permettant ainsi à lui seul de représenter le maximum de la variabilité dans le Pacifique Nord. Thompson et Wallace (1998) ont quant eux indiqué que l'indice NAO est une expression régionale de l'indice AO; d'où la forte corrélation qui existe entre ces deux oscillations. D'ailleurs, les analyses de Rossi et al. (2011) ont également relevé la similarité dans les compositions spectrales de ces deux derniers indices.

La bonne compréhension de ces indices et de leurs impacts sur des processus hydrologiques, comme celui des précipitations, est donc un atout pour l'amélioration de la caractérisation de cette composante indispensable du cycle de l'eau. L'intérêt est davantage grandissant avec le défi des CC en cours et la nécessité de disposer d'outils efficaces pour des aménagements hydrauliques durables. Par exemple, Kenyon et Hegerl (2010) ont montré que l'ENSO a des effets sur les précipitations extrêmes à l'échelle mondiale tandis que l'influence de NAO et AO sur ce type d'événements est plus significative respectivement en Europe et au-delà des latitudes moyennes de l'hémisphère Nord. D'après Zhang et al. (2010), l'influence des indices ENSO et PDO est très significative sur les précipitations journalières maximales durant l'hiver en Amérique du Nord. Au Canada, Shabbar et al. (1997) ont également indiqué que les épisodes El Nino et La Nina affectent les précipitations depuis la Colombie-Britannique jusque dans la région des Grands Lacs en passant par les Prairies. Stone et al. (2000) soulignent aussi une influence significative de l'indice PNA sur la fréquence des extrêmes de la précipitation totale journalière en Ontario et au sud du Québec durant l'automne et l'hiver et de façon significative durant la phase négative de PNA. Dans les régions du Nord-Est, Stone et al. (2000) indiquent que c'est plutôt l'intensité des précipitations extrêmes qui est affectée par l'indice NAO dont la phase positive est associée à une diminution des événements de précipitation extrême au Nord du Québec et du Labrador ainsi que dans la baie d'Hudson en hiver, puis au sud du Québec durant l'été.

Coulibaly (2006) associe le changement de régime des précipitations saisonnières au Canada observé autour de l'année 1940, à celui de l'indice AO qui avait montré une tendance d'occurrence de phases négatives plutôt extrêmes. Quant à Bonsal et Shabbar (2011), ils ont indiqué qu'à des épisodes El Nino (La Nina), sont associés des précipitations inférieures (supérieures) à la moyenne durant l'hiver depuis le Sud de la Colombie Britannique jusque dans les Prairies mais aussi dans la région des Grands Lacs. La même réponse des précipitations hivernales est observée durant les phases positives (négatives) de PDO dans l'Ouest Canadien, celles de PNA en Alberta, puis celles ce NAO au Nord-Est du Canada.

Ainsi, analyser statistiquement les interactions entre les indices climatiques et des variables hydro-climatologiques comme les précipitations, puis comprendre et décrire les mécanismes physiques qui les génèrent, c'est disposer d'outils robustes pour améliorer la prédiction des quantiles d'évènements extrêmes d'intérêt.

2.2 Distributions asymptotiques

L'approche POT et celle des MA, offrent des méthodes de modélisation des extrêmes asymptotiquement équivalentes. En effet, Naveau et al. (2005) expliquent qu'à mesure que la taille d'un échantillon augmente, la distribution des MA converge vers une distribution de Fréchet, de Weibull ou de Gumbel. La distribution des valeurs extrêmes généralisée (GEV-Generalized Extreme Value en anglais) formulée par Jenkinson (1955), incorpore chacune de ces lois. Sa fonction de distribution cumulative (fdc) est donnée par l'équation (1).

$$F(y, \xi, \beta, \psi) = \begin{cases} \exp \left[- \left(1 - \frac{\xi}{\beta} (y - \psi) \right)^{\frac{1}{\xi}} \right] & \xi \neq 0; \beta > 0 \\ \exp \left[- \exp \left(- \frac{(y-\psi)}{\beta} \right) \right] & \xi = 0; \beta > 0 \end{cases} \quad (1)$$

avec y représentant le vecteur des MA, ξ , β et ψ les paramètres de forme, d'échelle et de position respectivement.

En suivant le même raisonnement, les dépassements de seuil indépendants ont comme limite asymptotique la distribution de Pareto généralisée (*GPD-Generalized Pareto Distribution* en anglais) dont la fdc est formulée par Pickands (1975) comme suit à l'équation (2).

$$G(x, \xi, \sigma, u) = \begin{cases} 1 - \left[1 - \xi \left(\frac{x-u}{\sigma} \right)^{\frac{1}{\xi}} \right] & (x-u) \geq 0; \xi \neq 0; \sigma > 0 \\ 1 - \exp \left[- \left(\frac{x-u}{\sigma} \right) \right] & (x-u) \geq 0; \xi = 0; \sigma > 0 \end{cases} \quad (2)$$

avec x et $x - u$ désignant respectivement l'intensité du dépassement et sa magnitude; ξ, σ et u représentent les paramètres de forme, d'échelle et de position (c'est-à-dire le seuil) respectivement. Ainsi, le quantile stationnaire des extrêmes associé à un risque d'occurrence donné, est obtenu à partir de l'inverse de la fdc de la distribution d'ajustement utilisée.

Dans cette thèse, le terme intensité est utilisé pour caractériser la pointe de chaque dépassement de seuil, tandis que le nombre de dépassements de seuil observés par année forme la distribution de la fréquence des dépassements de seuil. La modélisation des extrêmes par la méthode des MA demeure la plus populaire. Cependant, cette approche consiste à ne retenir que la plus grande valeur annuelle de la série chronologique à analyser. Les observations étant déjà limitées dans le cadre des valeurs extrêmes, cette approche peut donc entraîner une perte importante d'information. En effet, les MA donnent lieu à un échantillon réduit surtout dans le cas de séries de courtes durées d'enregistrement, sans oublier qu'ils ne constituent pas forcément des événements extrêmes, ou que plus d'un événement extrême peuvent être observés par année (Bezak et al. 2014). De plus, dans le cas spécifique des précipitations qui se caractérisent par un cycle annuel, l'AF avec la méthode des MA pourrait essentiellement refléter celle d'une saison particulière (Zhang et al. 2001). Ainsi, toute autre approche permettant d'augmenter la taille de l'échantillon des extrêmes et de réduire les incertitudes dans les estimations, est à privilégier. La méthode POT se présente alors comme une alternative intéressante à l'approche des MA.

En effet, avec l'approche POT, tous les évènements dont la valeur est supérieure ou égale à un certain seuil peuvent être considérés (Ribatet et al. 2009; Khaliq et al. 2006) ou seulement les évènements indépendants. Pour assurer l'hypothèse d'indépendance de la série des extrêmes ainsi constituée, la technique du « *declustering* », qui consiste à imposer un temps de séparation entre des dépassements de seuil consécutifs, est généralement utilisée. Par exemple, Roth et al. (2012) ont suggéré un temps de séparation de 1 à 2 jours entre les dépassements de seuil pour un jeu de données constitué par des précipitations journalières. Une des conséquences du « *declustering* » de ces évènements extrêmes, est qu'elle rend la GPD et la GEV équivalente (Chavez-Demoulin et Davison 2012). Néanmoins, l'échantillon POT est généralement plus réaliste et de plus grande taille que celui des MA, et deux processus, l'intensité et la fréquence des dépassements de seuil, peuvent être examinés séparément ou conjointement (Katz et al. 2002; Lang et al. 1999), d'où une compréhension plus complète des phénomènes étudiés.

Ainsi, de même que la distribution de l'intensité des dépassements de seuil a la GPD comme limite asymptotique, de même, la distribution de la fréquence (m) de ces dépassements est généralement considérée comme suivant asymptotiquement une distribution de Poisson (PD) (Lang et al. 1999). Sinon, les distributions Binomiale (BD) et Binomiale Négative (BND) sont des alternatives à considérer lorsque le processus de Poisson ne peut être théoriquement validé (Bhunya et al. 2013; Eastoe et Tawn 2010; Önöz et Bayazit 2001; Lang et al. 1999). Les équations (3), (4) et (5) indiquent les fonctions de masse des lois PD, BD et BND (Önöz et Bayazit 2001).

$$PD(m = k) = \frac{e^{-\lambda} \lambda^k}{k!}, \quad k \in \mathbb{N} \quad (3)$$

$$BD(m = k) = \binom{r}{k} p^k (1-p)^{r-k}, \quad k \in \mathbb{N} \quad (4)$$

$$BND(m = k) = \binom{r + k - 1}{k} p^k (1-p)^{r-k}, \quad k \in \mathbb{N} \quad (5)$$

avec $\binom{r}{k} = [\frac{r!}{k!(r-k)!}]$ le nombre de combinaisons étant donné r et k ; λ désigne l'espérance de m ; p représente la probabilité d'avoir un dépassement de seuil (succès); r équivaut au nombre total des dépassements de seuil.

En effet, la PD n'est pas toujours la plus appropriée pour la modélisation statistique de la fréquence des dépassements de seuil même si le critère de leur indépendance indispensable à la convergence asymptotique, est vérifié. Cunnane (1979) a proposé une approche de validation statistique du processus de Poisson par le calcul de l'indice de dispersion I_t et d'une statistique t correspondant à cet indice. Les expressions de ces coefficients sont formulées à l'équation (6) (Önöz et Bayazit 2001) :

$$I_t = \frac{\text{Variance (m)}}{\text{Moyenne (m)}} \quad \text{et } t = (n - 1)I_t \quad (6)$$

La statistique t suit asymptotiquement une distribution du Khi-Carré à $(n-1)$ degrés de liberté, avec n le nombre d'années dans le cas d'analyse à une échelle annuelle. En partant de l'hypothèse « la vraie distribution suit une loi de Poisson » équivaut à « $I_t = 1$ », le processus de Poisson n'est pas rejeté si t appartient à la zone d'acceptation ($\chi^2_{\alpha/2}, \chi^2_{1-\alpha/2}$) avec α , le niveau de signification du test. Si $t < \chi^2_{\alpha/2}$, la distribution B est suggérée, et si $t > \chi^2_{1-\alpha/2}$, la distribution BN est plus appropriée.

La modélisation de l'intensité et de la fréquence des dépassements de seuil peut être faite séparément ou de façon couplée. Dans ce dernier cas, on fait appel à la théorie du processus ponctuel (PP) qui permet de combiner la GPD et la PD en un processus de Poisson non-homogène (Khaliq et al. 2006; Katz et al. 2002). De cette façon, une correspondance peut être établie entre ce processus et la distribution GEV (Lang et al. 1999). Une telle combinaison permet d'obtenir des quantiles interprétables en termes de période de retour, un concept très utilisé dans l'ingénierie appliquée. Pour ce faire, les paramètres de la distribution GEV sont exprimés en fonction des paramètres de PD et GPD, permettant ainsi d'obtenir des quantiles de MA mais à partir d'un échantillon POT. L'équation (7) présente cette reformulation des paramètres d'échelle et de position de la GEV, le paramètre de forme étant le même que celui de la GPD (Silva et al. 2016) :

$$\begin{cases} \beta = \sigma/\lambda^\xi \\ \psi = u + \beta \left(\frac{\lambda^{\xi-1}}{\xi} \right) \end{cases} \quad (7)$$

On peut également montrer que si la valeur du seuil est suffisamment grande, les équations (1) et (2) peuvent être combinées autrement en une seule distribution H (Coles, 2001), pour tout $x > u$ telle que présentée par l'équation (8) :

$$H(x) = \begin{cases} 1 - \left(1 - \frac{\xi(x-u)}{\beta}\right)^{\frac{1}{\xi}} & \text{si } \xi \neq 0 \\ 1 - \exp\left(-\frac{(x-u)}{\beta}\right) & \text{si } \xi = 0 \end{cases} \quad (8)$$

avec $\beta = \sigma + \xi(u - \psi)$

Les approches POT et AM sont ainsi asymptotiquement équivalentes.

2.3 Détermination du seuil en approche POT

La détermination de la valeur du seuil joue un rôle capital dans la validité des résultats de l'AF. Dans l'approche POT traditionnelle qui consiste à fixer le seuil à l'avance et à effectuer la modélisation à partir de la distribution des dépassements de seuil indépendants, le choix du seuil est très souvent fait de façon arbitraire comme c'est le cas lorsqu'on évalue différents percentiles généralement entre la médiane et le 99^{ème} percentile (Serinaldi et Kilsby 2014; Osman et al. 2013; Begueria et al. 2011; Sugahara et al. 2009; Pujol et al. 2007). Le 90^{ème} percentile est par exemple un choix conservateur mais, quand on travaille avec des données de précipitations journalières, on est amené à utiliser des valeurs supérieures au 95^{ème} percentile comme seuil et souvent, le 99^{ème} percentile est le plus adapté (Osman et al. 2013). Pujol et al. (2007) ont par exemple choisi la médiane des observations comme seuil pour l'étude des précipitations en France et ont testé la sensibilité des résultats à d'autres valeurs de seuil pour pallier aux éventuelles erreurs d'estimation liées à la valeur du seuil fixé. Pour ce faire, ils ont créé de nouvelles séries de dépassements de seuil, en abaissant et en augmentant les seuils de 10% et 20%. Cette analyse de la sensibilité des estimations à la variation du seuil, a en effet montré des différences significatives au niveau du calcul des quantiles.

Une autre façon de valider le choix du seuil est d'utiliser des techniques graphiques (Coles 2001; Lang et al. 1999 ; Davison et Smith 1990), en représentant d'une part l'évolution de la magnitude moyenne des dépassemens de seuil en fonction d'un vecteur de valeurs croissantes du seuil (*mean excess plot* en anglais) et d'autre part, l'évolution des paramètres de forme et d'échelle de la GPD en fonction du même vecteur. Quant à Caissie et El-Jabi (1992), ils suggèrent une estimation du seuil sur la base de l'indice de dispersion. Les valeurs de seuil pour lesquelles ce rapport sera très proche ou égale à 1, peuvent ainsi être choisies comme intervalle de seuils optimaux. Zoglat et al. (2014) propose à leur tour une gamme de méthodes statistiques combinée à des techniques graphiques pour le choix du seuil. Lang et al. (1999) recommandent d'identifier d'abord un intervalle de valeurs de seuil qui donnent de bon résultats avec le *mean excess plot* et l'indice de dispersion; ensuite, ils suggèrent de choisir dans cet intervalle la plus grande valeur du seuil pour laquelle le nombre moyen de pics par période est supérieur à 2 ou 3 (si on est intéressé à avoir un nombre moyen de pics sur une période donnée comme l'année par exemple).

Ainsi, il n'y a pas un consensus pour le choix du seuil. Alors, comme l'ont suggéré Davison et Smith (1990), il faut varier les valeurs de plus en plus grandes du seuil jusqu'à obtenir le seuil optimal. Plusieurs seuils peuvent ainsi répondre aux conditions de régularité requises pour la modélisation statistique, alourdisant l'étape des analyses préliminaires surtout avec de grandes bases de données. Néanmoins, cette étape est importante car la valeur du seuil contrôle la taille de l'échantillon des extrêmes, peut affecter le critère d'indépendance des dépassemens (si la valeur est trop faible) mais aussi la convergence vers la GPD et donc la qualité des estimations (Bezak et al. 2014 ; Scarrott et MacDonald 2012 ; Coles 2001 ; Lang et al. 1999).

Les modèles "mixtes" avec une composante GPD, peuvent permettre une estimation explicite du seuil (Behrens et al. 2004 ; MacDonald et al. 2011). Leur développement est encore timide mais prometteur. En effet, avec ce type de modèles, il est possible d'estimer le seuil tout comme les autres paramètres du modèle GPD, ce qui permet de quantifier également l'incertitude liée à l'estimation du seuil le long du processus de l'inférence.

Les modèles mixtes ont deux composantes, la distribution centrale constituée par les observations en-dessous du seuil et fréquemment ajustée par une loi de probabilité paramétrique, et la distribution des extrêmes pour les observations au-dessus du seuil (Scarrott et MacDonald 2012). À la faveur de la EVT, cette deuxième composante est généralement modélisée par la GPD ou le modèle couplé PD-GPD. Le seuil est alors déterminé au point de jonction des deux composantes du modèle "mixte". L'intérêt dans l'utilisation d'un tel modèle est qu'il permet de considérer toutes les observations disponibles et d'estimer le seuil de façon objective. Toutefois, sa robustesse dépend essentiellement de la qualité du choix des distributions pour la partie centrale. L'approche la plus simple dans le développement de ce type de modèle consiste à considérer des distributions paramétriques pour la 1^{ère} composante.

Des travaux antérieurs comme ceux de Frigessi et al. (2002), Behrens et al. (2004), Carreau et Bengio (2009) ont ainsi suggéré l'utilisation des distributions Weibull, Gamma et Normale pour l'ajustement sous le seuil. Un des problèmes avec ce genre de paramétrisation est la discontinuité marquée au point du seuil avec un risque significatif d'affecter la position du seuil en raison du degré de liberté supplémentaire qui est associé à ce paramètre (MacDonald et al. 2011). Frigessi et al. (2002) avaient ainsi introduit l'utilisation d'une fonction de transition entre ces deux composantes, la distribution de Cauchy, afin d'assurer la continuité au point du seuil. Toutefois, cette fonction ne garantit pas la continuité au niveau pratique et en plus, elle peut conduire à une dominance de la GPD sur la distribution centrale choisie (MacDonald 2012). Carreau et Bengio (2009) ont proposé une autre approche pour corriger la discontinuité au point du seuil en imposant la continuité de la fonction de densité du modèle "mixte" ainsi que celle de sa dérivée. Cependant, la formulation de ces deux contraintes entraîne une dépendance entre les paramètres du modèle "mixte", ce qui pourrait affecter la robustesse du modèle tout comme avec l'utilisation de la distribution Cauchy (Scarrott et MacDonald 2012).

Le choix a priori des distributions centrales n'apportant pas de flexibilité, d'autres études ont proposé des approches semi-paramétriques et non-paramétriques. Déjà, le modèle de Carreau et Bengio (2009) se démarque de l'approche paramétrique avec les contraintes de continuité formulées. Il peut donc être considéré comme un modèle "mixte" semi-paramétrique tout comme celui de Do Nascimento et al. (2012). Ces derniers ont proposé une généralisation du modèle de Behrens et al. (2004) en utilisant comme distribution de la première composante du modèle, n fonctions de densité de Gamma suivant une méthode de pondération. L'approximation avec $n=2$ a donné la meilleure performance dans l'étude de Do Nascimento et al. (2012). Des approches non-paramétriques sont également proposées pour davantage de flexibilité dans la définition de la distribution centrale. C'est le cas de Tancredi et al. (2006) qui ont utilisé une combinaison d'un nombre indéfini a priori de densités uniformes et de MacDonald et al. (2011) qui ont considéré la fonction de densité de Kernel pour la partie non-extrême de leurs modèles "mixtes" respectifs. Toutefois, autant l'approche paramétrique peut conduire au développement de modèles mixtes peu ou pas robustes, autant les approches semi-paramétrique et non-paramétrique entraînent généralement davantage de complexité au niveau de l'implémentation avec une absence de solution explicite lorsque des méthodes numériques sont utilisées. D'où la nécessité d'améliorer déjà la performance des modèles simples existants.

2.4 Méthodes d'estimation des paramètres

En AF, la méthode du maximum de vraisemblance (MV) est la plus populaire pour l'estimation des paramètres en raison de sa simplicité et de sa flexibilité (Coles 2001). Toutefois, ses estimateurs sont asymptotiquement efficients sous certaines conditions de régularité surtout en ce qui concerne le paramètre de forme des distributions extrêmes (El Adlouni et Ouarda 2008). D'où sa généralisation par la définition d'une loi a priori permettant de restreindre les valeurs du paramètre de forme des distributions comme la GEV (Martins et Stedinger 2000 ; El Adlouni et al. 2007) et la GPD (Martins et Stedinger 2001) dans l'intervalle [-0.5 ; +0.5]. Ces valeurs sont en effet les plus admissibles en hydrologie (Martins et Stedinger 2000).

Cette méthode dite du maximum de vraisemblance généralisée (MVG) est ainsi une approche Bayésienne où l'information a priori est apportée seulement pour le paramètre de forme de la distribution des extrêmes. Elle a montré son efficacité par rapport aux méthodes classiques d'estimation des paramètres comme la méthode du MV, celle des moments et des L-moments (El Adlouni et Ouarda 2008). L'inférence Bayésienne se présente alors comme une alternative à toutes ces méthodes, lorsque des informations a priori sont disponibles sur un ou plusieurs paramètres d'un modèle (Chu et Zao 2011). De plus, cette approche d'estimation est généralement sophistiquée grâce à la puissance des méthodes de Monte Carlo par Chaînes de Markov (MCMC) dont l'algorithme de Metropolis-Hastings ou de Gibbs (Hastings 1970 ; Metropolis et al. 1953) sont les plus utilisés. Le cadre Bayésien permet également d'obtenir la distribution a posteriori de chacun des paramètres estimés et d'évaluer l'incertitude associée aux estimations le long du processus de l'inférence statistique. D'ailleurs, cet avantage conduit généralement dans le cas des modèles "mixtes", à avoir une fonction de densité de probabilité a posteriori continue, tous les seuils a posteriori possibles étant naturellement considérés (Scarrott et MacDonald 2012 ; Do Nascimento et al. 2012).

2.5 Analyse fréquentielle non-stationnaire des extrêmes

Dans le contexte actuel des CC, le développement des modèles d'AF tente d'intégrer le maximum d'informations possible pour réduire les erreurs d'estimation des quantiles. En effet, cette estimation est d'une grande utilité pour l'étude des phénomènes hydrologiques, la gestion adéquate des ressources en eau et la conception des ouvrages hydrauliques. C'est dans ce contexte que l'hypothèse de stationnarité des variables aléatoires de différents processus hydrologiques peut très souvent être révoquée en raison de la présence de tendances dans les séries de données ou encore de la dépendance entre une variable d'étude et une ou plusieurs covariables. L'incorporation de ces éléments de non-stationnarité, est à la base du passage de l'AFs à l'AFns (Coles 2001).

L'AFns permet de tenir compte des changements dans la distribution des données, conditionnels au temps, à des indices climatiques ou d'autres variables explicatives (Katz et al. 2002). Les fonctions de dépendance les plus utilisées entre variable et covariables sont les fonctions polynomiales (El Adlouni et al. 2007). Chacun des paramètres d'une distribution ajustée comme la GEV ou la GPD, peut ainsi devenir une variable à laquelle est associée une ou plusieurs covariables au moyen d'une fonction linéaire, quadratique, cubique, etc. De nombreux développements de ce type d'approche non-stationnaire sont proposées avec par exemple la GEV (El Adlouni et Ouarda 2009; El Adlouni et Ouarda 2008; El Adlouni et al. 2007; Katz et al. 2002; Coles 2001), la GPD (Mondal et Mujumdar 2015; Osman et al. 2013; Begueria et al. 2011; Pujol et al. 2007), et la Log-Normale (Aissaoui-Fqayeh et al. 2009). Des variations de type cyclique sont également étudiées au moyen des fonctions trigonométriques comme c'est le cas avec les travaux de Mailhot et al. (2013) et de Sugahara et al. (2009). Toutes ces fonctions de dépendance fixées a priori sont de type linéaire offrant ainsi moins de flexibilité. D'où l'intérêt grandissant pour les combinaisons linéaires de fonctions polynomiales qui permettent d'évaluer aussi bien les dépendances linéaires que non-linéaires.

C'est le cas de l'utilisation des fonctions splines dites naturelles avec la GPD dans l'étude de Chavez-Demoulin et Davison (2005) et des fonctions splines de base appelées B-splines avec la distribution GEV dans les travaux de Padoan et Wand (2008), puis Nasri et al. (2013). En effet, de telles structures de dépendance semi-paramétriques se sont montrées plus flexibles et adaptées pour la modélisation non-stationnaire des extrêmes (Chavez-Demoulin et Davison 2012) que celles paramétriques qui sont toutefois encore très populaires. Par exemple, des études de cas ont révélé que la réponse de certaines variables hydro-climatologiques aux oscillations atmosphériques et océaniques de basse fréquence, peut être différente selon les phases négatives et positives des indices climatiques qui les caractérisent (Chandran et al. 2015; Canon 2015; Shabbar et al. 1997). De telles fonctions semi-paramétriques permettraient alors de capturer ce type de dépendance complexe.

3 MATÉRIELS ET MÉTHODES PROPOSÉS

3.1 Matériels d'application

3.1.1 Variable hydrologique

L'application des théories mathématiques et statistiques en hydrologie est un atout pour la compréhension physique de plusieurs processus hydrologiques (Katz et al. 2002). La caractérisation de la précipitation qui constitue la donnée d'entrée principale du cycle de l'eau et donc des processus hydrologiques, est à cet effet capitale d'autant plus que cette variable est affectée par une variabilité spatiale et temporelle naturelle (Anctil et al. 2005). Sa prévision peut être complexe et plus encore dans le contexte actuel des CC (IPCC 2012). D'où le choix des précipitations comme variable d'étude.

De façon plus précise, l'intérêt s'est porté sur la précipitation totale journalière associée au code 012 dans la base de données des relevés officiels des stations du Service Météorologique du Canada. Cette variable représente la somme de la pluie totale et de l'équivalent en eau de la neige totale enregistrée quotidiennement. Ces données climatiques historiques sont disponibles pour différentes variables au canada (http://climat.meteo.gc.ca/historical_data/search_historic_data_f.html) et avaient été mises à la disposition des chercheurs dans le cadre du projet FACE via un portail virtuel (<http://loki.qc.ec.gc.ca/DAI/login-e.php>) à partir duquel la base de données pour notre recherche a été constituée en 2013. Toutefois, ces données sont différentes des données climatiques canadiennes ajustées et homogénisées qui ont été corrigées et qui sont disponibles au pas de temps mensuel.

Les évènements extrêmes des séries temporelles de la précipitation totale journalière inventoriées, sont constitués par les dépassements de seuil indépendants. Deux variables ont fait l'objet d'intérêt à savoir l'intensité des dépassements de seuil (V1) et leur fréquence annuelle (V2).

3.1.2 Covariables climatiques

Les indices climatiques peuvent permettre de décrire l'influence des CC sur plusieurs processus hydro-climatologiques. Ils sont alors très utilisés comme covariables dans plusieurs approches d'analyse de régression, d'où le choix de ce type de covariable à la place du temps. Ce dernier est naturellement un facteur explicatif de la présence de variabilité dans les séries d'observations de nombreuses variables stochastiques.

En raison de l'impact reconnu des indices ENSO / SOI, NAO, AO, AMO, PDO, PNA sur le climat canadien (Bonsal et Shabbar 2011), ces six indices ont été considérés dans le cadre de cette recherche. Toutefois, l'humidité étant à l'origine de la formation des précipitations, un septième indice climatique a été considéré, celui du *Tropical Western Hemisphere Warm Pool* en anglais (WHWP). En effet, il existe dans la région du Mexique, ce qu'on peut appeler une « piscine chaude naturelle » (la température de l'eau y étant au-dessus de 28.5°C), désignée comme étant la deuxième plus grande source d'humidité sur la Terre après celle du Pacifique (Wang et Enfield 2001). Cette source s'étend depuis l'Est du Pacifique Nord aux Caraïbes en passant par le Golfe du Mexique. Drumond et al. (2011) ont montré que cette source d'humidité apporte une contribution significative aux régimes de précipitation de l'Amérique du Nord surtout à partir du mois de Juin.

Les séries chronologiques de chacun de ces sept indices consistent en des valeurs mensuelles standardisées et sont proposées par la Division des Sciences Physiques de la NOAA (*National Oceanic and Atmospheric Administration*) puis téléchargeables à partir de leur site web (<https://www.esrl.noaa.gov/psd/data/climateindices/list/>). Dans le cadre de notre recherche les périodes qui avaient été considérées en 2013 sont 1856-2013 pour AMO, 1948-2013 pour PDO et WHWP, 1950-2013 pour AO, NAO et PNA, puis 1951-2013 pour SOI. Toutefois, une période commune a été définie selon les besoins d'analyse des variables V1 et V2 avec les indices. Cette période variait donc souvent d'un indice à un autre et en fonction des stations d'observations de la précipitation totale journalière considérées.

3.1.3 Région d'étude

L'application des modèles en hydrologie statistique est généralement faite avec des données réelles relevées à une ou plusieurs stations. En vue d'évaluer les modèles développés dans le cadre de cette recherche, à une échelle canadienne qui ne soit ni trop spécifique (exemple : Assani et al. 2008), ni trop générale (exemple : Coulibaly 2006), toute la région du Sud-Est du Canada a été initialement considérée comme étude de cas. Cette région comprend les Maritimes (c'est-à-dire les provinces du Nouveau-Brunswick (NB), de L'Île-du-Prince-Édouard (IPE ou PEI pour *Prince-Edward-Island*) et de la Nouvelle Écosse (NE ou NS pour *Nova Scotia*), ainsi que les provinces de Terre-Neuve (TN ou NF pour *New-Foundland*) & Labrador (L), du Québec (QC) et de l'Ontario (ON). La figure 1 illustre le cadre géographique de cette zone d'étude.

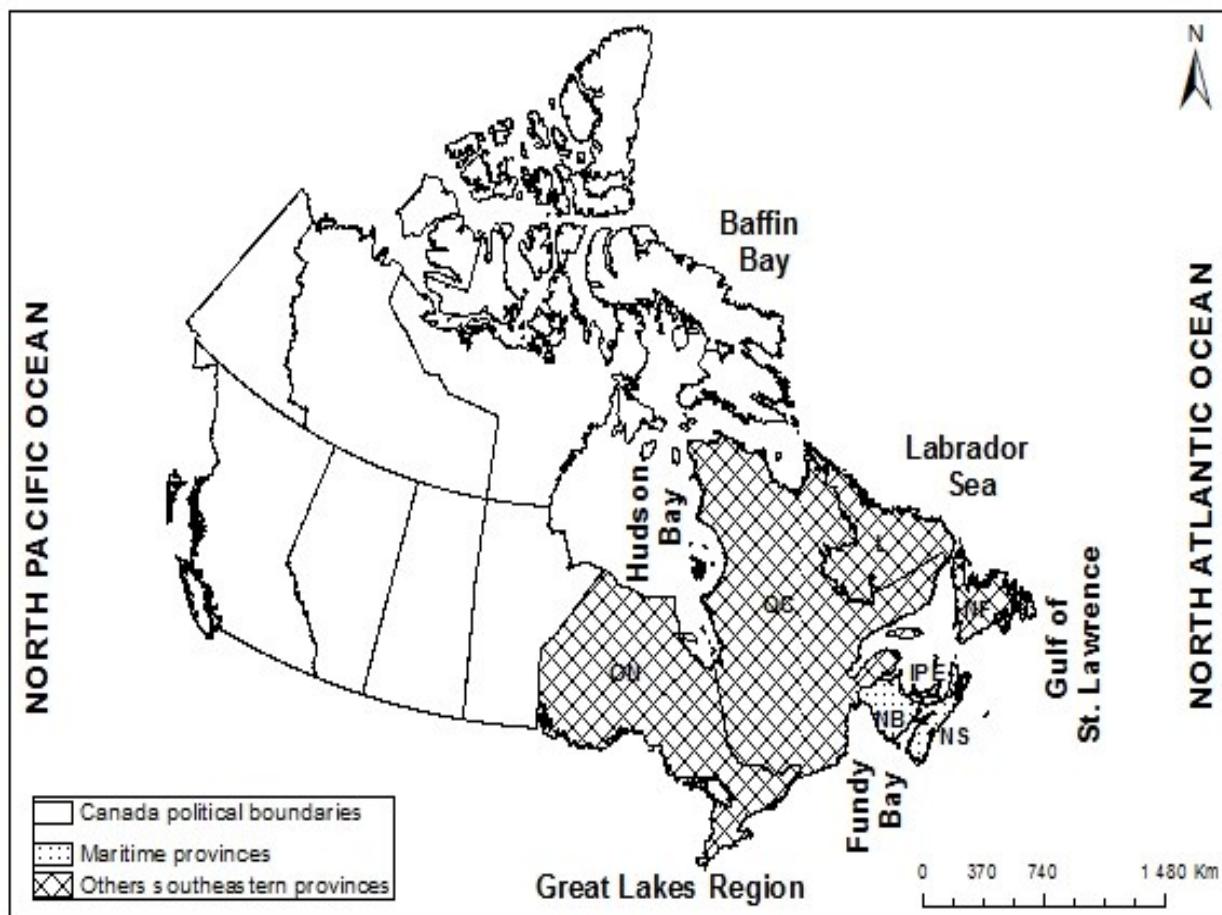


Figure 1: Cadre géographique de la zone d'étude

L'influence des indices climatiques sur les variables hydro-climatologiques (températures, précipitations, débits) est très bien documentée pour la région Ouest-Canadienne car leurs effets y sont plus significatifs et spatialement cohérents (Tremblay et al. 2011 ; Bonsal et Shabbar 2011 ; Bonsal et Shabbar 2008 ; Coulibaly 2006 ; Anctil et Coulibaly 2004 ; Coulibaly et Burn 2004 ; Bonsal et Shabbar 2004 ; Bonsal et al. 2001 ; Shabbar et al. 1997 ; Shabbar et Khandekar 1996), surtout en ce qui concerne la température. D'où la nécessité et l'intérêt de la présente recherche à l'échelle du Sud-Est du Canada.

Ainsi, un inventaire de toutes les stations d'observations de la précipitation totale journalière a été effectué permettant l'extraction de 1057 séries de données sur la période 1945-2013. Toutefois, la longueur des séries est différente d'une station à une autre pour des raisons liées entre autres, à la date d'implémentation des stations, à leur maintien mais aussi à la qualité des données qui y sont relevées (présence de données manquantes et/ou aberrantes). Afin de disposer d'une base de données qui réponde aux recommandations de l'Organisation Météorologique Mondiale (OMM), le critère selon lequel chaque série de données doit avoir au moins 30 années complètes d'observations, que ces années soient consécutives ou non, a été formulé dans cette thèse. Avec ce prétraitement des données originales, le nombre de stations est passé de 1057 à 329.

3.2 Développements méthodologiques

3.2.1 Interactions entre les précipitations et les covariables climatiques

L'analyse de corrélation non-paramétrique et l'analyse par ondelettes sont les deux approches adoptées dans cette recherche pour évaluer l'influence des sept indices climatiques identifiés (AO, NAO, AMO, PDO, PNA, SOI, WHWP), sur l'intensité et la fréquence des extrêmes de la précipitation totale journalière à l'échelle du Sud-Est du Canada. Le tau de Kendall est calculé pour chaque paire de jeu de données avec un niveau de signification de 5%. Une telle analyse de corrélation basée sur les rangs est une alternative à la régression linéaire surtout lorsqu'on travaille avec des valeurs extrêmes d'observations hydro-climatologiques qui ne sont généralement pas normalement distribuée (Chen et al. 2012 ; Yue et al. 2002). Toutefois, étant donné que cette méthode d'analyse détecte seulement des corrélations de types monotones, la transformée en ondelette qui permet d'en étudier la périodicité également (Mabille et al. 2012 ; Grinsted et al. 2004 ; Torrence et Combo 1998), a été utilisée comme outil d'analyse complémentaire. En effet, l'analyse par ondelettes est devenue très populaire pour étudier la variabilité dans les séries de données elles-mêmes (avec le spectre d'échelle ou *continuous wavelet transform-CWT* en anglais) ou entre deux séries (avec la transformée en ondelettes croisée ou *cross wavelet transform-XWT* et la cohérence de la transformée en ondelettes ou *wavelet transform coherence-WTC*). D'où sa popularité grandissante dans plusieurs études (Chandran et al. 2015 ; Ouachani et al. 2011 ; Rossi et al. 2011 ; Coulibaly 2006 ; Anctil et Coulibaly 2004 ; Coulibaly et Burn 2004).

Les indices climatiques décrivent la dynamique climatique à grande échelle tandis-que les variables hydrologiques caractérisent des évènements à plus petite échelle. Le lissage des observations est alors souvent employé pour maximiser les résultats des corrélations entre variables d'intérêt (Rossi et al. 2009 ; Assani et al. 2008). Afin d'optimiser les résultats des interactions entre chaque indice climatique Z et les variables d'intensité (V1) et de fréquence annuelle (V2) des dépassements de seuil journaliers indépendants, une agrégation des différents de jeu de données a été faite.

- Moyenne annuelle de V1;
- Moyenne annuelle de Z;
- Moyenne annuelle mobile aux 3 mois de Z avec comme fenêtre de départ, la période de Décembre à Février (DJF) et comme dernière fenêtre, celle de Novembre à Janvier (NDJ);

Etant donné que plusieurs stations d'observation de la précipitation sont impliquées, les indices climatiques ont été ordonnés à chaque station selon la valeur significative du tau de Kendall résultant de la corrélation de leurs différentes fenêtres de données (au nombre de 13) avec les séries V1 et V2. La finalité était d'identifier à chaque station l'indice le plus corrélé aux variables V1 et V2. Ainsi, pour chaque indice, on a retenu d'abord seulement la fenêtre à laquelle était associée la valeur maximale du tau de Kendall significatif. Ensuite, à chaque station, le rang 1 (donné ici au meilleur indice) a été attribué à l'indice avec la plus grande valeur significative du tau de Kendall et ainsi de suite. A cette première classification basée sur les plus fortes corrélations significatives, a été ajoutée une seconde classification où on n'a tenu compte que de la significativité de la corrélation détectée entre les indices et les variables V1 et V2. En effet, la finalité dans l'évaluation de ces interactions entre variables et covariables à l'échelle de la zone d'étude était de voir s'il y a un patron climatique régional qui se démarquait pour permettre de proposer des modèles d'AFns spatialement cohérents.

3.2.2 Modèle GPD non-stationnaire

Le premier modèle qui a été développé dans le cadre de cette thèse est la GPD avec l'incorporation de la non-stationnarité au niveau de son paramètre d'échelle. En effet, la forme classique d'incorporation de la non-stationnarité en approche POT consiste à garder le seuil constant et à modéliser la fréquence des dépassements de seuil indépendants par un processus de Poisson non-homogène et leur intensité par une GPD aux paramètres non-stationnaires. Toutefois, le paramètre de forme comme d'autres étude a été gardé constant afin de ne pas augmenter les incertitudes dans son estimation (Pujol et al. 2007 ; Sugahara et al. 2009 ; Coles 2001), étant donné que c'est un paramètre très déterminant dans la caractérisation du changement dans les variables comme la précipitation (Osman et al. 2013).

Pour tenir compte des récents développements au niveau des structures de dépendance utilisées en analyse non-stationnaire (Nasri et al. 2013 ; Padoan et al. 2008), les fonctions B-splines ont été considérées pour établir la relation de dépendance entre le paramètre d'échelle (σ) de la GPD (Equation 2) et la meilleure covariable climatique Z (c'est-à-dire la plus corrélée) choisie parmi les sept indices climatiques sélectionnés. En effet, ces fonctions sont des combinaisons linéaires de K (avec $K=k+m$) fonctions B-splines de degré d (ou d'ordre $m=d+1$) avec k nœuds internes (Racine 2014). L'application de la spline nécessite la définition d'un vecteur de points de contrôles t_i (de Boor 2001) tels que $t_0 \leq t_1 \leq \dots \leq t_{k+1}$, avec t_0 et t_{k+1} les noeuds d'extrémité additionnels. La fonction de dépendance B_s entre le paramètre d'échelle σ de la GPD et la covariable Z peut alors être définie comme à l'équation (9) :

$$\sigma_z = B_s(Z) = \sum_{i=0}^{k+j} \beta_i B_{i,j}(z) \quad z \in [t_0, t_{k+1}] \quad (9)$$

avec β_i les hyperparamètres de σ_z ; i désigne les points de contrôle avec une dimension de 1 si l'ordre j ($j=0, 1, 2, \dots, d$) de la fonction spline de base égale 0, et K dans les cas où $j>0$.

La fonction B-spline $B_{i,j}(z)$ d'ordre j et de dimension i , peut s'écrire comme donnée à l'équation (10) :

$$\begin{cases} B_{i,0}(z) = & \begin{cases} 1 & \text{si } t_i \leq z \leq t_{i+1} \\ 0 & \text{sinon} \end{cases} \\ B_{i,j+1}(z) = & P_{i,j+1}(z)B_{i,j}(z) + [1 - P_{i+1,j+1}(z)]B_{i+1,j}(z) \\ \text{où } P_{i,j}(z) = & \begin{cases} \frac{z-t_i}{t_{i+j}-t_i} & \text{si } t_{i+j} \neq t_i \\ 0 & \text{sinon} \end{cases} \end{cases} \quad (10)$$

Neuf modèles GPD (modèles GPD1 à GPD9) avec une seule covariable ont été évalués dans cette thèse au moyen de différentes combinaisons de paires de valeurs (k, d) comme le résume le tableau 1.

Tableau 1: Modèles GPD non-stationnaires évalués

$d \backslash k$	2	3	4
1	GPD1 (2, 1)	GPD4 (3, 1)	GPD7 (4, 1)
2	GPD2 (2, 2)	GPD5 (3, 2)	GPD8 (4, 2)
3	GPD3 (2, 3)	GPD6 (3, 3)	GPD9 (4, 3)

L'estimation des paramètres des modèles GPD non-stationnaires (GPDns) présenté au tableau 1, est faite avec la méthode MVG en utilisant des méthodes numériques comme celle de Newton-Raphson (Hosking et Wallis 1987). En effet, dans le cadre d'une approche bayésienne, il est préférable de réaliser toute l'analyse en tenant compte de toutes les composantes de la distribution a posteriori. Cependant, avec la méthode MVG utilisée dans cette thèse, la seule loi a priori informative est celle du paramètre de forme de la GPD. Cette loi a priori est la même pour tous les modèles GPD. Par conséquent, le rapport des lois a posteriori est égale au rapport des vraisemblances (par simplification du terme commun correspondant à la loi a priori). Le mode a posteriori dans ce cas de l'approche MVG correspond à une estimation du maximum de vraisemblance avec contrainte sur le paramètre de forme. De même, le facteur bayésien est égale au rapport de vraisemblance qui est lui-même équivaut au critère d'information de Akaïke (AIC). En effet, pour la même taille de l'échantillon, le logarithme du rapport de vraisemblance est égal à la différence entre les log-vraisemblances qui est égale à la différence des AIC des modèles comparés. D'où la comparaison des modèles GPDns entre eux mais aussi avec le modèle stationnaire GPDs (nommé ici GPD0) en se basant sur la valeur du AIC obtenue avec la formule donnée à l'équation (11).

$$AIC = -2L + 2n \quad (11)$$

avec L et $n=k+d+1$ désignant respectivement la log-vraisemblance et le nombre de paramètres de chaque modèle évalué.

Le modèle ayant la plus faible valeur de l'AIC peut être considéré comme le meilleur modèle (Katz 2013).

Les modèles GPDns évalués ont permis de calculer des quantiles conditionnels aux valeurs de la meilleure fenêtre de données de l'indice climatique régionalement dominant. De tels quantiles non-stationnaires ont été comparés avec le quantile stationnaire résultant du modèle GPD0 définie à l'équation (2). L'approche et les résultats de ces développements sont associés au premier objectif spécifique de cette thèse et font l'objet du 1^{er} article présenté dans la deuxième partie de cette thèse.

3.2.3 Modèle couplé PD-GPD non-stationnaire

Le modèle couplé PD-GPD non-stationnaire (PD-GPDns) a été développé comme une extension au modèle GPDns afin d'évaluer l'effet combiné de l'incorporation d'une covariable dans la modélisation des variables V1 et V2. Un tel modèle permet une interprétation pratique des quantiles comme celle associée à la notion de période de retour. Le développement du modèle PD-GPD avec covariable a été précédé par la validation de l'hypothèse que la fréquence annuelle des dépassements de seuil (V2) suit un processus de Poisson. En conséquence, un processus de Poisson non-homogène a été utilisé en faisant varier le paramètre d'intensité de Poisson en fonction du même indice climatique corrélé à la fois à V1 et V2. Un tel choix visait à évaluer la possibilité qu'un même indice climatique puisse aider à expliquer la variabilité à la fois dans les séries V1 et V2. L'estimation des paramètres de ce processus de Poisson non-homogène est faite au moyen des modèles linéaires généralisés (GLM).

L'incorporation de la non-stationnarité se fait séparément dans chacun des modèles PD et GPD (Katz 2013 ; Katz et al. 2002), d'où une évaluation indirecte des changements au niveau de la distribution GEV reparamétrisée (Équation 7). En effet, cette distribution est la résultante de la correspondance entre les distributions des maxima annuels et des dépassements de seuil (Lang et al. 1999). Les quantiles obtenus par inversion de la fdc de cette GEV non-stationnaire (GEVns) reparamétrisée, ont alors une grandeur de maxima annuels et non de dépassement de seuil, le taux moyen annuel des dépassements ayant été incorporé (Silva et al. 2016).

Les modèles développés suivant cette approche PD-GPD sont nommés GEV0 (tous les paramètres des modèles PD et GPD sont constants), GEV1 (le paramètre d'intensité de Poisson est constant tandis que celui d'échelle de la GPD varie en fonction d'un indice climatique Z) et GEV2 (le paramètre d'intensité de Poisson et celui d'échelle de la GPD varie en fonction du même indice climatique Z mais pas forcément avec la même fenêtre de données de cet indice). Pour chacun de ces modèles, l'AIC est calculé pour des fins de choix du meilleur modèle. L'approche et les résultats de ces développements sont également associés au second objectif spécifique de cette thèse et font l'objet du 2^{ème} article présenté dans la deuxième partie de cette thèse.

3.2.4 Modèle hybride GPD

Les deux approches (GPD et PD-GPD) de développement précédemment présentées sont basées sur l'approche POT classique où le seuil est fixé à l'avance et l'échantillon des dépassemens de seuil indépendants est utilisé pour l'estimation des paramètres du modèle. Afin d'estimer les quantiles à partir d'un modèle qui incorpore le seuil dans le vecteur des paramètres à estimer, trois modèles "hybrides" GPD sont développés. Le premier modèle "hybride" (modèle M1) est une combinaison de la GPD avec la distribution Gamma sans aucune contrainte de continuité au point du seuil de la fonction de densité de probabilité (fdp) résultante. Cette contrainte est ensuite imposée avec le second (M2) et troisième (M3) modèle, où les distributions centrales considérées sont respectivement la Gamma et la Normale.

Les distributions Gamma et Normale ont été choisies car elles sont très souvent utilisées dans la littérature (exemple : Behrens et al. 2004; Carreau et Bengio 2009; Do Nascimento et al. 2012) pour ce genre de développement.

Les fdp des modèles M1 et M2 (f_{G-GPD}) puis M3 (f_{N-GPD}), sont définies à l'équation (12).

$$\begin{cases} f_{G-GPD}(y_i) = [h(y_i) \cdot 1_{\{y_i < u\}}] + \left[((1 - H(u)) g(y_i - u)) \cdot 1_{\{y_i \geq u\}} \right] \\ f_{N-GPD}(y_i) = [f(y_i) \cdot 1_{\{y_i < u\}}] + \left[((1 - F(u)) g(y_i - u)) \cdot 1_{\{y_i \geq u\}} \right] \end{cases} \quad (12)$$

avec h et f les fdp de la Gamma et de la Normale puis H et F leur fdc respectives; g désigne la fdp de la GPD; y_i est l'observation i d'une variable aléatoire Y ; u représente le seuil et $(1 - H(u))$ dans le cas du modèle M1, est la condition d'observation d'un dépassement de seuil.

Les fonctions de vraisemblance L_{G-GPD} associée à M1 et M2, puis L_{N-GPD} pour M3 sont présentées à l'équation(13).

$$\begin{cases} L_{G-GPD}(\underline{y} | \theta_1; \theta_2) = \prod_{y_i \in C_1} h(y_i | \theta_1) \cdot \prod_{y_i \in C_2} (1 - H(u | \theta_1)) g(y_i - u | \theta_2) \\ L_{N-GPD}(\underline{y} | \theta_3; \theta_2) = \prod_{y_i \in C_1} f(y_i | \theta_3) \cdot \prod_{y_i \in C_2} (1 - F(u | \theta_3)) g(y_i - u | \theta_2) \end{cases} \quad (13)$$

avec θ_1 , θ_2 et θ_3 les vecteurs respectifs des paramètres de la Gamma, la GPD et la Normale; C_1 et C_2 représentent les sous-ensembles des observations en-dessous et au-dessus du seuil respectivement.

L'obtention de la continuité de la fdp au point du seuil avec par exemple le modèle M2, implique que $h(u|\theta_1) = (1 - H(u|\theta_1)) \cdot g(u|\theta_2)$; d'où $h_{\theta_1}(u) = (1 - H_{\theta_1}(u))g_{\theta_2}(0)$.

La continuité est formulée de façon similaire pour le modèle M3. Une solution explicite à cette équation est l'expression du paramètre d'échelle de la GPD en fonction des paramètres de la Gamma comme suit (équation 14) :

$$\begin{cases} \sigma_{G-GPD} = \frac{(1-H(u|\theta_1))}{h(u|\theta_1)} \\ \sigma_{N-GPD} = \frac{(1-F(u|\theta_3))}{f(u|\theta_3)} \end{cases} \quad (14)$$

L'inférence avec les modèles M1, M2 et M3 est faite dans un cadre Bayésien avec la construction d'un algorithme MCMC de Metropolis-Hasting (Hastings 1970; Metropolis et al. 1953) pour obtenir la distribution a posteriori de chacun des paramètres estimés ainsi que des quantiles (Ouarda et El Adlouni 2011; El Adlouni et Ouarda 2009; de Zea Bermudez et Turkman 2003). La moyenne et/ou le mode de chacune de ces distributions a posteriori peut constituer l'estimateur Bayésien des paramètres et des quantiles. Des distributions a priori non-informatives ont été utilisées par défaut en raison de l'absence d'information régionale ou historique. La loi normale tronquée a été utilisée comme loi a priori pour le seuil dans les études de Behrens et al. (2004) puis dans celle de MacDonald et al. (2011). Nous l'avons donc choisi dans le cadre de notre analyse et nous avons aussi considéré une loi uniforme discrète pour des fins d'analyse de comparaison entre ces deux lois a priori. Une distribution a priori conjointe basée sur l'approche de Jeffrey (Liseo 1993) a été choisie en raison de sa simplicité, pour les paramètres de forme et d'échelle de la Gamma. La loi Beta de Martins et Stedinger (2001) qui permet de restreindre les valeurs du paramètre de forme de la GEV ou de la GPD dans l'intervalle [-0.5; 0.5] a été considérée pour le paramètre de forme de la GPD et une loi uniforme discrète a été retenue pour le paramètre d'échelle de la GPD. La loi uniforme discrète, qui donne le même poids à chacune des valeurs possibles du paramètre d'intérêt, est également considérée pour tous les paramètres du modèle M3.

Le calcul des quantiles avec les modèles "hybrides" diffère de l'AF classique. En effet, dans ce dernier cas on travaille directement avec l'échantillon des extrêmes alors que c'est l'ensemble des observations qui est considéré pour l'inférence statistique avec les modèles M1, M2 et M3. Une adaptation de la probabilité annuelle au non-dépassement P_a a été proposée avec P_e , en se basant sur le nombre total de dépassements de seuil nd comme formulée à l'équation (15).

$$P_e = P_a^{\frac{1}{nd}} \quad (15)$$

À partir de la probabilité ajustée P_e , on a utilisé une approche de raisonnement basée sur les probabilités conditionnelles pour décider à partir de quelle composante de chacun des modèles hybrides M1, M2, M3, le quantile sera estimé. La figure 2 est une illustration de notre approche, avec H et G les fonctions de répartition respectives des distributions Gamma et GPD dans le cas des modèles M1 et M2. La probabilité de ne pas excéder la valeur y est P_d (avec $P_d=P_e$); P_u désigne la probabilité associée au seuil estimé \hat{u} et équivaut à $P_u=H(\hat{u}, \hat{a}, \hat{b})$, avec a et b les paramètres de forme et d'échelle de la distribution Gamma dans le cas des modèles M1 et M2. Si $P_d < P_u$, alors le quantile associé à P_d équivaut à $Q_{P_d} = H^{-1}(P_d, \hat{a}, \hat{b})$. Si $P_d > P_u$, alors $Q_{P_d} = G^{-1}(PP, \hat{\kappa}, \hat{\beta})$ avec $PP=(P_d-P_u)/(1-P_u)$. La même approche s'applique dans le cas du modèle M3.

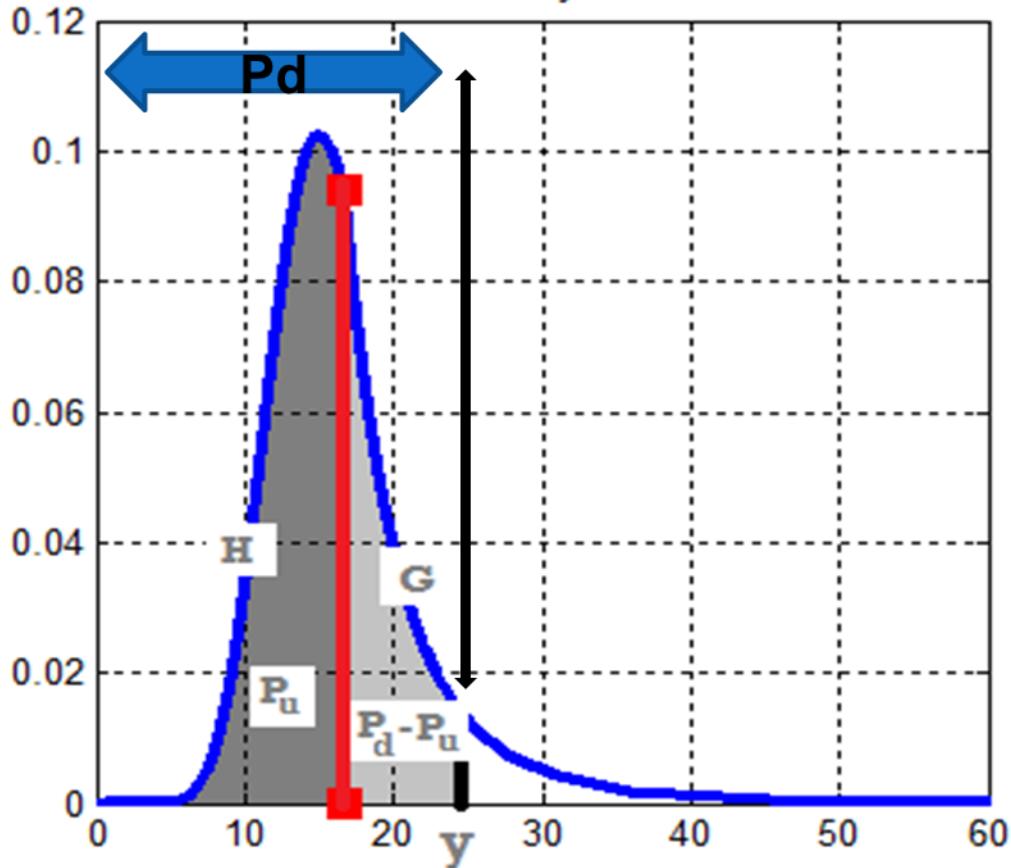


Figure 2: Schémas illustratif du calcul des probabilités au non-dépassement pour l'estimation des quantiles à partir des modèles hybrides M1, M2 et M3

L'évaluation des modèles M1, M2 et M3 a été faite par une étude de simulation. Deux types d'erreurs d'estimation ont été spécifiquement examinés : les erreurs liées à la méthode d'estimation des paramètres (type I), et celles associées au modèle de génération des données de simulation (type II). Dans le 1^{er} cas, un plan de génération des modèles M1 et M2 composé de neuf scénarios de simulation a été proposé en définissant tout d'abord trois probabilités ($p(u_0) = 0.7 ; 0.8 ; 0.9$) pour déterminer des valeurs théoriques du seuil u_0 à évaluer, de même que trois valeurs de coefficient d'asymétrie ($C_s=3 ; 2 ; 1$) pour obtenir des valeurs théoriques du paramètre de forme (a_0) de la Gamma sachant que $C_s = 2/\sqrt{a_0}$. Le choix de ces valeurs de C_s vise à explorer des distributions de type symétrique et avec une asymétrie positive.

Pour chacun des scénarios, la valeur théorique du paramètre d'échelle de la GPD (σ_0) a été déterminée à l'aide de l'équation (14), tandis que deux (02) valeurs théoriques (0.05 et 0.2) ont été fixées et évaluées pour le paramètre de forme (ξ_0) de la GPD. Quant au paramètre d'échelle de la Gamma, sa valeur théorique (b_0) a été fixée à 5. Par contre, seulement trois scénarios liés à $p(u_0)$ ont été considérés pour le modèle M3 et les valeurs théoriques 0 et 1 ont été choisies pour la moyenne (μ_0) et l'écart-type (a_0) de la distribution Normale utilisée. La détermination du seuil théorique u_0 pour les modèles M1 et M2 a été obtenu par inversion de la cdf de la Gamma sachant $p(u_0)$, a_0 et b_0 . De façon similaire, le seuil u_0 associé au modèle M3 a été également calculé. Ainsi, avec cette première approche, on a généré et on a estimé à partir du même modèle.

Dans le 2^{ème} cas d'évaluation des erreurs d'estimation, il a été question de générer les données à partir d'une même distribution, différente de celle développée, et ensuite de procéder à l'inférence statistique à partir des modèles M1, M2 et M3. Le modèle GEV a été considéré à cet effet. Trois valeurs théoriques ($\xi_0=0.01 ; 0.1 ; 0.2$) ont été évaluées pour son paramètre de forme tandis que celles de ses paramètres d'échelle (β_0) et de position (ψ_0) ont été fixées respectivement 100 et 1000.

Dans le cadre de l'inférence Bayésienne avec les modèles M1, M2 et M3, une analyse de sensibilité à la valeur du seuil initial u_1 , des résultats de tous les différents scénarios proposés, a été également proposée en explorant des valeurs u_1 équivalentes à la médiane ainsi qu'aux 90^{ème} et 99^{ème} percentiles. Pour les autres paramètres, les estimateurs obtenus avec la méthode du MVG (méthode du MV) ont été considérés pour définir les valeurs initiales des modèles M1 et M2 (modèle M3).

Pour des fins de comparaison avec l'approche POT classique, ces valeurs u_1 ont été considérées successivement comme seuil adéquat pour définir l'échantillon des extrêmes à partir des données générées avec la Gamma-GPD et la GEV. Chacun de ces échantillons a été ensuite ajusté avec le modèle PD-GPD classique (M0).

Les quantiles résultant des modèles M0, M1, M2 et M3 ont été finalement comparés suivant les scénarios de simulation formulés, en calculant le biais relatif (BR) et la racine de l'erreur quadratique moyenne relative (REQMR) des quantiles estimés.

Pour chaque scénario, la taille de l'échantillon des données générées vaut n=5000 et une chaîne de Markov de longueur N=10000 avec un temps de chauffe N0=5000 a été utilisée de sorte que la distribution a posteriori est formée par les N-N0 valeurs. La répétition du processus d'estimations pour chaque scénario a été faite 100 fois.

Afin d'évaluer l'application pratique des modèles M1, M2 et M3, des observations de la précipitation totale journalière enregistrée à la station 7025250 sur la période 1945-1992 ont été considérées. En effet, on a des données en continue sur cette période d'analyse choisie. La station est située à l'aéroport international de Montréal (Québec), et de coordonnées géographiques (latitude : 45.47; longitude : -73.75; altitude : 36). Toutefois, en raison de la présence de nombreux jours secs (jours avec 0mm) dans ce type de données, il est indispensable de considérer seulement les observations strictement positives afin d'éviter des estimations absurdes avec les modèles M1 et M2. Cette information n'a cependant pas été ignorée dans le processus de l'inférence statistique car, la proportion (p_0) associée à ces observations nulles a été déterminée et intégrée dans le calcul d'une nouvelle probabilité $P_{ja}=(P_j-p_0)/(1-p_0)$ en se basant sur l'étude de El Adlouni et al. (2011).

La finalité du développement de ces modèles hybrides est de répondre au second objectif spécifique de cette thèse. La démarche méthodologique et les résultats des analyses ainsi que leur discussion, font l'objet du 3^{ème} article qui est présenté dans la deuxième partie de cette thèse.

4 RÉSULTATS DES APPLICATIONS

4.1 Sélection des stations

Au départ de cette recherche, une base de données sur la précipitation totale journalière a été constituée avec 1057 stations à l'échelle de la zone d'étude définie. Après une analyse de la qualité des données, 329 stations ont été retenues pour la suite des analyses préliminaires. La figure 3 illustre leur localisation géographique.

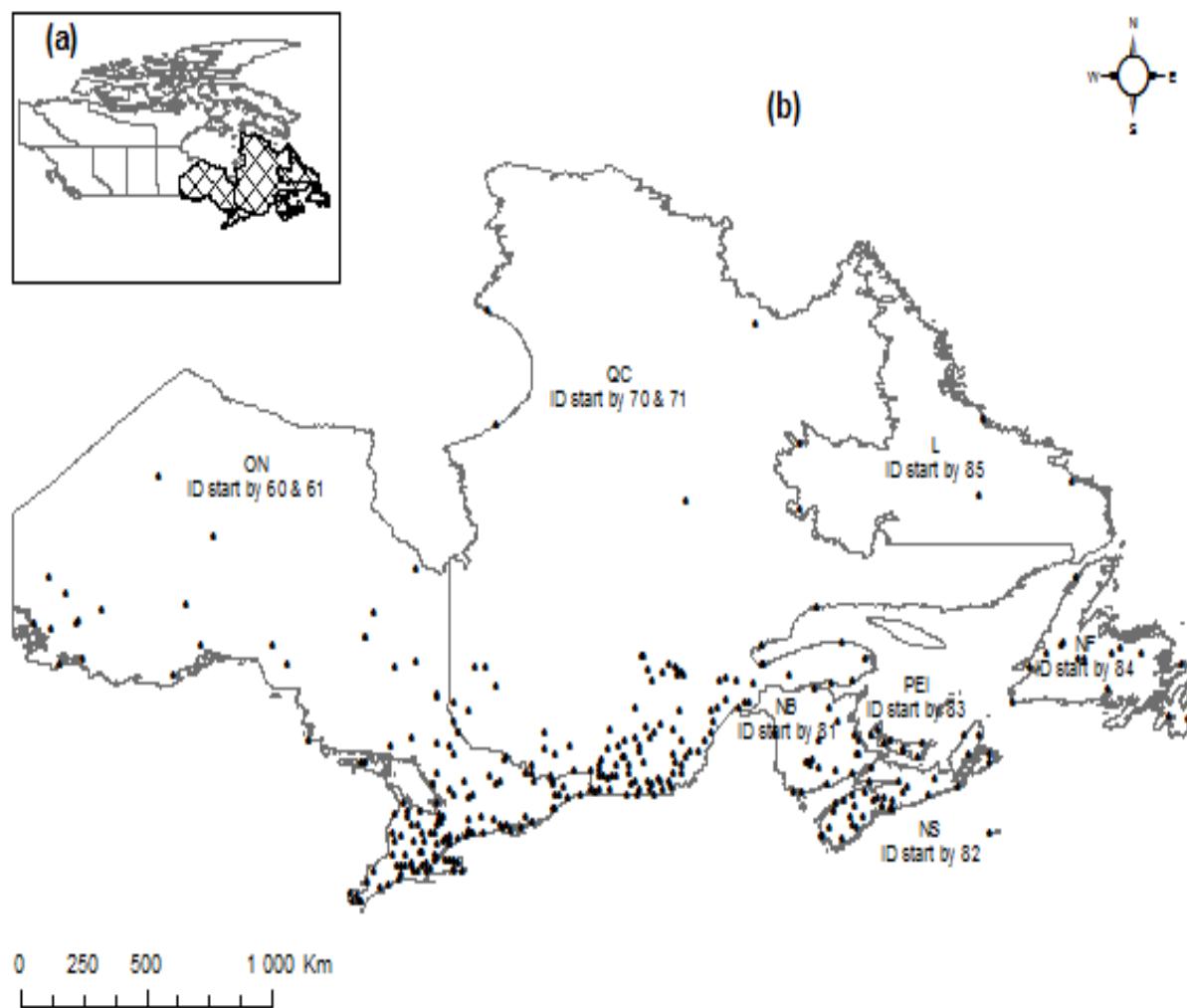


Figure 3: Région Sud-Est du Canada (a) et localisation des 329 stations comportant au moins 30 années de données journalières complètes (b)

Pour chacune de ces stations, la validité des valeurs de seuil comprises entre le 90^{ème} et 99^{ème} percentile a été évaluée au moyen du graphique des excédents moyens, celui de l'évolution des paramètres de forme et d'échelle de la GPD en fonction d'une séquence croissante de valeurs du seuil. Le minimum des maxima annuels a été aussi exploré comme une valeur potentielle de seuil adéquat. La fonction de l'autocorrélation a été ensuite appliquée à la série des dépassements de seuil pour vérifier leur indépendance selon le choix du seuil. De même, la validité de la GPD a été évaluée au moyen du test statistique du Chi-carré. Les illustrations en annexes 1, 2, 3, 4 et 5 sont un exemple de résultats obtenus avec les observations enregistrées à la station de l'aéroport de Miramichi au NB (ID 8101000). A cette station, le 99^{ème} percentile des observations vaut 33 mm et l'examen des annexes 3 à 5 montre une inflexion nette des courbes pour cette valeur du seuil. Pour chacune des 329 stations analysées, de tels résultats sont ainsi disponibles et ont permis de sélectionner 173 stations pour lesquels le 99^{ème} percentile apparaît comme le seuil le plus adéquat pour valider le critère d'indépendance et l'usage de la GPD à un maximum de stations en comparaison des autres valeurs de seuil. A partir de ces 173 stations, l'analyse de corrélation et celle par ondelettes ont été faites de même que l'évaluation du processus de Poisson. Le tableau 2 synthétise la répartition des stations suivant huit critères distincts :

- 1) Inventaire exhaustif des stations d'observation de la précipitation totale journalière dans le Sud-Est du Canada ;
- 2) Rétention à partir de (1), des séries comportant au moins 30 années complètes de données sur une base continue ou non ;
- 3) Validation à partir de (2), du critère d'indépendance des dépassements de seuil échantillonnes ;
- 4) Rétention à partir de (3), des stations ayant montré une corrélation significative maximale des séries V1 avec un des sept indices climatiques ;
- 5) Rétention à partir de (3), des stations ayant montré une corrélation significative maximale des séries V2 avec un des sept indices climatiques ;
- 6) Validation à partir de (3), de PD pour les séries V2 ;
- 7) Non validation à partir de (3), de PD mais validation de BND pour les séries V2 ;
- 8) Non validation à partir de (3), de PD mais validation de BD pour les séries V2.

Tableau 2: Répartition du nombre de stations par province et en fonction de huit critères

Provinces Critères \	PEI	NB	NS	NF&L	QC	ON	TOTAL
1	16	70	89	98	398	386	1057
2	8	19	35	22	109	136	329
3	6	9	18	8	54	78	173
4	4	8	16	6	42	60	136
5	4	6	14	7	41	66	138
6	4	8	17	6	49	72	156
7	2	1	1	2	4	4	14
8	0	0	0	0	1	2	3

L'annexe 6 illustre les résultats d'application des 3 derniers critères. La distribution de Poisson est adéquate à presque toutes les stations évaluées (156/173). La validité des résultats à venir est étroitement associée à ces huit critères énoncés.

4.2 Sélection de la covariable climatique régionale

Les résultats de l'analyse de corrélation ont permis d'examiner l'influence de chacun des indices climatiques considérés comme covariable potentielle à l'échelle du Sud-Est du Canada. Ainsi, lorsqu'on se base sur les indices qui montrent les plus fortes corrélations avec les séries de V1, on a par ordre d'importance décroissante AO (31 stations), PNA (30 stations), NAO (22 stations), SOI (16 stations), WHWP (15 stations), PDO (13 stations) et AMO (09 stations), pour un total de 136 stations. L'ordre de position et de grandeur sont presque les mêmes dans le cas des séries de V2 avec respectivement AO (27 stations), NAO (24 stations), PNA (23 stations), SOI (23 stations), WHWP (18 stations), PDO (16 stations) et AMO (07 stations), pour un total de 138 stations. Les annexes 7 à 13 présentent la spatialisation de ces résultats de l'analyse de corrélation pour chacun des indices climatiques considérés. On peut d'ores et déjà remarquer à l'aide de ces figures, une concurrence d'influence de plusieurs indices climatiques dans la région d'étude.

Quant aux tableaux 3 et 4, ils résument la répartition des stations ayant validé le critère n3 et pour lesquelles une corrélation significative (maximale ou non) a été détectée entre les indices climatiques et respectivement les séries V1 et V2.

Tableau 3: Répartition des stations décrivant V1, en fonction de leurs corrélations significatives avec les indices climatiques

Indices Provinces \ Indices	AO	NAO	PNA	SOI	WHWP	PDO	AMO
PEI	2	1	3	2	0	2	0
NB	3	3	2	4	4	3	2
NFL	3	4	2	1	3	0	1
NS	8	4	4	5	5	9	3
ON	22	27	19	20	19	16	15
QC	19	13	20	5	6	5	7
Total	57	52	50	37	37	35	28

Considérant la corrélation qui existe entre les indices AO et NAO, les indices AO et PNA apparaissent respectivement comme les covariables climatiques dominantes dans le Sud-Est du Canada, avec une influence plus importante que celle des autres indices, sur la variabilité dans l'intensité des précipitations totales journalières extrêmes.

Tableau 4: Répartition des stations décrivant V2, en fonction de leurs corrélations significatives avec les indices climatiques

Indices Provinces \ Indices	PNA	AO	NAO	PDO	AMO	WHWP	SOI
PEI	3	1	1	2	1	2	0
NB	5	1	0	0	0	1	1
NFL	3	2	4	3	4	2	2
NS	6	10	4	4	2	0	3
ON	24	24	25	20	20	18	9
QC	14	13	13	10	10	10	6
Total	55	51	47	39	37	33	21

Les mêmes indices AO et PNA se démarquent également quant à leur influence sur la fréquence des précipitations totales journalières extrêmes dans la zone d'étude, avec toutefois une légère dominance de PNA sur AO. Le tableau 5 propose une agrégation de l'influence de chacun des sept indices climatiques sur les précipitations totales journalières extrêmes du Sud-Est Canadien. En effet, en se basant sur le nombre de stations où un tau de Kendall significatif a été détecté entre V1 (puis V2) et un indice, versus le nombre de stations où aucune corrélation significative n'a été trouvée, quatre types d'influence ont été définis. Ainsi, le déterminant « influence forte » est attribué à l'indice avec le maximum de stations où une corrélation significative a été détectée avec V1 et/ou V2. Dans certaines provinces de la région d'étude, on n'a pas dénombré de stations avec des corrélations significatives avec V1 (puis V2) et un indice ; d'où le terme « pas d'influence ».

Tableau 5: Influence régionale des indices climatiques sur la variabilité dans l'intensité et la fréquence des précipitations extrêmes

Indices climatiques Provinces	AO	NAO	AMO	PNA	PDO	SOI	WHWP
NF							
L							
PEI							
NS							
NB							
QC							
ON							
Légende							
Influence forte	Influence moyenne	Influence faible	Pas d'influence				

Le croisement des tableaux 3, 4 et 5 permet de sélectionner deux covariables climatiques régionales, les indices AO et PNA, pour la modélisation de l'intensité et de la fréquence des dépassements de seuil indépendants de la précipitation totale journalière au Sud-Est du Canada. Ainsi, des 57 et 50 stations associées respectivement à AO et PNA tel que présenté dans le tableau 3, on a identifié 15 stations où les deux indices AO et PNA étaient significativement corrélés à V1 avec une dominance de PNA à 11 stations. D'où un groupe AO et un groupe PNA de 46 stations chacun, constitués pour l'AFns de V1.

L'annexe 14 donne une description de ces 96 stations. Il faut noter que le seuil moyen à ces stations est de 30 mm avec un écart-type de 5 mm, un seuil minimum de 21 mm et un seuil maximum de 46 mm. Les valeurs de seuil supérieures à la moyenne s'observent surtout au sud-est du Québec et dans les Maritimes.

Les résultats de l'analyse de corrélation à ces stations montrent une dominance de corrélations positives (et négatives) entre l'intensité des précipitations extrêmes et l'indice AO dans les Maritimes (dans les régions plus centrales du Québec et de l'Ontario). La corrélation est positive au niveau de la fréquence des précipitations extrêmes dans les Maritimes. Toutefois, dans les provinces du Québec et de l'Ontario, on observe un mélange de corrélations positives et négatives entre AO et V2.

En ce qui concerne l'indice PNA par contre, les corrélations sont à majorité positives au niveau de l'intensité et de la fréquence des précipitations extrêmes et à l'échelle du Sud-Est du Canada.

Les neuf modèles GPDns présentés au tableau 1, ont été ainsi évalués pour chacun des deux groupes de stations formés. Toutefois, en s'intéressant aux stations où le même indice est significativement corrélé aussi bien aux séries de V1 que celles de V2 et où le processus de Poisson est vérifié, 10 et 15 stations associées respectivement à AO et PNA ont été retenues. Ces deux derniers groupes de stations sont respectivement décrites aux annexes 15 et 16 et sont celles explorées dans le cadre du développement du modèle PD-GPD présenté dans l'article 2.

4.3 Variabilité et périodicité

L'analyse par ondelettes a révélé une variabilité interannuelle dans les séries de données des variables V1 et V2. Les périodicités les plus observées dans les 173 stations analysées, concernent les bandes d'énergie de fréquence 2-3 ans, 2-4 ans et 3-4 ans. Ainsi, on a résumé ces cycles de variabilité dans un premier groupe d'amplitude 2-4 ans. Par ailleurs, des cycles significatifs de 4-5 ans, 4-6 ans, 5-7 ans, 2-6 ans, 3-5 ans et 3-6 ans ont été également relevés. Toutefois, en raison de la longueur des séries de données (courtes séries temporelles), des variabilités de fréquence décennale et multi-décennale ont été aussi observées (6-7 ans, 6-8 ans, 7-10 ans, 8-9 ans, 8-14 ans, 8-16 ans, 10-16 ans, 14-16 ans), mais généralement en dehors du cône d'influence significative de la transformée en ondelette. Un portrait général de ces résultats est présenté en annexes 17 et 18 pour les variables V1 et V2 respectivement, avec six stations dans chaque cas. Ces stations ont en effet de longues séries temporelles de données (1945-2013).

La même analyse a été appliquée aux différentes séries de moyennes mobiles annuelles aux trois mois, constituées pour chacun des indices AO et PNA respectivement sur les périodes 1951-2012 et 1951-2013. La même variabilité interannuelle apparaît dans ces séries avec des fréquences localisées dans la période 1970-1990. En effet, les séries analysées pour l'indice AO montrent des variabilités de fréquence 2-3 ans et 3-5 ans. Au niveau de l'indice PNA, ce sont les bandes 2-4 ans et 4-6 ans qui se démarquent clairement. Comme au niveau des précipitations, des variabilités sont également identifiées dans les bandes 2-6 ans, 6-10 ans, 8-16 ans et 16-20 ans. Les résultats de la transformée en ondelette pour chacune de ces séries sont présentés en annexes 19 et 20 respectivement pour les indices AO et PNA.

Des bandes de variabilités communes apparaissent ainsi entre les séries de V1 (et V2) et celles des indices AO et PNA. D'ailleurs les résultats des analyses de la transformée en ondelette croisée (XWT) entre variables et covariables, confirment cette concordance des interactions dans les cycles de variabilité commune.

Les résultats des analyses de la cohérence de la transformée en ondelettes (WTC) révèlent la cohérence de ces interactions mais aussi confirment les résultats de l'analyse de corrélation. En effet, lorsque les interactions sont en phase (ou en antiphase), cela implique une corrélation positive (et négative) entre les variables analysées.

4.4 Analyse fréquentielle non-stationnaire avec l'approche POT classique

Les neufs modèles GPDns définis au tableau 1 ont été appliqués aux séries V1 du groupe des 46 stations de l'indice AO puis de celui des 46 stations de PNA (annexe 14). Les paramètres de chacun des modèles ont été estimés avec la méthode MVG et les quantiles ont été calculés pour trois probabilités annuelles au non-dépassement $np=0.5$; 0.9 et 0.99. Dans une hypothèse de stationnarité, ces probabilités correspondent à des périodes de retour de 2, 10 et 100 ans respectivement, et impliquent que le quantile peut être observé en moyenne une fois tous les 2, 10 et 100 ans respectivement. Cette interprétation est directe lorsqu'on travaille avec des maxima annuels ou quand la modélisation est faite avec un modèle PD-GPD.

Le 99^{ème} percentile choisi comme seuil (u) pour définir l'échantillon des extrêmes, décroît du Sud au Nord et de l'Est vers l'Ouest avec les plus grandes valeurs de seuil observées dans la province de la Nouvelle Ecosse. L'estimation du paramètre d'échelle constant (σ_0) varie entre 7 et 16 et son ordre de grandeur est spatialement proportionnel à celui du seuil. Les hyperparamètres (s_1, s_2, s_3) du paramètre d'échelle devenu variable (σ_1), sont au nombre de 3 et varient d'une station à une autre. Quant aux quantiles stationnaires, ils varient entre 25 et 55 mm, 45 et 75 mm puis 55 et 115 mm respectivement pour les périodes de retour 2 (Q2), 10 (Q10) et 100 ans (Q100).

L'agrégation des dépassements de seuil journaliers indépendants à une échelle annuelle a permis de générer des séries avec au moins 30 observations pour conduire l'AF. Celle-ci a par exemple fourni à partir des séries V1 des 96 stations, une estimation du paramètre de forme de la GPD stationnaire (ξ_0) et non-stationnaire (ξ_1) autour de la valeur 0.01 spatialement cohérente à l'échelle de la région d'étude.

Au niveau du calcul des quantiles conditionnels à AO, les mêmes probabilités au non-dépassement ont été considérées mais seuls les résultats associés à np=0.9 sont mis en lumière pour des fins de comparaison avec les quantiles stationnaires de période de retour 10 ans.

Le modèle GPDns avec la fonction B-splines à 2 nœuds et de degré 1 (GPD1), a montré la plus faible valeur du AIC en comparaison avec les huit autres modèles GPDns. Il a donc été retenu pour la modélisation des séries V1 à toutes les 46 stations associées à AO. La comparaison de ce modèle avec la GPD stationnaire (GPD0) indique que 18 stations sur les 46 présentent une plus faible valeur du AIC avec GPD0 même si les valeurs du AIC à ces 18 stations sont comparables pour les modèles GPD0 et GPD1.

Par ailleurs, une analyse de régression entre les phases positives (et négatives) de AO et leurs quantiles correspondants, a été effectuée. En faisant ensuite le ratio des pentes de la régression avec ces deux phases, les résultats ont permis de tester si les différentes tendances entre les séries de V1 et celles de AO étaient significativement linéaires ou non. En effet, des valeurs du ratio très proches ou égales à 1 confirmeraient une tendance linéaire significative. De plus, un ratio négatif justifierait l'hypothèse de la non-linéarité des interactions entre ces deux variables en raison de la structure physique des indices climatiques (phases négative et positive). La linéarité significative a été ainsi observée à 10 des 46 stations, ce qui implique qu'aux 36 restantes, l'hypothèse d'une tendance linéaire est rejetée. De ces 36 stations, le signe du ratio a permis d'isoler 11 stations où la non-linéarité est statistiquement significative. Trois valeurs de AO ont été considérées à chaque station pour l'analyse du quantile associé à np=0.9, les valeurs négative et positive les plus extrêmes (elles peuvent varier entre +/- 0.5 à +/- 3.5 selon la fenêtre de données retenue pour l'indice) et la valeur centrée en 0. Les quantiles associés à ces valeurs sont respectivement notés Qg (extrême gauche), Qd (extrême droite) et Qc (centre). La comparaison de ces différents quantiles indique qu'il peut y avoir une différence de l'ordre de 5 à 30 mm entre le quantile de période de retour 10 ans et les quantiles conditionnels à AO selon la phase de cet indice.

Au niveau du calcul des quantiles conditionnels à PNA, la même approche d'analyse a été adoptée. Le modèle GPD1 se présente également comme le meilleur choix. Toutefois, à deux stations (ID 6158255 et 7060400), le modèle avec 2 nœuds et de degré 2 (GPD2), a montré une meilleure performance (plus petite valeur du AIC) que les autres modèles GPDns ainsi que le modèle GPD0. Les annexes 21 et 22 illustrent les résultats de l'AF à ces deux stations. Par ailleurs, le modèle stationnaire GPD0 apparaît quant à lui très approprié pour la station 8300100 (annexe 23). Néanmoins, la comparaison des modèles GPD1 et GPD0 indique que 15 stations sur les 46 présentent une plus faible valeur de AIC avec GPD0. Toutefois, la différence au niveau du AIC de ces deux modèles à ces stations est également insignifiante. L'application du test statistique de la linéarité significative pour ce groupe de stations a permis d'identifier seulement 5 stations où la tendance linéaire est significative. Ainsi, l'hypothèse d'une tendance linéaire est également rejetée pour les 41 stations restantes. Plus encore, le signe du ratio a permis d'isoler parmi ces 41 stations, 12 stations où la non-linéarité est statistiquement significative. Quant aux quantiles stationnaires et conditionnels à PNA, une comparaison similaire à celle du groupe AO a été observée.

Les annexes 24 et 25 résument les résultats de l'AF pour ces deux groupes de stations respectifs. Il apparaît à la lumière de ces deux tableaux de synthèse (annexes 24 et 25), que les quantiles stationnaires et non-stationnaires sont très comparables pour une probabilité au non-dépassement $np=0.5$, mais la différence dans les valeurs augmentent proportionnellement avec np .

Différentes formes de quantiles ont été identifiées avec les résultats de modélisation des précipitations totales journalières extrêmes à ces 96 stations. En effet, la forme des quantiles non-stationnaires varie généralement selon que l'intensité des précipitations extrêmes change durant les deux phases de l'indice (annexes 26 et 27), ou durant seulement sa phase positive (annexes 28 et 29) ou négative (annexes 30 et 31), ou qu'elle change linéairement avec les valeurs de l'indice, de façon significative (annexe 32) ou non (annexe 33).

La combinaison des modèles PD et GPD a également permis de développer des modèles PD-GPD dont les quantiles sont déterminés à partir de l'inverse de la fdc de la GEV. Dans le cadre de ce développement, ce sont tous les dépassements de seuil indépendants qui sont considérés et non leur agrégation à une échelle annuelle. Ainsi, s'il y a plusieurs dépassements pour une année donnée, la même valeur annuelle de la covariable climatique est utilisée pour chacun des dépassements. L'adoption de cette technique a eu pour finalité la validation de la similarité des résultats de la modélisation à partir d'un échantillon des extrêmes basé sur tous les dépassements de seuil journaliers indépendants et sur leurs moyennes annuelles. L'approche PD-GPD a permis d'avoir une interprétation des quantiles en termes de probabilité d'occurrence d'annuelle. Toutefois, la maximisation de l'information par l'incorporation de la non-stationnarité dans le paramètre d'intensité de Poisson et celui du paramètre d'échelle de la GPD (modèle nommé PD1-GDP1), n'améliore pas les estimations plus que dans le cas où le premier paramètre est gardé constant (PD0-GDP1). Ces modèles donnent cependant de meilleures performances que le modèle avec tous les paramètres constants (PD0-GPD0).

Les quantiles obtenus avec les modèles couplés PD-GPD sont calculés avec l'inverse de la fonction de densité cumulative de la GEV reparamétrisée ; D'où leur grandeur de maxima annuels. Toutefois, l'estimation des paramètres de ces modèles PD-GPD est basée sur l'échantillon des dépassements de seuil et non des MA. Néanmoins, on observe les mêmes formes de quantiles conditionnels au niveau des modèles GPD et PD-GPD avec covariable.

Lorsque les corrélations de l'indice (AO ou PNA) avec V1 et V2 sont positives, et même lorsque la corrélation de l'indice est négative avec V1, mais positive avec V2, la courbe des quantiles obtenue à partir du modèle PD1-GPD1, est en-dessous (et au-dessus) de celle du modèle PD0-GPD1 durant la phase négative (et positive) de l'indice climatique (annexe 34). Toutefois, une relation inverse est observée lorsque la corrélation de l'indice avec les séries V1 et V2 est négative, et même lorsqu'elle est positive avec V1, et négative avec V2 (annexe 35).

4.5 Modèles hybrides GPD et estimation Bayésienne des quantiles

Le développement des modèles hybrides Gamma-GPD (modèles M1 et M2) et Normale-GPD (modèle M3) a permis de définir le seuil comme un paramètre à estimer, ce qui implique que l'incertitude qui s'y associe est quantifiée dans le processus de la modélisation. A la différence du modèle GPD classique (M0) où le seuil a été choisi de façon arbitraire comme équivalent successivement à la médiane, au 90^{ième} percentile puis au 99^{ième} percentile (et ce pour des fins d'analyse de sensibilité), l'utilisation des modèles hybrides proposés a permis une détermination objective et automatique de la valeur du seuil. La figure 4 illustre les modèles M1 et M2. Les valeurs théoriques utilisées pour leurs différents paramètres y sont présentées avec en abscisse les observations générées à partir de chaque mélange de modèles.

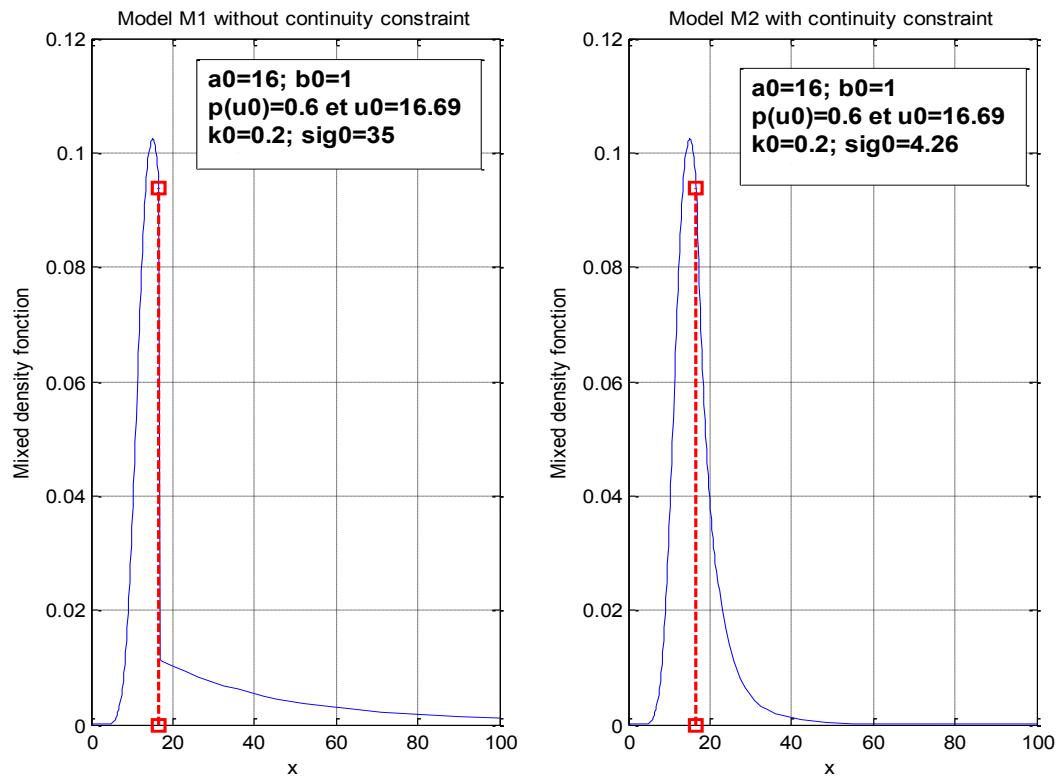


Figure 4: Exemple illustratif des modèles M1 (gauche) et M2 (droite) avec la fonction de densité de probabilité en bleu et la location du seuil estimé en rouge

Avec de tels modèles hybrides, c'est l'ensemble des observations disponibles qui est considéré pour l'estimation de tous les paramètres le seuil y compris, tandis qu'avec l'approche POT traditionnelle, c'est l'échantillon des dépassements de seuil qui est utilisé pour l'inférence statistique.

La comparaison des modèles M0, M1, M2 et M3 au moyen d'une étude de simulations a montré que les modèles M1 et M2 permettent d'estimer des quantiles avec des biais plus faibles que ceux obtenus avec les modèles M3 et M0, quand on compare les valeurs estimées aux valeurs théoriques choisies. L'évaluation des erreurs d'estimation des quantiles lorsqu'on génère l'échantillon à partir d'une Gamma asymétrique ($Cs=3$ et $a_0=0.44$) et qu'on donne peu de poids à la GPD (scénario 1 avec $K_0=0.05$), montre qu'il y a un risque de surestimation des quantiles important avec M0, et faible avec M1 et M2. Toutefois, lorsqu'on donne plus de poids à la GPD (scénario 2 avec $k_0=0.2$), il n'y a pas de changement avec M0, mais on remarque une sous-estimation importante des quantiles avec M1 et M2. Lorsqu'on génère et qu'on estime à partir du modèle M3 (Normale-GPD), il y a une grande sous-estimation des quantiles peu importe le poids qu'on donne à la GPD. Le tableau 6 résume ces résultats de comparaisons avec le quantile de période de retour 10 ans (Q10) lorsqu'on génère les données et qu'on les ajuste à partir du même modèle.

Tableau 6 : Comparaison des valeurs du RB et du RRMSE des quantiles pour l'erreur de type I

Modèles	M0	M1	M2	M3
Scenario 1		u0=3.58; k0=0.05 ; β0=3.42		u0=0.84; k0=0.05 ; β0=0.71
Résultats	RB(Q10)=53% RRMSE(Q10)=8%	RB(Q10)=2% RRMSE(Q10)=3%	RB(Q10)=12% RRMSE(Q10)=3%	RB(Q10)=-30% RRMSE(Q10)=1%
Scenario 2		u0=3.58; k0=0.2 ; β0=4.21		u0=0.84; k0=0.2 ; β0=0.71
Résultats	RB(Q10)=51% RRMSE(Q10)=11%	RB(Q10)=-30% RRMSE(Q10)=10%	RB(Q10)=-25% RRMSE(Q10)=5%	RB(Q10)=-57% RRMSE(Q10)=1%

Lorsque les données sont générées à partir d'un modèle beaucoup plus homogène (exemple de la GEV considérée dans cette analyse) que celui des modèles hybrides M1, M2 et M3, les biais des quantiles sont plus faibles avec le choix d'une GEV à queue légère (scénario 1 avec $k_0=0.1$) que celle à queue un peu plus lourde (scénario avec $k_0=0.2$). Toutefois, les quantiles sont sous-estimés avec les modèles M0, M1, M2 et M3. De plus, il y a une similarité dans les estimations des trois derniers modèles. Le tableau 7 résume ces résultats de comparaisons avec le quantile de période de retour 10 ans (Q10) lorsqu'on génère les données avec la GEV et qu'on les ajuste avec les modèles M0, M1, M2 et M3.

Tableau 7 : Comparaison des valeurs du RB et du RRMSE des quantiles pour l'erreur de type II

Modèles	M0	M1	M2	M3	
Scenario 1		$k_0=0.1; \delta_0 = 100; \psi_0 = 1000$			
Résultats	RB(Q10)=-12%	RB(Q10)=-7%	RB(Q10)=-7%	RB(Q10)=-6%	
	RRMSE(Q10)=2%	RRMSE(Q10)=1%	RRMSE(Q10)=1%	RRMSE(Q10)=1%	
Scenario 2		$k_0=0.2; \delta_0 = 100; \psi_0 = 1000$			
Résultats	RB(Q10)=-20%	RB(Q10)=-25%	RB(Q10)=-25%	RB(Q10)=-24%	
	RRMSE(Q10)=3%	RRMSE(Q10)=1%	RRMSE(Q10)=1%	RRMSE(Q10)=1%	

La comparaison des valeurs de seuil estimées avec les modèles M1, M2 et M3 montre qu'elles sont deux fois plus petites en moyenne que la valeur équivalente au 90^{ème} percentile (M0). Toutefois, cela n'affecte pas beaucoup l'ordre de grandeur de la moyenne a posteriori des paramètres de forme et d'échelle de la GPD lorsque les données sont ajustées à M0, M1, M2 et M3. En effet, les valeurs des estimations sont comparables.

L'ajustement des modèles M1, M2 et M3 aux observations de la précipitation totale journalière (station 7025250) pour la période 1945-1992, a montré leur application pratique. Les figures 5 et 6 illustrent respectivement l'ajustement des modèles M1 et M2 à cette série d'observations journalières, avec en abscisse les valeurs journalières et en ordonnées la densité de probabilité. La fonction de densité de probabilité (fdp) est restée continue au point du seuil avec l'étude de simulation autant pour les modèles M2 et M3 où cette contrainte a été imposée, qu'avec le modèle M1 dont tous les paramètres étaient libres. Cependant, avec le cas d'étude pratique, la discontinuité au point du seuil de la fdp est bien marquée lorsque l'ajustement des données réelles est fait avec le modèle M1 (Figure 5).

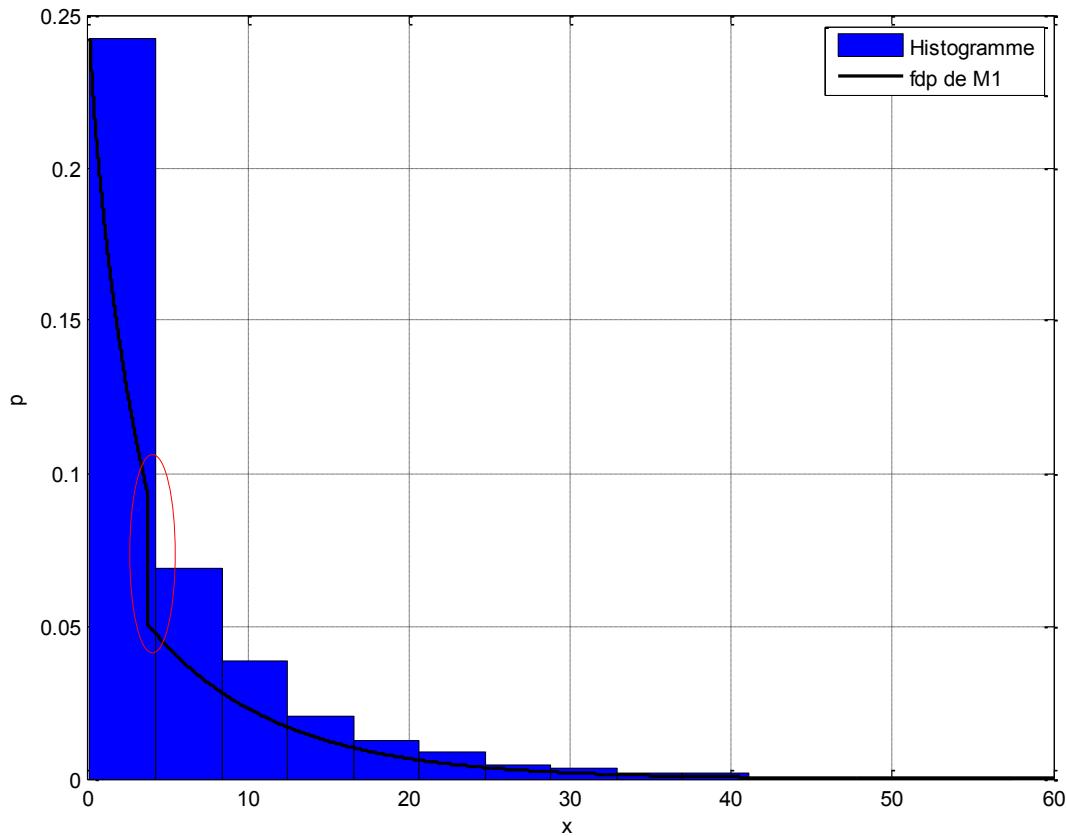


Figure 5: Ajustement du modèle M1 à la précipitation totale journalière de la station 7025250

Cette discontinuité marquée pourrait constituer une limite à l'utilisation de distribution paramétrique pour les données en-dessous du seuil à estimer. Néanmoins, la discontinuité peut être corrigée avec l'approche du modèle M2 (Figure 6).

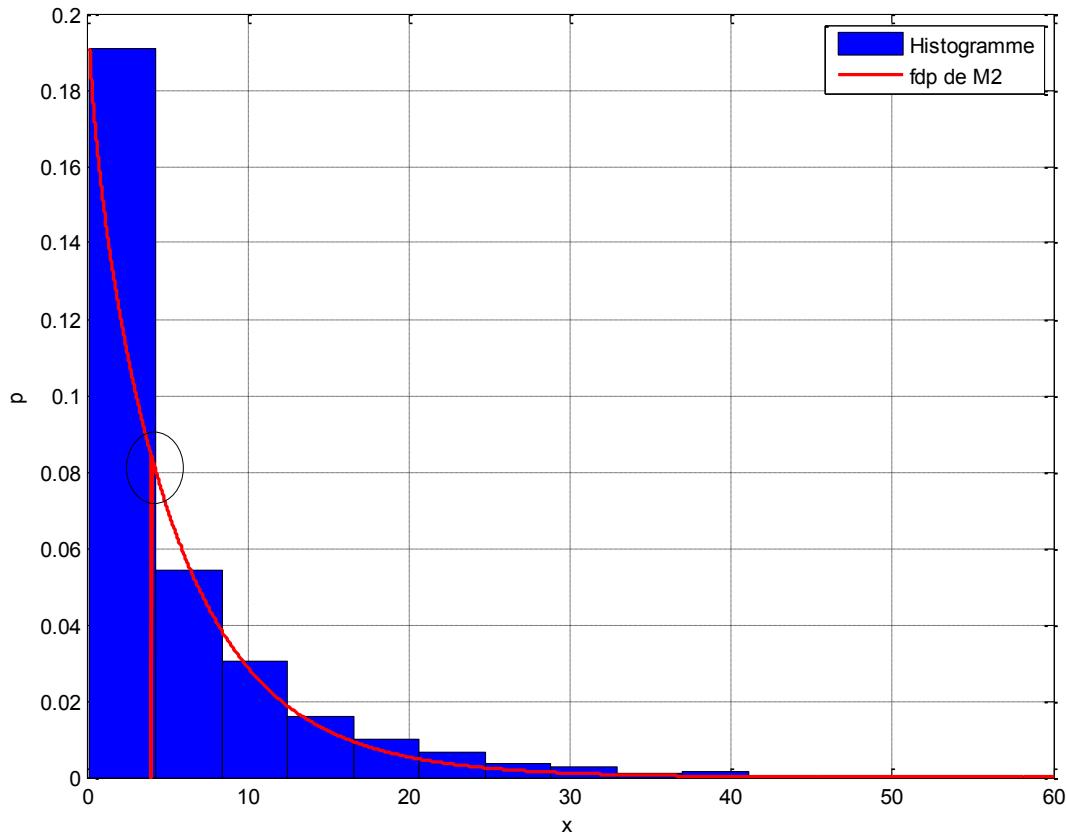


Figure 6: Ajustement du modèle M2 à la précipitation totale journalière de la station 7025250

Avec le modèle M3, des valeurs négatives peuvent être estimées, ce qui serait absurde pour des variables hydrologiques. La combinaison d'une Gamma-GPD comme modèle hybride paramétrique est donc à privilégier au sortir de cette analyse.

5 CONCLUSION GÉNÉRALE

5.1 Synthèse

La majorité des développements en analyse fréquentielle non-stationnaire est basée sur l'approche des maxima annuels que celle des dépassements de seuil. Etant donné l'équivalence asymptotique de ces deux méthodes d'analyse statistique des événements extrêmes, les modèles d'analyse fréquentielle développés dans le cadre de cette thèse ont utilisé l'approche des dépassements de seuil. Le premier modèle proposé est celui de Pareto Généralisé dont le paramètre d'échelle varie en fonction d'une covariable climatique. Pour déterminer cette covariable, le Sud-Est du Canada a été choisi comme région d'étude et la précipitation totale journalière comme variable d'intérêt. Au moyen d'analyse de corrélation des rangs de Kendall, la dépendance entre successivement les indices climatiques AMO, AO, NAO, PDO, PNA, SOI et WHWP, et des séries temporelles d'intensité de la précipitation extrême d'une part, puis celles de leur fréquence d'autre part, a été statistiquement testée. Cette analyse a été conduite à l'échelle de la région d'étude définie et a permis d'identifier les indices AO et PNA comme des patrons régionaux dominants ayant une influence significative sur la variabilité des observations de la précipitation. Afin de soutenir cette détection de relation de dépendance significative entre les précipitations extrêmes et les indices AO d'un côté et PNA d'un autre, l'analyse par ondelettes a été également effectuée. Celle-ci a permis d'observer une variabilité interannuelle dans les séries de données explorées, et de détecter également des périodicités de variabilités communes.

En utilisant des fonctions de dépendance semi-paramétriques, les fonctions B-splines, neuf modèles GPD-B-splines ont été évalués et ont permis d'identifier le modèle GPD1 comme le meilleur modèle en se basant sur la valeur du critère d'information d'Akaïke. La fonction B-spline de ce modèle comporte 2 nœuds internes et est une combinaison linéaire de polynômes de degré 1. Les quantiles estimés à partir de ce modèle GPD1 ont montré sa flexibilité en comparaison à celle du modèle GPD0 où tous les paramètres sont gardés constants.

Sur la base des résultats d'analyse des interactions entre les précipitations extrêmes et les indices, le second modèle PD-GPD a été développé afin de ramener l'analyse des quantiles à une échelle annuelle tout en étudiant l'effet de l'incorporation de la non-stationnarité (PD1) dans le paramètre d'intensité de la distribution de Poisson. Les résultats d'analyses avec le modèle PD1-GPD1, ont indiqué des changements dans la valeur des quantiles conditionnels entre les deux phases de la covariable climatique (AO ou PNA) selon l'étude de cas. Ces variations dépendent des signes de la corrélation détectée entre la covariable climatique et les variables V1 et V2 de la précipitation extrême.

Les modèles GPD1, PD1-GPD1 et PD0-GPD1, définis comme des modèles locaux à partir d'étude de cas, ont permis de détecter des interactions aussi bien linéaires que non-linéaires entre les précipitations extrêmes et les indices AO et PNA dans le Sud-Est du Canada. La flexibilité de tels modèles font qu'ils pourraient faire l'objet d'une extension régionale afin de fournir un outil robuste pour la prédiction des quantiles d'intérêts dans le domaine de l'hydrologie. A cet effet, les processus physiques sous-jacents à ces interactions doivent être étudiés. Par ailleurs, en raison de la concurrence remarquée de plusieurs indices dans la région d'étude, l'effet de plus d'une covariable climatique pourrait également être exploré.

Les modèles GPD et PD-GPD proposés dans le cadre stationnaire et non-stationnaire ont été développés sur la base de l'approche POT classique qui consiste à fixer le seuil à l'avance avant de procéder à l'estimation des paramètres du modèle sur la base de l'échantillon des dépassements de seuil. Le troisième modèle développé dans cette thèse a permis de proposer une approche d'analyse fréquentielle où le seuil est estimé de façon explicite et objective. Il s'agit d'un modèle hybride à deux composantes paramétriques dont la seconde est ajustée par la GPD. L'ajustement de la première composante avec une distribution Gamma puis une distribution Normale a permis de suggérer la première pour plus de robustesse. Ainsi, avec le modèle Gamma-GPD proposé, c'est l'ensemble des observations disponibles qui est utilisé pour estimer tous les paramètres de ce modèle hybride, le seuil y compris.

Toutefois, la justification asymptotique complète de ces types de modèle n'est pas encore établie en raison de la prise en compte des observations non-extrêmes (valeurs en-dessous du seuil). Ainsi, la multiplication des études de simulation et des cas d'application à des données réelles permettrait de mieux les analyser et les comprendre, afin d'améliorer leur performance.

5.2 Retombées pratiques

Les travaux de recherche effectués dans le cadre de cette thèse ont permis d'une part de proposer une première évaluation de l'utilisation des fonctions B-splines pour développer des modèles POT non-stationnaires flexibles. En effet, les dépendances linéaires et non-linéaires ont été automatiquement explorées en faisant varier les paramètres de la fonction B-spline. De plus, un programme Matlab de sélection automatique du meilleur modèle sur la base du critère AIC a été développé. Ce sont donc deux outils d'analyse fréquentielle des extrêmes définis avec l'approche des dépassements de seuil, qui sont proposés. L'échantillonnage des extrêmes avec la méthode des blocs de maxima pouvant conduire à une perte en information, de tels développements sont une contribution innovante dans le domaine de l'analyse fréquentielle des extrêmes hydro-climatologiques avec l'approche des dépassements de seuil. Par ailleurs, les indices climatiques AO et PNA ont été identifiés comme covariables potentielles pour la modélisation des extrêmes de la précipitation totale journalière dans le Sud-est du Canada. Comprendre les mécanismes physiques à l'origine des relations de dépendance détectées entre les variables étudiées, serait donc un atout pour la prédiction des précipitations extrêmes dans la région mentionnée. D'autre part, le développement théorique de modèles hybrides proposé pour apporter une solution à la détermination du seuil, est une approche qui en présence de distributions ayant un comportement de mélange de modèles, donnerait des résultats intéressants.

Toutes les démarches méthodologiques proposées offrent également l'avantage d'étudier différentes variables hydrologiques et d'explorer d'autres régions.

5.3 Limites et perspectives

Les modèles POT non-stationnaires développés dans cette thèse ont considéré l'utilisation d'un seul indice climatique. Ceci pourrait être une limite au développement de modèles réalistes. Les indices climatiques ayant des interactions entre eux-mêmes, l'effet conjugué de deux ou plusieurs indices climatiques pourrait alors être modélisé pour évaluer la performance des modèles associés. L'analyse en ondelette qui permet de caractériser la variabilité dans un espace temps-échelle, est un outil intéressant qui pourrait être utilisé pour tenir compte de ce genre d'interactions.

L'analyse de corrélation effectuée pour la sélection de la covariable a montré certaines incohérences spatiales. C'est le cas de l'observation de corrélations positive et négative à des stations proches les unes des autres. La longueur des séries de même que les périodes d'analyse différentes d'un site à un autre pourraient en être la cause. L'utilisation d'une base de données beaucoup plus homogène pourrait donc montrer de meilleurs résultats.

La discussion de la physique climatique entre les indices et les processus hydrologiques est limitée dans cette thèse. Des efforts sur ce point aideraient à donner un poids décisionnel à l'utilisation de covariables climatiques comme AO et PNA pour la modélisation des précipitations dans le sud-est du Canada.

Les modèles hybrides proposés dans le cadre de cette thèse reposent sur un fondement purement théorique. Les résultats de l'étude de simulation peuvent donc être moins réalistes que ce qui est souhaité, comme par exemple l'estimation de valeurs de seuil très faibles, pouvant remettre en cause la validité de l'indépendance des dépassements de seuil. Il serait donc adéquat de procéder à une analyse préliminaire de différentes séries de données hydrologiques pour proposer des valeurs empiriques des paramètres des modèles hybrides. Ces valeurs correspondraient ainsi à ce qui est observé en réalité pour des études de simulation réalistes.

Avec le modèle M1, on observe une discontinuité marquée au point du seuil. Cette situation peut s'expliquer par le fait qu'on combine deux phénomènes qui sont complètement différents (exemple de la Gamma-GPD). Toutefois, lorsqu'on impose une contrainte de continuité comme c'est le cas avec le modèle M2, on peut affaiblir la flexibilité de la GPD (exemple des biais négatifs observés au niveau de la comparaison des quantiles). Pour mieux évaluer la performance des modèles hybrides adaptés dans cette thèse, l'utilisation de distributions empiriques pouvant être représentées par un mélange de modèles, pourrait faire l'objet de cas d'application intéressants. L'évaluation des modèles hybrides dans un cadre non-stationnaire serait alors un sujet de recherche à explorer en hydrologie.

6 BIBLIOGRAPHIE GÉNÉRALE

1. Anctil F, Coulibaly P (2004) Wavelet analysis of the interannual variability in southern-Quebec streamflow. *American Meteorological Society*, 163-173.
2. AghaKouchak A et al. (eds.) (2013) *Extremes in a changing climate: detection, analysis and uncertainty*. Water Science and Technology Library 65. 423p.
3. Alexander LV et al. (2006) Global observed changes in daily climate extremes of temperature and precipitation. *Journal of Geophysical Research* 111, D05109. DOI :10.1029/2005JD006290.
4. Anctil F, Rousselle J, Lauzon N (2005) *Hydrologie: cheminements de l'eau*. Presses Internationales Polytechniques. 288p.
5. Aissaoui-Fqayey I, El Adlouni S, Ouarda TBMJ, St-Hilaire A (2009) Développement du modèle log-normal non-stationnaire et comparaison avec le modèle GEV non-stationnaire. *Hydrological Sciences Journal* 54(6) :1141-1156.
6. Assani AA, Lajoie F, Vadnais M-E, Beauchamp G (2008) Analyse de l'influence de l'oscillation Arctique sur la variabilité interannuelle des précipitations dans le bassin versant de la rivière Saint-François (Québec, Canada) au moyen de la méthode des corrélations canoniques. *Revue des sciences de l'eau* 21(1):19-33
7. Barrow E, Maxwell B, Gachon P (Eds) (2004) *La variabilité et le changement climatique au Canada: le passé, le présent et le futur*. DSCA Séries d'évaluation scientifique n2, Service de Météorologie du Canada, Environnement Canada, Toronto, Ontario, 114p.
8. Bates BC, Kundzewicz ZW, Wu S, Palutikof JP, Eds. (2008) *Climate Change and Water*. Technical Paper of the Intergovernmental Panel on Climate Change, IPCC Secretariat, Geneva. 200p.
9. Begueria S, Angulo-Martinez M, Vicente-Serrano SM, Lopez-Moreno JI, El-Kenawy A (2011) Assessing trends in extreme precipitation events intensity and magnitude using non-stationary peaks-over-threshold analysis: a case study in northeast Spain from 1930 to 2006. *Int J Climatol* 31:2102-2114.
10. Behrens Cibele N, Lopes Hedibert F, & Gamerman Dani (2004) Bayesian analysis of extreme events with threshold estimation. *Statistical modeling* 4:227-244.
11. Beranova R, Kysely J (2015) Links between circulation indices and precipitation in the Mediterranean in an ensemble of regional climate models. *Theor Appl Climatol*. DOI: 10.1007/s00704-015-1381-6
12. Bezak N, Brilly M, Sraj M (2014) Comparison between the peaks-over-threshold method and the annual maximum method for flood frequency analysis. *Hydrological Sciences Journal* 59(5):959-977.
13. Bonsal B, Shabbar A (2011) *Large-scale climate oscillations influencing Canada, 1900-2008*. *Canadian Biodiversity : Ecosystem Status and Trends 2010*. Technical Thematic Report No.4. Canadian Councils of Resource Ministers. Ottawa. 15p.

14. Bonsal BR, Shabbar A (2008) Impacts of large-scale circulation variability on low streamflows over Canada: a review. *Canadian Water Resources Journal* 33:137-154.
15. Bonsal B, Shabbar A (2004) Associations between low frequency variability modes and winter temperature extremes in Canada. *Atmosphere-Ocean* 42 (2):127-140. DOI: 10.3137/ao.420204
16. Bonsal BR, Shabbar A, Higuchi K (2001) Impacts of low frequency variability modes on Canadian winter temperature. *International Journal of Climatology* 21:95-108.
17. Caissie D, El-Jabi N (1992) Étude sur le choix du seuil de troncature en analyse des séries de durées partielles: application au Canada. *Revue des Sciences de l'eau* 5(2):291-307.
18. Cannon AJ (2015) Revisiting the nonlinear relationship between ENSO and winter extreme station precipitation in North America. *International Journal of Climatology*. DOI: 10.1002/joc.4263
19. Carreau J, Bengio Y (2009) A hybrid Pareto model for asymmetric fat-tailed data: the univariate case. *Extremes* 12:53-76.
20. Chandran A, Basha G, Ouarda T (2015) Influence of climate oscillations on temperature and precipitation over the United Arab Emirates. *International Journal of Climatology*. DOI:10.1002/joc.4339
21. Chavez-Demoulin V, Davison AC (2012) Modeling time series extremes. *REVSTAT-Statistical Journal* 10(1):109-133.
22. Chavez-Demoulin V, Davison AC (2005) Generalized additive modeling of sample extremes. *Appl. Statist* 54, Part 1: 207-222.
23. Chen L, Singh VP, Guo S (2012) Measure of correlation between River flows using the copula-entropy method. *J. Hydrol. Eng.* 18 (12):1591-1606.
24. Chu Pao-Shin, Zhao Xin (2011) Bayesian analysis for extreme climatic events: a review. *Atmospheric Research* 102:243-262.
25. Coles SG (2001) *An introduction to statistical modeling of extreme values*. Springer Series in Statistics. 208p.
26. Coulibaly P (2006) Spatial and temporal variability of Canadian season precipitation (1900-2000). *Advances in Water Resources* 29:1846-1865. DOI: 10.1016/j.advwatres.2005.12.013
27. Coulibaly P, Burn DH (2004) Wavelet analysis of variability in annual Canadian streamflows. *Water Resources Research* 40. W03105. DOI:10.1029-2003WR002667
28. Cunnane C (1979) A note on the Poisson assumption in partial duration series models. *Water Resour. Res.* 15(2):489-493.
29. Davison AC, Smith RL (1990) Models for exceedances over high thresholds. *Journal of the Royal Statistical Society. Series B (Methodological)* 52(3):393-442.
30. de Boor C (2001) *A practical guide to spline*. Springer, London, 208p.

31. de Zea Bermudez P, Turkman MAA (2003) Bayesian approach to parameter estimation of the generalized pareto distribution. *Test* 12: 259-277.
32. Do Nascimento FF, Gamerman D, Lopez HF (2012) A semi-parametric Bayesian approach to extreme value estimation. *Stat Comp* 22:661-675.
33. Dörte J (2013) *Nonstationarity in extremes and engineering design*. In : AghaKouchak A et al. (eds.) *Extremes in changing climate : detection, analysis and uncertainty*. Springer, pp 363-417.
34. Drumond A, Nieto R, Gimeno L (2011) On the contribution of the Tropical Western Hemisphere Warm Pool source of moisture to the Northern Hemisphere precipitation through a Lagrangian approach. *Journal of Geophysical Research* 116. D00Q04. DOI:10.1029/2010JD015397
35. Easterling DR, Evans JL, Groisman PY, Karl TR, Kunkel KE, Ambenje P (2000) Observed variability and trends in extreme climate events: a brief review. *Bulletin of the American Meteorological Society* 81(3): 417-425.
36. El Adlouni S, Ezzahid E, Mouatassim Y (2011) Mixed distribution for loss severity modeling with zeros in the operational risk loss. *International Journal of Applied Mathematics & Statistics* 21(J11). ISSN 0973-7545 (Online).
37. El Adlouni S, Ouarda TBMJ (2009) Joint Bayesian model selection and parameter estimation of the generalized extreme value model with covariates using birth-death Markov chain Monte Carlo. *Water Resources Research* 45. DOI:10.1029/2007WR006427
38. El Adlouni S, Ouarda TBMJ (2008) Comparaison des méthodes d'estimation des paramètres du modèle GEV non stationnaire. *Revue des Sciences de l'Eau* 21(1):35-50.
39. El Adlouni S, Ouarda TBMJ, Zhang X, Roy R, Bobée B (2007) Generalized maximum likelihood estimators for the nonstationary generalized extreme value model. *Water Resources Research* 43:1-13.
40. Fisher RA, Tippett LHC (1928) Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Math Proc Cambridge Philos Soc* 24(02):180-190.
41. Frigessi A, Haug O, Rue H (2002) A dynamic mixture model for unsupervised tail estimation without threshold selection. *Extreme* 5(3):219-235.
42. GIEC (2007) *Bilan 2007 des changements climatiques*. Contribution des Groupes de travail I, II et III au quatrième Rapport d'évaluation du Groupe d'experts intergouvernemental sur l'évolution du climat (équipe de rédaction principale, Pachauri R.K, & Reisinger A. (publié sous la direction de ~)). GIEC, Genève.103p.
43. Grinsted A, Moore JC, Jevrejeva S (2004) Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear Processes in Geophysics* 11:561-566.

44. Hastings W (1970) Monte Carlo sampling method using Markov Chains and their applications. *Biometrika* 55:97-109.
45. Hosking JRM, Wallis TR (1987) Parameter and quantile estimation for the Generalized Pareto distribution. *Technometrics* 29(3):339-349.
46. IPCC (2012) *Managing the risk of extreme events and disasters to advance climate change adaptation*. Special Report of the Intergovernmental Panel on Climate Change (IPCC). A special report of working groups I, II of the IPCC.... Cambridge University Press, Cambridge, UK, and New York, NY, USA. 582p.
47. Jenkinson AF (1955) The frequency distribution of the annual maximum (or minimum) of meteorological elements. *Quart J Roy Meteor Soc* 81:158-171.
48. Katz RW (2013) *Statistical methods for nonstationary extremes*. In : AghaKouchak A et al. (eds.) Extremes in changing climate : detection, analysis and uncertainty. Springer, pp 15-37.
49. Katz RW, Parlange MB, Naveau P (2002) Statistics of extremes in hydrology. *Advances in Water Resources* 25:1287-1304.
50. Kenyon J, Hegerl GC (2010) Influence of modes of climate variability on global precipitation extremes. *Journal of Climate* 23:6248-6262.
51. Khalil MN, Ouarda TBMJ, Ondo J-C, Gachon P, Bobée B (2006) Frequency analysis of a sequence of dependent and/or non-stationary hydro-meteorological observation: a review. *Journal of Hydrology* 329:534-552.
52. Lang M, Ouarda TBMJ, Bobée B (1999) Towards operational guidelines for over-threshold modeling. *Journal of Hydrology* 225:103-117.
53. Liseo B (1993) Elimination of nuisance parameters with reference non-informative priors. *Biometrika* 80: 295-304.
54. Lee T, Ouarda TBMJ, Li J (2013) An orchestrated climate song from the Pacific and Atlantic oceans and its implication on climatological processes. *International Journal of Climatology* 33:1015-1020.
55. Mabille G, Fettweis X, Erpicum M, Nicolay S (2012) Étude fréquentielle de données via la transformée en ondelette : application aux cycles climatiques. *BSGLg* 58 :5-15.
56. MacDonald A (2012) *Extreme value mixture modeling with medical and industrial applications*. PhD thesis, university of Canterbury, New Zealand.
57. MacDonald A, Scarrott CJ, Lee D, Darlow B, Reale M, Russell G (2011) A flexible extreme value mixture model. *Computational Statistics and Data Analysis* 55:2137-2157.
58. Mailhot A, Lachance-Cloutier S, Talbot G, & Favre A-C (2013) Regional estimates of intense rainfall based on the peak-over-threshold (POT) approach. *Journal of Hydrology* 476: 188-199.

59. Mailhot A, Duchesne S (2005) Impacts et enjeux liés aux changements climatiques en matière de gestion des eaux en milieu urbain. *VertigO - la revue électronique en sciences de l'environnement*, Hors-série 2. URL : <http://vertigo.revues.org/1931>. DOI: 10.4000/vertigo.1931
60. Martins ES, Stedinger JR (2001) Generalized maximum likelihood Pareto-Poisson estimators for partial duration series. *Water Resources Research* 37(10):2551-2557.
61. Martins Eduardo S, Stedinger Jery R (2000) Generalized maximum likelihood generalized extreme value quantile estimators for hydrological data. *Water Resources Research* 36(3):737-744.
62. Mekis E, Vincent L (2011) An overview of the second generation adjusted daily precipitation dataset for trend analysis in Canada. *Atmosphere-Ocean* 49(2):163-177.
63. Metropolis N, Rosenbluth A, Teller A, Teller E (1953) Equations of state calculation by fast computing machines. *Journal of Chemical Physics* 21:1087-1092.
64. Meylan P, Favre A.-C, & Musy A (2008) Hydrologie fréquentielle: une science prédictive. Presses Polytechniques et Universitaires Romandes. 173p.
65. Mondal A, Mujumdar P.P (2015) Modeling non-stationarity in intensity, duration and frequency of extreme rainfall over India. *Journal of Hydrology* 521:217-231.
66. Nasri B, El Adlouni S, Ouarda TBMJ (2013) Bayesian estimation for GEV-B-Spline model. *Open Journal of Statistics* 3:118-128.
67. Naveau P. et al. (2005) Statistical methods for the analysis of climate extremes. *C.R. Geoscience* 337:1013-1022.
68. Onibon H, Ouarda TBMJ, Barbet Marc, St-Hilaire A, Bobée B, Bruneau P (2004) Analyse fréquentielle régionale des précipitations journalières maximales annuelles au Québec, Canada. *Journal des Sciences Hydrologiques* 49(4):717-735.
69. Önöz B, Bayazit M (2001) Effect of the occurrence process of the peaks over threshold on the flood estimated. *Journal of Hydrology* 244(1-2):86-96.
70. Osman YZ, Rowan F, Sweeney JC (2013) Downscaling extreme precipitation in Ireland using combined peak-over-threshold generalised Pareto distribution model of varying parameters. *Journal of Water and Climate Change* 4(4):348-363.
71. Ouarda T.B.M.J, El Adlouni S (2011) Bayesian nonstationary frequency analysis of hydrological variables. *Journal of the American Water Resources Association* 47(3): 496-505.
72. Ouachani R, Bargaoui Z, Ouarda T (2011) Power of teleconnection patterns on precipitation and streamflow variability of upper Medjerda basin. *Int J Climatol.* doi: 10.1002/JOC.3407
73. Padoan SA, Wand MP (2008) Mixed model-based additive models for sample extremes. *Stat Probab Lett* 78(17):2850-2858.

74. Pickands J III (1975) Statistical inference using extreme order statistics. *Ann Stat* 3(1):119-131.
75. Pujol N, Neppel L, Sabatier R (2007) Approche régionale pour la détection de tendances dans les séries de précipitations de la région méditerranéenne française. *C.R. Geoscience* 339:651-658.
76. Racine JS (2014) A primer on regression splines. https://cran.r-project.org/web/packages/crs/.../spline_primer.pdf.
77. Rannie W (2016) The 1997 flood event in the Red river basin: causes, assessment and damages. *Canadian Water Resources Journal* 41(1-2):45-55. DOI: 10.1080/07011784.2015.1004198.
78. Ribatet M, Ouarda T.B.M.J, Sauquet E, Gresillon J.-M (2009) Modeling all exceedances above a threshold using an extremal dependence structure: inferences on several flood characteristics. *Water Resources Research* 45, W03407, doi:10.1029/2007WR006322.
79. Rootzén H, Katz RW (2013) Design life level: quantifying risk in a changing climate. *Water Resource Research* 49:5964-5972. DOI:10.1002/wrcr.20425
80. Rossi A, Massei N, Laignel B (2011) A synthesis of the time-scale variability of commonly used climate indices using continuous wavelet transform. *Global and Planetary Change* 78:1-13.
81. Rossi A, Massei N, Laignel B, Sebag D, Copard Y (2009) The response of the Mississippi River to climate fluctuations and reservoir construction as indicated by wavelet analysis of streamflow and suspended-sediment load, 1950-1975. *Journal of Hydrology* 377:237-244
82. Roth M, Buishand TA, Jongbloed G, Klein Tank AMG, van Zanten JH (2012). A regional peaks-over-threshold model in a nonstationary climate. *Water resources research* 48, W11533. Doi:10.1029/2012WR012214.
83. Salas JD, Obeysekera J (2014) Revisiting the concepts of return period and risk for nonstationary hydrological extreme events. *Journal of Hydrological Engineering* 19:554-568.
84. Scarrott C, MacDonald A (2012) A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT-Statistical Journal* 10(1):33-60.
85. Serinaldi F (2015) Dismissing return periods! *Stoch Environ Res Risk Assess* 29:1179-1189. DOI: 10.1007-s00477-014-0916-1
86. Serinaldi F, Kilsby CG (2015) Stationarity is undead: uncertainty dominates the distribution of extremes. *Adv What Resour* 77:17-36.
87. Shabbar A, Bonsal B, Khandekar M (1997) Canadian precipitation patterns associated with the Southern oscillation. *Journal of Climate* 10:3016-3027.
88. Shabbar A, Khandekar M (1996) The impact of El Nino-Southern Oscillation on the temperature field over Canada. *Atmosphere-Ocean* 34:401-416.

89. Silva, A.T., Naghettini, M., Portela, M.M., 2016. On some aspects of peaks-over-threshold modeling of floods under nonstationarity using climate covariates. *Stoch Environ Res Risk Assess* 30, 207-224. doi: 10.1007/s00477-015-1072-y
90. Soulis ED, Sarhadi A, TINEL M, Suthar M (2016) Extreme precipitation time trends in Ontario, 1960-2010. *Hydrological Processes*. DOI: 10.1002/hyp.10969
91. Stone DA, Andrew JW, Zwiers FW (2000) Trends in Canadian precipitation intensity. *Atmosphere-Ocean* 38(2):321-347.
92. Sugahara S, da Rocha RP, Silveira R (2009) Non-stationary frequency analysis of extreme daily rainfall in Sao Paulo, Brazil. *International Journal of Climatology* 29:1339-1349.
93. Tancredi A, Anderson C, O'Hagan A (2006) Accounting for threshold uncertainty in extreme value estimation. *Extremes* 9:87-106.
94. Teegavarapu Ramesh SV (2012) *Floods in a changing climate*: extreme precipitation. International Hydrology Series. Cambridge University Press. 270p.
95. Thistle ME, Caissie D (2013) Trends in air temperature, total precipitation and streamflow characteristics in Eastern Canada. *Can Tech Rep Fish Aquat Sci* 3018.
96. Thompson DWJ, Wallace JM (1998) The Arctic Oscillation signature in the wintertime geopotential height and temperature fields. *Geophysical Research Letters* 25 :1297-1300.
97. Torrence C, Compo GP (1998) A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society* 79:61-78.
98. Tremblay L, Larocque M, Anctil F, Rivard C (2011) Teleconnection and interannual variability in canadian groundwater levels. *Journal of Hydrology* 410: 178-188.
99. Trenberth KE, Jones PD, Ambenje P, Bojariu R, Easterling D, Tank AK, Parker D, Rahimzadeh F, Renwick JA, Rusticucci M, Soden B, Zhai P (2007) *Observations: Surface and Atmospheric Climate Change*. In: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller (ed)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. 336p.
100. Vincent L, Mekis E (2006) Changes in daily and extreme temperature and precipitation indices for Canada over the twentieth century. *Atmosphere-Ocean* 44(2):177-193.
101. Wang C, Enfield DB (2001) The tropical Western Hemisphere warm pool. *Geophysical Research Letters* 28:1635-1638.
102. Wazney L, Clark SP (2016) The 2009 flood event in the Rev river basin: causes, assessment and damages. *Canadian Water Resources Journal* 41(1-2):56-64. DOI: 10.1080/07011784.2015.1009949

103. Yue S, Pilon P, Cavadias G (2002) Power of the Mann-Kendall and Spearman's rho tests for detecting monotonic trends in hydrological series. *Journal of Hydrology* 259:254-271.
104. Zhang X, Wang J, Zwiers FW, Ya Groisman P (2010) The influence of large-scale climate variability on winter maximum daily precipitation over North America. *Journal of Climate* 23:2902-2915.
105. Zhang X, Hogg WD, Mekis E (2001) Spatial and temporal characteristics of heavy precipitation events over Canada. *Journal of Climate* 14:1923-1936.
106. Zhang X, Vincent LA, Hogg WD, Niitsoo A (2000) Temperature and precipitation trends in Canada during the 20th century. *Atmosphere-Ocean* 38(3):395-429. DOI: 10.1080/07055900.2000.9649654
107. Zoglat A, El Adlouni S, Badaoui F, Amar A, Okou CG (2014) Managing hydrological risks with extreme modeling: application of peaks over threshold model to the Loukkos watershed, Morocco. *J Hydrol Eng* 05014010:1-9.

PARTIE 2 : ARTICLES SCIENTIFIQUES

7 ARTICLE 1: NONSTATIONARY FREQUENCY ANALYSIS OF EXTREME DAILY PRECIPITATION AMOUNTS IN SOUTHEASTERN CANADA USING A PEAKS-OVER-THRESHOLD APPROACH

Alida N Thiombiano (@)⁽¹⁾, Salaheddine El Adlouni⁽²⁾, André St-Hilaire⁽¹⁾, Taha B.M.J Ouarda⁽¹⁾ and Nassir El-Jabi⁽²⁾

⁽¹⁾ Centre Eau–Terre–Environnement, Institut National de la Recherche Scientifique, 490, de la Couronne Street, Québec, Québec, G1K 9A9, Canada

⁽²⁾ Université de Moncton, 18 Antonine-Maillet Avenue, Moncton, New-Brunswick, Canada E1A 3EA.

E-mail de l'auteur de correspondance : alida_nadege.thiombiano@ete.inrs.ca

Article Publié dans: ***Theoretical and Applied Climatology* (2016): 124(1-2), doi: 10.1007/s00704-016-1789-7**

Acknowledgments: The Authors are grateful to the International Center for Research and Development (ICRD) and to the Natural Sciences and Engineering Research Council of Canada (NSERC) for the financial support through the FACE (*Faire Face Aux Changements Ensemble*) project grant.

Abstract

In this paper, statistical modeling of extreme daily precipitation amounts for Southeastern Canada is proposed using a nonstationary peaks-over-threshold approach. The Generalized Pareto Distribution (GPD) is fitted to excesses time series derived from annual averages of independent precipitation amounts events above a fixed threshold, the 99th percentile. The scale parameter of the GPD is allowed to vary as a function of a covariate. This association is modelled using B-spline functions. Nonlinear correlation and cross-wavelet analysis allowed identification of two dominant climate patterns as covariates for the study area, Arctic Oscillation (AO) and Pacific North American (PNA). The nonstationary frequency analysis showed an east-west behavior of the effect of the AO index on extreme daily precipitation amounts in the study area. Indeed higher quantiles of daily precipitation amounts are associated with the AO positive phase in Atlantic Canada, while those in more Southeastern part of Canada, especially in Southern Quebec and Ontario, are associated with the negative phase of the AO index. The negative phase of PNA also suggests significant correlation for these latter regions. Moreover, a regression analysis between AO (PNA) index and conditional quantiles provided slope values for on the one hand, the positive phase of the index and on the other hand, the negative phase. This allows computing a slope ratio which permits to sustain the nonlinear relation assumption between climate indices and precipitation, and the development of nonstationary GPD model for Southeastern Canada.

1 Introduction

Characterization of precipitation for present and future climates is very important for hydraulic engineering design and water management. As the main driver of the hydrologic cycle, precipitation is characterized by natural spatial and temporal inter-annual variability which can be amplified in the context of global warming (IPCC 2012), because weather and climate variables, particularly their extremes, are very likely to be impacted by climate change (CC) (IPCC 2012; Khaliq et al. 2006). For example, Bates et al. (2008) indicated that robust correlations between temperature and precipitation in many regions provide evidence that processes controlling the hydrological cycle are closely coupled. Déry and Wood (2004), and Stone et al. (2000) also notified that rising surface air temperature in response to anthropogenic forcing will lead to warmer and moister atmosphere which will lead to an intensification of the global hydrologic cycle, resulting in more extreme events.

In IPCC (2012), « *an extreme (weather or climate) event is generally defined as the occurrence of a value of a weather or climate variable above (or below) a threshold value near the upper (or lower) ends (tails) of the range of observed values of the variable* ». In fact, the threshold concept is widely used in the definition of extremes. Particularly in hydrology, many studies presented peaks-over-threshold (POT) approach as a good alternative to the most popular block maxima (BM) approach. The POT approach generally allows a larger number of extreme events than the BM approach (Begueria et al. 2011; Bezak et al. 2014; Coles 2001; Davison and Smith 1990; Katz 2013; Katz et al. 2002; Osman et al. 2013; Roth et al. 2012; Zoglat et al. 2014). More than one extreme value can be considered from each year. Particularly in the case of precipitation characterization, the annual cycle of precipitation can be better described with this approach. The BM approach usually leads to the characterization of extremes mostly originating from one particular season (Zhang et al., 2001).

It has then been shown that heavy daily precipitation (e. g. greater than the 95th percentile) events have become more intense and frequent in many regions of the world since 1950, even in places where total amounts have decreased (Bates et al. 2008; IPCC 2012). Moreover, some other studies indicated that the most consistent trends toward these heavier precipitation events were found in North America and that more precipitation now falls as rain rather than snow in Northern regions (IPCC 2012; Trenberth et al. 2007). Trenberth et al. (2007) have even shown that it has become significantly wetter in Eastern parts of North America. Shabbar et al. (1997) have also concluded that Eastern Canada is the second wettest region of Canada after the Pacific coast. Recent studies provide more evidence about these statements, by concluding that the Canadian climate is getting gradually wetter and warmer (Mekis and Vincent 2011; Thistle and Caissie 2013; Zhang et al. 2011; Zhang et al. 2000). Indeed, many studies have shown evidence of precipitation events increasing in Canada and more substantially in Southern and Eastern Canada than in the rest of the country (Mekis and Vincent 2011; Stone et al. 2000; Thistle and Caissie 2013; Vincent and Mekis 2006; Zhang et al. 2011; Zhang et al. 2001; Zhang et al. 2000). Zhang et al. (2000) showed that annual precipitation amounts have increased from 5% to 35% in Canada over the 1950-1998 period. According to Mekis and Vincent (2011), in Canada, annual rainfall and snowfall increased by about 12% and 4% from 1950 to 2009 respectively. These percentages are respectively more than 8% and 6% in Southern Canada. Zhang et al. (2001) indicated that the number of extreme rainfall and snowfall events have increased respectively in the spring over Eastern Canada, and in autumn and winter over Northern Canada. According to Vincent and Mekis (2006), the number of days with precipitation amounts and with rain significantly increased by about 20 to 30 days at the majority of locations across Canada, and substantially everywhere in Southern Canada.

All these studies indicate that precipitation in Southeastern Canada shows significant generally increasing trends and a shift to more intense precipitation events. One of the consequences of this significant increase in extreme precipitation is related to flood risk. Indeed, heavy rain events and/or snowmelt can generate flash and urban floods (Bates et al. 2008) among others risks.

Therefore, in the context of CC, it is important to study variability of extremes precipitation in order to provide more reliable estimations for design purposes. Many studies have in the same way tackled frequency analysis of hydroclimatic variables in a nonstationary manner (Begueria et al. 2011; Cannon 2015; Coles 2001; El Adlouni et al. 2007; Katz 2013; Khaliq et al. 2006; Osman et al. 2013; Ouachani et al. 2011, Salas and Obeysekera 2014; Sugahara et al. 2009) and concluded that nonstationary models allow improving statistical modeling of extreme values compared to the stationary models. Moreover, when distribution parameters depend on a climatic pattern rather than time as a covariate, further improvements have been noted (Begueria et al. 2011; Osman et al. 2013). Indeed, for a long time, most of this type of studies used time as a covariate for computing conditional quantiles, but now, low-frequency large-scale atmospheric and oceanic oscillations described through indices are getting great attention. In the Northern Hemisphere, the El Nino/ Southern Oscillation (ENSO), which is often measured by the Southern Oscillation Index (SOI), the Pacific Decadal Oscillation (PDO), the Pacific North American (PNA), the North Atlantic Oscillation (NAO) and the Arctic Oscillation (AO) indices, are among the most popular covariates used for statistical modeling. Bonsal and Shabbar (2011) showed that these popular indices, as well the Multi-decadal Oscillation (AMO), significantly impact Canadian climate. Indeed, according to Trenberth et al. (2007), changes with pattern of some of the atmospheric circulations are associated with CC and, indices involve and describe these variabilities. Therefore, many studies on such climatic patterns indicated that they may have a profound influence on surface climate and hydrological variables such as precipitation and streamflow, in many parts of the world (Anctil and Coulibaly 2004; Assani et al. 2008; Bates et al. 2008; Bonsal and Shabbar 2011, 2008; Cannon 2015; Déry and Wood 2005, 2004; Dörte 2013; Lee and Ouarda 2010; Ning and Bradley 2015; Ouachani et al. 2011; Teegavarapu 2013; Tremblay et al. 2011).

In light of this review, the present study has three objectives: ⁽¹⁾ identify regional climate indices showing significant relationships with extreme precipitation amounts events in Southeastern Canada, in order to use the best suited ones as covariate for precipitation extremes modeling across this region; ⁽²⁾ develop a nonstationary GPD model for that modeling; ⁽³⁾ propose a characterization of Southeastern Canada precipitation extremes conditional on the dominant climatic patterns.

In the next section of this paper, a description of the methodology adopted is provided, followed by data presentation in the 3rd section. Results are shown in the 4th section, and finally a discussion of the results with concluding remarks are presented in the last section.

2 Methods

2.1 Correlation analysis

Assuming the existence of a nonlinear relationship between precipitation and low-frequency indices of large-scale climate oscillations, nonlinear correlation analysis was performed between extreme precipitation amounts (as the dependent variable) and a set of indices hypothesized to be the potential covariates. The main reason for assuming such a hypothesis is related to the physical structure of the climate indices that typically shows a positive phase and a negative phase, often alternating at relatively low-frequency and generally related to pressure differences in atmospheric circulations. Indeed, many studies provided some evidence in such way as it is the case with Chandran et al. (2015) who found that the negative phase of the SOI play a significant role in the increase pattern of precipitation over the United Arab Emirates, while Canon (2015) showed that El Nino is associated with a decreased likelihood of extreme precipitation over the Great Lake (Ohio River valley) and Western Canada (Alaska). Shabbar et al. (1997) have also shown that winter precipitation in the upper St. Lawrence valley is enhanced during La Nina while no significant response occurring during El Nino years.

Another reason of assuming nonlinear association between variable of interest and indices is attributed to the possible combined influence of several climate indices operating at different temporal scales on hydroclimatic variables. For example, some studies showed that there are some strong associations between climate patterns such as NAO and AO, ENSO and PDO, ENSO and PNA (Assani et al. 2008; Bates et al. 2008; Bonsal and Shabbar 2011; Thompson and Wallace 1998; Trenberth et al. 2007). Given the fact that records of variable and covariates time series are different, a concomitant period of available data for both the variable and each covariate was first defined. Then, using Kendall's Tau, a non-parametric correlation measure based on the ranks of data, statistical association between extreme precipitation and each climate index is computed at a significant level of 5%.

2.2 Wavelet analysis

Wavelet analysis was demonstrated to be useful in the analysis of time series affected by nonstationarity (Grinsted et al. 2004; Rossi et al. 2011; Torrence and Combo 1998) and in the assessment of relationships between hydroclimatic variables and climate indices (Anctil and Coulibaly 2004; Chandran et al. 2015; Coulibaly 2006; Ouachani et al. 2011). In this study, continuous wavelet transform (CWT) was applied to both precipitation extremes and different climate indices of order to examine the dominant modes in their variability. To better assess the dependence between these variable and covariables, cross-wavelet transform (XWT) is computed and wavelet transform coherence (WTC) permits to highlight the significant coherence of common features between them.

2.3 Nonstationary Generalized Pareto model

Extreme values theory (EVT) formulated by Fisher and Tippett (1928) provides asymptotic arguments for extreme values modeling purpose. This theory can be applied in hydro-climatology when data behave like the generalized extreme value (GEV) distribution (Jenkinson 1955) or the GPD (Pickands 1975). Both distributions are respectively used for annual maxima and excesses over a high threshold (Coles 2001). Moreover, the Balkema-de Haan-Pickands theorem (Balkema and de Haan 1974; Pickands 1975) demonstrates the asymptotic convergence of the excesses over a high threshold to a GPD. In this paper, the emphasis is on the POT approach and particularly the statistical modeling of the excesses distribution over a high threshold (herein called intensity). Indeed, the POT approach allows modeling two components, the intensity of POT independent events (difference between exceedance and threshold) and their frequency (usually annual rate of exceedance). These processes are generally approximated by a one dimensional Poisson process coupled with the GPD and usually named the Poisson-GP approach (Coles 2001; Katz 2013).

Let (X_1, X_2, \dots, X_n) be the independent and identically exceedances distribution of observed daily precipitation amounts, and (Y_1, Y_2, \dots, Y_n) be the corresponding series of threshold excesses. According to Chavez-Demoulin and Davison (2005), for a sufficiently high threshold u , the variable $Y_i = X_i - u$ is asymptotically GPD distributed, with $X_i > u$ and $i = 1, 2, \dots, n$, n being the total number of independent POT events. For $y > 0$ and $\sigma_u > 0$, $F(y)$ is the cumulative distribution function which can be expressed as follows in its stationary form:

$$F(y) = \begin{cases} 1 - \left[1 + \xi \frac{y}{\sigma_u}\right]^{-\frac{1}{\xi}}, & \xi \neq 0 \\ 1 - \exp\left(-\frac{y}{\sigma_u}\right), & \xi = 0 \end{cases} \quad (1)$$

Where y is a random variable, ξ and σ_u are respectively the shape and the conditional (to the threshold u) scale parameters of the GPD.

In the nonstationary case, the classical way to model change in the exceedances distribution is to keep the threshold constant and construct covariate-dependent parameters (Coles 2001; Roth et al. 2012) as shown in many studies (Pujol et al. 2007, Begueria et al. 2011, Osman et al. 2013). Therefore, the threshold has been fixed herein to a value equal to the 99th percentile of the observed daily precipitation amounts. Moreover, the GPD shape parameter was also kept constant, given the fact that there is a great risk of uncertainty in its estimation (Osman et al. 2013, Pujol et al. 2007, Sugahara et al. 2009). Only the GPD scale parameter was allowed to change conditional to the best selected covariate. The dependence structure between this parameter and the covariate will be represented by a B-splines function. It is a parametric curve derived from a linear combination of polynomial pieces, also called basis splines of degree d with k knots. B-splines functions have the advantages of being more rigorous and flexible and allow representing nonlinear dependence structures (De Boor 2001; Nasri et al. 2013). In the present study the GPD model with B-splines is considered to model the effect of covariates on precipitation extremes.

Let (Z_1, Z_2, \dots, Z_n) be the values of the covariate (a climate index in the present study). Assuming that the scale parameter of the GPD is a B-splines function of the covariate denoted $g(z)$, then:

$$\sigma_{uz} = g(z) = \sum_{i=0}^{k+d} \beta_i B_{i,d}(z), \quad z \in [t_0, t_{k+1}] \quad (2)$$

for $j = 0, 1, \dots, d$

$$B_{i,j} = \begin{cases} \begin{cases} 1, & \text{if } t_i \leq z < t_{i+1} \\ 0, & \text{otherwise} \end{cases} & \text{when } j = 0 \\ h_{i,j+1}(z)B_{i,j}(z) + [1 - h_{i+1,j+1}(z)]B_{i+1,j}(z) & \text{when } j = 0, 1, \dots, d \end{cases} \quad (3)$$

$$\text{With, } h_{i,j} = \begin{cases} \frac{z-t_i}{t_{i+j}-t_i} & \text{if } t_{i+j} \neq t_i \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

Where, considering $k + 2$ real values $t_i | i \in \mathbb{Z}$, t is a non-decreasing real number of the knot sequence; the $B_{i,j}$ are the i^{th} B-spline basis functions of order j ; the β_i are fitted parameters corresponding to control points i .

In this paper, the number of knots k varies from 2 to 4 in order to take into account the physical phases of climate indices, and the degree d of the B-spline basis on each interval j was set from 1 to 3. This prior selection allows more flexibility and respects the parsimony criterion.

So in the nonstationary case, equation (1) becomes,

$$F_z(y) = \begin{cases} 1 - \left[1 + \xi \frac{y}{\sigma_{u_z}}\right]^{-\frac{1}{\xi}}, & \xi \neq 0 \\ 1 - \exp\left(-\frac{y}{\sigma_{u_z}}\right), & \xi = 0 \end{cases} \quad (5)$$

Inversion of each expression of $F_z(y)$ in (5) allows estimation of p-return level x_p with exceedance probability $(1 - p)$ where $0 < p < 1$, and/or T-return period for a quantile x_T [$T = \frac{1}{1-p}$] years (Khaliq et al. 2006; Mondal and Mujumdar 2015; Osman et al. 2013; Sugahara et al. 2009). The first metric (x_p) is more appropriate in a nonstationary case according to the meanings of the T-return period (Khaliq et al. 2006; Salas and Obeysekera 2014). However, T-return period estimation with the POT approach needs the annual average number of exceedances, often called the rate parameter λ . This parameter was computed as a ratio between the total number of exceedances and the number of years of record (Khaliq et al. 2006). Different expressions of the quantile function are as follow:

$$\begin{cases} x_p = u + \frac{\sigma_{u_z}}{\xi} [(1 - p)^{-\xi} - 1], & \xi \neq 0 \\ x_T = u + \frac{\sigma_{u_z}}{\xi} [(\lambda T)^\xi - 1], & \xi \neq 0 \\ x_p = u + [\sigma_{u_z} \ln(1 - p)], & \xi = 0 \\ x_T = u + [\sigma_{u_z} \ln(\lambda T)], & \xi = 0 \end{cases} \quad (6)$$

2.4 Estimation of GPD parameters

Even if the maximum likelihood (ML) method is most commonly used to estimate parameters in nonstationary frequency analysis because of its asymptotic properties, the generalized maximum likelihood (GML) method is more suitable since it can include prior information on the shape parameter of the extremal distribution (El Adlouni and Ouarda 2008). Such conditions help to avoid absurd estimates of the shape parameter since the expression of the GEV distribution or GPD is different depending on the value of their shape. Therefore, Martins and Stedinger (2000) proposed as prior distribution for the shape of the GEV distribution, a Beta distribution centered on a value of - 0.1 with a restricted range of values [- 0.5; + 0.5]. This range includes the most probable values observed in hydrology and has also been used for the GPD (Martins and Stedinger 2001a, 2001b). It was also used by El Adlouni et al. (2007) in their approach to extend the GML to a nonstationary framework. The same approach was used in Canon (2015). The GML is a special case of the Bayesian approach. Therefore, the generalized likelihood function Gl related to the vector of parameters $\theta = (\sigma_{u_z}, \xi)$ has the following form:

$$Gl(\theta|y) = l(y|\theta) * \pi(\xi) \quad (7)$$

Where $l(y|\theta)$ and $\pi(\xi)$ are respectively the log-likelihood function and the prior distribution of the GPD shape parameter as shown below:

$$l(y|\theta) = \begin{cases} -\sum_{i=1}^n \left\{ \ln(\sigma_{u_z}) - \left[(1 + 1/\xi) \ln \left(1 + \xi \left(\frac{y_i}{\sigma_{u_z}} \right) \right) \right] \right\}, & \xi \neq 0 \\ -\sum_{i=1}^n \left\{ \ln(\sigma_{u_z}) - \left(\frac{y_i}{\sigma_{u_z}} \right) \right\}, & \xi = 0 \end{cases} \quad (8)$$

$$\pi(\xi) \sim Beta(u = 6; v = 9) \quad (9)$$

So, the GML estimators are the mode of empirical marginal distributions of the parameters resulting from the following posterior distribution:

$$\pi(\theta|y) = \frac{l((y|\theta))\pi(\theta)}{\int l(y|\theta)\pi(\theta)d\theta} \quad (10)$$

In the present study, the GML approach is considered to estimate the parameters of the GPD model with covariates using the Newton-Raphson iteration method (Hosking and Wallis 1987).

2.5 Selection of the best GPD model with B-splines

For developing GPD model with B-splines, nine models resulting from the combination of knots and degree (k, d) pairs, with $k = 2, 3, 4$ and $d = 1, 2, 3$, were tested: (2,1), (2,2), (2,3); (3,1), (3,2), (3,3); (4,1) (4,2) (4,3). The number of knots allows to control the change in the form of dependency (linear or nonlinear), while the choice of the degree indicates the type of the dependence (i.e. linear, quadratic, cubic, etc.). So, for each station, the optimal model was chosen based on the Akaike Information Criteria (AIC). Its general form is given by:

$$AIC(M) = -2l(M) + 2m \quad (11)$$

where m designates the number of parameters for each model M tested, and l is the value of the corresponding log-likelihood function.

For a GPD model with B-splines, the number of parameters can be computed as $m = [d + k] + 1$ and the equation (11) becomes:

$$AIC = -2l + 2[(d + k) + 1] \quad (12)$$

The best model is the one with the smallest AIC value (Katz 2013; Mondal and Mujumdar 2015).

All computations and figures reported in this paper were prepared using the MATLAB Toolbox, and map illustrations were created with the ArcMap Toolbox.

3 Data

3.1 Extreme daily precipitation amounts

Southeastern Canada daily precipitation extremes (in mm) originating from both snow and rain events were extracted from all stations located in Ontario (ON), Quebec (QC), Newfoundland and Labrador (NFL), Nova Scotia (NS), New Brunswick (NB) and Prince Edward Island (PEI). With such variable, liquid and solid precipitation events are considered and the daily scale can be useful for urban hydraulic management. (. Moreover, the use of full daily or hourly time series allows to maximize the advantages of using the POT approach (Eastoe and Tawn 2010). Precipitation data were obtained through Data Access Integration (DAI), a portal of Environment Canada. For the entire study domain, time series of daily precipitation amounts from 1057 stations with different sample size were obtained. Only stations with record length of at least 30 full years were considered. This criterion was chosen following World Meteorological Organization recommendations, but also to tackle missing data problem, because POT analysis requires complete time series (Easterling 2013; Serinaldi and Kilsby 2014). Only 329 stations met this criterion (Fig. 1). Daily precipitation amounts from these stations were used in the analysis.

From long series of daily precipitation amounts, extreme values were defined using the POT approach (Lang et al. 1999). The sampling POT events consist in all events with values equal or greater than a given high threshold. In the present study, the 99th percentile was used as the threshold. This percentile is recommended in the literature as the most suitable threshold when dealing with daily precipitation (Osman et al. 2013). Nevertheless, other high percentiles (i.e. 90th, 95th to 98th), which are often used in the literature (Begueria et al. 2011; Osman et al. 2013; Pujol et al. 2007; Serinaldi and Kilsby 2014; Sugahara et al. 2009), were also evaluated in a preliminary analysis for the Southeastern region. The optimal threshold choice was based on graphical methods as the autocorrelation function, the mean excess plot and the plot of the shape and scale parameters of the GPD as a function of the threshold (Coles 2001; Davison and Smith 1990; Lang et al. 1999).

A chi-square of goodness-of-fit test for assessing the adequacy of GPD distribution was applied to excesses over the 99th percentile. The test suggested that the null hypothesis (i.e. these data come from a GPD population) was not rejected at the 5% significance level. Indeed, one of the major limitations to the use of the POT approach concerns the optimal threshold choice, because the reliability of all results is conditional to the threshold choice. The threshold value controls the size of the POT sample, independence of events and the adequacy of the GPD (Begueria et al. 2011; Coles 2001; Lang et al. 1999; Osman et al. 2013; Scarrott and MacDonald 2012; Zoglat et al. 2014). Nevertheless, when working with daily precipitation data, a declustering approach, which consists of applying a minimum temporal separation between events, is recommended to systematically satisfy the independence criterion. Roth et al. (2012) suggest using 1 or 2 days as the separation interval. In this study, a temporal separation of 1 day was applied. In the case of two consecutive days, the event with the maximum value was retained. Indeed, POT classical approach requires that the events be independent and identically distributed and follow a GPD (Bezak et al. 2014; Lang et al. 1999; Osman et al. 2013; Pujol et al. 2007; Sugahara et al. 2009). After declustering, the autocorrelation function was evaluated from the new series, which led to the selection of 173 stations from the 329 originally considered. Finally, due to the high local variability in precipitation observations, extreme daily precipitation amounts series were defined as the annual average of independent POT events of daily precipitation amounts. This definition is certainly smoothed out the extreme events but not the extreme nature of data. Indeed, it allows better optimizing of correlation signals (Assani et al. (2008) between low-frequency climate indices and extreme daily precipitation amounts. Smoothed time series also help to identify spatial coherence in the signal given the fact that hydroclimatic variables are observed at small scales whereas climate indices are large scales patterns (Rossi et al. 2009).

3.2 Large scale climate indices

To characterize extreme precipitation amounts, climate indices were considered. These indices *quantify the temporal evolution of the climate process in a particular region* (Lee et al. 2013). Indeed, they represent the state of global atmospheric circulation patterns also known as teleconnections (Bates et al. 2008). In this paper, seven (07) indices were considered: ENSO measured by the Southern Oscillation Index (SOI), PDO, PNA, NAO, AO, AMO and the Western Hemisphere Warm Pool (WHWP). Data are monthly time series from the National Oceanic and Atmospheric Administration (NOAA). In this study, availability of data was from 1856 to 2013 for AMO, 1948 to 2013 for PDO and WHWP, 1950 to 2013 for AO, NAO, and PNA, and 1951 to 2013 for SOI. In order to optimize the assessment of correlation between extreme precipitation and these teleconnections, twelve 3-month moving annual average values, starting in December (e.g. December-January-February, January-February-March, and so on), were constructed (Chandran et al. 2015, Ouachani et al. 2011). Additionally, the 12-month average of the climate index was also tested as a covariate. Correlation analysis of each series of extreme precipitation and sub-series of climate indices led to the selection of 136 stations out of the 173. The selected stations are those showing significant correlation with at least one climate index (Table 1). Ultimately, the dominant regional climate pattern, as quantified by the climate index, was selected for this group of 136 stations.

4 Results

4.1 Extreme daily precipitation amounts impacted by regional dominant climatic patterns

4.1.1 Correlation analysis and covariate selection

The computation of Kendall's Tau at a significant level of 5% allows identifying 136 stations from which a significant nonlinear correlation was found between the annual average of independent daily precipitation amounts events above the 99th percentile and at least one covariate. From these 136 stations, the successive steps are followed for regional dominant climate patterns identification:

- (1) For each station, the seven covariates are classified from 1st importance to last one given significant Kendall's Tau values.
- (2) Keeping only the best covariate per station (climate index from which the highest Kendall's Tau is obtained), stations are sorted according to the most correlated climate index identified, as highlighted by Table 2.
- (3) Regional dominant patterns are then defined as those with the greatest number of stations, herein AO and PNA showed by the first two columns (gray cells) of Table 2. As we can see, NAO can also be identified but, many studies showed that there is a high significant correlation between AO and NAO (Assani et al. 2008; Bonsal and Shabbar 2011; Thompson and Wallace 1998) and moreover, NAO is a regional expression of AO (Thompson and Wallace 1998).
- (4) Stations for which the top index was not AO or PNA, were sorted by the rank of these two aforementioned indices in the list of correlated covariates, allowing then to add respectively 15 and 16 stations to primary groups of the 31 AO stations and the PNA 30 stations.

Fig. 2 illustrated then nonlinear correlation results for the 46 stations selected with AO index and the other group of 46 stations with PNA index.

In light of this figure, significant Kendall's Tau varies on average between ± 20 and 30% and there is an east-west behavior of dependence relation between extreme daily precipitation amounts and AO. Indeed, more positives correlations are observed in the Maritime Provinces (NB, NS, and PEI) while, these are mostly negative in Ontario and Southern Quebec territories. PNA index seems to have more significant and representative effects on extreme daily precipitation amounts in Ontario and Southern Quebec provinces. So, extreme daily precipitation amounts modeling results will be presented for each case by an illustrated example at highlighted stations in Fig. 1.

4.1.2 Wavelet and cross-wavelet analysis with AO pattern

Correlation analysis identified the AO climate index as the principal regional dominant pattern. Using extreme daily precipitation amounts data from 1951 to 2005 at station 8206200 located in Nova Scotia (as a representative case), with annual average window of AO index values from September to November of the same period (1951-2005), continuous wavelet transform (CWT) from each variable show that there are clearly common features in the wavelet power of these two time series such as the significant peak in the $\sim 5\text{-}7$ year band around 1980 (Fig. 3).

From these two CWTs, a Cross Wavelet Transform (XWT) graph and a Wavelet Coherence (WTC) plot are then constructed (Fig. 4) in order to explore for the first one the common power and relative phase in time-frequency space of the two time series, and to infer from the second the significant coherence of this common power.

The XWT plot between extreme daily precipitation amounts and AO index allows highlighting that the common features from the individual CWT stand out as being significant at the 5% level, given that significant common power is observed in the 2-7 year band around 1980. Cause and effect relation between the phenomena recorded in the time series are expected when the oscillations are phase locked.

Therefore, it is comforting that the XWT graph shows that precipitation and AO are in phase in all the sectors with high significant common power, confirming that there is a strong link between AO and precipitation. The squared WTC of AO and precipitation also shows that a section stands out as being significant in the range of 2-4 year.

4.2 AO index: a principal climatic pattern for Southeastern extreme precipitation modeling

As previously explained in the methodology section, nine (09) GPD models with B-splines function were developed using firstly the index AO as covariate for the GPD scale parameter at 46 stations. The AIC criterion was then used to choose the optimal model. The lowest AIC value was found for a B-spline function with 2 knots and 1 degree (2, 1) for more than 90% of this group of stations.

4.2.1 Extreme precipitation behavior in Maritime Provinces

Fig. 5 shows the conditional quantiles (Q_{ns}) compared to stationary quantiles (Q_s), at the station 8206200 located in Nova Scotia, respectively for no-exceedance probabilities $p_1= 0.5$, $p_2= 0.9$ and $p_3= 0.99$. In a stationary case, these probabilities are respectively equivalent to return time of 2, 10 and 100 years.

This graph is a representative behavior of predicted precipitation extremes in the targeted Maritime Provinces. It shows that precipitation extremes of this region are in general positively correlated to AO index with a significant Kendall tau of almost 25%. Moreover, the nonlinear dependence is highlighted given that the most extreme quantiles (above 60 mm) are conditional to positive phase of AO index while the lowest ones (between 40 and 60 mm) are observed during its negative phase. Finally, when assuming Stationarity in quantiles estimation, there is a significant risk to overestimate this quantity during negative phase of AO or underestimate it over AO positive phase.

4.2.2 Extreme precipitation behavior in Ontario and Quebec Provinces

Fig. 6 indicates through the illustrated example at station 6057592 located in Ontario, with precipitation data record extended from 1961 to 2011 and AO window being annual average values from February to April of the same period, that there is an opposite behavior of precipitation extremes with AO index phase, in more Southeastern territories of Canada like Ontario and Southern Quebec provinces. Indeed, precipitation extremes are overall negatively correlated to AO with a significant Kendall tau of 25%. So, the most extreme quantiles (above 50 mm) are conditional to negative phase of the index while the lowest ones (between 30 and 40 mm) are observed during its positive phase. For stationary quantiles estimation, there is a significant risk to generally underestimate or to overestimate them respectively during negative and positive phase of AO.

4.3 PNA: a climatic factor of Ontario and Southern Quebec extreme precipitation behavior

PNA index has instead a heterogeneous impact on precipitation extremes in more Southern parts of the study area. Indeed, precipitation extremes have a more significantly negative dependence with PNA in Southern part of Quebec and in Ontario, except in its Southern tip area. The station 6101494 which borders these two regions is used for a case illustration. Precipitation data record are from 1951 to 1989 and PNA window consisted of annual average of values from January to March for the same record period. Fig. 7 shows a nonlinear negative dependence between PNA and extreme precipitation with a Kendall tau of almost 30%. So, prediction of the most extreme precipitation there would be observed during negative phase of PNA as indicated by Qnsp2 and Qnsp3. Furthermore, nonstationary quantiles permit to avoid an underestimation of the most precipitation extremes (Qnsp3).

In the Southern tip of Ontario, precipitation extremes show a more significant positive correlation with PNA index. At the station 6158255 with PNA index window values being those of March to May annual average from 1965 to 2006, the most precipitation extremes are conditional to positive values of this index (Fig. 8).

4.4 Regression analysis between conditional quantiles and AO (PNA) phases values

To motivate the choice of nonlinear models developed in this study, ratio between positive and negative slopes derived from regression analysis between AO conditional quantiles and AO in one case, and PNA conditional quantiles and PNA in another, respectively for positive and negative phase of each index, was computed. Results indicate that the GPD nonstationary model with 2 knots and 1 degree (2,1) was significant at more than 80% of stations related to AO and PNA too (Fig. 9). Indeed, the ratio was significantly different than a value of 1, suggesting that the nonstationary model was overall better than the stationary ones.

Moreover, the sign of these slope ratio values was also analyzed. Indeed, when the ratio value is positive, this means that slope during positive and negative phase of the index has the same sign; in this case a monotonous dependence trend is generally observed. However, if the ratio's sign is negative, this trend is related to index's phase and therefore is nonlinear. This last behavior is detected at almost 30% AO related stations and more than 35% for PNA stations. Therefore, significant nonlinear effect (case of negative sign of slope's ratio) of AO or PNA was highlighted at almost 25% of the stations for each index (Fig. 10). This relatively small percentage can probably be explained in part by the length of data. Indeed, in the study of Cannon (2015), the stationary model was proposed for almost 50% of stations analyzed. Nevertheless, he mentioned that the evidence of a spatially coherent nonlinear association between ENSO and winter precipitation was greatest especially for longer precipitation lengths. Therefore, the assumption of a nonlinear dependence between climate indices and precipitation extremes was shown to be significant in Canada for Southeastern precipitation characterization.

5 Discussion

Extreme daily precipitation amounts in Southeastern Canada were characterized through Peaks-Over-Threshold model with covariate. Moreover, the evaluation of potential dominant climate index in this region was most significant for Maritime and southern parts of Ontario and Quebec provinces. AO and PNA are the two dominant climate indices that can be used to improve the estimation of extreme precipitation in Southeastern Canada. Most previous studies suggest that ENSO, PDO, PNA, NAO and AO can explain the behaviour of some hydroclimatic variables like temperature and streamflow, and to a lesser extent precipitation. These studies were focused on western and Northern or Northeastern Canada (Bonsal and Shabbar 2011; 2008; Déry and Wood 2005; 2004; Shabbar and Bonsal 2004; Ning and Bradley 2015; Shabbar et al. 1997; Stone et al. 2000; Tremblay et al. 2011; Zhang et al. 2010). In these parts of Canada, strong relationships were highlighted particularly in winter time.

In this paper, it has been shown that the most precipitation extremes, defined as the annual average of independent extreme daily amounts, are observed during positive phase of the AO index in Maritime Provinces and over its negative phase in Southern Quebec and Ontario. These results provide element that can be take into account for a more optimal prediction of precipitation in these areas. For example, they are consistent with those of Coulibaly (2006) who used wavelet and cross-wavelet analysis to describe nonstationary relationships between seasonal amounts of daily total precipitation and the dominant modes of climate variability (PNA, NAO, AO and ENSO) in the Northern hemisphere throughout Canada. Indeed, he found that prior to 1940, AO pattern showed significant correlations in Eastern Canada mainly in the 2-6 year band, regardless of the season, and only with the winter season after 1940. He suggested that the shift observed in the precipitation pattern around 1940 is more likely related to the change in the AO towards extremely negative phase.

Assani et al. (2008) also found that precipitation events are positively correlated to AO index in the Saint-Francois River watershed, located in the south shore of the St. Lawrence River in Quebec. This basin shares some characteristics with the Maritime Provinces given to their geographical proximity. The Gulf of St. Lawrence and Fundy Bay bring warm and moist air masses and, it seems like the frequency of these masses increased with positive values of AO index, leading to more precipitation. Furthermore, an investigation of correlation between extreme precipitation amounts with some climate indices by Ning and Bradley (2015) over the Northeastern United States and Southeastern Canada showed that NAO, which is a regional expression of the AO (Thompson and Wallace 1998), is negatively correlated with precipitation extreme events over the Northern part of their study region which covered southern Ontario and Quebec.

The PNA index is another dominant pattern identified in this study that impacts precipitation extremes. Its effects are more significant in Ontario and Quebec where generally, the most extreme precipitation events are observed during the negative phase of this index. These results are in agreement with those of Stone et al. (2000) who indicated that during autumn and winter in Ontario and Southern Quebec, negative PNA generally leads to more precipitation extremes. Ning and Bradley (2015) also indicate that there is a negative correlation between precipitation extremes and PNA especially around the Great Lakes and more significantly over the Ohio valley.

This identification of AO and PNA as the dominant patterns in Southeastern Canada is consistent given that NAO and PNA are the two strongest and most robust modes of variability in atmospheric pressure and circulation in the extratropical Northern Hemisphere (Stone et al. 2000), and knowing that NAO is a regional expression of the AO. Indeed, Wallace and Gutzler (1981) noticed that NAO is always associated with a north-south-oriented dipolar structure in the pressure field over the Atlantic Ocean. It's also the case of AO but with an additional center of action over the Pacific Ocean. AO incorporates many of the NAO features and induces a stronger than usual zonal flow that affects Northern hemisphere meteorological conditions (Thompson and Wallace 1998; 2001).

However, it is important to notice that results presented and discussed herein remain related to aforementioned criteria defined for precipitation ground stations and the selected regional dominant patterns. Indeed, by selecting long records with no gaps, many stations were lost for subsequent analysis. Moreover, when looking for independent and identically generalized Pareto distributed series, other stations were abandoned given the results of the autocorrelation analysis. Finally, by considering the results of only significant correlation for identifying precipitation stations correlated to AO (PNA) index, modeling was achieved for a relatively small number of stations. Even if this can be seen as a limit to more representative regional characterization, it provided on the one hand robust results from the nonstationary frequency analysis made in this part of Canada. On the other hand, these results may lead to further studies where the combined effect of two or more regional pattern could be explored by also taking into account more stations.

6 Conclusion

Beginning with the assumption of a nonlinear relationship between extreme precipitation amounts and low frequency climatic patterns, this study explored correlation between several climate indices and specific measure of extreme precipitation. Two dominant patterns for Southeastern Canada were identified: AO and PNA. Hence, two groups of stations were constructed for extreme precipitation characterization of this region. The analysis consisted in the development of a GPD model with covariate for the entire region. A nonstationary GPD-B-spline model with 2 knots and 1 degree (2, 1) was then used and statistical modeling results allowed concluding that most extreme daily precipitation amounts in the Maritimes region of Canada are usually observed during the positive phase of the AO, while in more Southern regions of Canada specifically Southern Quebec and Ontario, this occur mostly during the negative phase of this index. Stations where the PNA was dominant show that the most extreme daily precipitation amounts are significantly conditional on negative phase of PNA in Ontario and Southern Quebec. These results are consistent with previous studies on the same subject.

The results of this study will contribute in the development of an approach for predicting hydroclimatic variables by integrating climate indices as explanatory covariates especially in the context of climate change. A perspective on this study would be to explore further the temporal and spatial behavior of AO and PNA indices in the context of climate change in order to highlight some specific characteristics for a more optimal use of these covariates as predictors for Southeastern Canadian precipitation.

References

- Anctil F, Coulibaly P (2004) Wavelet analysis of the interannual variability in Southern-Quebec streamflow. American Meteorological Society 163-173
- Assani AA, Lajoie F, Vadnais M-E, Beauchamp G (2008) Analyse de l'influence de l'oscillation Arctique sur la variabilité interannuelle des précipitations dans le bassin versant de la rivière Saint-François (Québec, Canada) au moyen de la méthode des corrélations canoniques. Revue des sciences de l'eau 21(1):19-33
- Balkema A.A, de Haan L (1974) Residual life time at great age. Ann Probab 2(5):792-804
- Bates BC, Kundzewicz ZW, Wu S, Palutikof JP, Eds. (2008) Climate Change and Water. Technical Paper of the Intergovernmental Panel on Climate Change, IPCC Secretariat, Geneva
- Begueria Santiago, Angulo-Martinez Marta, Vicente-Serrano Sergio M, Lopez-Moreno J. Ignacio, El-Kenawy Ahmed (2011) Assessing trends in extreme precipitation events intensity and magnitude using non-stationary peaks-over-threshold analysis: a case study in northeast Spain from 1930 to 2006. Int J Climatol 31:2102-2114
- Bezak N, Brilly M, Sraj M (2014) Comparison between the peaks-over-threshold method and the annual maximum method for flood frequency analysis. Hydrological Sciences Journal 59(5):959-977
- Bonsal B, Shabbar A (2011) Large-scale climate oscillations influencing Canada, 1900-2008. Canadian Biodiversity : Ecosystem Status and Trends 2010, Technical Thematic Report No.4. Canadian Councils of Resource Ministers. Ottawa, ON
- Bonsal B.R & Shabbar A (2008) Impacts of large-scale circulation variability on low streamflows over Canada. Canadian Water Resources Journal 33 :137-154
- Cannon A.J (2015) Revisiting the nonlinear relationship between ENSO and winter extreme station precipitation in North America. International Journal of Climatology. DOI: 10.1002/joc.4263

Chandran A, Basha G and Ouarda T (2015) Influence of climate oscillations on temperature and precipitation over the United Arab Emirates. International Journal of Climatology. DOI:10.1002/joc.4339

Chavez-Demoulin V and Davison AC (2005) Generalized additive modeling of sample extremes. *Appl. Statist* 54, Part 1: 207-222

Coles Stuart G (2001). An introduction to statistical modeling of extreme values. Springer Series in Statistics

Coulibaly P (2006) Spatial and temporal variability of Canadian season precipitation (1900-2000). *Advances in Water Resources* 29:1846-1865. Doi:10.1016/j.advwatres.2005.12.013

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

De Boor C (2001) A practical guide to spline. Springer, London

Déry SJ, Wood EF (2005) Decreasing river in Northern Canada. *Geophysical Research Letters* 32, L10401. doi: 10.1029/2005GL022845

Déry SJ, Wood EF (2004) Teleconnection between the Arctic Oscillation and Hudson Bay river discharge. *Geophysical Research Letters* 31, L18205, doi: 10.1029/2004GL020729

Dörte J (2013) Nonstationarity in extremes and engineering design. In : AghaKouchak A, Easterling D, Hsu K (ed) *Extremes in changing climate : detection, analysis and uncertainty*. Springer, pp 363-417

Eastoe EF, & Tawn JA (2010) Statistical models for overdispersion in the frequency of peaks over threshold data for a flow series. *Water Resources Research* 46:1-12

Easterling DR (2013) Global data sets for analysis of climate extremes. In : AghaKouchak A, Easterling D, Hsu K (ed.) *Extremes in changing climate : detection, analysis and uncertainty*. Springer, pp 347-361

El Adlouni S, Ouarda TBMJ, Zhang X, Roy R, Bobée B (2007) Generalized maximum likelihood estimators for the nonstationary generalized extreme value model. Water Resources Research 43:1-13

El Adlouni S & Ouarda T.B.M.J (2008) Comparaison des méthodes d'estimation des paramètres du modèle GEV non stationnaire. Revue des Sciences de l'Eau 21(1):35-50

Fisher RA, Tippett LHC (1928) Limiting forms of the frequency distribution of the largest or smallest member of a sample. Math Proc Cambridge Philos Soc 24(02):180-190

Grinsted A, JC Moore and Jevrejeva S (2004) Application of the cross wavelet transform and wavelet coherence to geophysical time series. Nonlinear Processes in Geophysics 11:561-566

IPCC (2012) Managing the risk of extreme events and disasters to advance climate change adaptation. Special Report of the Intergovernmental Panel on Climate Change (IPCC). A special report of working groups I, II of the IPCC.... Cambridge University Press, Cambridge, UK, and New York, NY, USA

Jenkinson AF (1955) The frequency distribution of the annual maximum (or minimum) of meteorological elements. Quart J Roy Meteor Soc 81:158-171

Katz R.W (2013) Statistical methods for nonstationary extremes. In : AghaKouchak A, Easterling D, Hsu K (ed.) Extremes in changing climate : detection, analysis and uncertainty. Springer, pp 15-37.

Katz RW, Parlange MB, Naveau P (2002) Statistics of extremes in hydrology. Advances in Water Resources 25:1287-1304

Khaliq MN, Ouarda TBMJ, Ondo J-C, Gachon P, Bobée B (2006) Frequency analysis of a sequence of dependent and/or non-stationary hydro-meteorological observation: a review. Journal of Hydrology 329:534-552

Lang M, Ouarda TBMJ, Bobée B (1999) Towards operational guidelines for over-threshold modeling. Journal of Hydrology 225:103-117

Lee T, Ouarda TBMJ, Li J (2013) An orchestrated climate song from the Pacific and Atlantic oceans and its implication on climatological processes. International Journal of Climatology 33:1015-1020

Lee T, Ouarda TBMJ (2010) Long-term prediction of precipitation and hydrologic extremes with nonstationary oscillation processes. Journal of Geophysical Research 115, D13107, doi: 10.1029/2009jd012801

Mekis E, Vincent L (2011) An overview of the second generation adjusted daily precipitation dataset for trend analysis in Canada. Atmosphere-Ocean 49(2):163-177

Martins Eduardo S, Stedinger Jery R (2001b) Historical information in a generalized maximum likelihood framework with partial duration and annual maximum series. Water Resources Research 37(10):2559-2567

Martins Eduardo S, Stedinger Jery R (2001a) Generalized maximum likelihood Pareto-Poisson estimators for partial duration series. Water Resources Research 37(10):2551-2557

Martins Eduardo S, Stedinger Jery R (2000) Generalized maximum likelihood generalized extreme value quantile estimators for hydrological data. Water Resources Research 36(3):737-744

Mondal A, Mujumdar P.P (2015) Modeling non-stationarity in intensity, duration and frequency of extreme rainfall over India. Journal of Hydrology 521:217-231

Nasri B, El Adlouni S, Ouarda TBMJ (2013) Bayesian estimation for GEV-B-Spline model. Open Journal of Statistics 3:118-128

Ning L, Bradley RS (2015) Winter climate extremes over the Northeastern United States and Southeastern Canada and teleconnections with large-scale modes of climate variability. Journal of Climate 28: 2475-2493. Doi:10.1175/JCLI-D-13-00750.1

Osman YZ, Rowan F, Sweeney JC (2013) Downscaling extreme precipitation in Ireland using combined peak-over-threshold generalised Pareto distribution model of varying parameters. Journal of Water and Climate Change 4(4):348-363

Ouachani R, Bargaoui Z, Ouarda T (2011) Power of teleconnection patterns on precipitation and streamflow variability of upper Medjerda basin. *Int J Climatol.* doi: 10.1002/JOC.3407

Pickands J III (1975) Statistical inference using extreme order statistics. *Ann Stat* 3(1):119-131

Pujol N, Neppel L, Sabatier R (2007) Approche régionale pour la détection de tendances dans les séries de précipitations de la région méditerranéenne française. *C.R. Geoscience* 339:651-658

Rossi A, Massei N, Laignel B, Sebag D, Copard Y (2009) The response of the Mississippi River to climate fluctuations and reservoir construction as indicated by wavelet analysis of streamflow and suspended-sediment load, 1950-1975. *Journal of Hydrology* 377:237-244

Rossi A, Massei N, Laignel B (2011) A synthesis of the time-scale variability of commonly used climate indices using continuous wavelet transform. *Global and Planetary Change* 78:1-13

Roth M, Buishand TA, Jongbloed G, Klein Tank AMG, van Zanten JH (2012) A regional peaks-over-threshold model in a nonstationary climate. *Water resources research* 48, W11533. doi:10.1029/2012WR012214

Salas JD, Obeysekera J (2014) Revisiting the concepts of return period and risk for nonstationary hydrological extreme events. *Journal of Hydrological Engineering* 19:554-568

Scarrott C, MacDonald A (2012) A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT-Statistical Journal* 10(1):33-60

Serinaldi F, Kilsby CG (2014) Rainfall extremes: towards reconciliation after the battle of distributions. *Water Resources Research* 50:336-352

Shabbar A, Bonsal B (2004) Associations between low frequency variability modes and winter temperature extremes in Canada. *Atmosphere-Ocean* 42 :127-140

Shabbar A, Bonsal B, Khandekar M (1997) Canadian precipitation patterns associated with the Southern oscillation. *Journal of Climate* 10:3016-3027

Stone DA, Andrew JW, Zwiers FW (2000) Trends in Canadian precipitation intensity. *Atmosphere-Ocean* 38(2):321-347

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

Teegavarapu Ramesh SV (2013) Floods in a changing climate : extreme precipitation. *International Hydrology Series*. Cambridge University Press

Thistle ME, Caissie D (2013) Trends in air temperature, total precipitation and streamflow characteristics in Eastern Canada. *Can Tech Rep Fish Aquat Sci* 3018

Thompson DWJ, Wallace JM (2001) Regional climate impacts of the Northern hemisphere annular mode. *Science* 293 :85-89

Thompson DWJ, Wallace JM (1998) The Arctic Oscillation signature in the wintertime geopotential height and temperature fields. *Geophysical Research Letters* 25 :1297-1300

Torrence C and Compo GP (1998) A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society* 79(1):61-78.

Tremblay L, Larocque M, Anctil F, Rivard C (2011) Teleconnection and interannual variability in canadian groundwater levels. *Journal of Hydrology* 410: 178-188

Trenberth KE, Jones PD, Ambenje P, Bojariu R, Easterling D, Tank AK, Parker D, Rahimzadeh F, Renwick JA, Rusticucci M, Soden B, Zhai P (2007) Observations: Surface and Atmospheric Climate Change. In: *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller (ed)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA

Vincent L, Mekis E (2006) Changes in daily and extreme temperature and precipitation indices for Canada over the twentieth century. *Atmosphere-Ocean* 44(2):177-193

Wallace JM, Gutzler DS (1981) Teleconnections in the geopotential height field during the Northern Hemisphere winter. *Monthly Weather Review* 109:784-812

Zhang X, Brown R, Vincent L, Skinner W, Feng Y, Mekis E (2011) Canadian climate trends 1950-2007. Canadian biodiversity: ecosystem status and trends 2010. Technical thematic report n5, Ottawa

Zhang X, Wang J, Zwiers FW, Ya Groisman P (2010) The influence of large-scale climate variability on winter maximum daily precipitation over North America. *Journal of Climate* 23:2902-2915

Zhang X, Hogg WD Mekis E (2001) Spatial and temporal characteristics of heavy precipitation events over Canada. *Journal of Climate* 14:1923-1936

Zhang X, Vincent LA, Hogg W.D, Niitsoo A (2000) Temperature and precipitation trends in Canada during the 20tn century. *Atmosphere-Ocean* 38(3):395-429. DOI: 10.1080/07055900.2000.9649654

Zoglat A, El Adlouni S, Badaoui F, Amar A, Okou CG (2014) Managing hydrological risks with extreme modeling: application of peaks over threshold model to the Loukkos watershed, Morocco. *J Hydrol Eng* 05014010:1-9

Tables illustration

Table 1 Southeastern Canadian stations selection for nonstationary modeling of extreme precipitation

Criteria	Provinces	ON	QC	NS	NB	NFL	PEI	Total
Full inventory		386	398	89	70	98	16	1057
Length ≥ 30 years		136	109	35	19	22	08	329
Independence criterion		78	54	18	09	08	06	173
Correlation with at least one covariate		60	42	16	08	06	04	136

Table 2 Repartition of stations with the best climate index and regionally dominant pattern (gray cells)

Covariates Provinces	AO	PNA	NAO	SOI	WHWP	PDO	AMO	Total
PEI	1	2	1	0	0	0	0	4
NB	1	1	2	1	2	0	1	8
NFL	3	1	0	1	1	0	0	16
NS	4	2	2	2	0	3	3	16
ON	9	11	11	9	11	6	3	60
QC	13	13	6	3	1	4	2	42
Total	31	30	22	16	15	13	9	136

Figures illustration

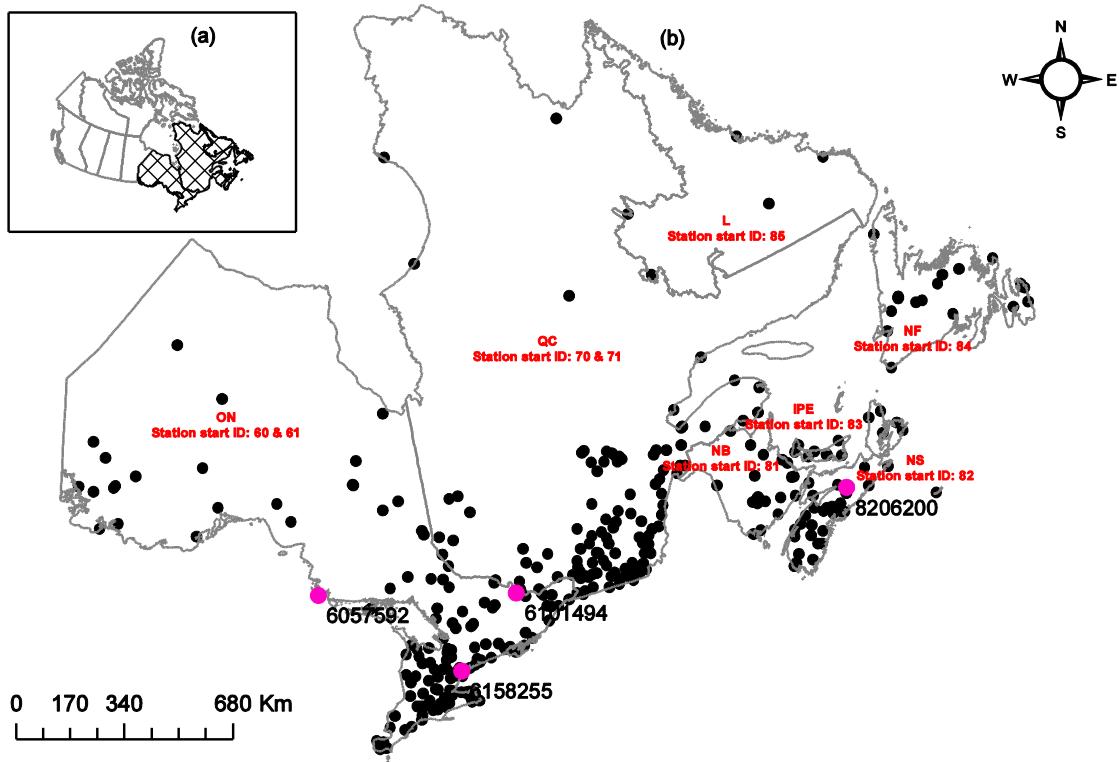


Fig. 1 Study region in Southeastern Canada (a) and location of the 329 stations considered in the study (b). Violet highlighted stations are targeted for illustration

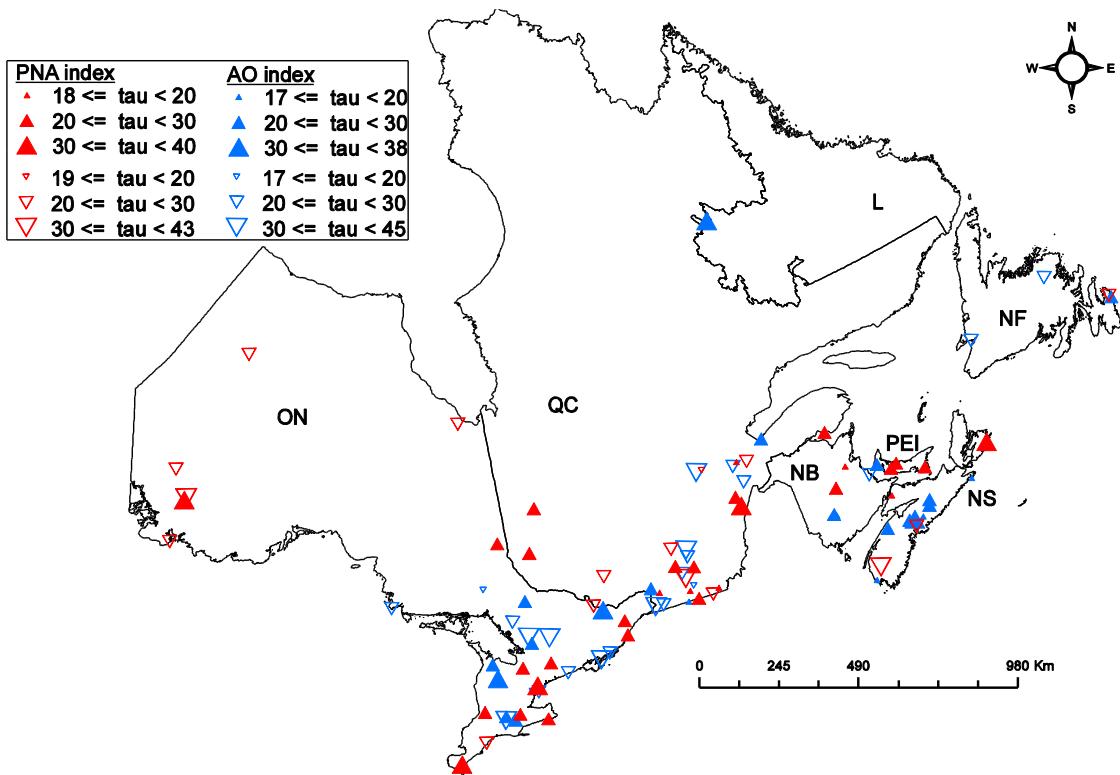


Fig. 2 Kendall's tau values (in percentage) are shown in red for PNA and in blue for AO. Ascendant and descendant symbols respectively indicate positive and negative correlation

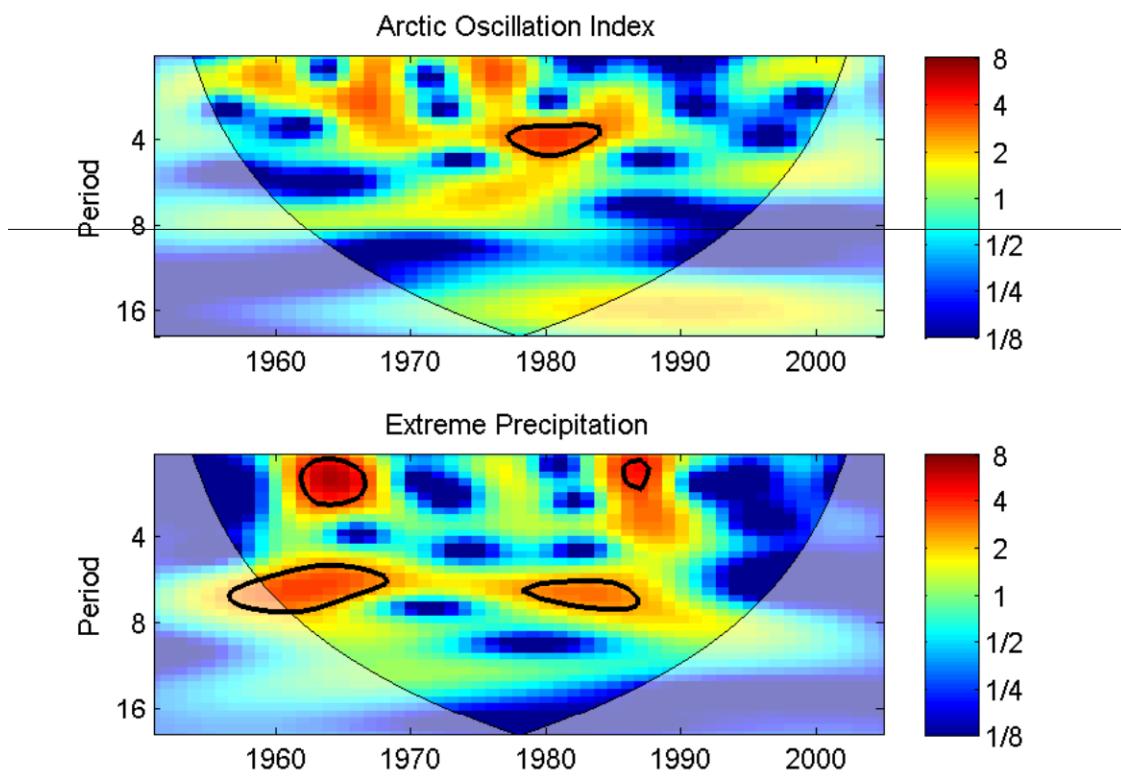


Fig. 3 CWT of extreme precipitation and AO index presented respectively in bottom and top graph. The thick black contour designates the 5% significance level against red noise and the cone of influence (COI) where edge effects might distort the picture is shown in a lighter shade.

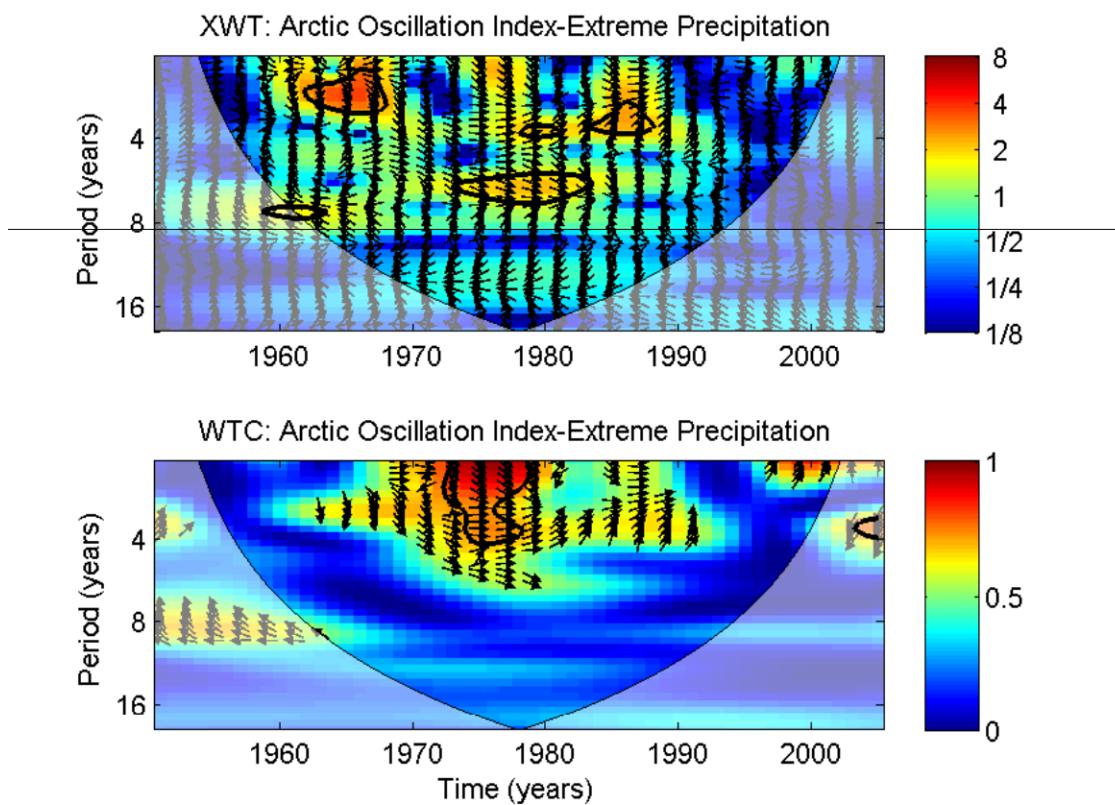


Fig. 4 XWT (top) and WTC (bottom) between AO and extreme precipitation. The 5% significance level against red noise is shown as a thick contour. The relative phase relationship is shown as arrows with in-phase pointing right.

Note: Phase arrows indicate the relative phase relationship between the series (pointing right: in-phase; left: anti-phase; down: series 1 leading; up: series 2 leading). Series 1 represents the index and series 2 represents precipitation.

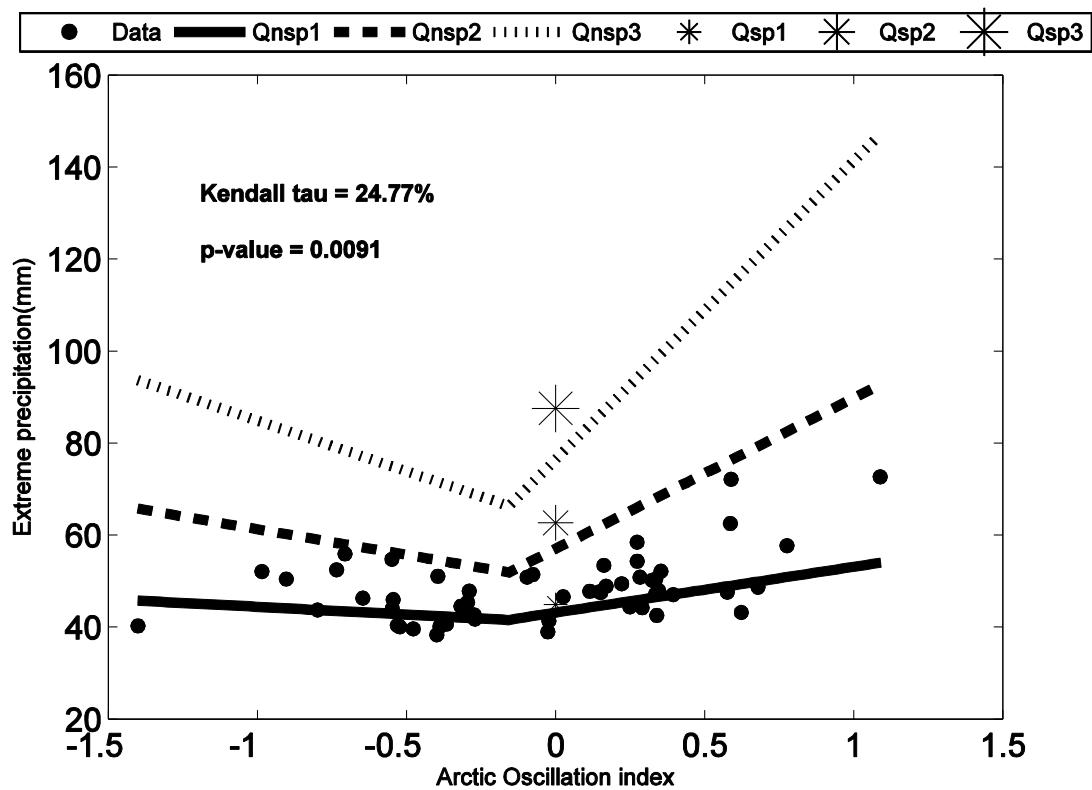


Fig. 5 Station 8206200: Estimated quantiles conditional to the AO index for $p_1= 0.5$, $p_2= 0.9$ and $p_3= 0.99$ illustrated by solid line for Qnsp1, dashed line for Qnsp2 and dotted line for Qnsp3, respectively. Stationary quantiles are represented by graduated asterisk symbols

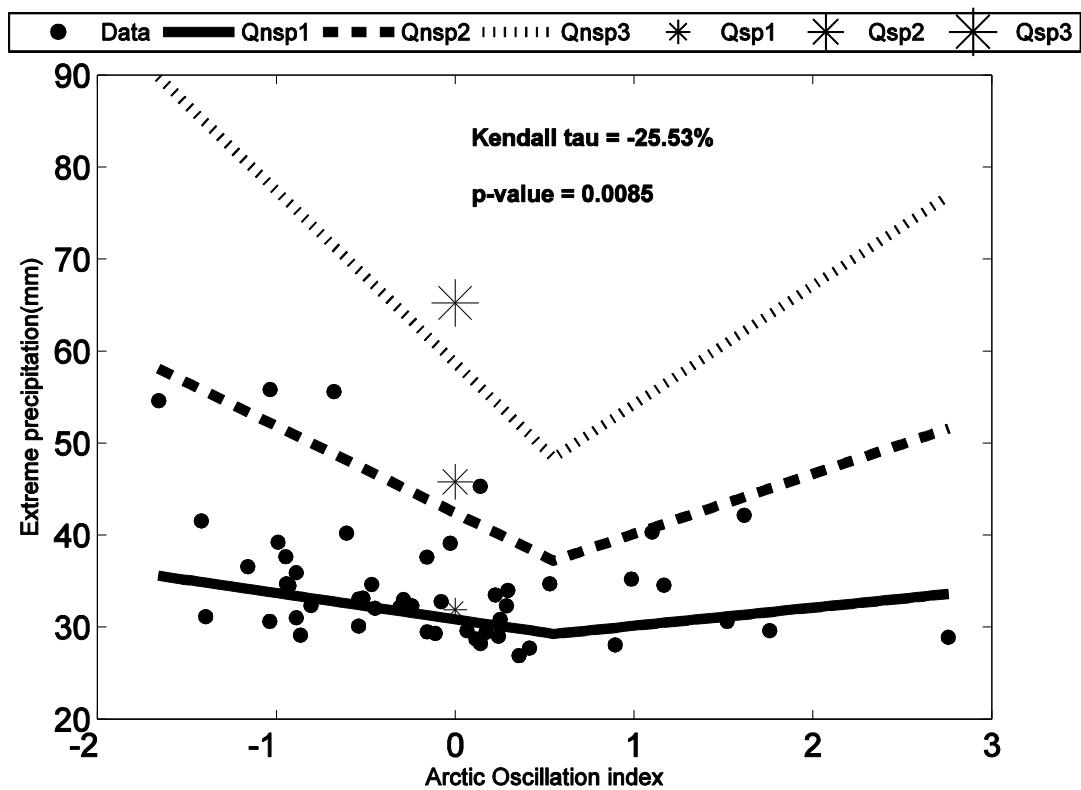


Fig. 6 Station 6057592: Estimated quantiles conditional to the AO index for $p_1= 0.5$, $p_2= 0.9$ and $p_3= 0.99$ illustrated by solid line for Qnsp1, dashed line for Qnsp2 and dotted line for Qnsp3, respectively. Stationary quantiles are represented by graduated asterisk symbols

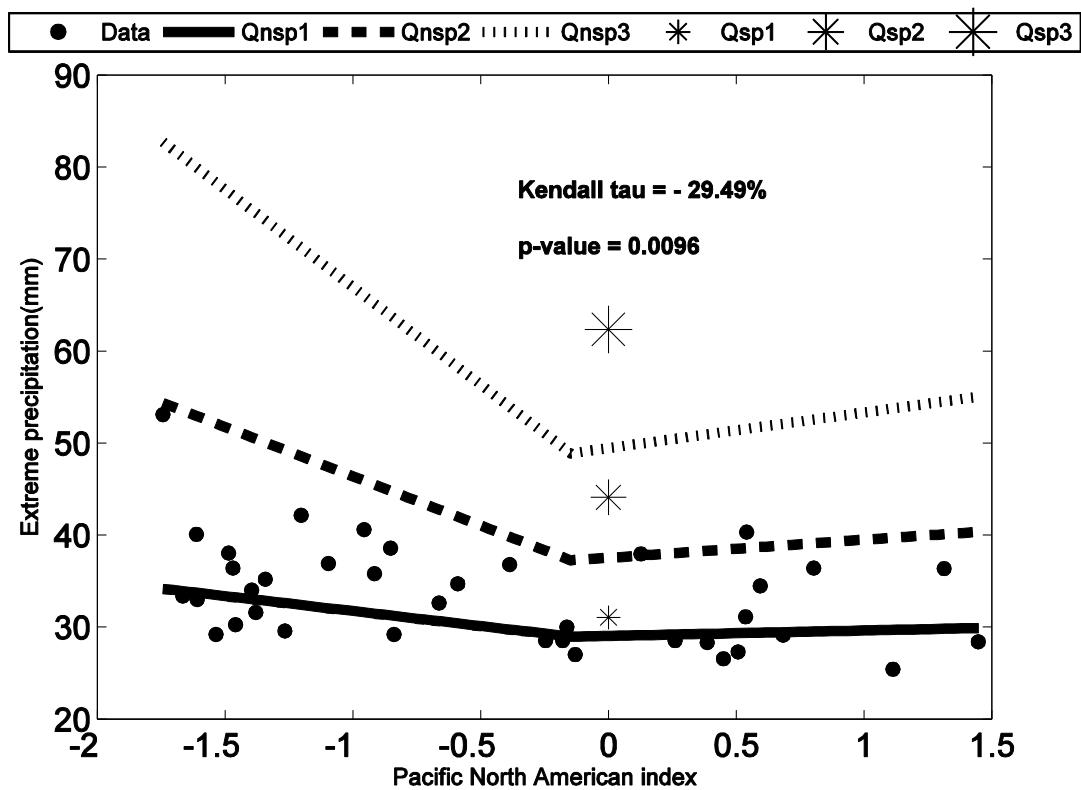


Fig. 7 Station 6101494: Estimated quantiles conditional to the PNA index for $p_1= 0.5$, $p_2= 0.9$ and $p_3= 0.99$ illustrated by solid line for Qnsp1, dashed line for Qnsp2 and dotted line for Qnsp3, respectively. Stationary quantiles are represented by graduated asterisk symbols

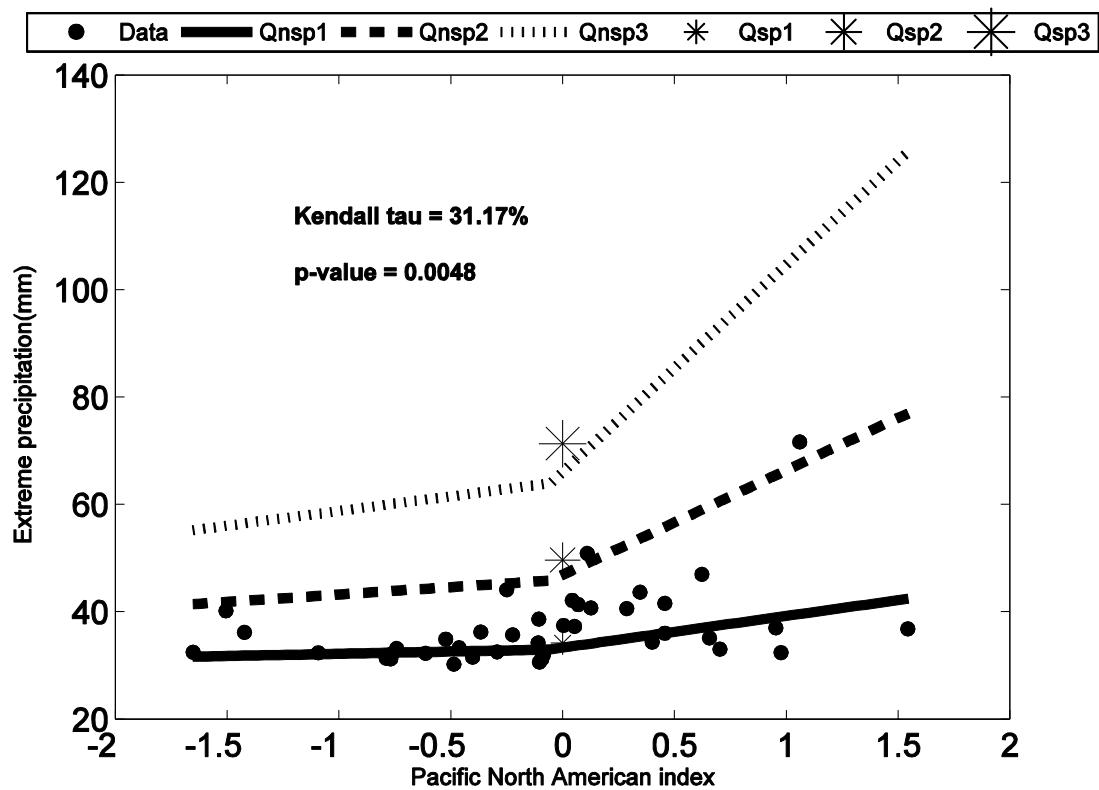


Fig. 8 Station 6158255: Estimated quantiles conditional to the PNA index for $p_1= 0.5$, $p_2= 0.9$ and $p_3= 0.99$ illustrated by solid line for Qnsp1, dashed line for Qnsp2 and dotted line for Qnsp3, respectively. Stationary quantiles are represented by graduated asterisk symbols

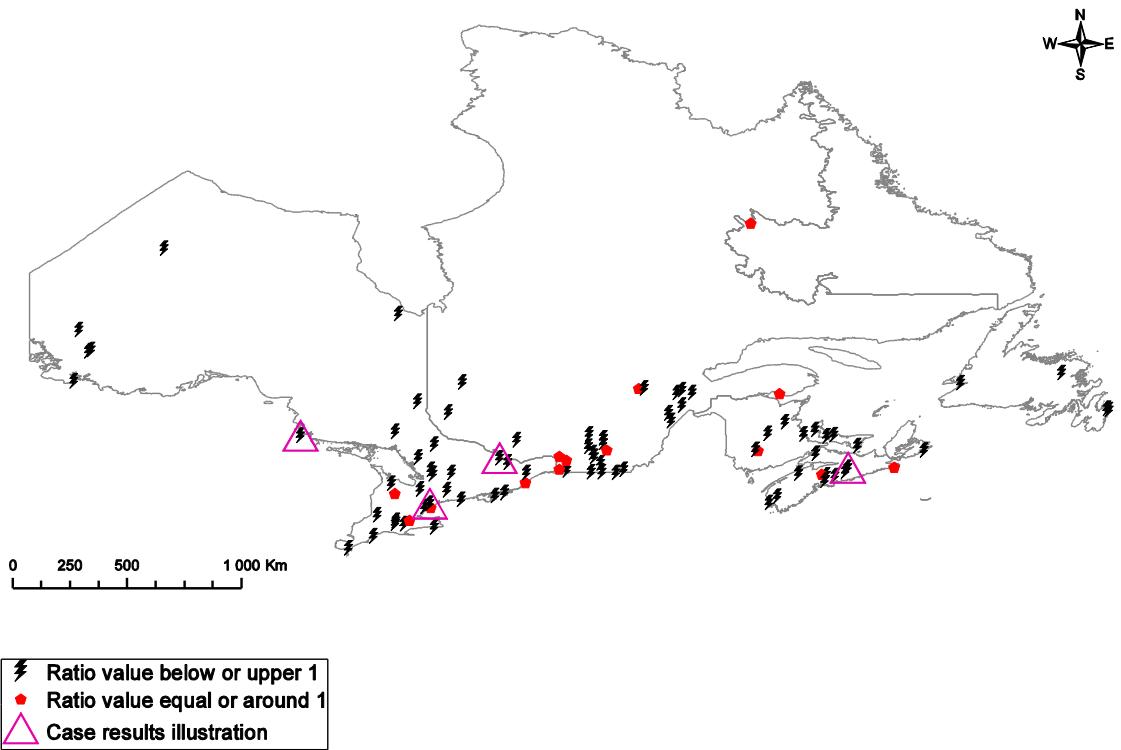


Fig. 9 Comparison between GPD-B-spline models. Stationary suggested models are represented by red pentagon while black's bolt show nonstationary ones. Triangles indicate targeted stations used as examples in results section.

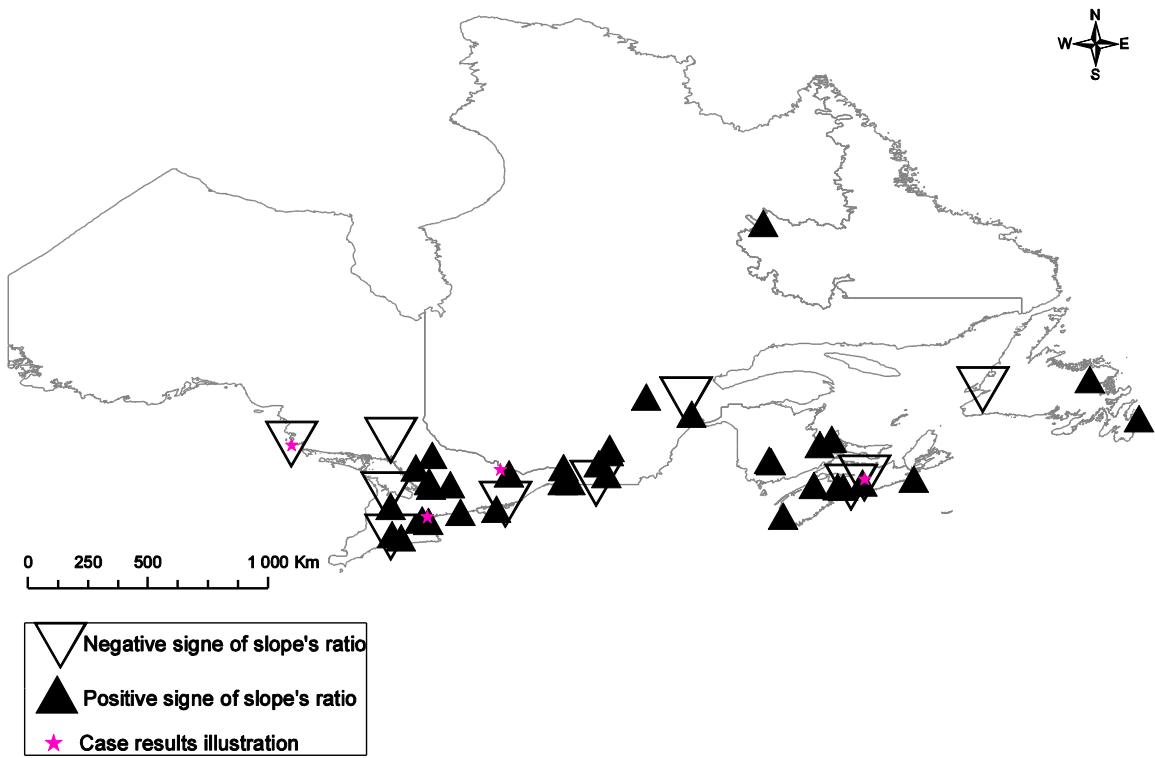


Fig. 10 Comparison between GPD-B-spline models. Downward and upward triangles indicate respectively nonlinear and linear dependency between AO index and precipitation extremes. Violet star indicate stations used in the results section for illustrations.

8 ARTICLE 2: NONLINEAR RESPONSE OF PRECIPITATION TO CLIMATE INDICES USING A NONSTATIONARY POISSON-GENERALIZED PARETO MODEL: A CASE STUDY OF SOUTHEASTERN CANADA

Short title: Nonlinear and nonstationary Poisson-Generalized Pareto model

Alida N Thiombiano (@)⁽¹⁾, André St-Hilaire⁽¹⁾, Salaheddine El Adlouni⁽²⁾ and Taha B.M.J Ouarda⁽¹⁾

⁽¹⁾ Centre Eau–Terre–Environnement, Institut National de la Recherche Scientifique, 490, de la Couronne Street, Québec, Québec, G1K 9A9, Canada

⁽²⁾ Université de Moncton, 18 Antonine-Maillet Avenue, Moncton, New-Brunswick, Canada E1A 3EA.

E-mail de l'auteur de correspondance : alida_nadege.thiombiano@ete.inrs.ca

Article Soumis pour publication à : *International Journal of Climatology*

Acknowledgments: The Authors are grateful to the International Center for Research and Development (ICRD) and to the Natural Sciences and Engineering Research Council of Canada (NSERC) for the financial support through the FACE (*Faire Face Aux Changements Ensemble*) project grant.

Abstract

The teleconnections between a number of large scale climate patterns and hydro-climatological variables allow using these indices as covariates in the development of nonstationary extreme value models. In this paper, a Generalized Pareto distribution (GPD) and a Poisson distribution (PD) with one climatic covariate are studied separately using respectively the smoothing-based B-spline functions and the logarithmic link function. Both models are further combined, resulting in a Poisson-Generalized Pareto model (P-GP) which is then used to estimate quantiles for various non-exceedance annual probabilities. The proposed nonlinear and nonstationary P-GP model is applied to daily precipitation amounts from Upper Stewiacke (ID: 8206200) and “Grandes Bergeronnes” (ID: 7042840) stations, located respectively in Nova Scotia and Quebec provinces of Canada. The Arctic Oscillation (AO) and the Pacific North American (PNA) climate indices are used as covariates respectively at these stations because of the significant nonlinear correlation found with the intensity and the frequency time series of observed precipitation extremes. Modeling of Peaks-Over-Threshold (POT) values with and without covariates, shows significant differences in the estimated quantiles according to the Akaike Information Criterion (AIC), which is used to compare the proposed models. Moreover, a strong nonlinear response of precipitation extremes to the selected climatic covariate is observed at both stations. These interactions can be attributed to the variability in atmospheric circulation dynamics.

Compared to the annual maxima approach, the adopted POT methodology has the advantage of modeling a larger sample of extremes and interpretation of quantiles given the return period concept.

Keywords: Nonlinearity; Nonstationarity; Poisson-Generalized Pareto model; AO index; PNA index; Climatic Covariates; Precipitation; Canada.

1 Introduction

The extreme value theory (EVT) provides a solid justification to the use of the probabilistic distributions such as the Generalized Extreme Value (GEV) and the Generalized Pareto Distribution (GPD) for extreme events frequency analysis purposes (Fisher and Tippett, 1928; Jenkinson, 1955; Pickands, 1975; Coles 2001). Both distributions are widely used in hydrology to fit respectively the annual maxima (AM) and the exceedances over a fixed suitable threshold. However, the interpretation of the quantiles obtained from the GPD model in terms of return period, requires information of the yearly number of peaks which is generally assumed to be Poisson distributed. The Point Process (PP) theory allows to represent the Poisson distribution (PD) and the GPD as a two-dimensional nonhomogeneous Poisson process (Katz et al. 2002). Thus, a correspondence can be established between the combined distributions PD-GPD and the GEV distribution (Lang et al. 1999). The parameters of this latter distribution can be expressed in function of the parameters of the Poisson-Generalized Pareto (P-GP) model (Silva et al. 2016). With the Peaks-Over-Threshold (POT) approach, both processes (i.e. the intensity and frequency) of the extreme events are then captured.

Classical formulation of the EVT models assumes that the observations are independent and identically distributed (iid). However, with the mounting evidence concerning climate change (IPCC 2012), there is a growing interest in the development of nonstationary statistical methods (Khaliq et al., 2006; Katz, 2013; Dörte, 2013) which can lead to more reliable estimates of various quantiles compared to the stationary approach. A quantile is a design value associated with given probability of non-exceedance which is often expressed in terms of a return period, a concept often used in engineering applications.

Nonstationary modeling commonly uses dependence between the parameters of a given probabilistic distribution and covariates. The dependence could be expressed in the form of polynomial functions (El Adlouni et al., 2007), smoothing splines (Chavez-Demoulin and Davison, 2005) or smoothing based on B-spline functions (Nasri et al., 2013). The B-spline semi-parametric functions lead to more flexible nonstationary models than the polynomial parametric functions (Padoan and Wand, 2008).

Climate indices are widely used as covariates for the modeling of hydrological variables in a nonstationary framework. Indeed, strong relationships have been observed worldwide between large-scale atmospheric/oceanic climate indices and hydro-climatological variables thereby helping to understand the variability in such variables. The definition of climate indices from Lee et al. (2013), "*time series that allow quantifying the temporal evolution of climate process in a particular region*", can justify their use as explanatory variables for various hydro-climatological events. For example, Kenyon and Hegerl (2010) highlighted the worldwide effects of the El-Nino-Southern Oscillation (ENSO) on extreme precipitation events, and the significant response of the latter to the North Atlantic Oscillation (NAO) and the Arctic Oscillation (AO) respectively over the European continent and Northern Hemispheric midlatitude. Lee and Ouarda (2010) studied the future evolution of the regional scaled winter precipitation and the extreme hydro-climatological variables in Eastern Canada, by modeling and projecting a nonstationary oscillation process. The significant oscillation signal of the NAO winter index is then extended from the Empirical Mode Decomposition (EMD) process and the future pattern of extreme hydro-climatological variables is estimated using the NAO winter index as a covariate. Zhang et al. (2010) showed that winter daily precipitation maxima over North America are significantly influenced by the ENSO and the Pacific Decadal Oscillation (PDO) indices. Stone et al. (2000) found that the Pacific North American (PNA) climatic pattern has a stronger influence on the frequency of extreme precipitation events than their intensity in Ontario and southern Quebec during autumn and winter seasons. Coulibaly (2006) related the change around the year 1940 in the Canadian seasonal precipitation to AO, and also observed strong correlation between the former and the PNA index after 1970. Bonsal and Shabbar (2011) synthesized the spatial and seasonal effects of ENSO, PDO, PNA, NAO, AO and the Atlantic Multi-decadal Oscillation (AMO) on the Canadian climate. Thiombiano et al. (2016) identified the AO and PNA indices as the two dominant climatic patterns that influence precipitation extremes over Southeastern Canada, and found an East-West correlation sign shift between the AO index and the intensity of daily precipitation extremes.

The literature evidences about climatic teleconnections are dominated by a based-prior assumption of linear interactions between these low frequency climate indices and hydro-climatological variability. However, nonlinear dependences are also observed because of the physical structure (i.e. negative and positive phases) of the climate indices. For example, using composite and correlation analyses, Shabbar et al. (1997) showed that winter precipitation in the upper St. Lawrence valley is enhanced during the La Niña phase while no significant response occurred during the El Niño years. Quadratic response of precipitation to ENSO and AO indices have been found over the Northern Hemisphere by Wu et al. (2005) and Hsieh et al. (2006) based on artificial neural network analysis. Fleming and Dahlke (2014) highlighted similar nonlinear relationships with the annual total flow volume time series using the same indices for the same region. Indeed, they detected parabolic downward and upward interactions by applying an Akaike Information Criterion-based polynomial selection approach. Chandran et al. (2015) found through linear correlation and wavelet analyses, that the negative phase of the Southern Oscillation Index (SOI) is significantly associated with the increase in precipitation over the United Arab Emirates. Canon (2015) compared linear and nonlinear GEV models and showed that El Niño is associated with a decreased likelihood of extreme precipitation over the Great Lakes (Ohio River valley) and Western Canada (Alaska). Silva et al. (2016) also found that high frequency and magnitude of floods are not monotonically increasing or decreasing with the Niño3.4 index in the Itajaí river basin located in southern Brazil.

Nonlinear relationships remain complex but need to be explored and used because they can improve analysis of hydroclimatic variables. Hence, the present study aims at investigating such interactions and developing a flexible nonlinear and nonstationary P-GP model for the statistical modeling of hydro-climatological variables of interest, in this study precipitation. The methodology is described in Section 2 and the case study is presented in Section 3. A discussion follows in Section 4 and the concluding remarks in Section 5.

2 Methods

2.1 POT and PP frameworks

The extreme value sample with the POT approach is generally constituted of all events that exceed a suitable high threshold. From the data sample, the exceedances over the threshold have the GPD as an asymptotic limiting model (Pickands 1975) with cumulative distribution function (cdf) given by

$$G(x, \xi, \sigma, u) = \begin{cases} 1 - \left[1 - \xi \left(\frac{x-u}{\sigma} \right)^{\frac{1}{\xi}} \right] & (x-u) \geq 0; \xi \neq 0; \sigma > 0 \\ 1 - \exp \left[- \left(\frac{x-u}{\sigma} \right) \right] & (x-u) \geq 0; \xi = 0; \sigma > 0 \end{cases} \quad (1)$$

where x and $x - u$ are respectively the peak (i.e., intensity) and the exceedance values; ξ, σ and u are respectively the GPD shape, scale and threshold parameters.

From the POT sample, the occurrence rate of peaks can also be fitted by a PD whose probability mass function is given by

$$P(m = n) = \exp(-\lambda) \frac{\lambda^n}{n!}, \quad n \in \mathbb{N} \quad (2)$$

where m is a random variable representing the number of exceedances per year and λ is the PD rate parameter which is the expected annual frequency of peaks.

Quantile Q_p in the POT framework for a given non-exceedance probability p can be obtained directly by inverting the GPD cdf from equation (1) as follows

$$Q_p(p, \xi, \sigma) = \begin{cases} \left[\frac{\sigma}{\xi} \left[1 - (1-p)^{\xi} \right] \right] + u & \xi \neq 0 \\ [-\sigma \ln(1-p)] + u & \xi = 0 \end{cases} \quad (3)$$

However, to estimate the return level of exceedances (Q_T) associated with a T -years return period, the PD rate parameter needs to be incorporated as follows

$$Q_T(T, \lambda, \xi, \sigma) = \begin{cases} \left[\frac{\sigma}{\xi} \left[1 - \left(\frac{1}{\lambda T} \right)^{\xi} \right] \right] + u & \xi \neq 0 \\ \left[-\sigma \ln \left(\frac{1}{\lambda T} \right) \right] + u & \xi = 0 \end{cases} \quad (4)$$

In the case of a Poisson process, there is a correspondence between the POT and AM distributions (Lang et al. 1999) formulated as $F(y) = \exp[-\lambda(1 - G(x))]$ where y represents the AM values; $F(y)$ and $G(x)$ are respectively the GPD and GEV cdf; λ is the expected annual frequency of peaks. A straight estimate of Q_T is thus obtain directly by inverting the GEV cdf (equation 5), but with parameters expressed as function of the PD and GPD parameters.

$$F(y, \xi, \beta, \psi) = \begin{cases} \exp \left[- \left(1 - \frac{\xi}{\beta} (y - \psi) \right)^{\frac{1}{\xi}} \right] & \xi \neq 0; \beta > 0 \\ \exp \left[- \exp \left(- \frac{(y - \psi)}{\beta} \right) \right] & \xi = 0; \beta > 0 \end{cases} \quad (5)$$

where β and ψ represent respectively the GEV scale and location parameters and are reparameterized as given in equation (6) (Silva et al. 2016); the shape parameter ξ is the same as in equation (1);:

$$\begin{cases} \beta = \sigma/\lambda^{\xi} \\ \psi = u + \beta \left(\frac{\lambda^{\xi}-1}{\xi} \right) \end{cases} \quad (6)$$

Quantiles from F^{-1} are AM (Q_A) and are expressed in equation (7) with a non-exceedance probability p .

$$Q_A(p, \xi, \beta, \psi) = \begin{cases} \psi + \left(\frac{\beta}{\xi} \right) \left[1 - (-\ln(p))^{\xi} \right] & \xi \neq 0 \\ \psi - \beta \ln(-\ln(p)) & \xi = 0 \end{cases} \quad (7)$$

The estimation of the GEV parameters is based on all independent POT events and not on AM dataset.

2.2 Assumptions of the POT model

The use of the classical POT model implies that observations are iid (Lang et al. 1999). Hence, before the statistical inference, it is important to check independence of peaks. Moreover, the threshold is traditionally fixed so that peaks over such a value are Poisson distributed.

This assumption for the number of peaks needs to be verified. It is possible to work with all peaks (Ribatet et al. 2009) and sometimes, other distributions like the Binomial or Negative Binomial can be more suitable for modelling annual number occurrence of peaks (Bezak et al. 2014).

The PD can be verified with the test based on the dispersion index I proposed by Cunnane (1979). It is defined as follows (Önöz and Bayazit, 2001)

$$I = \frac{1}{N-1} \sum_{i=1}^N \frac{(m_i - \bar{m})^2}{\bar{m}} \quad (8)$$

where N is the number of years, in the case of annual time series; m_i is the number of exceedances in year i and \bar{m} is the mean of $(m_i)_{i=1,\dots,N}$. The test statistic d corresponding to this index is given by

$$d = \sum_{i=1}^N \frac{(m_i - \bar{m})^2}{\bar{m}} = (N - 1)I \quad (9)$$

The statistic d asymptotically follows a χ^2 distribution with $(N - 1)$ degrees of freedom. As the PD has a dispersion index $I=1$, the poissonian hypothesis is not rejected if the computed d value is in the range $(\chi^2_{\alpha/2}, \chi^2_{1-\alpha/2})$ where α is the significance level. If $d < \chi^2_{\alpha/2}$, the Binomial distribution must be preferred, and if $d > \chi^2_{1-\alpha/2}$, the Negative Binomial distribution is more appropriate. The case of $I>1$ corresponds to what is called the overdispersion phenomenon and is indicated to be normally more realistic. However, it is possible to take it into account with a simple non-homogeneous Poisson process (Eastoe and Tawn, 2010).

The independence criterion can be validated by assessing suitable thresholds (e.g., high percentiles) with the mean excess plot and the GPD shape and scale parameters stability given an increasing sequence of thresholds (Davison and Smith, 1990; Lang et al., 1999; Coles, 2001). To systematically meet this criterion, a declustering technique (Roth et al., 2012) is widely used and can be validated through the autocorrelation function and the Chi-square goodness-of-fit test for the GPD adequacy. All these graphical techniques were used to assess peaks independence given the selected threshold.

2.3 Nonstationary POT models

The nonstationarity is commonly incorporated separately in each component of the P-GP model and can thus be easily integrated indirectly into the reparameterized GEV distribution (Katz, 2013).

2.3.1 GPD B-spline model

To take into account nonstationarity in equation (1), the GPD threshold and shape parameters were kept constant, and only the scale parameter was allowed to vary as a function of the covariate Z , referred as σ_Z . To assure positive values of the scale parameter, the logarithm is usually applied to σ_Z . The link function h , defined by $\sigma_Z = h(Z)$, was assumed a linear combination of suitable basis spline (B-spline) functions $B_{k,d}$ having k knots and used to form a piecewise polynomial function of degree d (Nasri et al., 2013)

$$\sigma_Z = h(Z) = \beta_0 + \sum_{j=1}^k \beta_j B_{j,d}(Z) \quad (10)$$

where β_0 and β_j are the regression parameters.

This dependence structure is flexible since it allows exploring linear as well as nonlinear linkages between the variables. Moreover, such spline smooth functions are robust for extreme value modeling as the impact of the outliers and non-local effects are limited (Chavez-Demoulin and Davison, 2012). Some applications with the cubic splines and B-spline functions for a POT framework were provided respectively in Chavez-Demoulin and Davison (2005) and Thiombiano et al. (2016). The use of B-spline functions with the GEV distribution was also proposed in Padoan and Wand (2008) and Nasri et al. (2013). The nonstationary GPD is then obtained by replacing the stationary scale parameter in equation (1) by its expression from equation (10).

The Generalized Maximum Likelihood (GML) method is used to estimate the parameters of the proposed nonlinear and nonstationary GPD model. It is a particular case of the Bayesian approach (El Adlouni et al., 2007), where a Beta distribution ($u=6$; $v=9$) defined on the interval $[-0.5; +0.5]$ was used as prior only for the GPD shape parameter to avoid unfeasible estimates (Martins and Stedinger; 2001).

The GML estimators are obtained using a procedure based on the Newton-Raphson iteration (Hosking and Wallis, 1987). The selection of the best GPD B-spline model can then be obtained through the deviance statistic or more generally the AIC (Katz, 2013). The AIC for the proposed GPD B-spline model can be computed as

$$AIC(r) = -2\log L + 2r \quad (11)$$

where L is the likelihood function, and $r = [(d + k) + 1]$ is the number of parameters associated with each of the GPD B-spline tested models given a set of d and k values (Thiomiano et al., 2016).

2.3.2 Non-homogeneous Poisson process

One of the main objectives of the present study is to incorporate the effect of a covariate on the distribution of number of peaks. To this end, the PD parameter λ is also allowed to vary conditional to the same covariate Z in order to investigate its effect both on the intensity and frequency of peaks distributions. This non-homogeneous Poisson process with rate parameter λ_Z is herein investigated and given by

$$\log \lambda_Z = \lambda_0 + \lambda_1 Z \quad (12)$$

The hyperparameters λ_0 and λ_1 can be estimated through a generalized linear model (GLM) for the PD with the logarithmic link function.

2.3.3 Nonstationary P-GP model

The set of stationary parameters (σ and λ) should now be replaced by (σ_Z and λ_Z) as formulated in equations (10) and (12), allowing subsequently to take into account the effect of the covariate in reparameterized parameters of the GEV distribution expressed in equation (6).

2.3.4 Quantile estimation and uncertainty measure

In this paper, the quantiles are estimated under the GPD model (i.e., Q_p from (3)) and the P-GP model (i.e., Q_A from (6)) with the non-exceedance probability p in the stationary (without covariate Z) and nonstationary (with covariate Z) frameworks. The following models are thus used to obtain the quantiles:

- Stationary quantile Q_p from the model named GPD0 expressed by (1).
- Stationary quantile Q_A from the model named PD0-GPD0 formulated by combining (5) and (6).
- Nonstationary quantile Q_p from model named GPD1 formulated by combining (1) and (10).
- Nonstationary quantile Q_A from model named PD0-GPD1 formulated by combining (5), (6) and (10).
- Nonstationary quantile Q_A from model named PD1-GPD1 formulated by combining (5), (6), (10) and (12).

The parameters of the GPD models are the GML estimates and were obtained based on the Newton-Raphson procedure instead of the classical Bayesian approach with Monte Carlo Markov Chain methods. The AIC can thus be used to compare the models. The model with the lowest AIC value is considered as the best one. The measure of the uncertainty associated with estimated quantiles is provided as intervals of credibility (ICs) computed for the 95% probability of confidence. An adjusted asymptotic approach is used at this end. . This new approach was proposed by Ashkar and El Adlouni (2015), who showed that it allows improving the normality of GPD-based quantiles and leads to more accurate ICs for quantiles in the right-tail of the GPD than the ICs obtained from the traditional large-sample based theory (Ashkar and Ouarda, 1996).

2.4 Selection of the covariate

The correlation and wavelet analyses are useful tools for investigating interactions between the hydro-climatological variables. They are both used in this study as complementary methods.

For the climate indices of interest, a nonparametric correlation assessment can be appropriate, regardless of the physical structure (positive and negative phases) of these indices. The use of the Kendall's tau, appears to be more suitable than the popular linear correlation measures, specifically for non-normal distributions like those of the hydro-climatological extremes (Yue et al. 2002; Chen et al., 2012). However, given the fact that parametric or nonparametric correlation analysis identify monotonically dependence, wavelet analysis which allows to capture the time-scale changes in and between time series through the continuous wavelet transform (CWT), the cross-wavelet transform (XWT) and the wavelet transform coherence (WTC) plots (Torrence and Compo, 1998; Grinsted et al., 2004), is additionally used.

3 Case study

3.1 Data

3.1.1 Climate indices

The AO and PNA indices are separately considered as climatic covariates for the development of a univariate nonstationary P-GP model. The support for this choice is based on the work of Thiombiano et al. (2016) who identified both climatic patterns as the dominant modes of low-frequency variability of the precipitation intensity time series of over Southeastern Canada. The former index is the first empirical orthogonal function of the Northern Hemisphere (20°–90°N) winter sea level pressure data, while the latter index is the dominant climatic pattern of low-frequency variability in the Northern Hemisphere extratropics (Rossi et al. 2011).

The monthly standardized time series of these indices were obtained from the Physical Sciences Division of the National Oceanic and Atmospheric Administration (NOAA) [<http://www.esrl.noaa.gov/psd/data/climateindices/list/>]. The assessment of the interaction between each of these indices and the hydro-climatological variable of interest, is carried out by pooling the monthly time series in two ways: the annual average of monthly values from January to December (JD) on one hand, and the three-monthly moving average for December through February (DJF), January through March (JFM), etc., on the other hand.

3.1.2 Precipitation

The hydro-climatological variable analyzed herein is the observed daily precipitation amounts recorded at the Upper Stewiacke and “*Grandes Bergeronnes*” stations and obtained from Environment Canada. These meteorological stations are located in Southeastern Canada and respectively in Nova Scotia and Quebec provinces (Figure 1). This region has four Atlantic provinces [Newfoundland and Labrador (NFL), Prince Edward Island (PEI), New Brunswick (NB), Nova Scotia (NS)] and two central provinces [Quebec (QC) and Ontario (ON)].

The analysis period is 1951-2005 for the Upper Stewiacke station and 1951-2012 for the “*Grandes Bergeronnes*” station, respectively leading to 53 years (with 51 complete years) and 60 years (with 57 complete years) of daily precipitation observations. Extreme daily precipitation amounts are the peaks above the 99th percentile used herein as a suitable threshold. Each peak value represents the intensity of the precipitation event while the number of occurrence of peaks in a year defines their frequency. Henceforth, the intensity and frequency variables are respectively named V1 and V2. For the correlation analysis, the daily observations from V1 were averaged at an annual scale.

3.2 Results

The threshold values corresponding to the 99th percentile of the daily precipitation amounts for the entire record are 37 mm and 32 mm respectively at Upper Stewiacke and *Grandes Bergeronnes*, leading to a sample of one day declustered exceedances containing 210 and 207 independent peaks respectively. The average value of peaks is 49 mm and 46 mm respectively over the analyzed period, while the average annual number of occurrences is between 3 and 4. The dispersion index is 1.06 at the former station and 1.26 at the latter station and the computed statistic d validated the PD assumption, thus the P-GP developed model.

3.2.1 Assessment of the interactions

The correlation measure between V1 then V2 variables and each of the 3-monthly and annual averaged climate indices values, is assessed by computing the Kendall's Tau with 5% significance level. The results of these analyses at the studied stations are presented in Table I with significant correlations highlighted in bold.

Hence, the average of AO values from the months September to November (SON) and October to December (OND) constituted the covariate AO data respectively for V1 and V2 variables for nonstationary modeling at the Upper Stewiacke station. In the case of the *Grandes Bergeronnes* station, DJF and JD time windows were retained for V1 and V2 respectively.

The assessment of V1, V2, AO and PNA time series CWTs showed significant features of variability predominately in the range of periods spanning 2-8 years. This detection of variability has more physical meaningful than the simple correlation analysis, thus sustaining the nonstationary frequency analysis. Common and coherent significant features were also clearly highlighted by the XWT and WTC plots between the explored variable-covariate datasets. These wavelet results can confirm and complement the correlation results by comparing the direction of the interaction arrows. Arrows pointing right (or left) indicate a positive (negative) correlation.

An illustration of results for V1-AO and V2-PNA is given in Figure 2 and Figure 3 for the Upper Stewiacke and *Grandes Bergeronnes* stations respectively.

For the interpretation of Figures 2 and 3, the thick black contour designates the 5% significance level against the red noise and, the cone of influence where the edge effects might distort the picture is shown in a lighter shade. The darker the red colour is in the enclose feature, the stronger the variability is. The phase relationship between the time series (see XWTs and WTCs illustrative results) is represented as arrows with in-phase pointing right (positive correlation) and anti-phase pointing left (negative correlation).

Significant features are found from the Upper Stewiacke station V1 time series in the 2-3 year and the 5-8 year periods respectively around the decades 1960-1970 and 1980-1990 (Figure 2a). In its corresponding covariate AO data, a 3-5 year feature can also be observed around the year 1980 (Figure 2b). Common significant features effectively appear from 1960 to 1990 in the 2-7 year period between these two correlated datasets (Figure 2c), with a strong covariance between them in the 2-5 year period during the decade 1970-1980 (Figure 2d). Moreover, the XWT and WTC results highlighted that these datasets are in-phase, confirming the strong positive nonlinear correlation found between them (Table I). At the *Grandes Bergeronnes* station, significant features are also present in the different time series (Figure 3) with, however, less strong and synchronized enclose features. Nevertheless, arrows in the XWT (Figure 3c) and WTC (Figure 3d) plots show that these datasets are also phase-locked, sustaining the positive nonlinear correlation found between them (Table 1).

3.2.3 Modeling results

Nine combinations of degree d and interior knots k are examined to assess the fitting of the GPD1 to V1 time series. The results of the obtained AIC values are presented in Table II for both stations. The combination (k, d) resulting in the smallest AIC value is highlighted in bold and is considered as the best GPD1 model. At the Upper Stewiacke and *Grandes Bergeronnes* stations, a B-spline with 2 knots and 1 degree (2, 1) is the best combination with respectively an AIC value of 360.2259 and 433.7658. Those are smaller than the AIC associated with the GPD0 model (364.2293 and 434.1122 respectively).

The comparison of AIC values from GPD1 models and GPD0 shows that the proposed nonlinear GPD1 models are all better (except the model with combination (4, 3)) than GPD0 at the Upper Stewiacke station. However, only the nonlinear GPD1 with combination (2, 1) outperforms GPD0 at the *Grandes Bergeronnes* station. Nevertheless, the model GPD B-spline allowed capturing nonlinear relationships at both stations and showed its flexibility in the detection of simple to complex interactions.

Moreover, a regression analysis between conditional quantiles and climatic covariate values is carried out for both the positive and negative phases of the indices used (see details in Thiombiano et al. 2016). This allows computing further the ratio of slopes from negative and positive phase of the index. Negative ratios are obtained at both studied stations. When the ratio value is positive, this means that slope during the positive and negative phases of the index has the same sign; in this case a monotonous dependence trend is generally observed.

However, if the ratio's sign is negative, the trend is related to the phase of the index and is therefore nonlinear. The value of this ratio is -3.44 at the Upper Stewiacke station and -0.41 at the *Grandes Bergeronnes* station. These values being also different than 1, this effectively suggests that the nonstationary model (GPD1 with 2 knots and 1 degree) is overall better than the stationary one (GPD0).

The estimations of extreme precipitation event quantiles under the models defined in Section 2.3.4 (GPD0, GPD1, PD0-GPD0, PD0-GPD1, PD1-GPD1) are computed for three widely useful non-exceedance probabilities ($p=0.5$, 0.9 and 0.99), but only the second one is used to illustrate modeling results. The Figures 4 and 5 showed these results respectively at the Upper Stewiacke and *Grandes Bergeronnes* stations. The purpose of adding the annual maxima observations on Figures 4 and 5, is to understand the difference between estimated quantiles resulting from a simple GPD versus a combined P-GP model. It is also important to notice that POT observations in Figures 4 and 5, are independent peaks of daily precipitation amount. The same monthly value of the climatic covariate is therefore used for the peaks occurring in that same month, hence, the alignment of some observations.

The quantiles estimated under GPD asymptotic limit convergence using models GPD0 and GPD1, are systematically inferior to the T-years return quantiles obtained from models PD0-GPD0, PD0-GPD1 and PD1-GPD1. The comparison between PD0-GPD0 and PD0-GPD1 then PD1-GPD1, showed a difference in values at about around 40 mm and 70 mm respectively.

In the nonstationary case, the estimates depend on the value of the covariate and show a clear nonlinear association between the climate index and the precipitation extremes. In both concave (Figure 4) and convex (Figure 5) nonlinear structures, the AM quantiles obtained from the PD0-GPD1 model are superior (inferior) to those from the PD1-GPD1 model during the negative (positive) phase of the index. These results suggest that PD0-GPD1-based quantiles are not systematically above or below the PD1-GPD1-based quantiles for all covariate values.

At the Upper Stewiacke station (Figure 4), the quantile estimates increase with the increase in the absolute value of the AO index for both positive and negative values. The conditional quantile curve has a concave form, highlighting clearly the nonlinear response of precipitation extreme events to AO index at this station. Comparing models PD0-GPD0, PD0-GPD1 and PD1-GPD1 through AIC values, the second model outperforms the third model which is better than the first model. This result suggests that incorporation of nonstationarity in the estimation of the GPD scale parameter provided a better model.

At the *Grandes Bergeronnes* station (Figure 5), the nonlinear dependence between precipitation extremes and the PNA index takes a convex form in comparison to the concave form observed in Figure 4. Hence, the estimated quantiles decrease with the increase in the absolute value of the PNA index for both positive and negative values. However, the model PD1-GPD1 leads to the lowest AIC value. This model outperforms PD0-GPD1 which was found to be better than the PD0-GPD0 model. Thus, incorporation of nonstationarity both in the GPD scale and PD intensity parameter provides better estimates of quantiles at this station.

4 Discussion

The present study used the POT framework to develop on one hand a Generalized Pareto model with one climatic covariate by considering the B-spline link function, and on the other hand a non-homogeneous Poisson process using the GLM approach. The flexibility of the splines and B-spline smoothing functions was demonstrated in past studies of Chavez-Demoulin and Davison (2005, 2012), Padoan and Wand (2008), Nasri et al. (2013), and Thiombiano et al. (2016). In the present study, the two case studies allow illustrating the flexibility of the nonstationary P-GP models proposed, as simple to complex dependence between studied hydro-climatological variables and climatic patterns can be automatically explored and compared. The P-GP model adopted herein offers the advantage to evaluate the dependence between the GPD scale parameter and the threshold (Coles 2001) and to adjust the POT quantiles to an annual analysis time step (Katz et al. 2002; Silva et al. 2016). For the case study, the PD assumption has been validated, resulting in the use of a P-GP model instead for example of a negative Binomial-Poisson model in the case we need to address the overdispersion of peak occurrences (Eastoe and Tawn, 2010). However, when this overdispersion is statistically detected, a non-homogeneous Poisson process can simply be used in place of the negative Binomial distribution.

The nonlinear responses of extreme precipitation to the AO and PNA indices found at the studied stations show that there is a strong risk of underestimating or overestimating quantiles using the common assumption of linear interaction between variables and covariates. However, nonlinear relationships are less explored than linear ones because of their complexity.

For example, the linear and nonlinear responses of the atmospheric circulation and climatic patterns to a regional climatic dynamic are associated (Wu et al. 2004). Several internal and external factors can be involved in these large scales oscillations as in the local hydrological processes, impacting the significance and strength of some correlation signals.

In Southeastern Canada, Thiombiano et al. (2016) showed that there is a concurrence of indices AMO, AO, NAO, PDO, PNA, SOI and WHWP able to help explaining the behavior of extreme precipitation. This co-influence of climate indices leads to small but significant correlation between candidate climatic covariate and hydro-climatological variable of interest (e.g. Table I). As opening tasks to assess the combined effect of several climatic patterns, Tables III and IV provide an overview of the co-influence of the climate indices AMO, AO, NAO, PDO, PNA, SOI and WHWP on V1 and/or V2 distributions at several precipitation stations located in Southeastern Canada. For each index, an inventory of stations where significant correlation is found with the Kendall's Tau, is made. Climate indices are thus ranked with the top index having the maximum number of stations.

From Table III and IV, we can conclude that AO and PNA indices have the highest impact on V1 and V2 than the other indices. Indeed, NAO is known to be a regional expression of AO (Thompson and Wallace, 2001), while Rossi et al. (2011) suggested that the PNA index combines the activity of the ENSO (herein represented by SOI) and PDO indices. Nevertheless, a range of the level of influence of each climate index on V1 and V2 is resumed on Table V. It was defined for each province based on the number of stations with significant correlation. The highest level of the influence corresponds to the index with the largest number of correlated stations, while the absence of any correlation with stations is identified by "No" influence.

An investigation of the combined effects of more than one index on hydrological variables like precipitation and streamflows could provide an interesting research perspective. Furthermore, the concave and convex nonlinear responses illustrated in Figures 4 and 5, are also observed at other stations in Southeastern Canada, as illustrated in Figure 6, suggesting the possibility to extend the nonlinear and nonstationary P-GP local model proposed herein to a regional model. This could help improving precipitation forecasts in this region with a focus on understanding the underlying physical mechanisms behind the interactions between climate patterns and hydro-climatological processes in this region.

Indeed, Trenberth (1990) showed that the shift in atmospheric circulation constitutes the principal cause of regional variability in observed wind, temperature, precipitation and other climatic variables. For example, Wu et al. (2005) explained that anomalous northerlies from the Arctic area transfer colder air over Northeastern Canada, leading to negative temperature anomalies there. These authors also indicated that the anomalous alongshore flow along the West coast of North America brings the normal moist westerly flow farther North, generating negative precipitation anomalies from Oregon to Southern British Columbia. Climate indices involve large-scale atmospheric and oceanic variability; there is the teleconnection between the dynamic of dry / moisture air masses flux and hydro-climatological responses. Thompson and Wallace (2001) also explained the association between the sea level pressure variability, the winds (which are loaded with dry or moisture air masses) direction, and the warm-cool phases of AO, NAO and PNA.

5 Conclusion

The quantiles are often expressed in term of a return period for engineering application purpose. Their estimates within the GPD framework need to be adjusted by incorporating the information on the yearly occurrence of peaks. The combined P-GP approach was considered at this end. An assessment of the Poisson distributed assumption for the peaks frequency variable was firstly performed and the effect of incorporating the nonstationarity in the GPD and PD was evaluated. In the GPD0 and PD0-GPD0 models all parameters are constant, while in the GPD1 model, the scale parameter is allowed to vary conditional to a covariate through a B-spline dependence function. For the nonstationary models PD0-GPD1 and PD1-GPD1, the PD parameter was respectively assumed constant and followed a logarithmic function of the same covariate as used for the GPD1. The comparison of stationary ones with nonstationary models allows recommending the PD0-GPD1 and PD1-GPD1 models over the PD0-GPD0 in the case of a combined P-GP approach.

The same climatic pattern was used to evaluate the P-GP nonstationary models. For a further physical explanation of dependences found between the AO (or PNA) index and the variables V1 and V2, this choice will result in a simple task. But, in light of the potential other patterns identified, different indices can be considered.

The B-spline function used to develop GPD1 allowed comparing various linear and nonlinear nonstationary GPD models. Indeed, the polynomial and quadratic function commonly used in nonstationary frequency analysis is a special case of the GPD-B-spline model (i.e. $(k=1; d=1)$ and $(k=1; d=2)$ respectively). This flexibility of the B-spline function could improve predictions in hydro-climatology, specifically when climatic patterns are used as covariates.

The evaluation of a P-GP model can also offer a large sample of extremes which is an asset for a good estimation of the model parameters. It is important to validate the Poissonian hypothesis when using the P-GP approach, because Binomial or negative Binomial distributions may be more adequate than the common assumed Poisson distribution.

Future research efforts may focus on a better understanding of nonlinear relationships, and on the extension of the proposed nonstationary models to a regional scale.

References

- Ashkar F, El Adlouni S. 2015. Adjusting for small-sample non-normality of design event estimators under a generalized Pareto distribution. *Journal of Hydrology* **530**:384-391.
- Ashkar F, Ouarda TBMJ. 1996. On some methods of fitting the generalized Pareto distribution. *Journal of Hydrology* **177**: 117-141.
- Bezak N, Brilly M, Sraj M. 2014. Comparison between the peaks-over-threshold method and the annual maximum method for flood frequency analysis. *Hydrological Sciences Journal* **59**:959-977.
- Bonsal B, Shabbar A. 2011. Large-scale climate oscillations influencing Canada, 1900-2008. Canadian Biodiversity : Ecosystem Status and Trends 2010. *Technical Thematic Report No.4*. Canadian Councils of Resource Ministers. Ottawa, ON.
- Cannon AJ. 2015. Revisiting the nonlinear relationship between ENSO and winter extreme station precipitation in North America. *International Journal of Climatology*. doi: 10.1002/joc.4263
- Chandran A, Basha G, Ouarda T. 2015. Influence of climate oscillations on temperature and precipitation over the United Arab Emirates. *International Journal of Climatology*. doi:10.1002/joc.4339
- Chavez-Demoulin V, Davison AC. 2012. Modeling time series extremes. *REVSTAT-Statistical Journal* **10**:109-133.
- Chavez-Demoulin V, Davison AC. 2005. Generalized additive modeling of sample extremes. *Appl. Statist.* **54**:207-222.
- Chen L, Singh VP, Guo S. 2012. Measure of correlation between River flows using the copula-entropy method. *J. Hydrol. Eng.* **18**:1591-1606.
- Coles SG. 2001. An introduction to statistical modeling of extreme values. *Springer Series in Statistics*.

Coulibaly P. 2006. Spatial and temporal variability of Canadian season precipitation (1900-2000). *Advances in Water Resources* **29**:1846-1865. doi: 10.1016/j.advwatres.2005.12.013

Cunnane C. 1979. A note on the Poisson assumption in partial duration series models. *Water Resour Res.* **15**:489-493.

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

Dörte J. 2013. Nonstationarity in extremes and engineering design. In : AghaKouchak A et al. (eds.) *Extremes in changing climate : detection, analysis and uncertainty*. Springer, pp 363-417.

Eastoe EF, Tawn JA. 2010. Statistical models for overdispersion in the frequency of peaks over threshold data for a flow series. *Water Resources Research* **46**:1-12.

El Adlouni S, Ouarda TBMJ, Zhang X, Roy R, Bobée B. 2007. Generalized maximum likelihood estimators for the nonstationary generalized extreme value model. *Water Resources Research* **43**:1-13.

Fleming SW, Dahlke HE. 2014. Parabolic northern-hemisphere river flow teleconnections to El Nino-Southern Oscillation and the Arctic Oscillation. *Environmental Research Letters* **9**,104007. doi: 10.1088/1748-9326/9/10/104007

Grinsted A, Moore JC, Jevrejeva S. 2004. Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear Processes in Geophysics* **11**:561-566.

Hosking JRM, Wallis TR. 1987. Parameter and quantile estimation for the Generalized Pareto distribution. *Technometrics* **29**:339-349.

Hsieh WW, Wu A, Shabbar A. 2006. Nonlinear atmospheric teleconnection. *Geophysical Research Letters* **33**, L07714. doi: 10.1029/2005GL025471

IPCC. 2012. Managing the risk of extreme events and disasters to advance climate change adaptation. Special Report of the Intergovernmental Panel on Climate Change (IPCC). *A special report of working groups I, II of the IPCC.... Cambridge University Press*, Cambridge, UK, and New York, NY, USA.

Jenkinson AF. 1955. The frequency distribution of the annual maximum (or minimum) of meteorological elements. *Quart J Roy Meteor Soc* **81**:158-171.

Katz R.W. 2013. Statistical methods for nonstationary extremes. In : AghaKouchak A et al. (eds.) *Extremes in changing climate : detection, analysis and uncertainty*. Springer, pp 15-37.

Katz R.W, Parlange MB, Naveau P. 2002. Statistics of extremes in hydrology. *Advances in Water Resources* **25**:1287-1304.

Kenyon J, Hegerl GC. 2010. Influence of modes of climate variability on global precipitation extremes. *Journal of Climate* **23**:6248-6262.

Khaliq MN, Ouarda TBMJ, Ondo J-C, Gachon P, Bobée B. 2006. Frequency analysis of a sequence of dependent and/or non-stationary hydro-meteorological observation: a review. *Journal of Hydrology* **329**:534-552.

Lang M, Ouarda TBMJ, Bobée B. 1999. Towards operational guidelines for over-threshold modeling. *Journal of Hydrology* **225**:103-117.

Lee T, Ouarda TBMJ, Li J. 2013. An orchestrated climate song from the Pacific and Atlantic oceans and its implication on climatological processes. *International Journal of Climatology* **33**:1015-1020.

Lee T-S, Ouarda TBMJ. 2010. Long-term prediction of precipitation and hydrologic extremes with nonstationary oscillation processes. *Journal of Geophysical Research* **115**, D13107. doi:10.1029/2009JD012801

Martins ES, Stedinger JR. 2001. Generalized maximum likelihood Pareto-Poisson estimators for partial duration series. *Water Resources Research* **37**:2551-2557.

Nasri B, El Adlouni S, Ouarda TBMJ. 2013. Bayesian estimation for GEV-B-Spline model. *Open Journal of Statistics* **3**:118-128.

- Önöz B, Bayazit M. 2001. Effect of the occurrence process of the peaks over threshold on the flood estimated. *Journal of Hydrology* **244**:86-96.
- Padoan SA, Wand MP. 2008. Mixed model-based additive models for sample extremes. *Stat Probab Lett* **78**:2850-2858.
- Pickands J. 1975. Statistical inference using extreme order statistics. *Ann Stat* **3**:119-131.
- Rossi A, Massei N, Laignel B. 2011. A synthesis of the time-scale variability of commonly used climate indices using continuous wavelet transform. *Global and Planetary Change* **78**:1-13.
- Roth M, Buishand TA, Jongbloed G, Klein TAMG, van Zanten JH. 2012. A regional peaks-over-threshold model in a nonstationary climate. *Water resources research* **48**, W11533. doi: 10.1029/2012WR012214
- Shabbar A, Bonsal B, Khandekar M. 1997. Canadian precipitation patterns associated with the Southern oscillation. *Journal of Climate* **10**:3016-3027.
- Silva AT, Naghettini M, Portela MM. 2016. On some aspects of peaks-over-threshold modeling of floods under nonstationarity using climate covariates. *Stoch Environ Res Risk Assess* **30**:207-224. doi: 10.1007/s00477-015-1072-y
- Stone DA, Andrew JW, Zwiers FW. 2000. Trends in Canadian precipitation intensity. *Atmosphere-Ocean* **38**:321-347.
- Thiombiano AN, El Adlouni S, St-Hilaire A, Ouarda TBMJ, El-Jabi N. 2016. Nonstationary frequency analysis of extreme daily precipitation amounts in Southeastern Canada using a peaks-over-threshold approach. *Theoretical and Applied Climatology* **124**(1-2). doi: 10.1007/s00704-016-1789-7
- Thompson DWJ, Wallace JM. 2001. Regional climate impacts of the Northern hemisphere annular mode. *Science* **293**:85-89.
- Torrence C, Compo GP. 1998. A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society* **79**:61-78.

- Trenberth KE. 1990. Recent observed interdecadal climate changes in the Northern Hemisphere. *Bulletin of the American Meteorological Society* **71**:988-993
- Zhang X, Wang J, Zwiers FW, Ya Groisman P. 2010. The influence of large-scale climate variability on winter maximum daily precipitation over North America. *Journal of Climate* **23**:2902-2915.
- Wang C, Enfield DB. 2001. The tropical Western Hemisphere warm pool. *Geophysical Research Letters* **28**:1635-1638.
- Wu A, Hsieh WW. 2005. The nonlinear patterns of North American winter temperature and precipitation associated with ENSO. *Journal of Climate* **18**:1736-1752.
- Yue S, Pilon P, Cavadias G. 2002. Power of the Mann-Kendall and Spearman's rho tests for detecting monotonic trends in hydrological series. *Journal of Hydrology* **259**:254-271.

Tables illustration

Table I. Kendall's Tau (in %) with significant correlation in bold

Climate index windows													
	JD	DJF	JFM	FMA	MAM	AMJ	MJJ	JJA	JAS	ASO	SON	OND	NDJ
Upper Stewiacke station													
AO&V1	+20	-3	+2	+1	+10	+17	+15	+12	+15	+18	+25	+16	+7
AO&V2	+14	+7	+6	+2	-7	-9	-4	+6	+13	+5	+15	+23	+22
<i>Grandes Bergeronnes</i> station													
PNA&V1	+8	+18	+15	+5	-3	-4	+9	+3	+5	-1	+2	+1	+11
PNA&V2	+26	+6	+3	+14	+20	+24	+14	+12	+11	+16	+20	+22	+18

Table II. AIC values for different GPD B-spline models with the lowest AIC in bold

Knots k		2	3	4
Degree d		Upper Stewiacke station		
1		360,22	362,50	362,70
2		362,83	362,52	363,73-
3		363,17	363,78	365,92
Degree d		<i>Grandes Bergeronnes</i> station		
1		433,76	436,33	437,78
2		436,56	437,65-	439,58
3		437,10	439,51	440,65

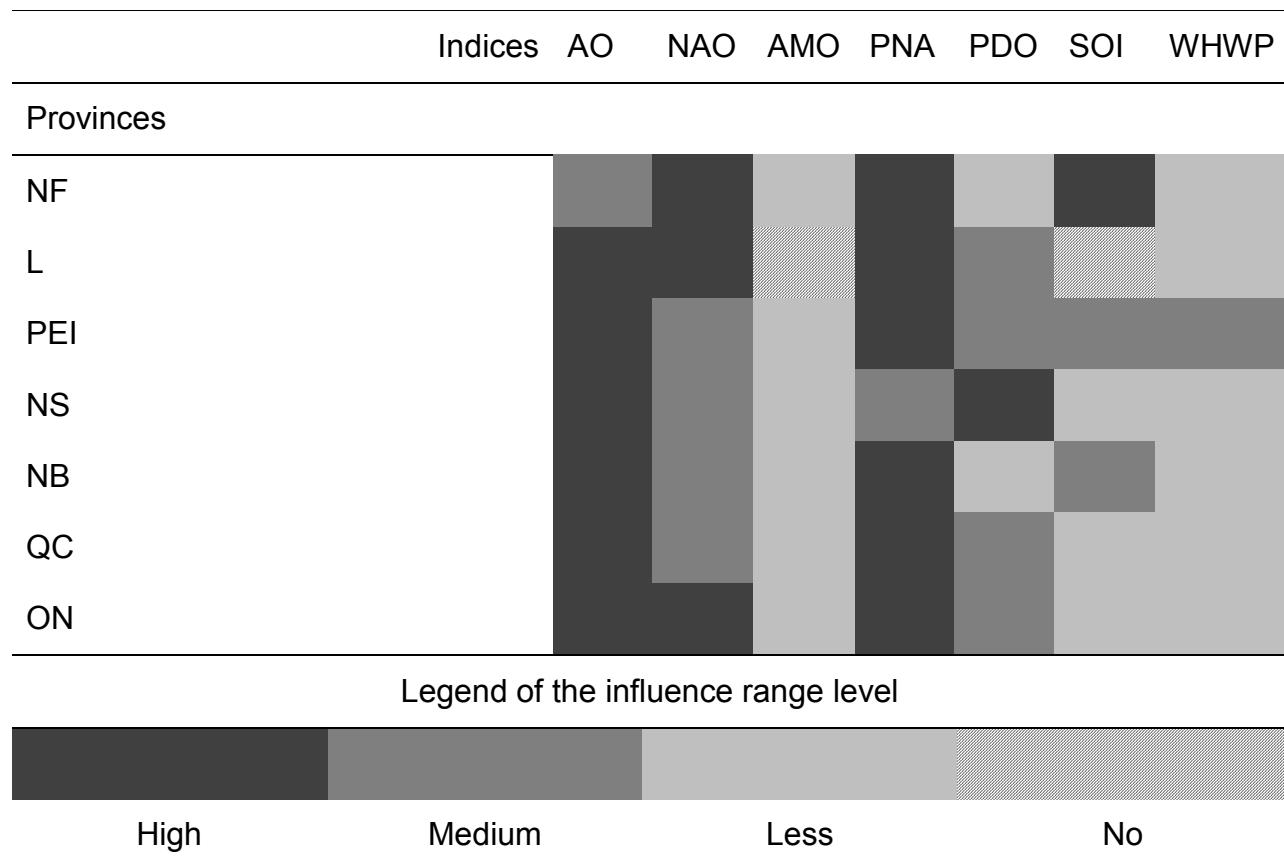
Table III Repartition of the number of stations by province given the significant correlation between the intensity of daily precipitation extremes and studied climate indices

Indices	AO	NAO	PNA	SOI	WHWP	PDO	AMO
Provinces							
PEI	2	1	3	2	0	2	0
NB	3	3	2	4	4	3	2
NFL	3	4	2	1	3	0	1
NS	8	4	4	5	5	9	3
ON	22	27	19	20	19	16	15
QC	19	13	20	5	6	5	7
Total	57	52	50	37	37	35	28

Table IV Repartition of the number of stations by province given the significant correlation between the frequency of daily precipitation extremes and studied climate indices

Indices	PNA	AO	NAO	PDO	AMO	WHWP	SOI
Provinces							
PEI	3	1	1	2	1	2	0
NB	5	1	0	0	0	1	1
NFL	3	2	4	3	4	2	2
NS	6	10	4	4	2	0	3
ON	24	24	25	20	20	18	9
QC	14	13	13	10	10	10	6
Total	55	51	47	39	37	33	21

Table V The level of influence of seven climate indices on daily precipitation extremes by province in Southeastern Canada



Figures illustration

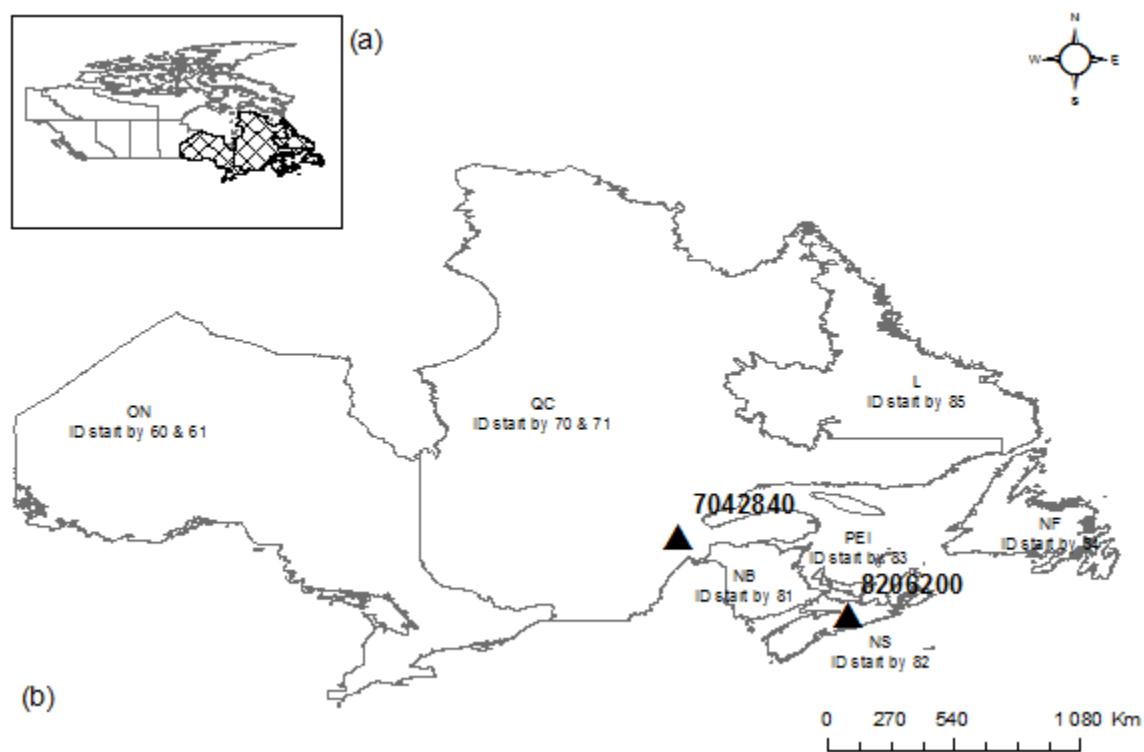


Figure 1. Southeastern Canada (a) and location of studied stations (b)

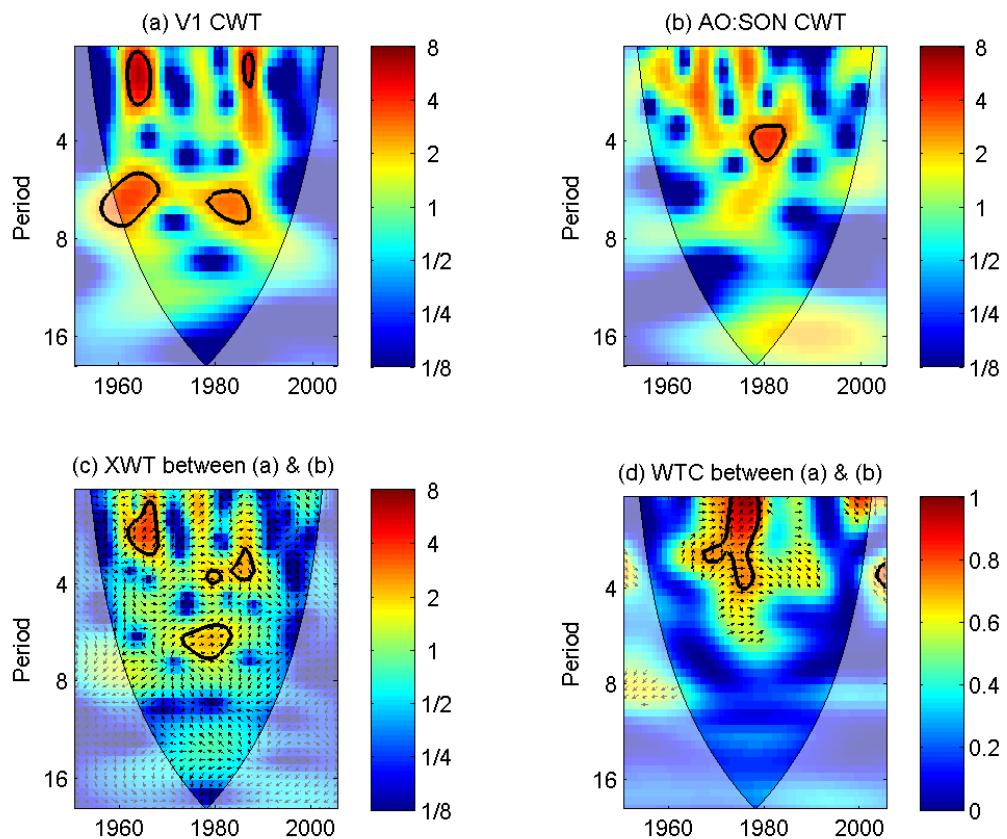


Figure 2. Wavelet analysis results with CWT illustration for V1 (a) and AO index: SON window (b). XWT (c) shows the common features between (a) & (b). WTC (d) highlights the covariance between (a) & (b)

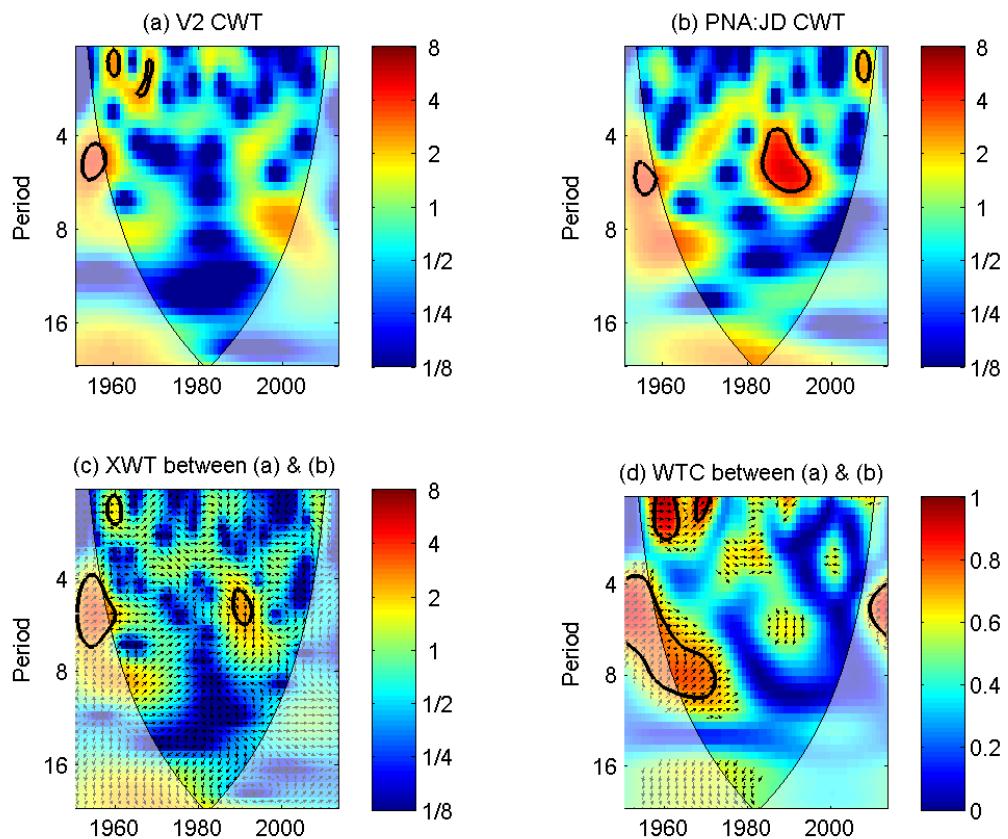


Figure 3. Wavelet analysis results with CWT illustration for V2 (a) and PNA index: JD window (b). XWT (c) shows the common features between (a) & (b). WTC (d) highlights the covariance between (a) & (b)

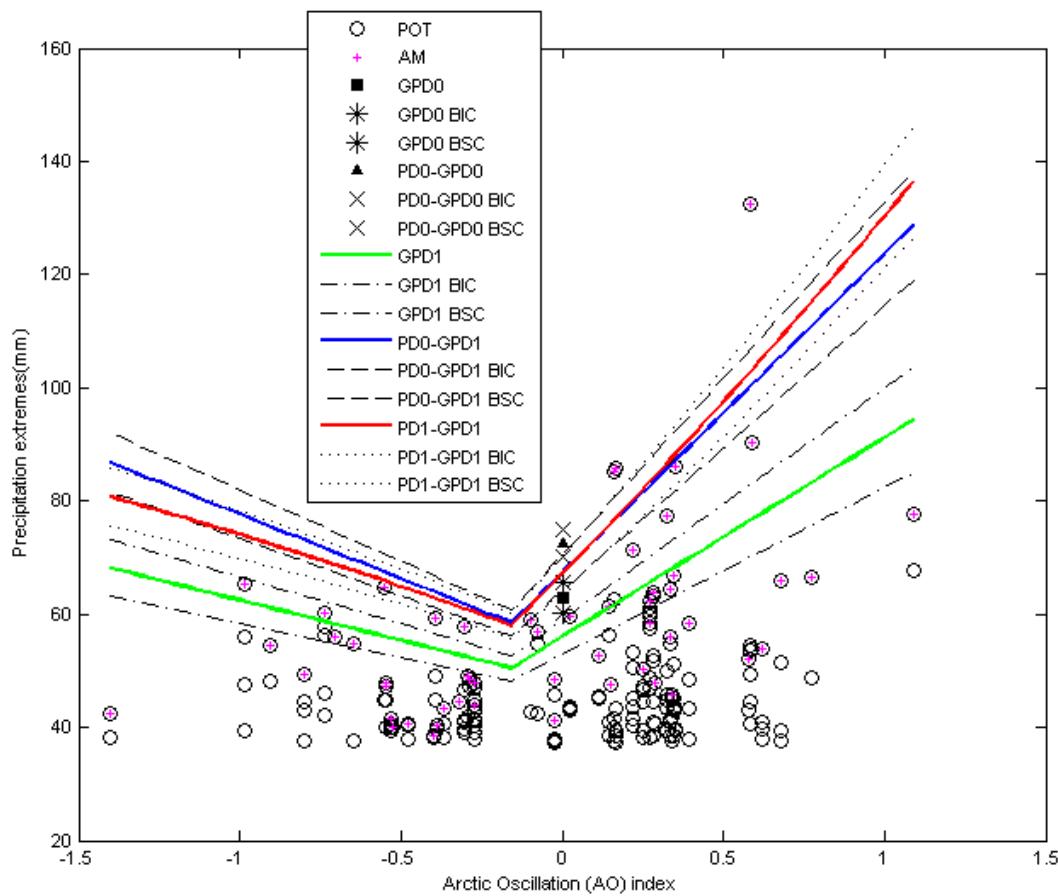


Figure 4. Estimated quantiles from stationary (GPD0, PD0-GPD0) and nonstationary (GPD1, PD0-GPD1, PD1-GPD1) models, using the Arctic Oscillation as a covariate, associated with a non-exceedance probability of 0.9. Lower (BIC) and upper (BSC) 5% confidence intervals are also shown. Illustrated observations are independent daily peaks (POT) and annual maxima (AM) precipitation values for the Upper Stewiacke station.

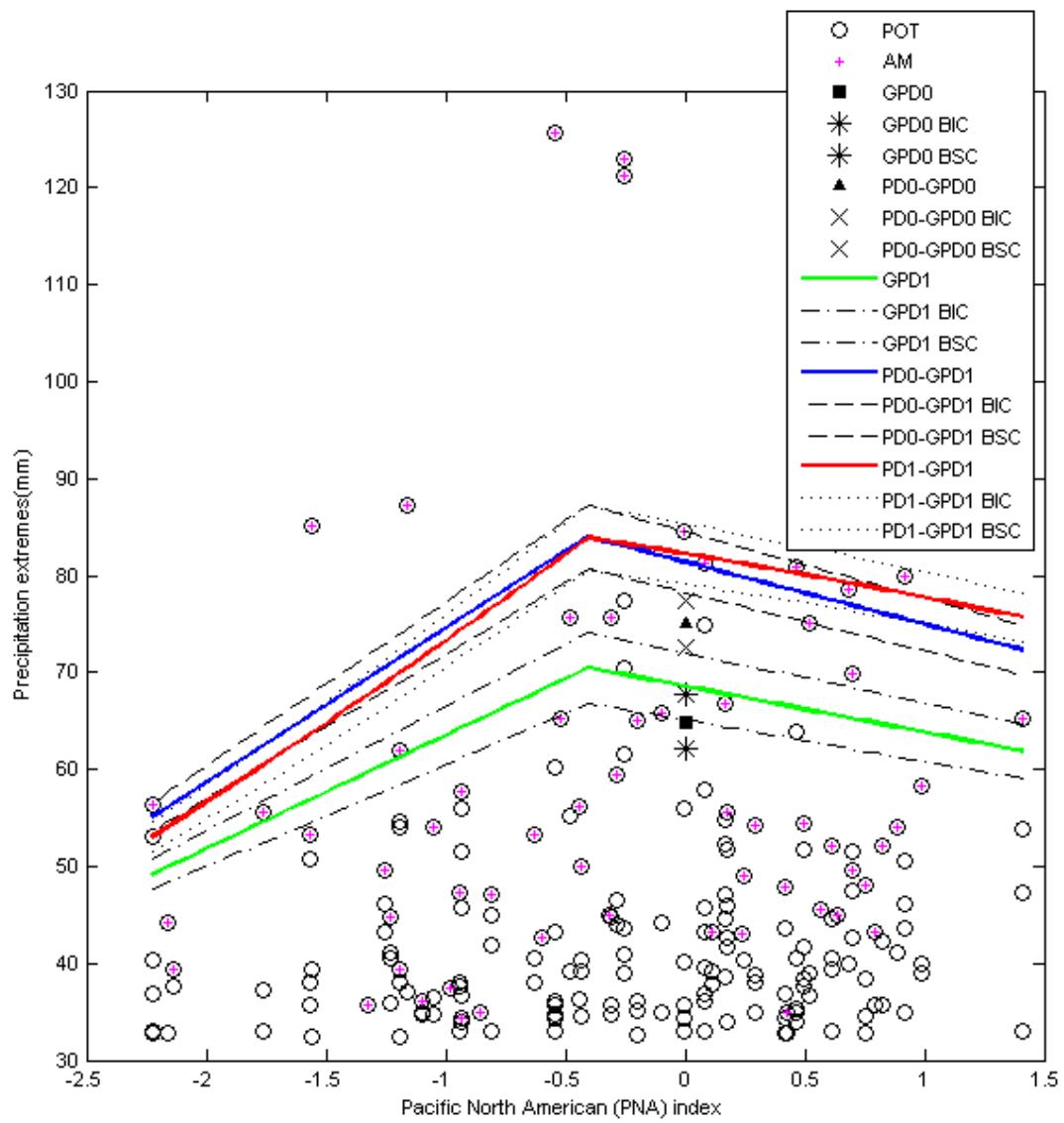


Figure 5. Same description as in Figure 4, but using the Pacific North American index as a covariate for the *Grande Bergeronnes* station.

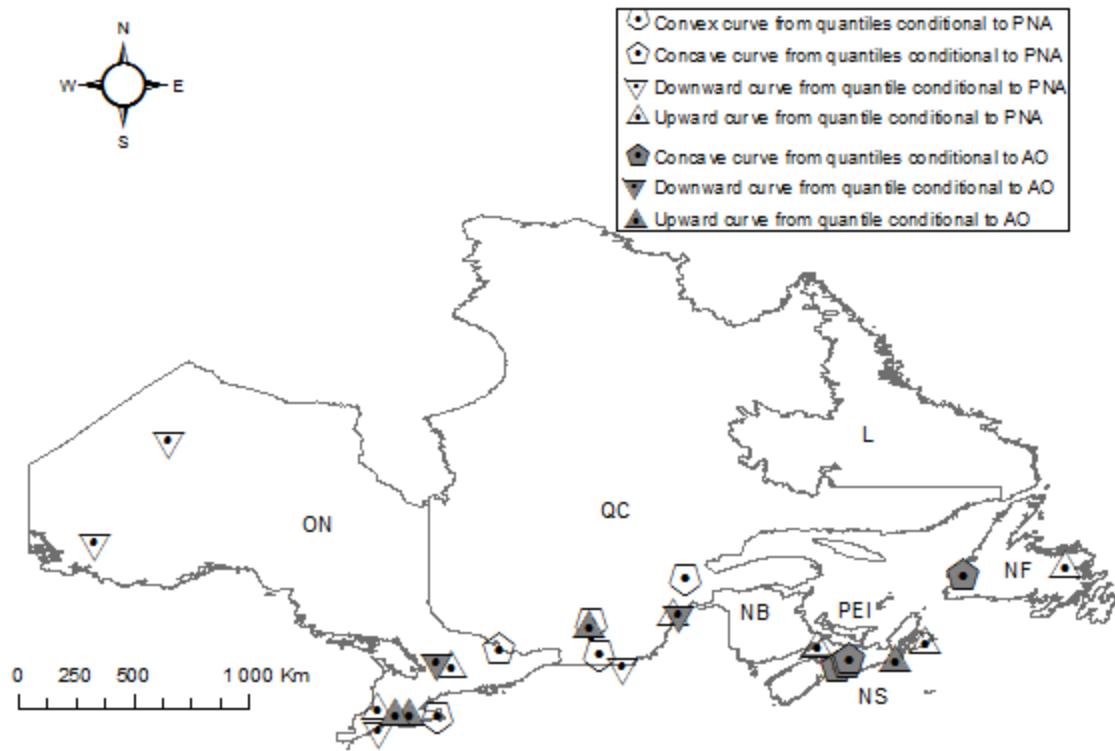


Figure 6. Linear and nonlinear response of precipitation extremes to AO and PNA indices at some stations in Southeastern Canada

9 ARTICLE 3: BAYESIAN INFERENCE WITH HYBRID GENERALIZED PARETO MODEL.

Alida N Thiombiano (@)⁽¹⁾, Salaheddine El Adlouni⁽²⁾, André St-Hilaire⁽¹⁾, and Taha B.M.J Ouarda⁽¹⁾

⁽¹⁾ Centre Eau–Terre-Environnement, Institut National de la Recherche Scientifique, 490, de la Couronne Street, Québec, Québec, G1K 9A9, Canada

⁽²⁾ Université de Moncton, 18 Antonine-Maillet Avenue, Moncton, New-Brunswick, Canada E1A 3EA.

E-mail de l'auteur de correspondance : alida_nadege.thiombiano@ete.inrs.ca

Article à Soumettre pour publication au *Stochastic Environmental Research and Risk Assessment* (SERRA).

Acknowledgments: The Authors are grateful to the International Center for Research and Development (ICRD) and to the Natural Sciences and Engineering Research Council of Canada (NSERC) for the financial support through the FACE (*Faire Face Aux Changements Ensemble*) project grant.

Abstract

The peaks-over-threshold (POT) approach is less popular than the annual maxima method because of the complexity associated with the threshold determination. The so-called “mixture models” applied in the POT framework offer a solution when the full range of dataset is available (e.g. daily observations). Indeed, the simplest form of these models consists of a Generalized Pareto distribution (GPD) limit model for the heavy tail component (i.e. observations above the threshold), and a parametric distribution (e.g., Gamma, Weibull, Normal, Student-t, ...) for the values below the threshold, which is the junction point of both components. Thus, the method aims at tackling the subjectivity and complexity associated with the choice of the threshold in the POT classical approach. In this paper, such hybrid Generalized Pareto model is developed with two separate bulk distributions: the Gamma and the Gaussian. The transition at the threshold point is also assessed from the resulting probability density function in two ways: (1) enforce one single continuity constraint at that location (models M2 and M3) and (2) let free all the model parameters (model M1). A non-informative Bayesian approach is proposed to estimate the parameters of the developed models as well as their associated quantiles. The estimates of the GPD classical model (M0) where the threshold is fixed before the inference process are also presented for comparison purposes. The assessment of M0, M1, M2 and M3 estimation errors, is done through a simulation study by evaluating type I and type II errors, using the relative bias and the relative root mean square error. Subsequently, an application of these four models is discussed for daily precipitation amounts from Pierre Trudeau International Airport station located in Montreal (Quebec, Canada).

Keywords: Hybrid model; Generalized Pareto distribution; Gamma distribution; Normal distribution; Bayesian inference; Precipitation; Canada.

1. Introduction

The frequency analysis of hydro-climatological extremes can be carried out by studying distributions of either the annual maxima (AM) or the peaks over threshold (POT). An asymptotic convergence result of both approaches with their limiting distributions, namely, the Generalized Extreme Value (GEV) and Generalized Pareto (GP) distributions is well established in the literature (Fisher and Tippett 1928; Jenkinson 1955; Pickands 1975). The sample size for these extreme value analyses is generally small mainly due to short historical records, whereas large samples can help reducing the estimation errors.

The POT approach has been considered as a viable and an interesting alternative to the AM method (Lang et al. 1999). However, it has not gained practical acceptance and remains mainly a topic of research in hydro-climatology. This may be due to the additional levels of complexity represented by the selection of the threshold and extraction of exceedances. The threshold is often fixed subjectively (e.g., using graphical tools) to achieve a minimal bias-variance in the model estimations (Davison and Smith 1990). With such an approach, there is no quantification of the uncertainty associated with the threshold selection. However, this uncertainty can be quantified with an objective determination of the threshold as a parameter to be estimated, through the so-called “mixture models” (Scarrott and MacDonald 2012).

The mixture models developed in the POT framework allow considering the entire time series of a variable of interest, by combining two components: the bulk (also named central) distribution composed of non-extreme events (i.e. those events which are below the threshold) and the right tail distribution of the time series (i.e. the events which are above the threshold), the threshold being their junction point. For the latter component, the GP distribution is automatically used because of its solid theoretical justification, hence the preferred expression “hybrid GP models” is used in this paper. For the bulk distribution, a sensitivity analysis should help for the choice of a flexible model in order to avoid a negative impact on the selection of the threshold and subsequently the fit of the events above it.

For example, to fit the bulk distribution, Frigessi et al. (2002) chose a Weibull distribution, Behrens et al. (2004) assessed the Gamma and Weibull distributions, Tancredi et al. (2006) tried a mixture of uniform densities, Carreau and Bengio (2009) considered the Normal distribution, MacDonald et al. (2011) proposed a Kernel density and Do Nascimento et al. (2012) tested a mixture of Gamma densities.

The probability density function (pdf) of the GP hybrid model constituted with a given parametric bulk distribution can show a discontinuity at the threshold point as it is the case with the hybrid Gamma-GP model proposed by Behrens et al. (2004). This discontinuity can lead to weak performance of the model in presence of realistic smooth transition. However, this lack of continuity can naturally be carried out through Bayesian inference, which provides a full posterior distribution of threshold estimates so that a continuous pdf can be observed (Scarrott and MacDonald 2012). Nevertheless, some constraints can be applied to enforce the pdf continuity at the threshold point. Frigessi et al. (2002) used a Cauchy cumulative density function (cdf) at this end, while Carreau and Bengio (2009) proposed two constraints for the pdf and its derivative continuity at the threshold point. However, in both cases, the enforced continuity fails in applications (Scarrott and MacDonald 2012). Furthermore, the use of nonparametric bulk distributions (e.g., Tancredi et al. 2006; MacDonald et al. 2011) can be an alternative for more flexible models but with some complexities to obtain explicit solutions mainly if numerical methods are used.

The hybrid GP models evaluated in this study are already developed in the literature (Behrens et al. 2004; Carreau and Bengio 2009), but they are less popular than block maxima or classic POT approaches because they sometimes had relatively poor performance in practice, as it is the case of the model proposed by Carreau and Bengio (2009). We then reviewed the theoretical approach of these models through Behrens et al. (2004) study and proposed an adaptation of Carreau and Bengio (2009) model from that former study, for better application purposes. The purpose of the present study is thus to develop a robust hybrid GP model in a default Bayesian framework by comparing two parametric bulk distributions.

In the next section, we recall the hybrid GP models and the Bayesian inference process. A simulation study is conducted to assess properties of extreme quantiles. A case study is then presented with the developed models. Conclusions follow with some recommendations.

2. Methodology

For a suitable high threshold u and a sequence of independent and identically distributed observations $Y = \{y_i, i = 1, \dots, n\}$, under some regularity conditions, exceedances above u (*i.e.*, $y_i - u$) can be approximated by a Generalized Pareto distribution (Pickands 1975; Davison and Smith, 1990) with its pdf given by:

$$g(y_i|u, \kappa, \beta) = \begin{cases} \frac{1}{\beta} \left(1 + \kappa \left(\frac{y_i-u}{\beta}\right)\right)^{-1-1/\kappa} & \kappa \neq 0 \\ \frac{1}{\beta} e^{-(y_i-u)/\beta} & \kappa = 0 \end{cases} \quad (1)$$

where $\beta > 0$ and $\kappa \in \mathbb{R}$ are the scale and the shape parameters, respectively.

For this model, called henceforth M0, the threshold is fixed first and only the exceedances sample will further be considered for the inference. In this way, the threshold determines the exceedances sample size as well as the quality of the estimation of GPD parameters and quantile estimations. Higher percentiles are widely used to set the threshold value (Thiombiano et al. 2016), and the suitability of such potential thresholds is usually investigated through graphical techniques (Coles 2001). To overcome this subjective choice, a hybrid GP model is proposed where the threshold is estimated jointly with the set of distribution parameters. In the present study, two bulk distributions (*i.e.*, distributions fitted to the values below the selected threshold) are investigated to develop a robust hybrid GP model, the Gamma and Normal distributions. The former distribution is bounded on the left at zero and unbounded on the right, while the latter distribution is symmetric.

2.1. Hybrid Gamma-GP model

Two hybrid GP models with the Gamma distribution are developed. The first model (M1) corresponds to Gamma-GP model as presented in Behrens et al. (2004). For the second model (M2), we impose a constraint on the scale parameter of the GP component, which is the heavy tail part of the proposed Gamma-GP model, to ensure continuity of the resulting pdf. The hybrid Gamma-GP model has the following pdf (f_{GPG}) and likelihood function (L) conditional to the parameters $\theta_1 = (a, b)$ and $\theta_2 = (\kappa, \beta)$ and a sample $\underline{y} = (y_1, \dots, y_n)$ for a given random variable Y as follows:

$$f_{GPG}(y_i) = [h(y_i) \cdot 1_{\{y_i < u\}}] + \left[((1 - H(u)) g(y_i - u)) \cdot 1_{\{y_i \geq u\}} \right] \quad (2)$$

$$L(\underline{y} | \theta_1, \theta_2) = \prod_{y_i \in C1} h(y_i | \theta_1) \cdot \prod_{y_i \in C2} (1 - H(u | \theta_1)) g(y_i - u | \theta_2) \quad (3)$$

where $C1$ and $C2$, respectively, represent the components of the observations below and above the threshold u ; H and h are the cdf and the pdf of the Gamma distribution respectively, with the vector of the parameters $\theta_1 = (a, b)$. The parametrisation considered for the Gamma distribution is $h(y_i | \theta_1) = \frac{1}{b^a \Gamma(a)} y_i^{a-1} e^{-y_i/b}$; $y_i > 0$, $a > 0$, $b > 0$; and the coefficient of skewness $C_s = \frac{2}{\sqrt{a}}$. The pdf of the GP distribution, $g(\cdot | \kappa, \beta)$, is given in equation (1).

In the case of model M2, the continuity of the function f_{GPG} at the threshold u , is guaranteed by equating

$$h(u | \theta_1) = (1 - H(u | \theta_1)) \cdot g(u | \kappa, \beta) \quad (4)$$

This is equivalent to fixing the scale parameter of the GP distribution:

$$\beta = \frac{(1 - H(u | \theta_1))}{h(u | \theta_1)} \quad (5)$$

which is a function of the threshold and the parameters of the bulk distribution of the hybrid model.

This formulation of model M2 pdf continuity constraint at the threshold point is based on the study of Carreau and Bengio (2009) which however enforced two constraints in the pdf of their model and up to its first derivative too. As highlighted by equations (5) and (6), we adopted a different approach. The mixture model proposed in their paper, does not account for the tail fraction ($1 - H(u)$). MacDonald et al. (2011) also indicated that by fixing the threshold and GPD scale parameters to ensure the continuity constraints Carreau and Bengio (2009) had imposed the location of the threshold to be estimated is somewhat influenced, and the constraints being two strong, they can lead to a dependency between parameters. This can finally affect the flexibility of the model (). Although, in our study, an application of the approach of Carreau and Bengio (2009) as they defined it, fails in practice because the continuity of the resulting pdf at the threshold is not guaranteed. Moreover, the weighting factor they added to the pdf expression to ensure that the density integrates to one is not verified.

2.2. Hybrid Normal-GP model

The hybrid Normal-GP model (M3) has the same expression as M2, but the pdf and cdf of the Gamma distribution, h and H , are replaced by those of the Normal distribution, noted f and F henceforth. The pdf of the Normal distribution is $f(y|\mu, \sigma) = \frac{1}{(\sigma\sqrt{2\pi})} \exp\left(\frac{-(y-\mu)^2}{(2\sigma^2)}\right)$, where μ and σ represent respectively the mean and standard deviation, and form the vector $\theta_3 = (\mu, \sigma)$. The continuity constraint is considered for this model for sensitivity analysis of the hybrid GP model fitting to the choice of different parametric bulk distributions. For the continuity at the threshold of model M3 pdf, the equations 2, 3, 4 and 5 become respectively 6, 7, 8 and 9.

$$f_{GP_G}(y_i) = [f(y_i) \cdot 1_{\{y_i < u\}}] + \left[((1 - F(u)) g(y_i - u)) \cdot 1_{\{y_i \geq u\}} \right] \quad (6)$$

$$L(y|\theta_3; \theta_2) = \prod_{y_i \in C1} f(y_i|\theta_3) \cdot \prod_{y_i \in C2} (1 - F(u|\theta_3)) g(y_i - u|\theta_2) \quad (7)$$

$$f(u|\theta_3) = (1 - F(u|\theta_3)) \cdot g(u|\kappa, \beta) \quad (8)$$

$$\beta = \frac{(1 - F(u|\theta_3))}{f(u|\theta_3)} \quad (9)$$

2.3. Bayesian inference

Estimation of parameters for the model M_j ($j=1, 2, 3$) is carried in a Bayesian framework in order to obtain a full conditional posterior distribution of each parameter estimates. The mean or the mode of this distribution can be considered as the Bayesian estimator of the parameter because they are associated with a quadratic and absolute loss functions respectively. These criteria are useful to assess the error in the statistical decision of the chosen estimator (Robert 2006). Moreover, uncertainty can be easily quantified along with the inference process, and credibility intervals deduced from the posterior distributions of the estimates. The first step is to select prior distributions of the parameters. This information is updated with the available observations using the Bayes theorem and, then, the posterior distribution (de Zea Bermudez and Turkman, 2003; Robert 2006; Ouarda and El Adlouni 2011) is given by $P(\theta|y) = \frac{L(y|\theta)\pi(\theta)}{\int L(y|\theta)\pi(\theta)d(\theta)}$; where $\underline{y} = (y_1, y_2, \dots, y_n)$ is the vector of observations with size n; $\pi(\cdot)$ represents the prior probability density of the parameters θ constituted by $(\theta_1; \theta_2)$ in the case of M1 and M2, and $(\theta_3; \theta_2)$ for M3; $L(y|\theta)$ is the likelihood of the observations knowing θ ; $P(\theta|\underline{y})$ is the posterior probability density of the parameters given the observations.

For models M1 and M2, two priors are investigated for the threshold, a truncated Normal distribution (Behrens et al. 2004; MacDonald et al. 2011; Do Nascimento et al. 2012), given by equation (10) and a non-informative Uniform discrete distribution.

$$\pi_u(u) = \frac{1}{\sqrt{2\pi\sigma_u^2}} \frac{\exp[-0.5(u-\mu_u)^2/\sigma_u^2]}{F[-(e_1-\mu_u)/\sigma_u]} \quad (10)$$

where the truncated parameter e_1 is set as the median (Q_{50}) of the sample \underline{y} , while the 70th percentile (Q_{70}) of \underline{y} is chosen for the mean parameter μ_u , and the standard deviation computed as $\sigma_u = \frac{Q_{99}-Q_{50}}{2}$ with Q_{99} the 99th percentile of \underline{y} . F is the cdf of the Normal distribution.

A joint non-informative Jeffrey's prior is considered for the Gamma distribution parameters using the “*Poly-Gamma*” function (Liseo, 1993). We also assume independence between the GP distribution parameters, by considering a Beta (6, 9) distribution on the interval $[-0.5; 0.5]$ for the GP shape parameter (Martins and Stedinger 2001), while a non-informative Uniform discrete distribution for the scale parameter. The joint prior distributions are

$$\pi_{a;b}(a, b) = \sqrt{aPG(1, a) - 1/b} \quad (11)$$

$$\pi_{\kappa;\beta}(\kappa, \beta) = f_{Beta}(\kappa, 6, 9)^{*}1 \quad (12)$$

Therefore, for the models M1 and M2, $\pi(\theta) = \pi_u(u) * \pi_{a;b}(a, b) * \pi_{\kappa;\beta}(\kappa, \beta)$

In the case of model M3 (GP-Normal), only a non-informative Uniform discrete distribution is considered for all parameters, which is equivalent to $\pi(\theta) = 1$. For the model M0, the generalized maximum likelihood (GML) approach (Martins and Stedinger 2000; El Adlouni and Ouarda 2008) is simply used as it is a Bayesian case where prior information is provided only for the shape parameter. The posterior distribution of this parameter is hence computed numerically using a procedure based on the Newton-Raphson iteration (Hosking and Wallis 1987).

The posterior distributions of the parameters of models M1, M2 and M3 are however determined using Monte Carlo Markov Chain (MCMC) simulation methods. The Metropolis-Hastings (M-H) MCMC algorithm (Metropolis et al. 1953; Hastings 1970) is commonly used for that purpose and is therefore implemented in this study. Initial values are set for each parameter of the vector θ . Firstly, the threshold u_1 is fixed successively as the median (Pth_{50}), the 90th (Pth_{90}) and 99th (Pth_{99}) percentiles of the generated data vector, in order to perform a sensitivity analysis of the inference results for low to high initial threshold values. This is important because the choice of the threshold value must allow dealing with a good compromise in the estimates bias and variance. Subsequently, the GML estimates of θ_{M1} and θ_{M2} given by $(a1, b1, \kappa1, \beta1)$, and the maximum likelihood (ML) estimates of $\theta_{M3} = (\mu1, \sigma1, \kappa1, \beta1)$, are obtained as initial values.

However, in the case of M2 and M3, the associated GPD scale parameters are given by

$$\beta_{M2} = \frac{(1-H(u_1|\theta_{M2}))}{h(u_1|\theta_{M2})} \quad (13)$$

$$\beta_{M3} = \frac{(1-F(u_1|\theta_{M3}))}{h(u_1|\theta_{M3})} \quad (14)$$

Let θ_1 be the current state of the generated Markov chain. For each parameter of the vector θ_1 , the corresponding random posterior distribution S_{θ_2} associated with the proposed candidates θ_2 , is generated using a Normal distribution centered on θ_1 with very small variance proportional to the ML estimate of each parameter. For each M-H iteration i , each candidate of θ_2 is accepted with probability ϕ (El Adlouni and Ouarda, 2009) as follows with an example using the GPD shape parameter

$$\phi(\kappa_1, \kappa_2) = \min \left\{ 1, \frac{\pi(\kappa_2) S_{\theta_2}(\kappa_1|\kappa_2)}{\pi(\kappa_1) S_{\theta_2}(\kappa_2|\kappa_1)} \right\} \quad (15)$$

The Markov chain length N and the burn-in period N_0 are fixed herein respectively to 10000 and 5000, so that the posterior distribution is the $N-N_0$ latest values of the estimates. The quadratic loss function has been considered in this study. The Bayesian estimator is the empirical mean of this posterior distribution of each parameter to be estimated.

3. Simulation study

3.1. Approach

3.1.1. Assessment of Type I estimation errors

The parameter values set in this section, and the next one, are called herein true values. To assess the M_j ($j=1, 2, 3$) models, two approaches are considered. Firstly, the estimation method was evaluated by generating data and fitting them using the same model (*i.e.* $\underbrace{M_j}_{j=1,2,3} \rightarrow \underbrace{M_j}_{j=1,2,3}$). A sample of size $n=5000$, assumed to represent 13.7 years of daily observations, are considered in each case.

For comparison purposes, this assessment is also made for model M0 by generating data similarly as with M1 and then modeling it with M0. The purpose is to evaluate the quantile estimates associated with the method of estimation in each case supposing that the fitted model is the right one (Type I estimation errors). The relative bias (RB) and the relative root mean square error (RRMSE) are used at this end as follows:

$$RB = \frac{1}{N_s} \sum_{t=1}^{N_s} \frac{\hat{Q}_{t,P_a} - Q_{P_a}}{Q_{P_a}} \quad (16)$$

$$RRMSE = \sqrt{\frac{1}{N_s} \sum_{t=1}^{N_s} \left(\frac{\hat{Q}_{t,P_a} - Q_{P_a}}{Q_{P_a}} \right)^2} \quad (17)$$

where \hat{Q}_{t,P_a} is the quantile estimates for the iteration t and given the non-exceedance annual probability P_a ; $N_s = 100$ is the number of samples evaluated; and Q_{P_a} is the true value of the quantile, with probability of non-exceedance P_a .

Moreover, with models M0, M1 and M2, nine scenarios presented in Table 1, are considered regarding three different probabilities $p(u_0)$ for true thresholds u_0 determination and three skewness coefficient (C_s) associated to the Gamma distribution shape. This latter choice looks at analyzing asymmetric ($C_s=3$) and symmetric ($C_s=2, 0.5$) Gamma distributions. The C_s values were selected to cover various skewness behavior commonly observed in hydrological variables, while u_0 value was evaluated assuming that 30% (i.e. $p(u_0)=0.7$) of the sample values are above the threshold.

Table 1. Data generation scenarios with a Gamma-GPD hybrid model

	C_s	3 ($a_0=0.44$)	2 ($a_0=1$)	0.5 ($a_0=16$)
$p(u_0)$				
0.7		Case 11	Case 12	Case 13
0.8		Case 21	Case 22	Case 23
0.9		Case 31	Case 32	Case 33

As we can see from Table 1, the small value of the Gamma shape parameter is associated with a large positive skewness. The Gamma scale parameter true value b_0 is fixed at 5. For the generation with model M3 (which is the Normal distribution case), only the probability $p(u_0)$ associated with the threshold u_0 is necessary, the true mean and standard deviation being respectively set as 0 and 1. Only three scenarios (henceforth named Plans 1, 2 and 3 respectively for $p(u_0)=0.7$, 0.8 and 0.9) were associated with model M3 instead of the nine scenarios defined for models M0, M1 and M2. For each model M1, M2 and M3, the true threshold is then the inverse of the cdf of the appropriate bulk distribution given the corresponding parameters $(p(u_0), a_0, b_0)$ for M1 and M2, and $(p(u_0), \mu_0, \sigma_0)$ for M3.

For generating second component of the models M_j ($j=1, 2, 3$), which is a GPD, two true values are investigated for the shape parameter, $\kappa_0 = 0.05$ and $\kappa_0 = 0.2$ in order to evaluate different shape behaviors, while the true scale value is determined from equation (13) or (14) because of the theoretical analytical equation obtained for model M2 and M3.

3.1.2. Assessment of Type II estimation errors

Here, we want to evaluate the error associated to the right model choice instead of the method of estimation. The GEV distribution, which is widely used in statistical hydrology for estimating AM-based quantiles, is used to generate data, which are subsequently fitted by the models M_j ($j=0, 1, 2, 3$) as $GEV \rightarrow \underbrace{M_j}_{j=0,1,2,3}$. The cdf of GEV is

given by

$$GEV_{\kappa, \delta, \psi}(y) = \begin{cases} \exp \left[- \left(1 + \kappa \left(\frac{y-\psi}{\delta} \right) \right)^{-\frac{1}{\kappa}} \right] & \kappa \neq 0 \\ \exp \left[- \exp \left(- \frac{y-\psi}{\delta} \right) \right] & \kappa = 0 \end{cases} \quad (18)$$

where κ, δ, ψ are respectively the GEV shape, scale and location parameters.

The GEV distribution is more homogeneous than the hybrid GP models. With the second assessment proposed in this section, the adequacy of the more heterogeneous hybrid GP models can be discussed through BR and RRMSE.

For generation of data with the GEV distribution, the true values 0.01, 0.1 and 0.2 are explored for the shape parameter, while those for the scale and position parameters are respectively chosen to be 100 and 1000. The same sensitivity analysis for different initial values of threshold u_1 (Q_{50} , Q_{90} ; Q_{99}), is also examined.

3.1.3. Quantile estimation

The quantiles are estimated for three annual non-exceedance probabilities $P_a=0.5$, 0.9 and 0.99 which correspond respectively to return period $T=2$, 10 and 100 years. The true quantiles Q_{P_a} and the estimate quantiles (\hat{Q}_{t,P_a} , $t = 1, \dots, N_s$) are computed from models M0, M1, M2 and M3, and estimation errors are compared based on Types I and II assessment approaches described in sections 3.1.1 and 3.1.2.

A Poisson-Generalized Pareto (P-GP) approach is considered for the estimation with the model M0. The proportion of data points above the threshold is obtained in term of probability p_λ . Which is associated to the Poisson distribution rate parameter λ . The resulting quantiles are estimated from a GEV distribution (F^{-1}_{GEV}), where the scale (δ_0) and location (ψ_0) parameters are expressed using Poisson and GPD parameters (Katz et al. 2002) as follows:

$$\begin{cases} M0Q_{P_a} = 1/F_{GEV}(P_a, \kappa_0, \delta_0, \psi_0) \\ M0\hat{Q}_{t,P_a} = 1/F_{GEV}(P_a, \hat{\kappa}, \hat{\delta}, \hat{\psi}) \end{cases} \quad (19)$$

where $\kappa_0 = 0.05$ for one test scenario and $\kappa_0 = 0.2$ for another;

$$\delta_0 = \frac{\beta_0}{p_{\lambda_0}^{\kappa_0}} \text{ and } \psi_0 = u_0 - \left[\delta_0 \left(\frac{p_{\lambda_0}^{\kappa_0} - 1}{\kappa_0} \right) \right] \text{ with } p_{\lambda_0} = 1 - H(u_0, a_0, b_0)$$

For the quantile estimates, $\hat{p}_\lambda = s/n$ must be used in place of p_{λ_0} . The size of the data above the fixed threshold is s and $n=5000$.

The non-exceedance probability P_e associated to the annual events is formulated as follows based on the annual non-exceedance probability P_a and on the expected number of peaks λ over the studied period:

$$P_e = P_a^{\frac{1}{\lambda}} \quad (20)$$

The true and estimated quantiles under models M1 and M2 (note that the same approach is used with M3), can be obtained as follows

$$\begin{cases} Q_{Pe} = \text{gaminv}(Pe, a, b) & \text{if } Pe < Pu \\ Q_{Pe} = [\text{gpinv}(pp, \kappa, \beta)] + u & \text{else} \end{cases} \quad (21)$$

With P_u the cdf of the Gamma distribution knowing the parameters (u, a, b) , and $pp = Pe - Pu$.

3.2. Results

3.2.1. Assessment of Type I estimation errors

The same model is used to simulate data and to estimate parameters. For the models M0, M1, M2 and M3, the simulation study results are presented for the GPD shape parameter $\kappa_0=0.05$ which showed the lowest RB and RRMSE in quantile estimates than when $\kappa_0=0.2$. The values of RB and RRMSE increase from less than 10% to more than 50% according to the simulation scenario, when the second component of the hybrid GPD model is generated by setting $\kappa_0=0.2$. Additionally, the performance of the estimated quantiles corresponding to three non-exceedance probabilities (0.5, 0.9 and 0.99) are summarized for all scenarios except the case where $C_s=2$ ($a_0=1$). The cases related to this C_s provided results which lie between the other two examined C_s values. The best scenarios for generating data are further provided for model M0, M1, M2 and M3 based on RB and RRMSE results.

In the following Tables, the RB and RRMSE (shown in brackets) are in %. The models based quantiles are associated with 2 years (Q2), 10 years (Q10) and 100 years (Q100) return periods.

3.2.1.1. Model M0

The results of simulation study show that, irrespective of the scenario of data generation, the quantiles are underestimated (negative RB) when the initial threshold u_1 (which correspond to the fixed threshold for M0) is equal to the median ($P_{th,50}$) of the data vector. However, an overestimation (positive RB) of the quantiles is noticed when this initial value of the threshold corresponds to the 90th and 99th percentiles ($P_{th,90}$ and $P_{th,99}$). The quantile estimates with $u_1=P_{th,99}$ are associated with higher RB (values vary between 54% to 456%) and RRMSE (range 13 to 63%) regardless of $p(u_0)$. With $u_1=P_{th,50}$ (and $u_1=P_{th,90}$), RB and RRMSE values increase (decrease) as $p(u_0)$ increases. Thus, $u_1=P_{th,90}$ seems to be adequate for data generation. Obviously, the probability $p(u_0)=0.9$ provided the lowest RB (less than 20% when using $C_s=3$ and less than 5% when $C_s=0.5$) and RRMSE values (up to 11% and 6% given the same respective C_s values) for quantile estimates. Table 2 presents the simulation study results for the assessment of Type I estimation errors when generating from a Gamma-GP hybrid model and modeling data with M0.

Table 2. Results of RB [and RRMSE] in % for various M0 based quantiles

Cs	u1=median		u1=90 th percentile		u1=99 th percentile	
	3	0.5	3	0.5	3	0.5
p(u0)	T=2 years					
0.7	-20 [1]	-10 [0.4]	149 [8]	39 [2]	456 [51]	139 [19]
0.8	-41[1]	-17 [0.3]	79 [6]	22 [2]	288 [41]	97 [16]
0.9	-65 [1]	-28 [0.3]	17 [3]	4 [1]	162 [28]	56 [13]
p(u0)	T=10 years					
0.7	-8 [2]	-11 [1]	91 [10]	32 [5]	279 [50]	122 [28]
0.8	-23 [1]	-14 [1]	53 [8]	17 [4]	191 [43]	90 [24]
0.9	-44 [1]	-22 [1]	12 [6]	1 [3]	125 [33]	54 [19]
p(u0)	T=100 years					
0.7	0.1 [3]	-13 [2]	66 [15]	29 [10]	208 [61]	115 [42]
0.8	-21 [3]	-19 [2]	40 [14]	13 [8]	145 [53]	89 [40]
0.9	-33 [3]	-24 [2]	9 [11]	-1 [6]	107 [44]	56 [31]

The best scenario to generate data is thus chosen to be with $u1=P_{th,90}$, $p(u0)=0.9$ and $Cs=0.5$ (meaning $a0=16$). This is the case 33 from Table 1. The RB (and RRMSE) associated with this scenario decreases (increases) as the return period T of quantile increases. From this scenario, the true values and the GML estimates of the GPD shape (κ) and scale (β) parameters are provided in Table 3. The GEV reparameterized scale (δ) and location (Ψ) parameters are computed given the GPD scale parameter, the threshold and the Poisson intensity probability of the proportion of data points above this threshold, are also presented. The 2 years (Q2), 10 years (Q10) and 100 years (Q100) return period quantiles are thus provided with their respective RB and RRMSE in percentage (%).

Given the fact that with model M0, the threshold is not estimated, $u_0=\text{gaminv}(0.9, 16, 5)$ which lead to the value 106.46, while the fixed threshold for the inference is $u_1=90^{\text{th}}$ percentile of the vector of simulated data. This value is equal to 115.20, thus a RB of 8% with a RRMSE of 0.5%. The GML estimates of κ and β are thus obtained from the sample of exceedances specifically above the u_1 value (115).

Table 3. Type I errors assessment results with case 33 for model M0

	True value	GML estimates	RB	RRMSE
κ	0,0500	0,0375	-25	87
β	13,7806	12,7976	-7	5
δ	15,4621	13,9788	-10	7
Ψ	140,0917	145,9894	4	1
Q2	145,8110	151,1512	4	1
Q10	176,9202	178,9746	1	3
Q100	220,0646	217,1299	-1	6

3.2.1.2. Models M1 and M2

The assessment of two priors, a truncated Normal distribution and a Uniform discrete distribution, for the threshold in the development of models M1 and M2, showed similar results for quantile estimates. So, the latter prior is used henceforth for presentation and discussion of results.

An observation of the quantile estimates with models M1 and M2, highlights that there are overall overestimated when $u_1=P_{\text{th}}_{50}$ whatever the values of C_s , and specifically with $C_s=3$ when $u_1=P_{\text{th}}_{90}$ as well as $u_1=P_{\text{th}}_{99}$. With $C_s=0.5$, the quantiles are underestimated when $u_1=P_{\text{th}}_{90}$ and $u_1=P_{\text{th}}_{99}$. Generation of data from models M1 and M2 also showed that the estimates RB and RRMSE of the quantiles are much higher with $C_s=3$ when $u_1=P_{\text{th}}_{99}$ than when $u_1=P_{\text{th}}_{50}$. The opposite observation is made when $C_s=0.5$. An initial threshold $u_1=P_{\text{th}}_{90}$ provided optimal results for both models. With this u_1 , the results showed that RB decreases as T increases overall when $C_s=3$, while it increases as T increases when $C_s=0.5$ regardless of the value of $p(u_0)$.

Given the fact that the results are overall comparable regardless of the value of $p(u_0)$ and that $p(u_0)=0.9$ showed larger RRMSE for quantile estimates when $C_s=0.5$, $p(u_0)=0.8$ can be used for generating data on both models M1 and M2. Tables 4 and 5 respectively present the results of the simulation study for M1 and M2.

Table 4. Results of RB [and RRMSE] in % for various M1 based quantiles

Cs	u1=median		u1=90 th percentile		u1=99 th percentile	
	3	0.5	3	0.5	3	0.5
p(u_0)	T=2 years					
0.7	17 [3]	12 [4]	-9 [3]	-5 [1]	20 [17]	-2 [3]
0.8	14 [3]	28 [2]	8 [3]	-3 [1]	15 [13]	-1 [3]
0.9	7 [4]	35 [2]	8 [3]	1 [8]	13 [11]	-0.2 [3]
p(u_0)	T=10 years					
0.7	12 [3]	10 [5]	3 [3]	-9 [1]	13 [18]	-6 [4]
0.8	9 [3]	28 [2]	2 [3]	-7 [2]	9 [15]	-4 [4]
0.9	2 [4]	37 [2]	2 [3]	-1 [10]	7 [12]	-3 [5]
p(u_0)	T=100 years					
0.7	5 [3]	6 [5]	-3 [3]	-14 [2]	7 [21]	-10 [5]
0.8	3 [3]	26 [3]	-4 [3]	-11 [2]	2 [17]	-8 [6]
0.9	-3 [4]	38 [3]	-4 [3]	-5 [11]	0.2 [15]	-7 [8]

Table 5. Results of RB [and RRMSE] in % for various M2 based quantiles

Cs	u1=median		u1=90 th percentile		u1=99 th percentile	
	3	0.5	3	0.5	3	0.5
p(u0)	T=2 years					
0.7	19 [3]	12 [4]	18 [2]	-5 [1]	26 [12]	-4 [2]
0.8	18 [3]	28 [2]	19 [3]	-4 [2]	26 [11]	-2 [2]
0.9	13 [3]	35 [1]	18 [3]	7 [10]	24 [8]	-1 [2]
p(u0)	T=10 years					
0.7	12 [3]	10 [4]	12 [3]	-9 [2]	20 [13]	-8 [2]
0.8	12 [3]	29 [2]	12 [3]	-8 [2]	20 [12]	-6 [3]
0.9	8 [3]	39 [2]	13 [3]	5 [12]	20 [9]	-6 [3]
p(u0)	T=100 years					
0.7	6 [3]	6 [5]	6 [3]	-14 [2]	14 [13]	-13 [2]
0.8	6 [4]	27 [2]	6 [4]	-12 [2]	14 [13]	-12 [4]
0.9	3 [4]	39 [2]	6 [4]	2 [13]	14 [10]	-11 [4]

The best scenario to generate data is thus chosen to be with $u1=P_{th90}$ and $p(u0)=0.8$ for both models M1 and M2. However, in the case of M1 (M2), $Cs=3$ ($Cs=0.5$) provided better estimates of quantiles than when using $Cs=0.5$ ($Cs=3$). The cases 21 and 23 from Table 1 are thus considered as best generation scenarios for models M1 and M2 respectively.

Tables 6 and 7 summarize respectively the modeling results with M1 and M2. The estimates of their parameters are the posterior mean (PM) resulting from the empirical posterior distribution of each estimated parameter and quantiles. The value of $u1$ is 8 (116) with model M1 (M2) best simulation scenario which is case 21 (case 23) from Table 1.

Table 6. Type I errors assessment results for model M1

	True value	PM estimates	RB (%)	RRMSE (%)
u	3.58	3.61	0.8	0.8
k	0.05	0.01	-77	9
β	3.42	3.68	7	3
Q2	21.52	23.40	8	3
Q10	29.88	30.58	2	3
Q100	41.44	39.87	-4	3

Table 6 shows that the quantiles are overestimated with model M1. The RB is positive for Q2 and Q10. There is an underestimation of Q100 (negative RB).

The empirical posterior distributions of model M1 parameters and quantiles with the case 21 are illustrated on Figure 1.

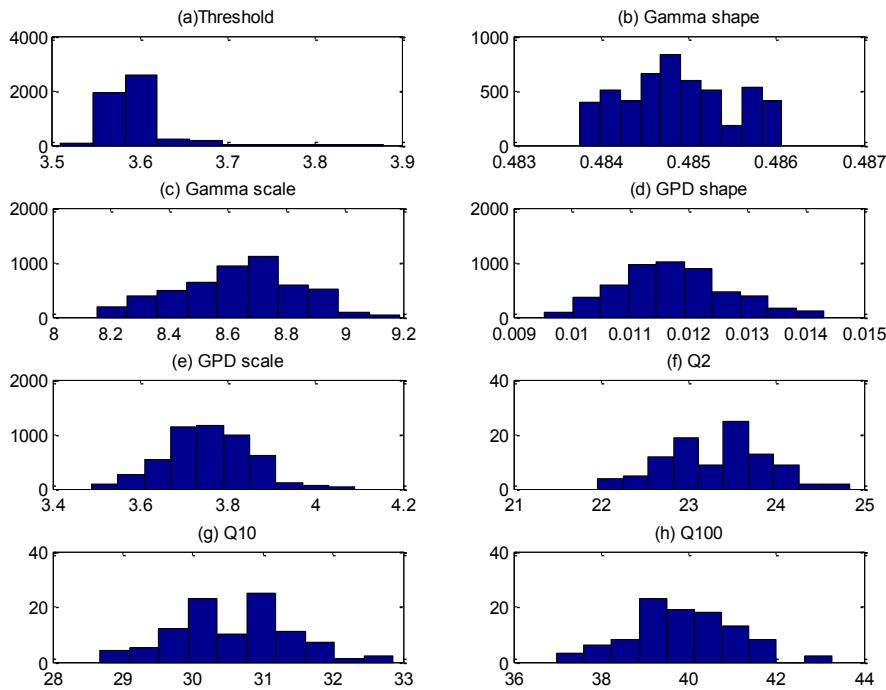


Figure 1. Empirical posterior distributions for model M1 parameters and based quantiles with Type I errors assessment

The empirical posterior distributions of the parameters are represented from 1.a to 1.e with the PM on horizontal axis and the N-N0 iteration of the MCMC algorithm on vertical axis. The empirical posterior distributions of quantiles Q2, Q10 and Q100 are illustrated from 1.f to 1.h where the vertical axis shows the number of sample Ns with the N-N0 iterations of the MCMC algorithm evaluated.

Table 7. Type I errors assessment results with for model M2

	True value	PM estimates	RB (%)	RRMSE (%)
u	96.17	109	13	3
k	0.05	0.01	-78	9
β	16.16	14.97	-7	4
Q2	180.92	174.44	-4	2
Q10	220.38	203.70	-8	2
Q100	274.99	241.05	-12	2

The model M2 based quantiles are all underestimated according to Table 7. Figure 2 presents the same results as with Figure 1 but for model M2 with the case 23.

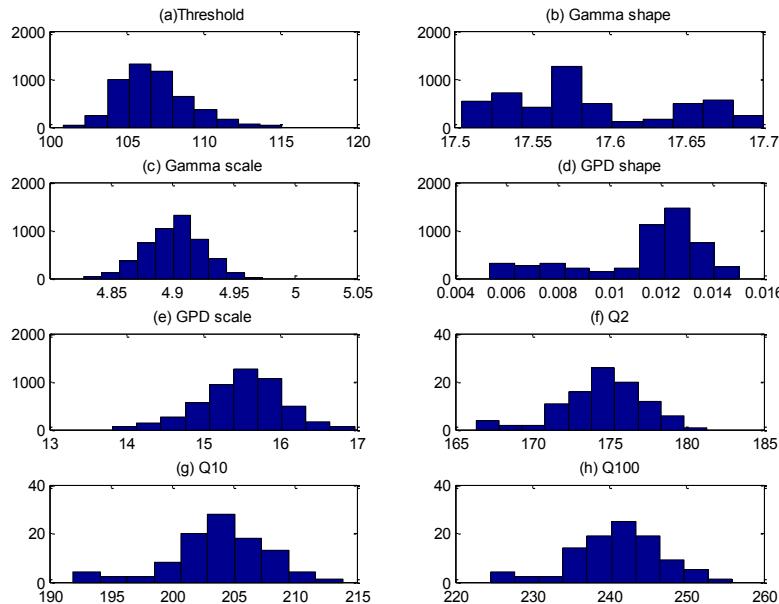


Figure 2. Empirical posterior distributions for model M2 parameters and based quantiles with Type I errors assessment

3.2.1.3. Model M3

The modeling results of model M3 in a default Bayesian framework showed that the Markov chain did not converge when the initial threshold u_1 is chosen to be equal to the median of the generated data, irrespective of the true threshold probability ($p(u_0)$) used. The best convergence of the Markov chain is associated with $u_1=99^{\text{th}}$ percentile regardless of all probabilities $p(u_0)$.

The quantiles are underestimated with $u_1=P_{\text{th},90}$ or $u_1=P_{\text{th},99}$, irrespective of the probability $p(u_0)$. Even if the results are comparable, the lowest RB and RRMSE are observed when $u_1=99^{\text{th}}$ percentile and $p(u_0)=0.9$. Thus, the Plan 3 defined in section 3.1.1 when we have to generate data from model M3, is of interest. Table 8 summarizes results for generating data and modeling them with M3.

Table 8. Results of RB [and RRMSE] in % for various M3 based quantiles

	$u_1=90^{\text{th}}$ percentile	$u_1=99^{\text{th}}$ percentile
$p(u_0)$		T=2 years
0.7	-37 [1]	-34 [1]
0.8	-25 [1]	-24 [1]
0.9	-22 [1]	-11 [1]
$p(u_0)$		T=10 years
0.7	-42 [1]	-44 [1]
0.8	-30 [1]	-33 [1]
0.9	-26 [1]	-20 [1]
$p(u_0)$		T=100 years
0.7	-47 [1]	-52 [1]
0.8	-35 [1]	-41 [1]
0.9	-31 [1]	-28 [1]

The best scenario for model M3 type I estimation error is with $u_1=99^{\text{th}}$ percentile and $p(u_0)=0.9$. Table 9 presents the results of the quantile estimates with this best scenario.

Table 9. Type I errors assessment results with for model M3

	True value	PM estimates	RB (%)	RRMSE (%)
u	1.28	2.81	1	5
k	0.05	0.01	-80	0.1
β	0.57	0.39	-31	1
Q2	3.78	3.35	-11	1
Q10	5.12	4.10	-20	1
Q100	6.98	5.04	-28	1

The model M3 based quantiles are all underestimated according to Table 9 results. The RB values are higher with this model than those associated with model M2. Figure 3 presents the same results as Figures 1 & 2 but for model M3 with Plan 3.

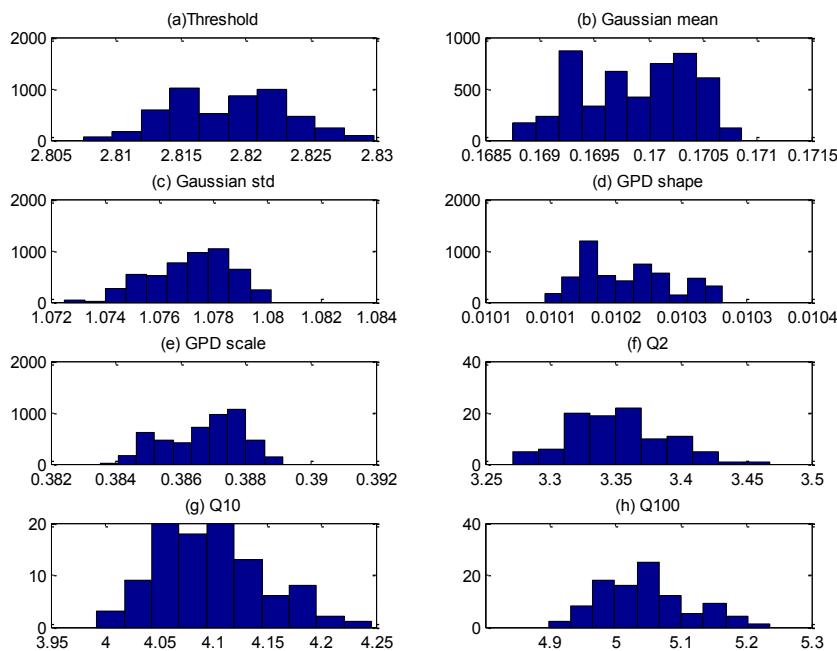


Figure 3. Empirical posterior distributions for model M3 parameters and based quantiles with Type I errors assessment

3.2.2. Assessment of Type II estimation errors

The GEV distribution is widely used for modeling extreme values. It is thus considered herein to generate data that will be modelled with M0, M1, M2 and M3 for assessing Type II estimation errors. The results of this assessment with the defined initial values of the threshold (u_1) and the GEV shape parameter $\kappa_0=0.1$, scale parameter $\delta_0 = 100$ and location parameter $\psi_0 = 1000$, are summarized in Table 10.

For this Table 10, the RB and RRMSE (shown in brackets) are in %. The models based quantiles are associated with 2 years (Q2), 10 years (Q10) and 100 years (Q100) return periods

Table 10: Type II errors assessment results for models M0, M1, M2 and M3

	u1=median			u1=90 th percentile			u1=99 th percentile		
	Q2	Q10	Q100	Q2	Q10	Q100	Q2	Q10	Q100
M0	-37 [0.2]	-36 [0.5]	-37 [1]	-12 [1]	-12 [2]	-12 [5]	40 [12]	42 [19]	46 [32]
M1	3 [1]	-0.4 [1]	-7 [1]	-5 [1]	-6 [1]	-8 [1]	-0.05 [2]	-4 [3]	-11 [3]
M2	3 [1]	-0.3 [1]	-7 [1]	-2 [1]	-7 [1]	-14 [1]	-3 [1]	-8 [1]	-15 [1]
M3	5 [1]	1 [1]	-5 [1]	-1 [1]	-6 [1]	-13 [1]	-2 [1]	-6 [1]	-13 [1]

Overall, the quantiles are underestimated when generating data from the GEV distribution and modelling it with either M0, or M1, or M2 or M3. The model M0 (models M1, M2, M3) quantile estimates show lower RB when $u_1=P_{th90}$ (with $u_1=P_{th50}$). The comparison of these results with those from section 3.2.1 allows concluding that the reduction in the bias of model M0 quantile estimates is more important when generating data from a heterogeneous hybrid Gamma-GP model than from a homogeneous GEV distribution. The estimates from model M1, M2 and M3 are similar with generated data from a GEV distribution. However, there is an important reduction in the estimation errors when generating data from a GEV and estimating quantiles with models M1, M2 and M3. This is specifically associated with the GEV and GPD shape parameter which is set at 0.05.

Indeed, as the shape value increase, the errors in both assessments are very comparable. This is the case with a value $\kappa_0=0.2$ for which results are not presented herein. The values of RB are higher with $\kappa_0=0.2$ than $\kappa_0=0.1$, the difference reaches up to 35%.

3.2.3. Comparison of models

The simulated data from the same generation scenario is considered at this point in order to compare each model to another given the RB and RRMSE of quantile estimates. For this purpose, the probability $p(u_0)=0.8$ is selected and the initial threshold u_1 is fixed at the 90th percentile. An overview of the estimates is additionally provided with the GPD shape parameter set at $\kappa_0=0.05$ and 0.2. In the case of models M0, M1 and M2, results are illustrated with an asymmetric Gamma distribution (i.e. $C_s=3$).

Results for Type I (Type II) error assessments are shown in Tables 11 (Table 12) for all models. These results are summarized for the 10-year return period quantile (Q10) because length of the generated data is of the order of 10 years ($n=5000$ values that result in 13.7 years).

Table 11: Theoretical and estimated values of the parameters of models M0, M1, M2 and M3. The RB and RRMSE are given for the 10 years return period quantiles Q10 (Type I errors assessment) associated with

Models	M0	M1	M2	M3
Scenario		$u_0=3.58; k_0=0.05; \beta_0=3.42$		$u_0=0.84; k_0=0.05; \beta_0=0.71$
Results	$u=u_1=8.25$	$\hat{u}=3.61$	$\hat{u}=2.23$	$\hat{u}=1.68$
	$\hat{\kappa}=0.04$	$\hat{\kappa}=0.01$	$\hat{\kappa}=0.01$	$\hat{\kappa}=0.01$
	$\hat{\beta}=3.80$	$\hat{\beta}=3.67$	$\hat{\beta}=4.22$	$\hat{\beta}=0.50$
	$RB(Q10)=53\%$	$RB(Q10)=2\%$	$RB(Q10)=12\%$	$RB(Q10)=-30\%$
	$RRMSE(Q10)=8\%$	$RRMSE(Q10)=3\%$	$RRMSE(Q10)=3\%$	$RRMSE(Q10)=1\%$
Scenario		$u_0=3.58; k_0=0.2; \beta_0=4.21$		$u_0=0.84; k_0=0.2; \beta_0=0.71$
Results	$u=u_1=8.50$	$\hat{u}=4.84$	$\hat{u}=3.89$	$\hat{u}=1.70$
	$\hat{\kappa}=0.14$	$\hat{\kappa}=0.01$	$\hat{\kappa}=0.01$	$\hat{\kappa}=0.01$
	$\hat{\beta}=4.21$	$\hat{\beta}=4.30$	$\hat{\beta}=4.56$	$\hat{\beta}=0.50$
	$RB(Q10)=51\%$	$RB(Q10)=-30\%$	$RB(Q10)=-25\%$	$RB(Q10)=-57\%$
	$RRMSE(Q10)=11\%$	$RRMSE(Q10)=10\%$	$RRMSE(Q10)=5\%$	$RRMSE(Q10)=1\%$

According to Table 11, when generating data from the GPD with $k_0=0.05$, model M1 provides lower RB than the other models. There is an important underestimation of the Q10 using model M3. The estimated threshold from this model is very low and can explain this result. The Q10 are overestimated using the classical GPD model (M0). When $K_0=0.2$, model M2 is better than model M1, M0 and M3. Because of the small estimated values of the threshold, models M1, M2 and M3 underestimated Q10. Nevertheless, the estimates of the GPD shape and scale parameters are very comparable irrespective of the model. As a weight is given to the GPD component (e.g. $k_0=0.2$), the value of the estimated threshold increases, and with him the RB except with model M0 (RB comparable for $k_0=0.05$ and 0.2).

Table 12: Theoretical values of the GEV distribution parameters and models M0, M1, M2 and M3 based estimates. The RB and RRMSE are given for the 10 years return period quantiles Q10 (Type II errors assessment)

Models	M0	M1	M2	M3
Scenario	k0=0.1; $\delta_0 = 100$; $\psi_0 = 1000$			
Results	$u=u_1=1243.63$	$\hat{u}=1017.74$	$\hat{u}=1016.31$	$\hat{u}=985.70$
	$\hat{\kappa}=0.02$	$\hat{\kappa}=0.01$	$\hat{\kappa}=0.01$	$\hat{\kappa}=0.002$
	$\hat{\beta}=131.74$	$\hat{\beta}=135.24$	$\hat{\beta}=134.01$	$\hat{\beta}=148.13$
	RB(Q10)=-12%	RB(Q10)=-7%	RB(Q10)=-7%	RB(Q10)=-6%
	RRMSE(Q10)=2%	RRMSE(Q10)=1%	RRMSE(Q10)=1%	RRMSE(Q10)=1%
Scenario	k0=0.2; $\delta_0 = 100$; $\psi_0 = 1000$			
Results	$u=u_1=1271.43$	$\hat{u}=996.31$	$\hat{u}=1001.35$	$\hat{u}=979.69$
	$\hat{\kappa}=0.18$	$\hat{\kappa}=0.01$	$\hat{\kappa}=0.01$	$\hat{\kappa}=0.01$
	$\hat{\beta}=163.39$	$\hat{\beta}=157.75$	$\hat{\beta}=156.89$	$\hat{\beta}=162.33$
	RB(Q10)=-20%	RB(Q10)=-25%	RB(Q10)=-25%	RB(Q10)=-24%
	RRMSE(Q10)=3%	RRMSE(Q10)=1%	RRMSE(Q10)=1%	RRMSE(Q10)=1%

In light of the results presented in Table 12, models M1, M2 and M3 lead to similar estimates of the parameters and quantiles when the data are generated from the same homogeneous GEV distribution. As in Table 11, the RB increase when a weight is given to the GPD component ($k_0=0.2$).

4. Case study

The precipitation time series from station 7025250 (latitude: 45.47; longitude: -73.75; altitude: 36) located at Pierre Elliott Trudeau International Airport (Montreal, Quebec, Canada) is used to conduct a case study. A record of 48 years of daily precipitation amounts from 1945 to 1992 (leading to 17532 values) is considered. However, only days with precipitation (i.e. values >0) is used to fit the developed models. This allows evaluating specifically the performance of models M1 and M2 because of the selected Gamma distribution to fit the data below the estimated threshold. In this way, the proportion of the days without precipitation (i.e. value=0), is assessed and found to be equal to 55% of the sample (leading to a size of 7869 non-zero values). A graphical overview of the series of the annual number of zeros showed a stationary time series with an annual average of 200 zeros and a standard deviation equal to 13. To account for this proportion of zeros (p_0) in the estimation of quantiles, the non-exceedance probability P_e is adjusted as $P_{ad}=(P_e-p_0)/(1-p_0)$. This is done following the method used in El Adlouni et al. (2011).

Both the 90th and 99th percentiles are considered for comparison purposes and to set the threshold value for M0 model and to initialize thresholds for models M1 and M2. The former percentile provided the best results in the simulation study, while the latter percentile is found adequate for modeling with the POT approach (Osman et al. 2013; Thiombiano et al. 2016). The parameter estimates are presented in Table 13 for models M0, M1, M2, M3 and GEV. The GEV distribution is used to fit the AM of the case study sample and is considered for comparison purposes between the GEV and M0 (which is a Poisson-GP model).

Table 13. Estimates of the parameters and the negative log-likelihood function of the model for an initial threshold u (in mm) equal to the 99th (90th) percentile

M0	M1	M2	M3	GEV
Estimates				
s=78 (687)	s=29 (3303)	s=549 (3276)	s=3714 (4905)	s=48
u=35 (15)	u1=35 (15)	u1=35 (15)	u1=35 (15)	$\hat{\xi}=-0.0380$
	$\hat{u}=44.30 (3.80)$	$\hat{u}=18.65 (4.03)$	$\hat{u}=3.27 (1.61)$	$\hat{\delta}=9.82$
	$\hat{a}=0.77 (0.95)$	$\hat{a}=0.78 (0.94)$	$\hat{a}=5.28 (3.72)$	$\hat{\psi}=41.66$
	$\hat{b}=7.47 (4.47)$	$\hat{b}=7.44 (5.99)$	$\hat{a}=3.75 (2.48)$	
$\hat{\kappa}=0.0398$ (-0.0159)	$\hat{\kappa}=0.0172$ (0.0096)	$\hat{\kappa}=0.0175$ (0.0454)	$\hat{\kappa}=0.0120$ (0.0082)	
$\hat{\beta}=9.93 (9.64)$	$\hat{\beta}=13.97 (7.98)$	$\hat{\beta}=6.95 (5.70)$	$\hat{\beta}=7.63 (7.14)$	
Negative log-likelihood				
253.4 (2233.3)	21457 (21268)	21459 (21553)	24465 (23323)	184

s is the sample size of extreme events

Moreover, the estimates of the threshold under the hybrid GP models (M1, M2 and M3) are above (and significantly below) u1 with model M1 when u1=99th percentile (u1=90th percentile). In the case of models M2 and M3, the estimated threshold is below u1 no matter which percentile is used. However, a very low threshold is specifically estimated by M3, suggesting that this model is not adequate for daily precipitation, the values below 10mm being more frequent in the dataset of this variable. The use of such distribution to model the non-extreme events can thus lead to absurd estimates when dealing with hydrological variables like precipitation.

There is an important difference in the threshold estimates from models M1 and M2 when $u_1=99^{\text{th}}$ percentile. However, the resulting GPD shape parameter estimates are comparable with an estimate that can represent a light tail, in contrary to what is obtained from M0 (i.e. heavier tailed distribution). When $u_1=90^{\text{th}}$ percentile, the estimated threshold from M1 and M2 are similar but not the estimated GPD shape parameter. With no surprise, the GPD scale parameter is different between M1 and M2, because of the pdf continuity constraint enforced at the threshold point in the latter model. In the case of M1 and M2, the estimates of the Gamma distribution shape parameter are associated with skewness coefficient of order 2.

Daily precipitation amounts with the location of the 90^{th} and 99^{th} percentiles are illustrated in Figure 4.

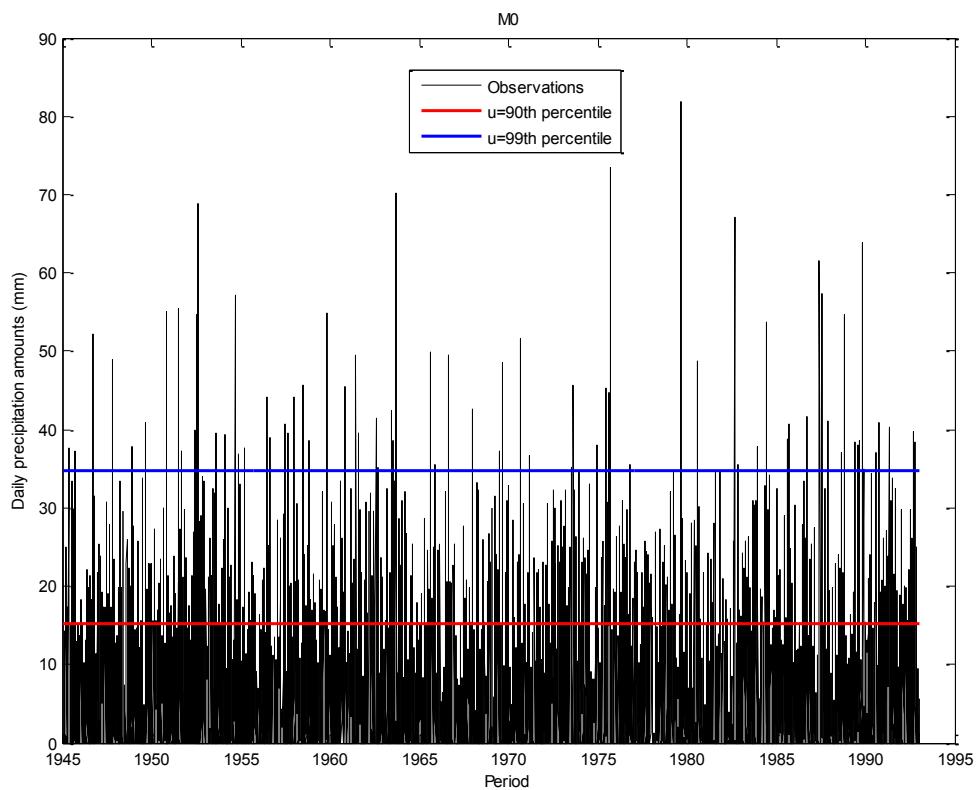


Figure 4. Time series of daily precipitation amounts at station 7025250

The application of models M1, M2 and M3 to strictly positive values of this case study dataset shows discontinuity at the threshold point with model M1, while models M2 and M3 show continuity at the threshold point. Figures 5, 6 and 7 give an overview of these results associated with an initial threshold equal to the 90th percentile. The x values represent the daily precipitation amounts and the estimated threshold is illustrated in red.

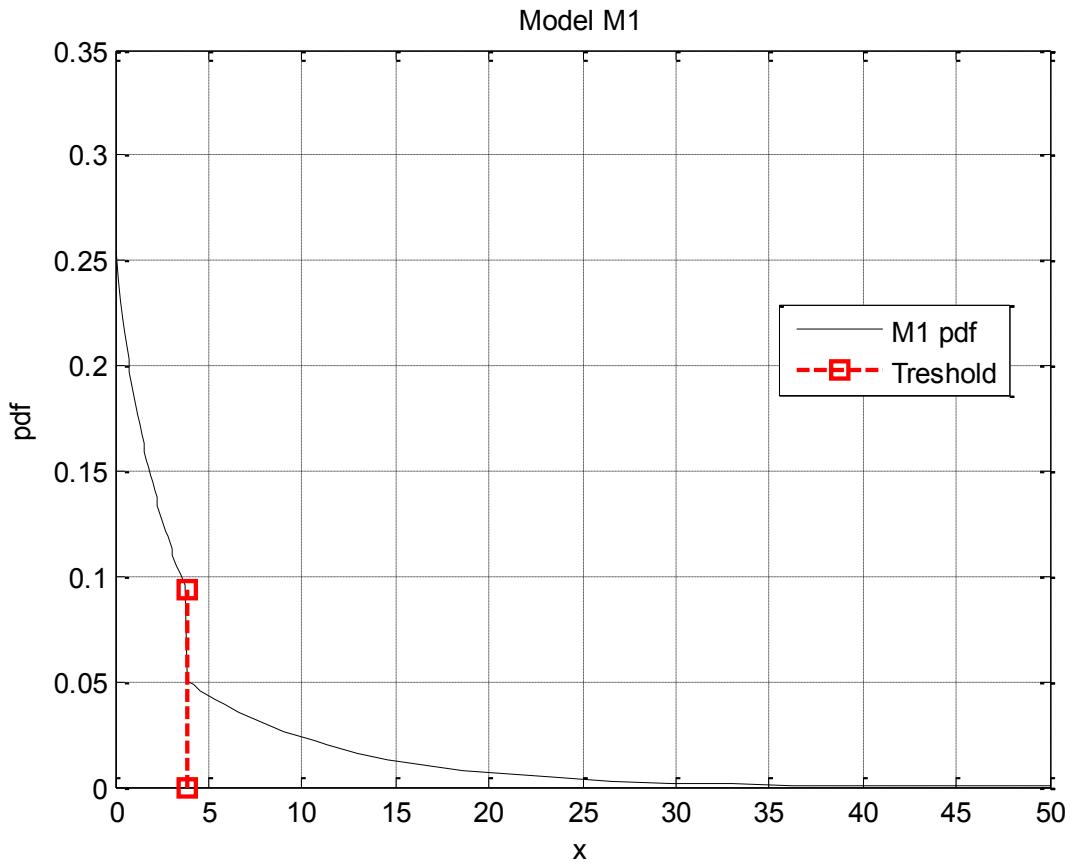


Figure 5. Probability density function of model M1 with observed discontinuity at the threshold point

Without a continuity constraint at the threshold point, the performance of model M1 (Figure 5) can be questionable as reported by Behrens et al. (2004) and discussed by Scarrott and MacDonald (2012). But for model M2 (Figure 6) where the continuity at the threshold point is enforced, it appears that the constraint performed very well in practice.

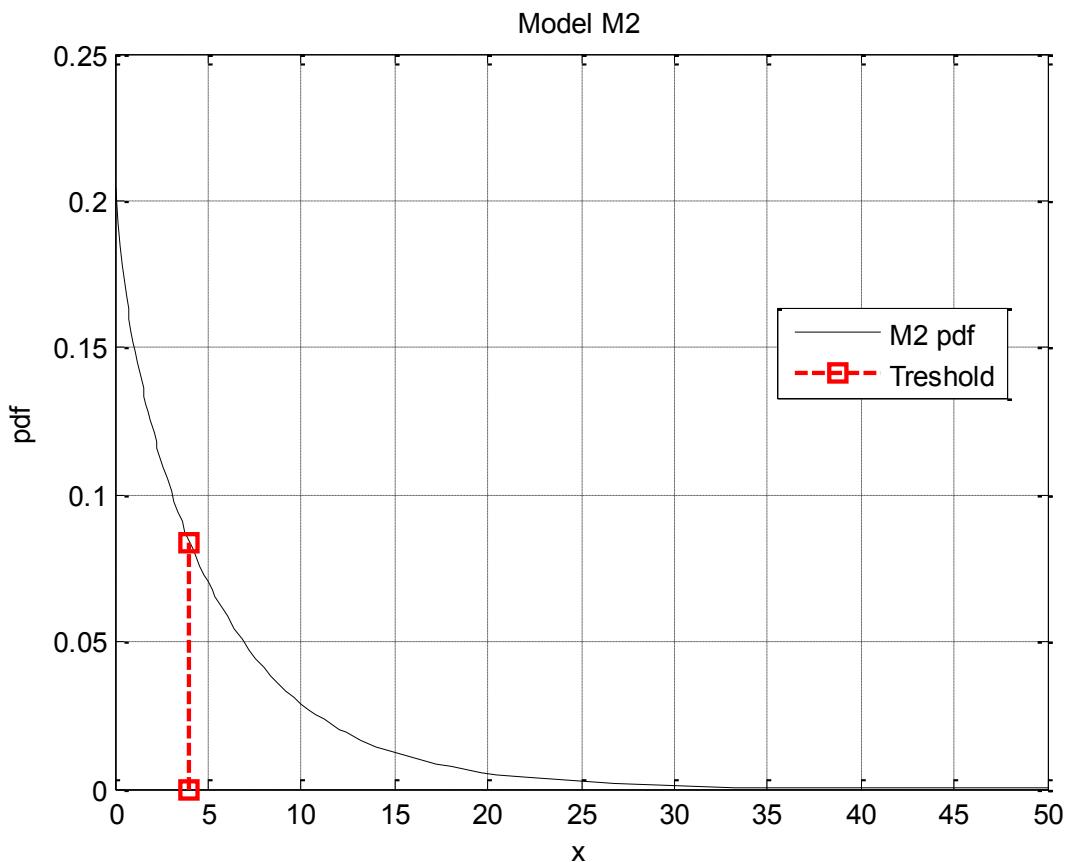


Figure 6. Probability density function of model M2 with correction of the discontinuity at the threshold point

In the case of modeling M3, the continuity is observed at the threshold point (Figure 7). Nevertheless, with such a low threshold, this model is not appropriate for modeling of precipitation extremes because the histogram of studied data shows that the most frequent values ranged between 0 and 10mm.

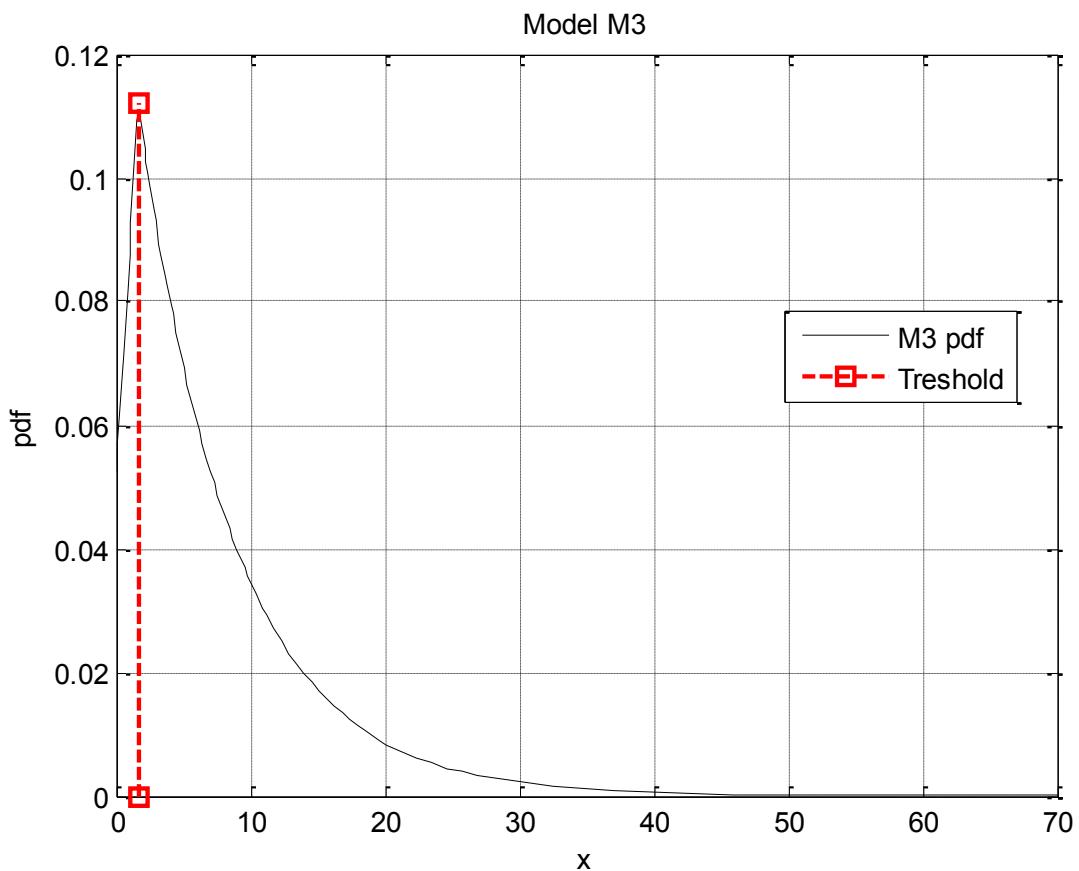


Figure 7. Probability density function of model M3 with correction of the discontinuity at the threshold point

The estimates of quantiles are presented in Table 14 for each of these models, in addition to the Annual Maxima (AM) quantiles resulting from the GEV model. To measure the uncertainty associated with the quantile estimates, the 95% confidence interval (CI) is computed using the adjusted asymptotic approach proposed by Ashkar and El Adlouni (2015). This method improves the assumption of normality of the GPD-based quantiles, often used in estimating confidence intervals.

Table 14. Models M0, M1, M2, M3 and GEV based quantile estimates with confidence intervals shown in brackets, for two initial values u_1 of the threshold

T	M0	M1	M2	M3
$u=u_1=35$ (99 th percentile)				
2 years	43 [41; 44]	36 [12; 61]	36 [34; 39]	43 [42; 45]
10 years	62 [58; 67]	55 [23; 87]	50 [46; 54]	59 [57; 60]
100 years	89 [80; 97]	88 [48; 129]	68 [63; 72]	78 [76; 80]
$u=u_1=15$ (90 th percentile)				
2 years	45 [44; 46]	41 [40; 42]	34 [33; 35]	40 [39; 41]
10 years	64 [62; 65]	56 [55; 58]	48 [47; 49]	54 [52; 55]
100 years	86 [82; 90]	76 [74; 79]	66 [65; 67]	71 [70; 73]
GEV				
2 years		45 [43; 47]		
10 years		63 [56; 70]		
100 years		83 [67; 90]		

The quantiles presented in Table 14 are comparable with sometimes a difference up to 20 mm between models.

5. Conclusion

The peaks-over-threshold (POT) approach is the most useful framework for statistical modeling OF extreme values after the annual maxima method. However, the subjectivity and complexity associated with the choice of the threshold make the POT technique less popular despite its advantages related, for example, to a potential larger sample size than with the annual maxima and to its ability to model both frequency and intensity processes.

With the ongoing development of mixture models in the POT modeling area, the objective estimation of the threshold and the quantification of its associated uncertainty are of particular interest. The focus of this paper was to tackle this deficiency of the standard POT approach by developing mixture models called herein the hybrid GP models. The GP distribution for extreme observations is combined with a Gamma distribution for relatively non-extreme observations so that the threshold is objectively estimated at the boundary of the two distributions. To ensure continuity at the convergence point, a constraint is formulated that lead to model M2, which is compared to M1 and subsequently to M3 and M0 models. All these models were developed and assessed using simulated and real data. To tackle the model pdf continuity issue at the threshold point, only one constraint is considered, the continuity of the resulting hybrid GP pdf by formulating the GPD scale parameter as a function of the threshold and the Gamma distribution shape and scale parameters. It is possible to enforce also the continuity of the derived pdf and thus to derive the GPD shape parameter in the same way. However, this second constraint is not investigated because it does not take into account the above-threshold component of the hybrid model. Additionally, the hybrid GP model could be more robust if the constraints are enforced so that it is the Gamma shape and scale parameters that are expressed as functions of the GPD parameters. But, the assessment of the inversion of the link function is not straightforward analytically. This point can thus be explored in future research. Moreover, the understanding of mixture models is still complex because of the lack of theoretical justification of the asymptotic statistical properties of non-extreme values. Empirical case studies can be helpful to improve the performance of hybrid GP models. The implementation with the parametric bulk distribution is simple. However, semi-parametric and non-parametric distributions have also some attention for developing more flexible models.

References

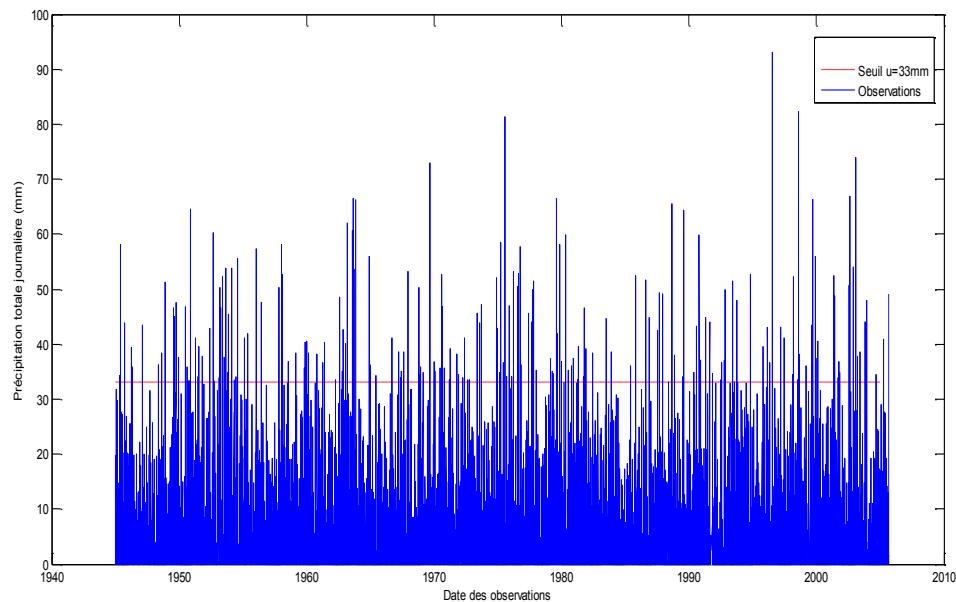
- 1) Ashkar F, El Adlouni S (2015) Adjusting for small-sample non-normality of design event estimators under a generalized Pareto distribution. *J. Hydrol.* 530, 384-391.
- 2) Behrens CN, Lopes Hedibert F, Gamerman D (2004) Bayesian analysis of extreme events with threshold estimation. *Statistical modeling* 4:227-244.
- 3) Carreau J, Bengio Y (2009) A hybrid Pareto model for asymmetric fat-tailed data: the univariate case. *Extremes* 12:53-76.
- 4) Coles SG (2001) An introduction to statistical modeling of extreme values. Springer Series in Statistics, 208p.
- 5) Davison AC, Smith RL (1990) Models for exceedances over high thresholds. *Journal of the Royal Statistical Society. Series B (Methodological)* 52(3): 393-442.
- 6) de Zea Bermudez P, Turkman MAA (2003) Bayesian approach to parameter estimation of the generalized pareto distribution. *Test* 12: 259-277.
- 7) Do Nascimento FF, Gamerman D, Lopez HF (2012) A semi-parametric Bayesian approach to extreme value estimation. *Stat Comp* 22:661-675.
- 8) El Adlouni S, Ezzahid E, Mouatassim Y (2011) Mixed distribution for loss severity modeling with zeros in the operational risk loss. *International Journal of Applied Mathematics & Statistics* 21(J11). ISSN 0973-7545 (Online).
- 9) El Adlouni S, Ouarda TBMJ (2009) Joint Bayesian Model Selection and Parameter estimation of the generalized extreme value model with covariates using birth-death Markov chain Monte Carlo. *Water Resources Research* 45, W06403. doi:10.1029/2007WR006427
- 10)El Adlouni S, Ouarda TBMJ (2008) Comparaison des méthodes d'estimation des paramètres du modèle GEV non stationnaire. *Revue des Sciences de l'Eau* 21(1):35-50.
- 11)Fisher RA, Tippett LHC (1928) Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Math Proc Cambridge Philos Soc* 24(02):180-190.

- 12)Frigessi A, Haug O, Rue H (2002) A dynamic mixture model for unsupervised tail estimation without threshold selection. *Extreme* 5(3):219-235.
- 13)Hastings W (1970) Monte Carlo sampling method using Markov Chains and their applications. *Biometrika* 55:97-109.
- 14)Hosking, J.R.M., Wallis, T.R., 1987. Parameter and quantile estimation for the Generalized Pareto distribution. *Technometrics* 29(3) :339-349.
- 15)Jenkinson AF (1955) The frequency distribution of the annual maximum (or minimum) of meteorological elements. *Quart J Roy Meteor Soc* 81:158-171.
- 16)Katz RW, Parlange MB, Naveau P (2002) Statistics of extremes in hydrology. *Advances in Water Resources* 25:1287-1304.
- 17)Lang, M., Ouarda, T.B.M.J., Bobée, B., 1999. Towards operational guidelines for over-threshold modeling. *J. Hydrol.* 225, 103-117.
- 18)Liseo B (1993) Elimination of nuisance parameters with reference non-informative priors. *Biometrika* 80: 295-304.
- 19)MacDonald A, Scarrott CJ, Lee D, Darlow B, Reale M, Russell G (2011) A flexible extreme value mixture model. *Computational Statistics and Data Analysis* 55:2137-2157.
- 20>Martins ES, Stedinger JR (2001) Generalized maximum likelihood Pareto-Poisson estimators for partial duration series. *Water Resources Research* 37(10): 2551-2557.
- 21>Martins ES, Stedinger JR (2000) Generalized maximum likelihood generalized extreme value quantile estimators for hydrological data. *Water Resources Research* 36(3):737-744.
- 22>Mekis E, Vincent L (2011) An overview of the second generation adjusted daily precipitation dataset for trend analysis in Canada. *Atmosphere-Ocean* 49(2):163-177.
- 23>Metropolis N, Rosenbluth A, Teller A, Teller E (1953) Equations of state calculation by fast computing machines. *Journal of Chemical Physics* 21:1087-1092.

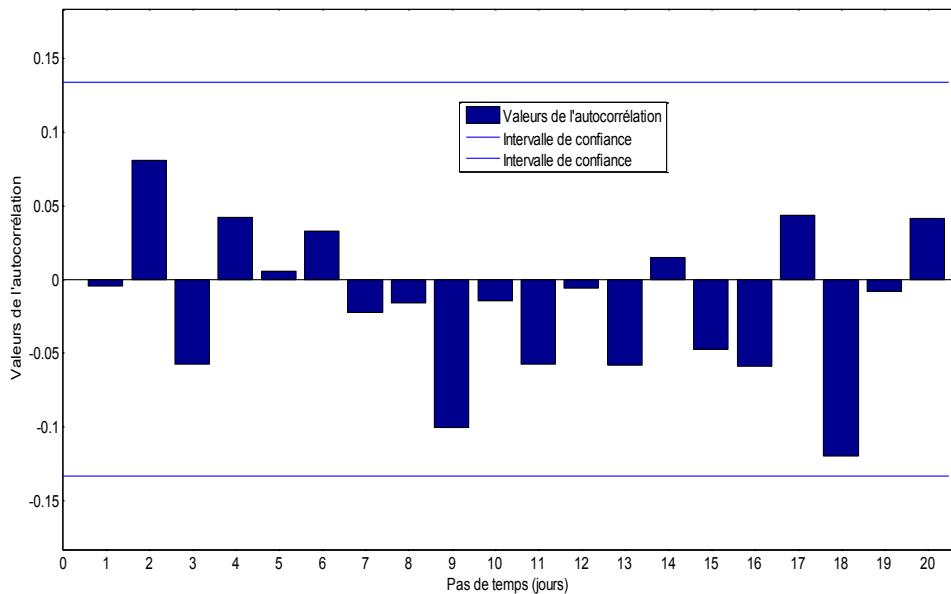
- 24)Osman YZ, Rowan F, Sweeney JC (2013) Downscaling extreme precipitation in Ireland using combined peak-over-threshold generalised Pareto distribution model of varying parameters. *Journal of Water and Climate Change* 4(4):348-363.
- 25)Ouarda TBMJ, El Adlouni S (2011) Bayesian nonstationary frequency analysis of hydrological variables. *Journal of the American Water Resources Association* 47 (3):496-505.
- 26)Pickands J (1975) Statistical inference using extreme order statistics. *Ann Stat* 3(1): 119-131.
- 27)Robert CP (2006) *Le choix bayesien: Principes et pratique*. Collection statistique et probabilités appliquées. Springer, 654p.
- 28)Scarrott C, MacDonald A (2012) A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT-Statistical Journal* 10(1):33-60.
- 29)Tancredi A, Anderson C, O'Hagan A (2006) Accounting for threshold uncertainty in extreme value estimation. *Extremes* 9:87-106.
- 30)Thiombiano AN, El Adlouni S, St-Hilaire A, Ouarda TBMJ, El-Jabi N (2016) Nonstationary frequency analysis of extreme daily precipitation amounts in Southeastern Canada using a peaks-over-threshold approach. *Theor Appl Climatol* 124(1-2). doi: 10.1007/s00704-016-1789-7

ANNEXES

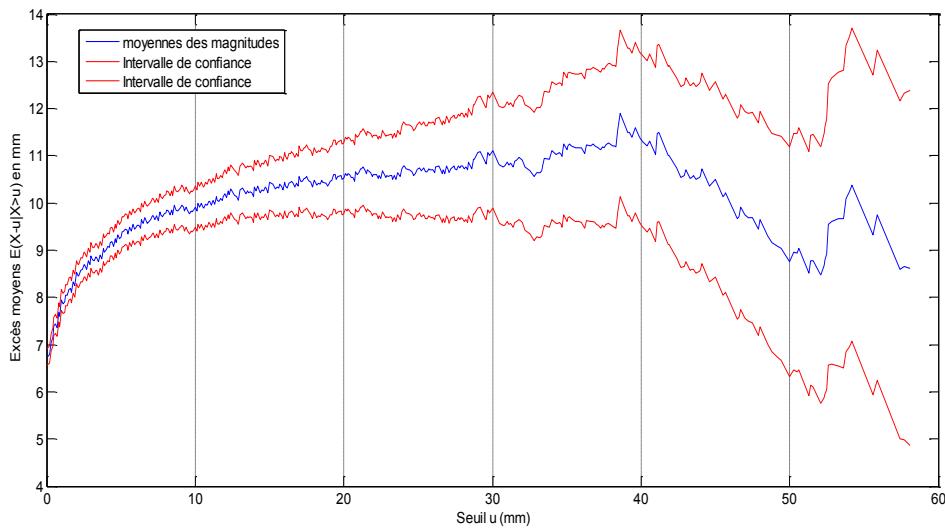
Annexe 1: Distribution de la précipitation totale journalière à l'aéroport de Miramichi (Station 8101000) dans la province du NB, sur la période 1945-2005



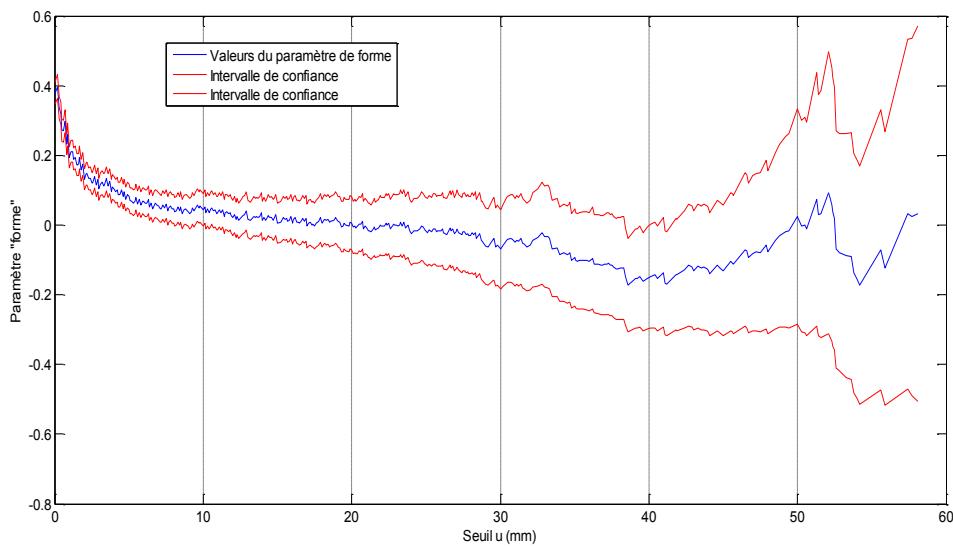
Annexe 2: Fonction de l'autocorrélation entre les dépassages de seuil à la station 8101000



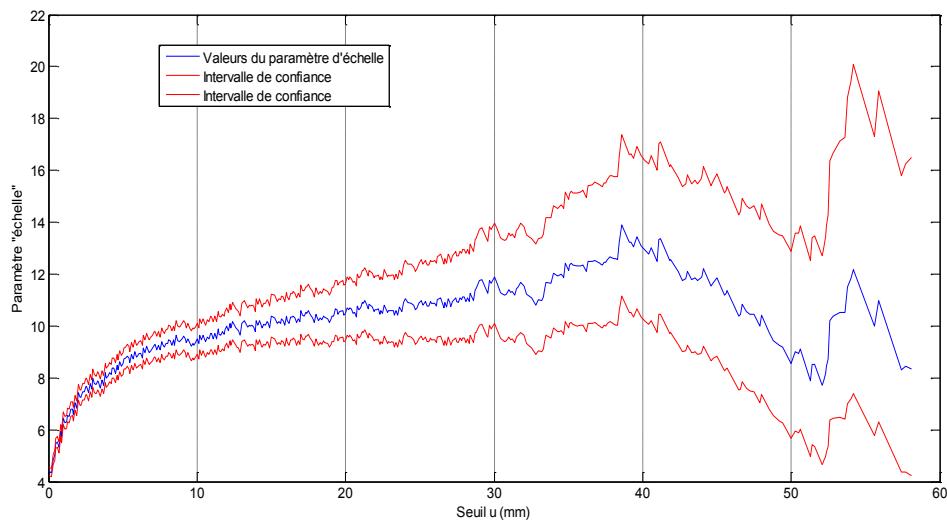
Annexe 3: Évolution de la magnitude moyenne des dépassemens de seuil à la station 8101000



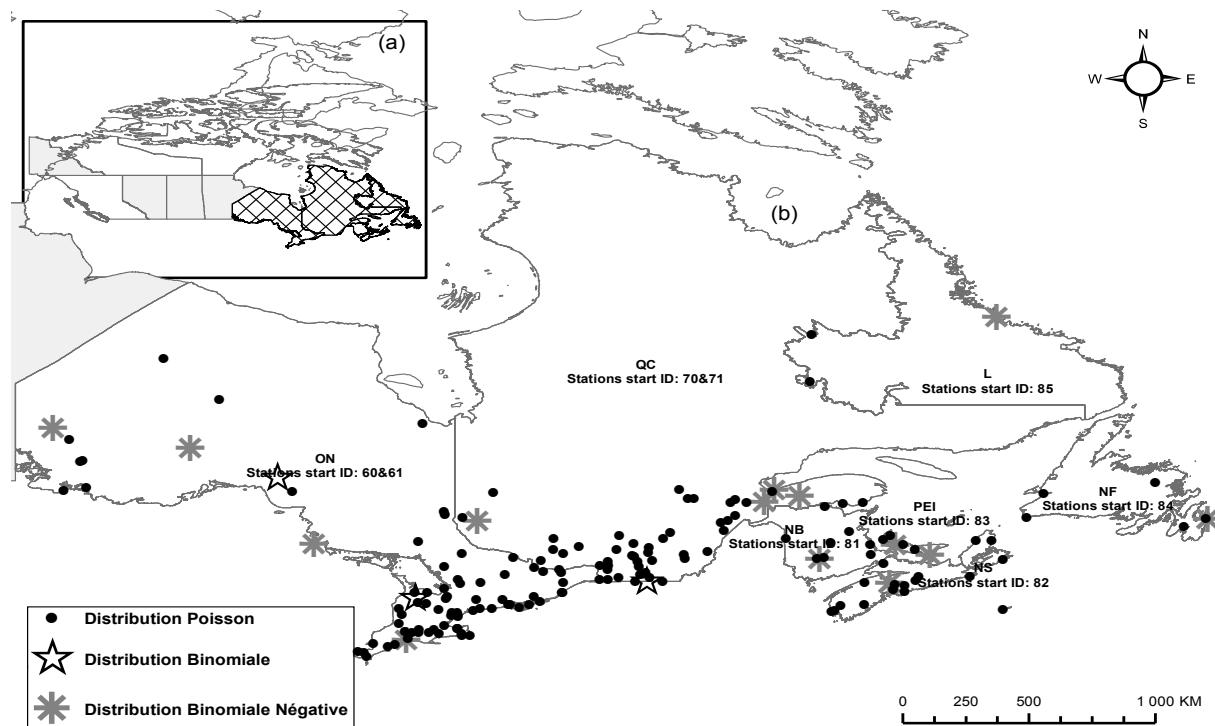
Annexe 4: Évolution du paramètre de forme de la GPD avec les dépassemens de seuil de la station 8101000



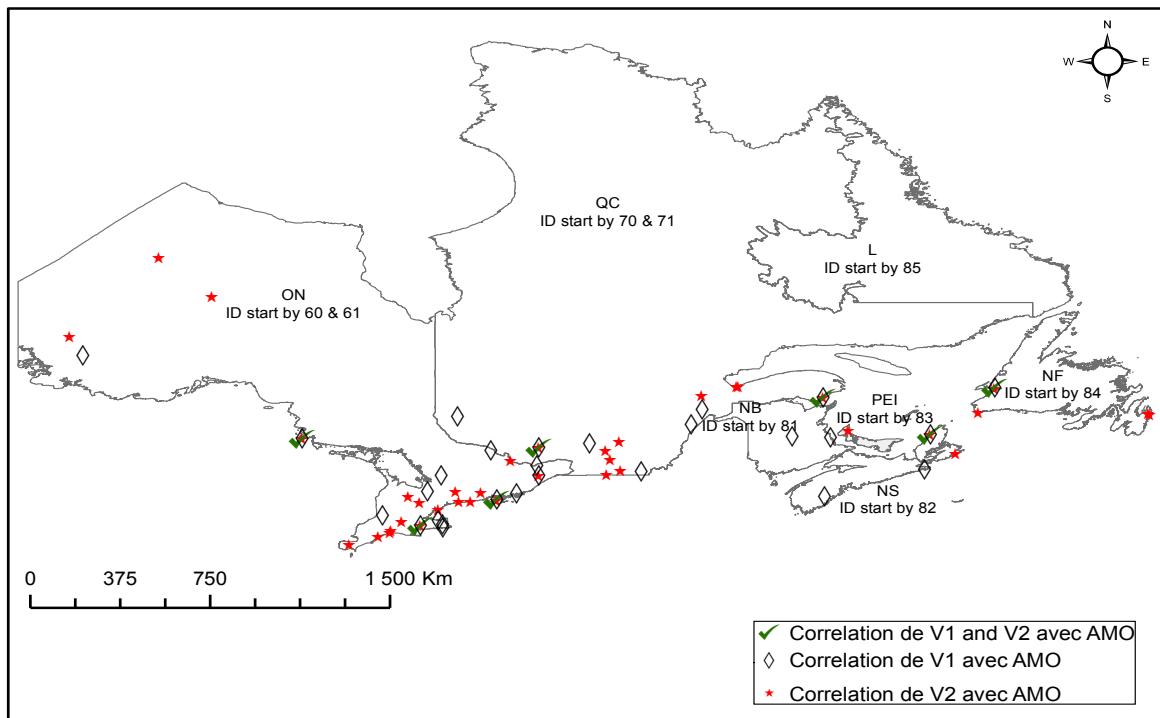
Annexe 5: Évolution du paramètre d'échelle de la GPD avec les dépassesments de seuil de la station 8101000



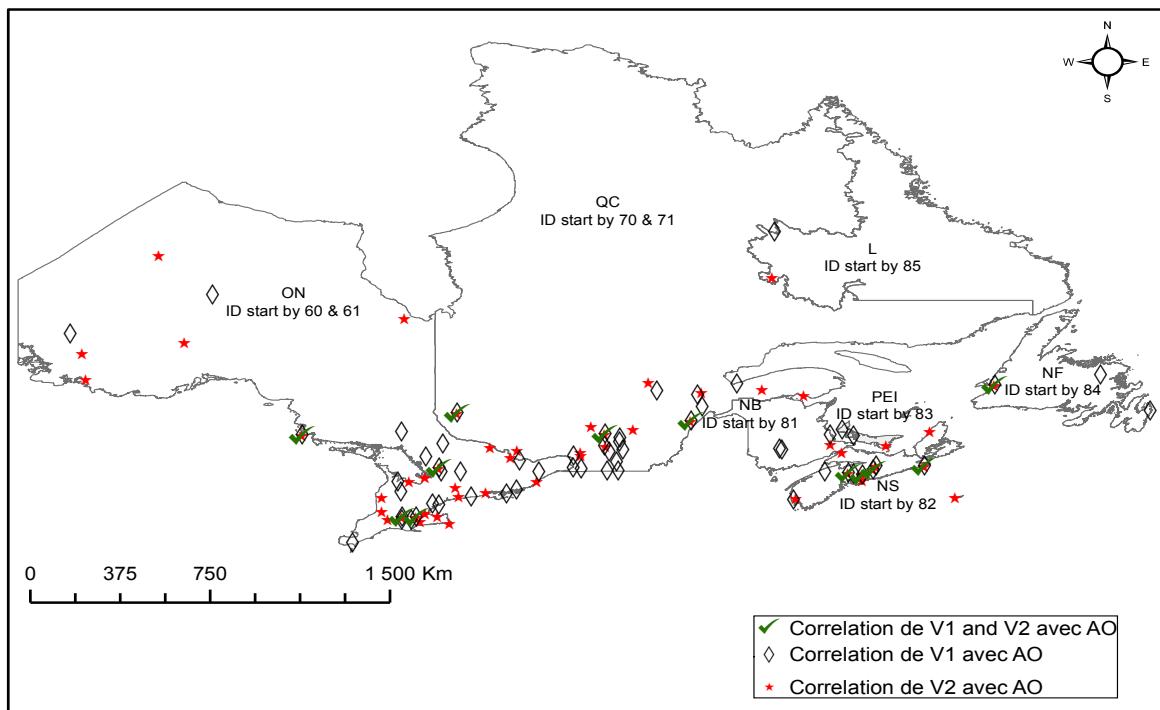
Annexe 6: Validité du processus de Poisson pour la modélisation de la fréquence annuelle des dépassesments de seuil indépendants à l'échelle du Sud-Est (b) du Canada (a)



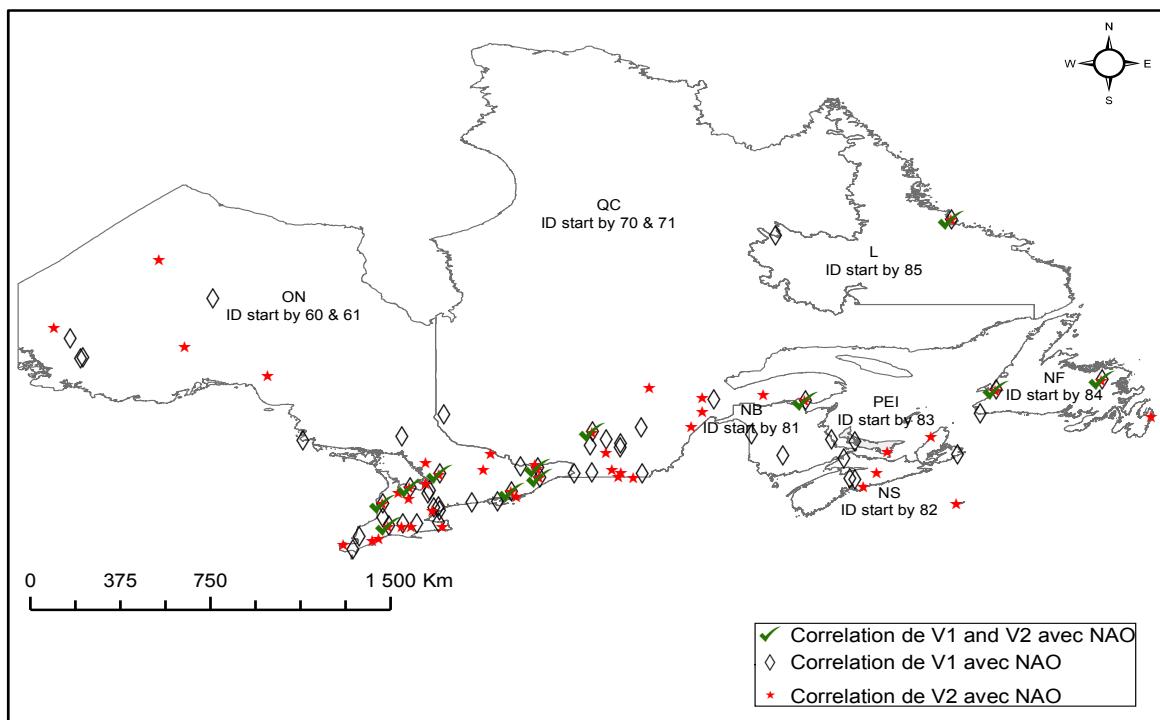
Annexe 7: Stations où V1 et V2 ont montré une corrélation significative avec l'indice AMO



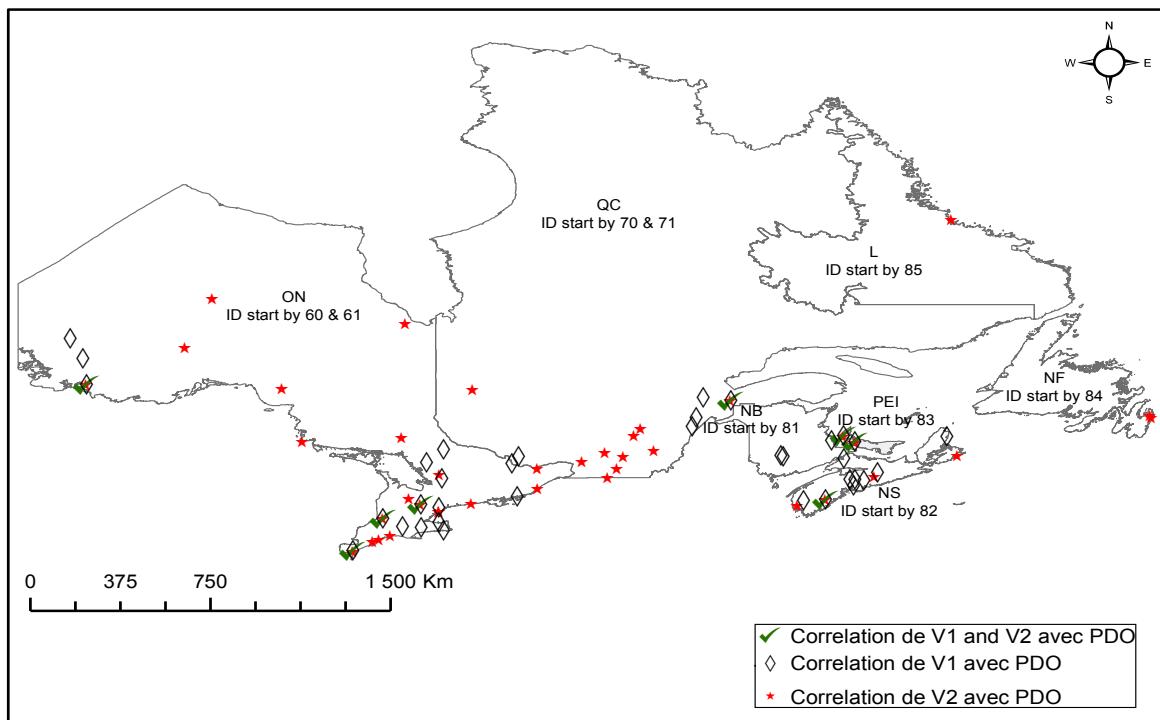
Annexe 8: Stations où V1 et V2 ont montré une corrélation significative avec l'indice AO



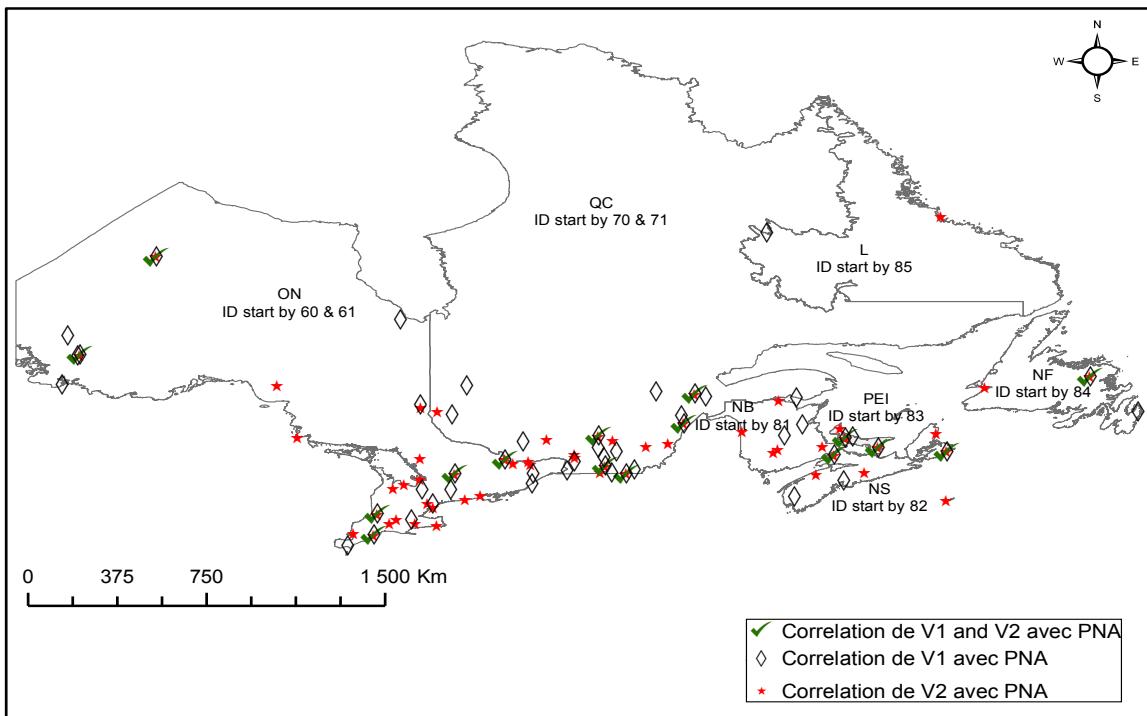
Annexe 9: Stations où V1 et V2 ont montré une corrélation significative avec l'indice NAO



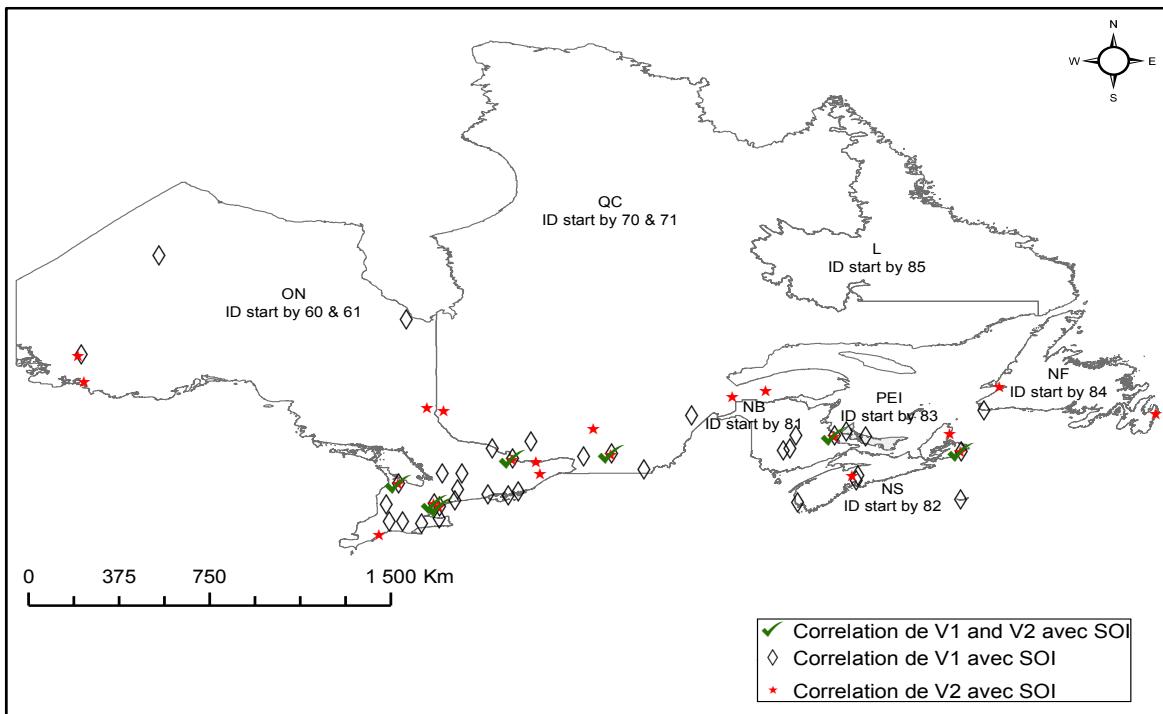
Annexe 10: Stations où V1 et V2 ont montré une corrélation significative avec l'indice PDO



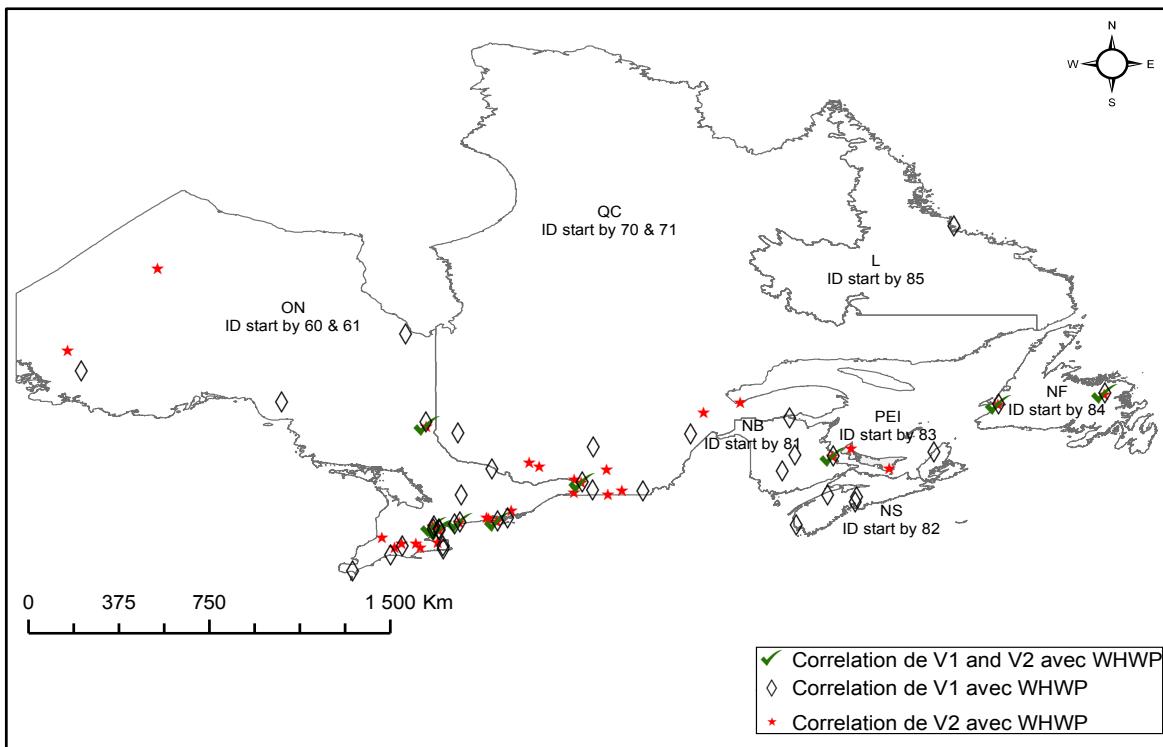
Annexe 11: Stations où V1 et V2 ont montré une corrélation significative avec l'indice PNA



Annexe 12: Stations où V1 et V2 ont montré une corrélation significative avec l'indice SOI



Annexe 13: Stations où V1 et V2 ont montré une corrélation significative avec l'indice WHWP



Annexe 14: Description des 96 stations des groupes AO (en gras) et PNA

Provinces	ID de la station	Nom de la station	Indices climatiques	Période d'enregistrement	Période d'analyse	Tau de Kendall (%)	p-value
PEI	8300100	ALLISTON	PNA	1945-2006	1951-2005	22,56	0,0409
	8300500	LONG RIVER	PNA	1957-2003	1957-2003	27,5	0,008
	8300525	O'LEARY	AO	1957-2004	1957-2004	29,61	0,0028
	8300700	SUMMERSIDE A	PNA	1945-2002	1951-1998	26,41	0,0162
NB	8100590	BUCTOUCHÉ	AO	1965-1999	1966-1999	-26,35	0,0325
	8101000	MIRAMICHI A	PNA	1945-2005	1951-2005	18,81	0,0456
	8101200	DOAKTOWN	PNA	1945-2006	1951-2005	20,53	0,0325
	8101500	FREDERICTON A	AO	1951-2012	1951-2011	19,74	0,0307
	8101600	FREDERICTON CDA	AO	1945-2000	1951-2000	26,72	0,0064
NFL	8401700	GANDER INT'L A	AO	1945-2012	1951-2011	-26,04	0,0037
	8402925	PETTY HARBOUR	AO	1955-1999	1955-1999	25,13	0,0224
	8403600	ST JOHN'S WEST CDA	PNA	1950-1996	1951-1995	-26,22	0,0119
	8403800	STEPHENVILLE A	AO	1945-2013	1951-2012	-26,92	0,0024

NS	8201410	DEMING	AO	1956-2006	1956-2005	19,85	0,0429
	8202000	GREENWOOD A	AO	1945-2013	1952-2012	20,27	0,0251
	8202250	HALIFAX STANFIELD INT'L A	AO	1953-2012	1953-2012	19,52	0,0381
	8202700	KEMPTVILLE	PNA	1950-1991	1951-1991	-33,6	0,0023
	8203161	LOUISBOURG	PNA	1972-2006	1972-2005	35,23	0,0036
	8203535	MIDDLE MUSQUODOBOIT	AO	1961-2006	1961-2005	24,64	0,0189
	8203600	MOUNT UNIACKE	AO	1945-2003	1951-1999	33,87	0,0009
	8203700	NAPPAN CDA	PNA	1945-2005	1951-2004	19,86	0,0407
	8204800	ST MARGARET'S BAY	PNA	1945-2006	1951-2006	-28,61	0,0029
	8205000	SALMON HOLE	AO	1945-2005	1951-2005	28,06	0,0042
	8206200	UPPER STEWIACKE	AO	1945-2005	1951-2005	24,77	0,0091
	8206500	YARMOUTH A	AO	1945-2012	1951-2012	17,68	0,0431

ON	61519JM	CRESSY	AO	1966-2002	1966-2001	-30,98	0,0082
	6010738	BIG TROUT LAKE	PNA	1945-1992	1951-1992	-26	0,0173
	6012198	EAR FALLS	PNA	1950-1999	1951-1996	-22,33	0,0314
	6022475	FORT FRANCES	PNA	1945-1995	1951-1995	-25,33	0,0187
	6032117	DRYDEN	PNA	1945-1997	1951-1997	30,43	0,0026
	6032119	DRYDEN A	PNA	1970-2005	1970-2004	-36,74	0,0023
	6057592	SAULT STE MARIE A	AO	1945-2012	1961-2011	-25,53	0,0085
	6068150	SUDBURY A	AO	1954-2013	1954-2012	-19,85	0,0283
	6072325	ENGLEHART	PNA	1948-2000	1951-2000	23,43	0,0194
	6075425	MOOSONEE UA	PNA	1945-2006	1951-2006	-20,48	0,0311
	6100345	ARNPRIOR GRANDON	AO	1959-1999	1959-1999	30,38	0,0086
	6100971	BROCKVILLE PCC	PNA	1965-2006	1966-2006	22,96	0,0381
	6101494	CHENAUX	PNA	1950-1990	1951-1989	-29,49	0,0096
	6104025	KEMPTVILLE	PNA	1945-1997	1951-1997	22,92	0,0254
	6104175	KINGSTON PUMPING STATION	AO	1960-2006	1961-2006	-21,53	0,0381
	6110605	BEATRICE	AO	1945-1979	1951-1979	-44,83	0,0004
	6112133	DUNCHURCH	AO	1973-2006	1973-2006	-29,97	0,0133
	6112171	DURHAM	AO	1947-2003	1951-2003	37,31	0,0001

ON	6112340	ESSA ONT HYDRO	PNA	1958-2001	1958-2000	25,85	0,0171
	6115525	MUSKOKA A	AO	1945-2006	1951-2005	24,05	0,0105
	6116132	OWEN SOUND MOE	AO	1964-2006	1965-2006	24,04	0,0277
	6116702	POWASSAN	AO	1974-2006	1974-2006	25,97	0,0351
	6122370	EXETER	PNA	1961-2006	1961-2006	22,94	0,027
	6134390	LEAMINGTON	PNA	1945-1978	1951-1978	34,47	0,0114
	6135583	NEW GLASGOW	PNA	1957-2006	1957-2006	-24,35	0,0179
	6136606	PORT COLBORNE	PNA	1964-2006	1964-2006	26,56	0,0164
	6140954	BRANTFORD MOE	PNA	1960-2006	1961-2006	25,4	0,0156
	6142420	FOLDENS	AO	1963-2006	1964-2006	-30,97	0,004
	6147664	SCOTLAND	AO	1971-2006	1971-2006	25,71	0,0276
	6149625	WOODSTOCK	AO	1945-2006	1951-2006	20,45	0,0331
	6151689	COBOURG STP	AO	1970-2006	1970-2006	-24,44	0,0365
	6158255	THORNHILL GRANDVIEW	PNA	1965-2006	1965-2006	31,17	0,0048
	6158665	TORONTO ISLAND A	AO	1957-2006	1957-2005	-27,45	0,009
	6158733	TORONTO LESTER B. PEARSON INT'L A	AO	1945-2013	1951-2012	-19,32	0,0314
	6163156	HALIBURTON A	AO	1945-1992	1951-1992	-31,6	0,0048
	6164433	LINDSAY FROST	PNA	1974-2006	1974-2006	25,38	0,0386

QC	7010720	BERTHIERVILLE	PNA	1945-1995	1951-1994	23,47	0,0294
	7015730	OKA	PNA	1945-2013	1952-2012	19,77	0,0275
	7016816	ST ALEXIS DES MONTS	PNA	1963-2013	1963-2012	-26,88	0,0061
	7018564	TROIS RIVIERES	AO	1945-1986	1951-1986	-31,61	0,007
	7020040	ABERCORN	PNA	1950-1985	1951-1984	29,06	0,0155
	7021840	COATICOOK	PNA	1949-2013	1951-2012	18,25	0,04
	7022160	DRUMMONDVILLE	PNA	1945-2013	1951-2012	21	0,0203
	7022320	FARNHAM	PNA	1952-2013	1952-2012	19,12	0,033
	7022375	FLEURY	AO	1967-2013	1967-2012	-23,31	0,0231
	7022800	GRANBY	AO	1948-2013	1951-2012	-17,93	0,0419
	7025440	NICOLET	AO	1945-2013	1951-2012	-28,95	0,0012
	7025745	ORMSTOWN	AO	1962-2013	1963-2012	-28,94	0,0035
	7026040	PHILIPSBURG	AO	1950-2013	1951-2012	18,31	0,0394
	7026836	ST ANICET	AO	1960-2013	1961-2012	-30,78	0,0017
	7027360	ST HYACINTHE	PNA	1945-1979	1951-1979	-42,53	0,0025
	7028280	STANSTEAD	PNA	1948-1991	1951-1990	-22,84	0,0421
	7033650	LACHUTE	AO	1963-2013	1963-2012	21,47	0,0322
	7034480	MANIWAKI UA	PNA	1953-2003	1953-1993	-26,62	0,0148

QC	7042840	GRANDES BERGERONNES	PNA	1951-2013	1951-2013	18,03	0,0425
	7048320	TADOUSSAC	AO	1945-2004	1951-2001	-20,07	0,0472
	7050455	BARRAGE LAC MORIN	AO	1945-1984	1951-1984	-26,03	0,0379
	7051120	CAPLAN	PNA	1947-2004	1951-2001	22,42	0,0271
	7054095	LA POCATIERE CDA	PNA	1945-1996	1951-1996	20,84	0,0474
	7055120	MONT-JOLI A	AO	1945-2013	1951-2012	24,47	0,0059
	7057650	STE PERPETUE	PNA	1963-1997	1963-1997	40	0,0011
	7058560	TROIS PISTOLES	PNA	1951-2004	1951-2001	-26,38	0,0095
	7060400	BAGOTVILLE A	PNA	1945-2013	1951-2012	-19,84	0,0256
	7066080	PORTAGE DES ROCHES	AO	1945-1984	1951-1983	-35,91	0,0042
	7080600	BELLETERRE	PNA	1951-2004	1951-2001	25,27	0,0138
	7090120	AMOS	PNA	1945-2000	1951-1999	23,61	0,0198
	7117825	SCHEFFERVILLE A	AO	1948-2010	1951-2009	30,83	0,0065

Annexe 15: Influence de l'indice AO sur V1 et V2

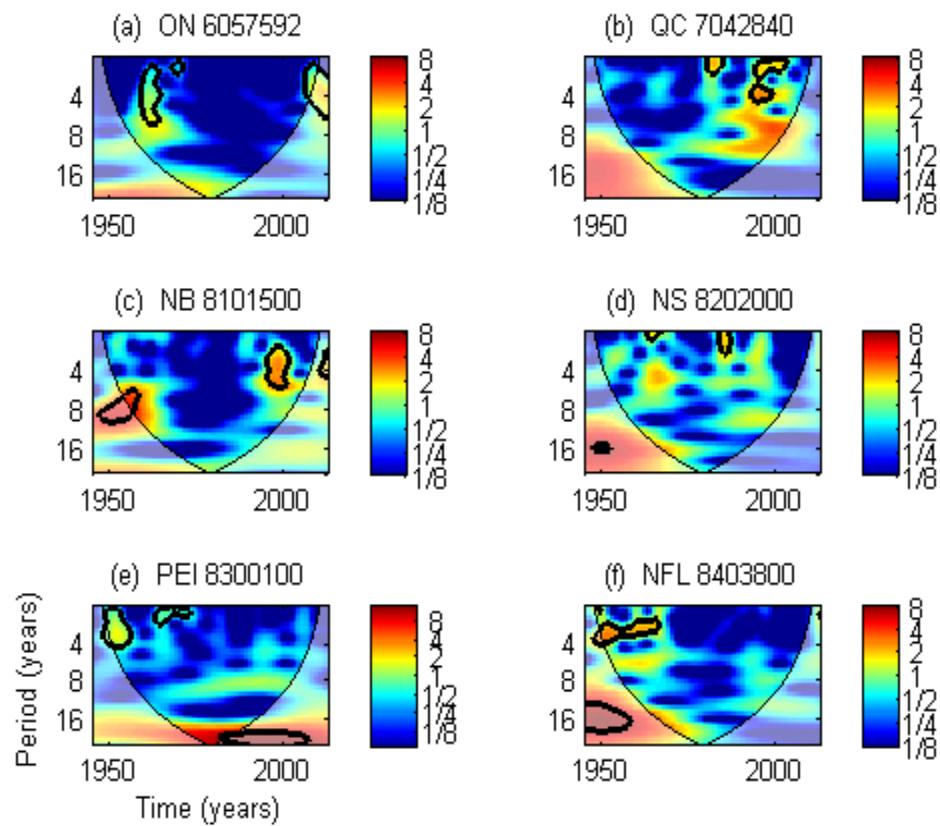
n°	ID	Nom	Prov	Période	u	S	N	λ	AO1	tau1	pv1	AO2	tau2	pv2
1	8403800	STEPHENVILLE A	NF	1951-2012	31	219	60	3,7	MAM	-26,92	0,0024	DJF	35,03	0,0002
2	8201410	DEMING	NS	1956-2005	41	174	50	3,5	MJJ	19,85	0,0429	JAS	21,35	0,0421
3	8202250	HALIFAX STANFIELD INT'L A	NS	1953-2012	46	192	54	3,6	SON	19,52	0,0381	SON	21,13	0,0262
4	8203535	MIDDLE MUSQUODOBOIT	NS	1961-2005	40	151	44	3,4	OND	24,64	0,0189	AMJ	-28,38	0,01
5	8206200	UPPER STEWIACKE	NS	1951-2005	37	190	53	3,6	SON	24,77	0,0091	OND	22,98	0,0202
6	6110605	BEATRICE	ON	1951-1979	27	103	29	3,6	SON	-44,83	0,0005	JAS	-35,76	0,0106
7	6140954	BRANTFORD MOE	ON	1961-2006	26	162	44	3,7	JJA	21,16	0,0441	AMJ	22,82	0,0364
8	6149625	WOODSTOCK	ON	1951-2006	29	189	52	3,6	AMJ	20,45	0,0331	FMA	23,19	0,0182
9	7016816	ST ALEXIS DES MONTS	QC	1963-2012	31	169	49	3,6	NDJ	25,70	0,0095	JJA	34,67	0,001
10	7057650	STE PERPETUE	QC	1963-1997	32	118	33	3,6	MAM	-37,50	0,0019	JJA	34,63	0,0062

u=seuil (mm); S=nombre total des dépassements de seuil sur la période d'analyse; N=nombre d'années; λ=taux de dépassement moyen annuel; tau1=tau de Kendall (%) pour V1 avec sa p-value pv1, tau2=tau de Kendall Tau (%) pour V2 avec sa p-value pv2, AO1 (AO2)= série de AO corrélée à V1 (V2); MAM=moyenne mobile annuelle de Mars à Mai; JD=moyenne annuelle de Janvier à Décembre.

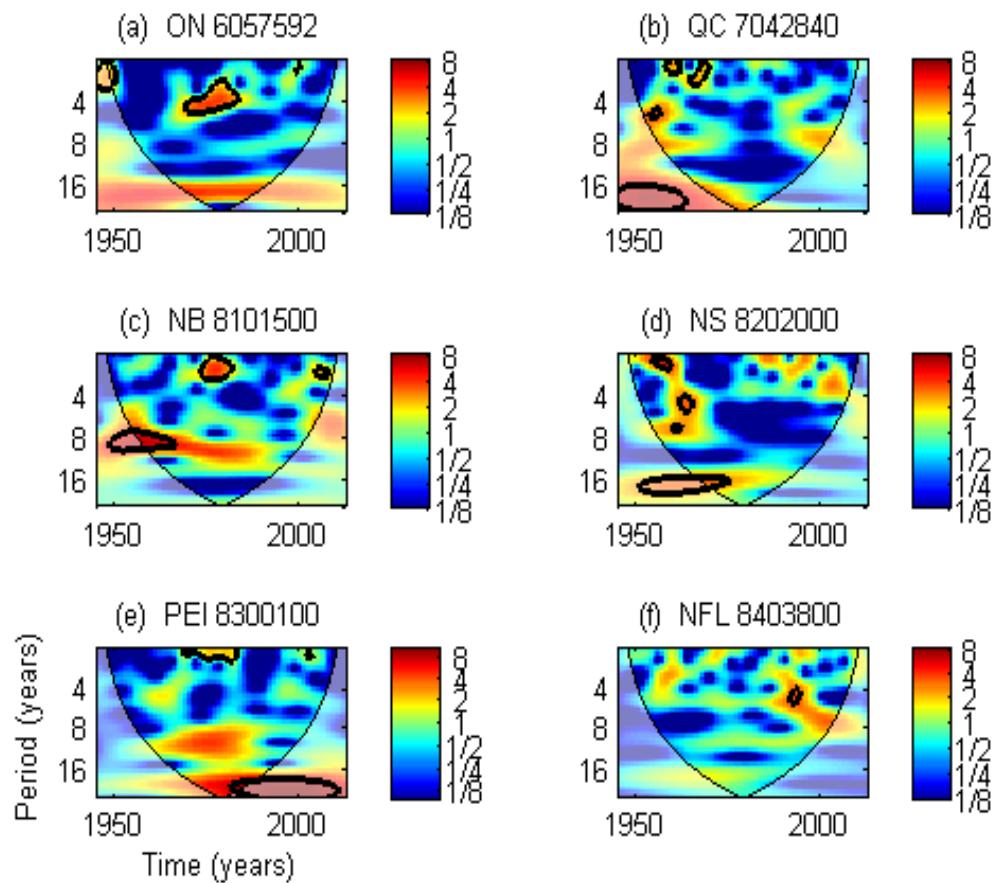
Annexe 16: Influence de l'indice PNA sur V1 et V2

n°	ID	Nom	Prov	Période	u	S	N	λ	PNA1	tau1	pv1	PNA2	tau2	pv2
1	8401700	GANDER INT'L A	NF	1951-2011	30	211	59	3,6	MAM	18,61	0,0382	JFM	21,26	0,0238
2	8203161	LOUISBOURG	NS	1972-2005	43	120	33	3,6	JFM	35,23	0,0036	MAM	30,67	0,0168
3	8203700	NAPPAN CDA	NS	1951-2004	32	203	51	4	ASO	19,86	0,0407	DJF	27,87	0,0278
4	6010738	BIG TROUT LAKE	ON	1951-1992	21	142	40	3,5	SON	-25,99	0,0173	ASO	-29,96	0,0087
5	6032119	DRYDEN A	ON	1970-2004	26	119	33	3,6	AMJ	-36,74	0,0023	OND	-26,87	0,0341
6	6101494	CHENAUX	ON	1952-1989	25	141	38	3,7	JFM	-29,49	0,0096	JJA	-40,96	0,0006
7	6122370	EXETER	ON	1961-2006	29	155	44	3,4	MAM	22,94	0,0270	JD	-29,43	0,0073
8	6135583	NEW GLASGOW	ON	1957-2006	29	173	46	3,8	NDJ	-24,35	0,0179	NDJ	22,64	0,0293
9	6136606	PORT COLBORNE	ON	1964-2006	29	148	41	3,7	JD	26,56	0,0164	NDJ	24,73	0,0291
10	6163156	HALIBURTON A	ON	1951-1992	28	150	39	3,9	OND	29,86	0,0078	SON	30,53	0,0074
11	7016816	ST ALEXIS DES MONTES	QC	1963-2012	31	170	50	3,4	JJA	-26,88	0,0061	MAM	22,06	0,0342
12	7022320	FARNHAM	QC	1952-2012	30	213	59	3,6	MJJ	19,12	0,0330	JAS	20,41	0,0288
13	7028280	STANSTEAD	QC	1951-1990	31	137	39	3,5	MJJ	-22,84	0,0421	MAM	25,64	0,0306
14	7042840	GRANDES BERGERONNES	QC	1951-2012	32	207	60	3,5	DJF	18,03	0,0425	JD	26,1	0,0045
15	7057650	STE PERPETUE	QC	1963-1997	32	118	33	3,6	MAM	40,00	0,0011	MAM	25,72	0,0425

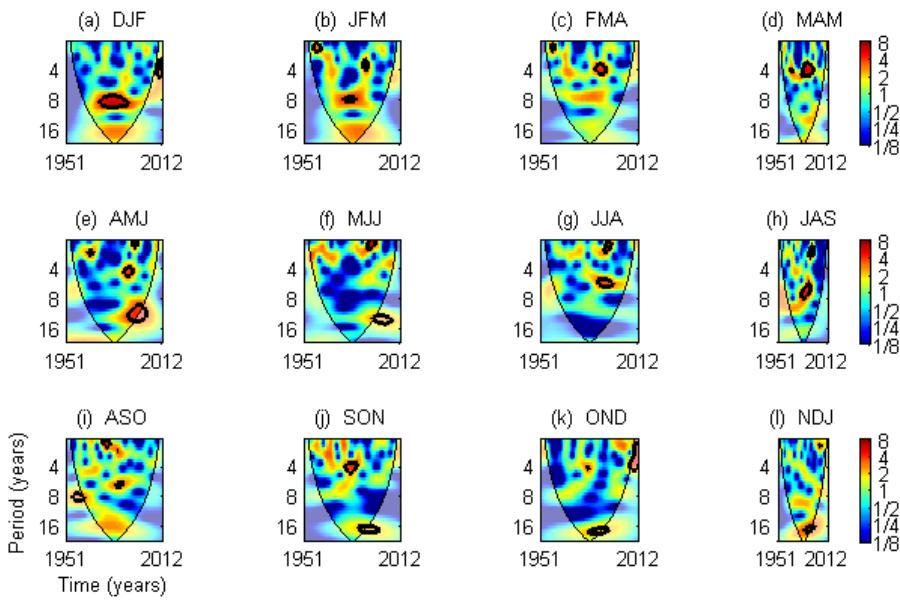
Annexe 17: Transformée en ondelette continue pour des séries de V1



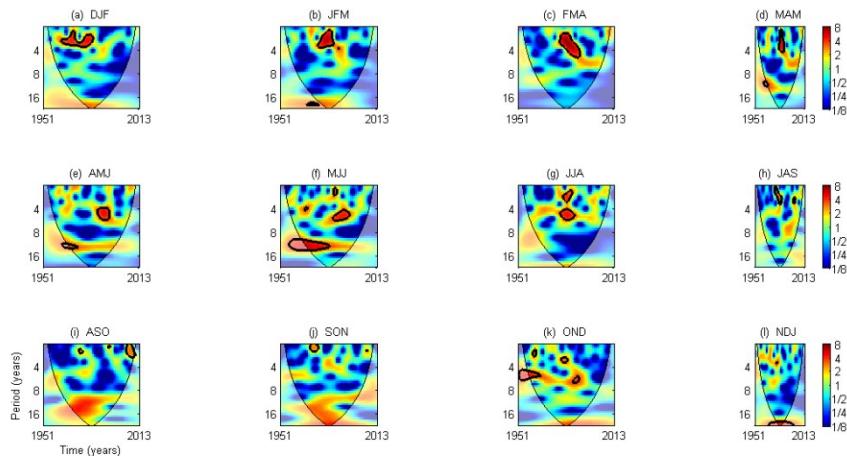
Annexe 18: Transformée en ondelette continue pour des séries de V2



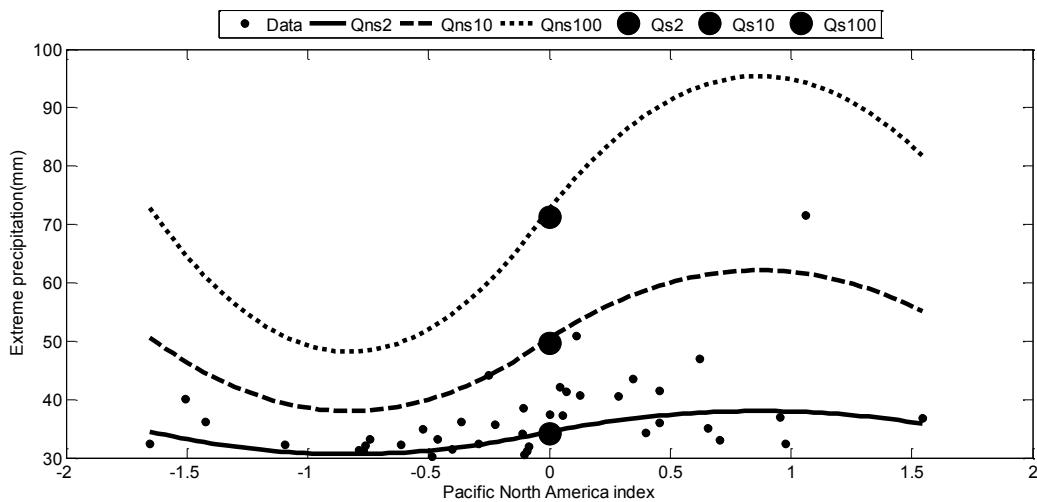
Annexe 19: Transformée en ondelette continue des 12 séries annuelles (a-l) constituées pour l'indice AO



Annexe 20: Transformée en ondelette continue des 12 séries annuelles (a-l) constituées pour l'indice PNA

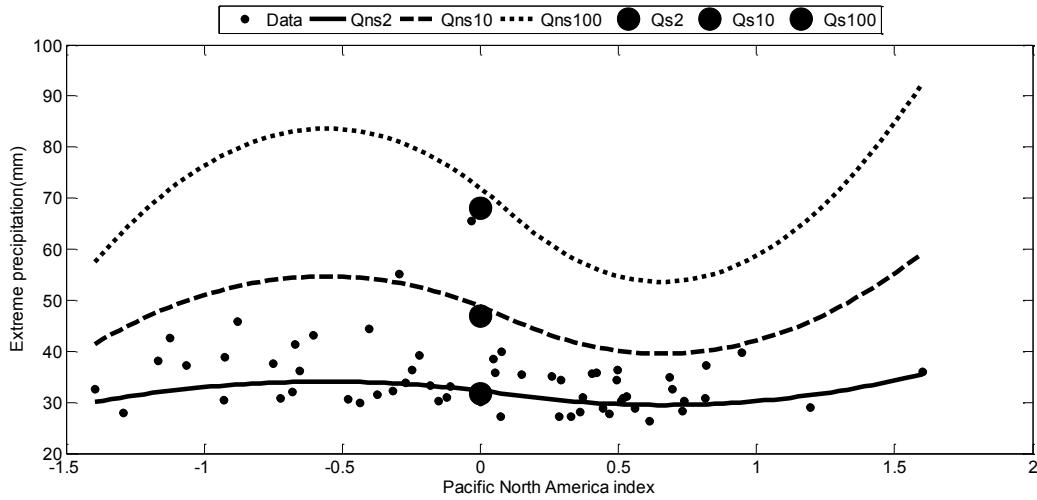


Annexe 21: Modèle GPD avec fonction B-spline à deux nœuds et de degré 2 : station 6158255

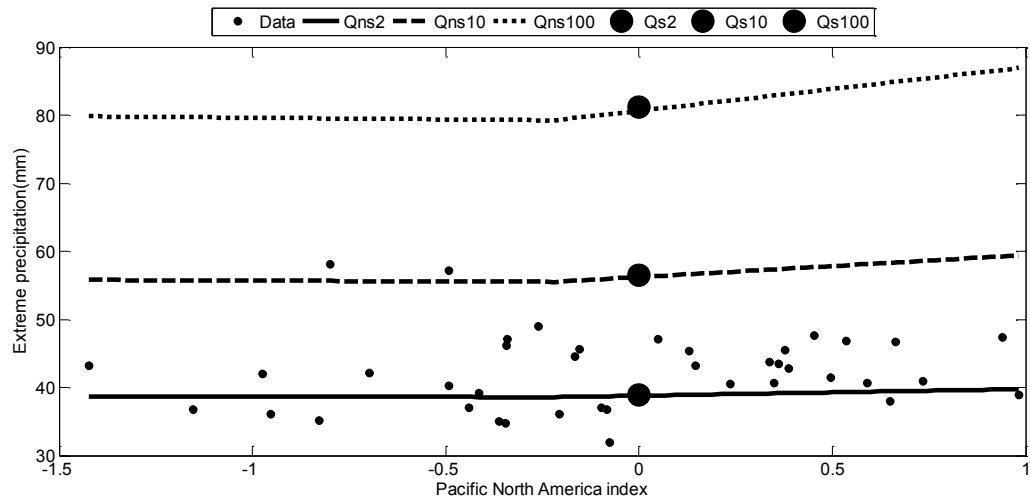


Qns2, 10 et 100 représente respectivement les quantiles non-stationnaires associés à $p=0.5; 0.9$ et 0.99 . Quant à Qs, il s'agit des quantiles stationnaires.

Annexe 22: Modèle GPD avec fonction B-spline à deux nœuds et de degré 2 : station 7060400



Annexe 23: Quasi-équivalence des quantiles stationnaires et non-stationnaires à la station 8300100



Annexe 24: Résultats de l'AF pour les 46 stations associées à l'indice AO

ID	Nom	u	N	ξ_0	σ_0	Q2	Q10	Q100	ξ_1	s1	s2	s3	Qg	Qc	Qd
8300525	O'LEARY	32	48	0,0119	10	39	54	76	0,0093	5,16	12,58	14,04	44	57	64
8100590	BUCTOUCHÉ	32	33	0,0111	14	41	63	94	0,0104	17,45	15,24	9,96	71	64	54
8101500	FREDERICTON A	34	57	0,0114	13	44	65	95	0,0105	10,42	13,45	17,38	58	65	74
8101600	FREDERICTON CDA	34	50	0,0108	11	42	60	85	0,012	1,77	12,59	17,92	39	63	75
8401700	GANDER INT'L A	30	59	0,011	10	39	53	75	0,0089	15,57	8,9	6,56	65	50	45
8402925	PETTY HARBOUR	40	40	0,0106	14	49	71	102	0,0118	6,47	18,31	22,71	55	75	100
8403800	STEPHENVILLE A	31	60	0,0117	12	39	58	83	0,0101	17,99	8,39	11,17	72	55	56
8201410	DEMING	41	50	0,0122	14	51	74	105	0,0083	9,89	13,76	18,36	64	74	83
8202000	GREENWOOD A	33	58	0,0113	13	42	61	89	0,0096	8,08	14,96	14,52	52	63	66
8202250	HALIFAX STANFIELD INT'L A	46	54	0,0117	12	54	73	100	0,0103	13,79	9	19,76	77	70	90
8203535	MIDDLE MUSQUODOBOIT	40	44	0,0111	11	47	65	89	0,0097	8,24	7,85	17,82	59	63	80
8203600	MOUNT UNIACKE	46	46	0,0114	12	55	74	101	0,0096	6,36	10,62	18,23	61	74	87
8205000	SALMON HOLE	42	49	0,0114	15	52	76	110	0,0093	7,06	17,08	27,51	58	80	104
8206200	UPPER STEWIACKE	37	53	0,0114	11	45	63	88	0,0098	12,69	6,49	24,79	66	58	93
8206500	YARMOUTH A	39	62	0,0113	16	50	74	109	0,01	9,34	16,57	18,57	60	77	81
61519JM	CRESSY	28	36	0,0122	9	34	48	68	0,0069	19,18	7,41	5,97	71	45	42

6057592	SAULT STE MARIE A	26	51	0,0112	9	32	46	65	0,0082	14,33	5	11,43	58	42	52
6068150	SUDBURY A	28	58	0,0118	9	34	48	68	0,01	14,94	7,76	8,26	62	46	47
6100345	ARNPRIOR GRANDON	25	37	0,0109	10	32	47	69	0,0105	6,94	9,49	28,26	41	45	88
6104175	KINGSTON PUMPING STATION	27	45	0,0127	10	34	50	71	0,0104	13,8	8,33	10,07	59	48	50
6110605	BEATRICE	27	29	0,0123	10	34	51	74	0,0088	23,33	9,46	4,17	80	47	37
6112133	DUNCHURCH	30	34	0,0108	11	38	56	80	0,0093	23,02	8,23	7,05	82	50	47
6112171	DURHAM	30	49	0,0114	12	38	57	84	0,0107	6,03	11,1	17,68	44	58	70
6115525	MUSKOKA A	27	54	0,0106	9	33	48	68	0,008	6,5	8,06	16,64	42	46	65
6116132	OWEN SOUND MOE	30	41	0,0116	10	36	52	74	0,0107	12,63	7,75	14,24	58	49	62
6116702	POWASSAN	27	33	0,0113	12	35	54	80	0,0044	4,72	13,6	13,81	38	58	59
6142420	FOLDENS	30	41	0,0118	13	39	59	87	0,0195	13,89	17,28	0,77	61	64	32
6147664	SCOTLAND	29	36	0,0104	15	39	62	94	0,0064	9,85	13,32	26,89	52	59	90
6149625	WOODSTOCK	29	52	0,0115	12	38	57	84	0,0101	8,83	12,03	17,45	49	56	69
6151689	COBOURG STP	26	36	0,0111	11	34	51	74	0,0085	16,28	8,87	7,19	63	49	43
6158665	TORONTO ISLAND A	26	44	0,0116	11	33	50	73	0,0101	16,68	10,13	4,24	63	47	36
6158733	TORONTO LESTER B. PEARSON INT'L A	27	59	0,0113	11	34	51	74	0,0109	12,96	10,9	2,09	56	51	32
6163156	HALIBURTON A	28	39	0,0119	9	34	49	69	0,0093	13,43	9,94	3,36	58	49	36
7018564	TROIS RIVIERES	31	36	0,0101	11	38	55	79	0,0044	13,12	12,22	4,63	61	53	42

7022375	FLEURY	32	46	0,0112	10	39	55	78	0,0085	21,86	8,33	6,41	81	51	47
7022800	GRANBY	33	61	0,011	11	40	57	81	0,0094	11,92	11,24	5,03	60	58	44
7025440	NICOLET	28	59	0,0112	10	34	50	71	0,008	17,31	7,59	4,95	67	45	39
7025745	ORMSTOWN	27	49	0,0117	12	36	55	82	0,0098	19,12	11,24	7,21	71	55	44
7026040	PHILIPSBURG	29	60	0,012	11	37	55	80	0,0101	7,69	13,59	11,6	47	57	56
7026836	ST ANICET	28	50	0,0107	11	35	52	75	0,0123	15,92	8,71	2,25	64	52	33
7033650	LACHUTE	31	48	0,0117	11	39	56	80	0,0107	5,83	10,03	15,37	45	56	66
7048320	TADOUSSAC	31	47	0,0106	10	38	55	78	0,0092	16,13	6,11	10,15	67	50	54
7050455	BARRAGE LAC MORIN	30	32	0,0124	9	37	52	72	0,0105	12,94	9,02	6,29	60	51	45
7055120	MONT-JOLI A	27	60	0,0109	9	33	47	67	0,01	-0,9875	11,22	8,6	27	52	46
7066080	PORTAGE DES ROCHES	25	31	0,0103	10	32	48	69	0,0113	16,37	9,51	4,27	62	45	35
7117825	SCHEFFERVILLE A	22	58	0,0111	7	27	38	54	0,01	4,8	8,39	12,25	33	39	49

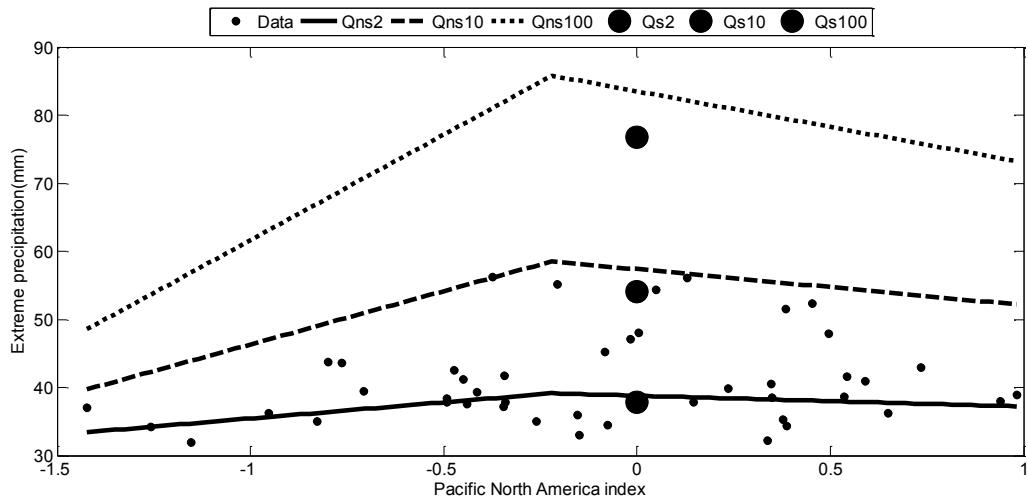
Annexe 25: Résultats de l'AF pour les 46 stations associées à l'indice PNA

ID	Nom	u	N	ξ_0	σ_0	Q2	Q10	Q100	ξ_1	s1	s2	s3	Qg	Qc	Qd
8300100	ALLISTON	31	40	0,0117	11	39	57	81	0,0102	10,8	10,68	12,42	56	56	59
8300500	LONG RIVER	31	45	0,0124	10	38	54	77	0,0087	3,78	12,21	9,36	40	37	52
8300700	SUMMERSIDE A	31	40	0,0125	9	37	51	70	0,0103	4,09	10,3	9,72	40	54	53
8101000	MIRAMICHI A	33	54	0,0121	10	40	57	79	0,0099	7,89	10,05	15,21	51	56	67
8101200	DOAKTOWN	34	52	0,0118	11	41	58	81	0,0107	5,15	12,38	11,7	46	61	60
8403600	ST JOHN'S WEST CDA	41	44	0,0112	13	50	71	101	0,0111	15,53	16,18	3,66	76	71	50
8202700	KEMPTVILLE	44	39	0,0117	14	54	75	106	0,0098	27,15	10,9	9,78	106	69	67
8203161	LOUISBOURG	43	33	0,0104	12	51	71	98	0,0102	6,98	10,94	21,1	59	69	91
8203700	NAPPAN CDA	32	51	0,0115	13	41	63	92	0,0105	8,05	12,48	20,85	51	60	79
8204800	ST MARGARET'S BAY	42	52	0,0113	13	51	71	99	0,0091	20,98	11,16	9,23	89	67	63
6010738	BIG TROUT LAKE	21	41	0,011	9	27	41	60	0,0095	11,84	8,33	6,64	48	40	36
6012198	EAR FALLS	23	45	0,0125	10	30	46	68	0,0095	15,19	8,79	9,13	57	43	44
6022475	FORT FRANCES	28	42	0,0101	16	39	64	100	0,011	19,75	18,23	6,51	72	63	43
6032117	DRYDEN	24	46	0,0111	11	31	48	72	0,009	11,38	7,47	20,99	50	48	71
6032119	DRYDEN A	26	33	0,0121	13	35	56	84	0,0095	16,11	12,54	2,92	63	56	33
6072325	ENGLEHART	25	48	0,0104	10	32	47	68	0,0102	7,29	8,52	15,09	42	45	59
6075425	MOOSONEE UA	22	53	0,0086	11	30	47	71	0,01	25,68	5,75	9,03	80	36	43

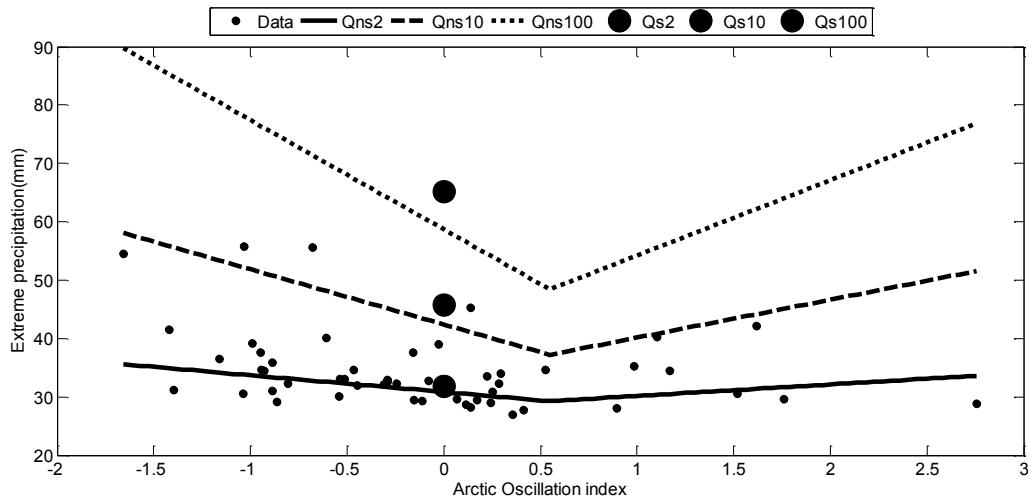
6100971	BROCKVILLE PCC	29	40	0,0114	12	38	56	82	0,0103	7,91	11,09	14,7	47	57	63
6101494	CHENEAUX	25	38	0,0109	8	31	44	62	0,01	12,89	5,23	6,62	54	38	40
6104025	KEMPTVILLE	29	46	0,012	9	35	50	70	0,0111	3,08	9,77	13,93	36	51	60
6112340	ESSA ONT HYDRO	26	41	0,0111	10	33	49	72	0,0122	1,92	10,45	15,32	31	52	61
6122370	EXETER	29	45	0,0109	13	38	58	86	0,0099	8,7	11,08	25,66	49	55	87
6134390	LEAMINGTON	29	27	0,0102	14	39	61	92	0,0088	9,19	10,83	39,39	50	54	118
6135583	NEW GLASGOW	29	46	0,0111	12	38	57	84	0,0094	24,15	10,04	8	83	51	47
6136606	PORT COLBORNE	29	40	0,0112	11	37	55	80	0,0095	5,02	13,54	12,22	41	60	57
6140954	BRANTFORD MOE	26	44	0,0127	11	34	52	77	0,0075	6,61	10,36	14,66	42	53	60
6158255	THORNHILL GRANDVIEW	27	40	0,0108	10	34	50	71	0,0076	6,11	8,08	22	41	47	77
6164433	LINDSAY FROST	25	33	0,0116	11	33	49	73	0,0099	7,81	9,41	17,91	43	47	66
7010720	BERTHIERVILLE	30	42	0,0108	11	38	55	79	0,0098	6,36	9,98	18,67	45	58	72
7015730	OKA	29	59	0,012	11	37	53	77	0,0098	5,36	11,27	17,92	42	54	70
7016816	ST ALEXIS DES MONTs	31	50	0,0099	12	40	59	87	0,0076	11,72	13,91	8,98	58	61	52
7020040	ABERCORN	31	34	0,012	11	38	55	78	0,0098	7,39	7,98	21,31	48	56	80
7021840	COATICOOK	30	60	0,0118	12	38	57	83	0,0099	8,57	12,38	13,92	50	58	61
7022160	DRUMMONDVILLE	30	58	0,0112	11	37	54	77	0,0084	4,89	11,43	17,37	41	55	69
7022320	FARNHAM	30	59	0,0111	12	38	56	83	0,01	-1,24	14,73	12,03	30	63	57
7027360	ST HYACINTHE	29	26	0,0117	10	37	53	76	0,0087	18,85	11,02	5,2	72	49	42

7028280	STANSTEAD	31	39	0,012	9	36	50	69	0,0097	13,07	7,81	7,02	60	48	47
7034480	MANIWAKI UA	27	41	0,0119	10	33	48	70	0,0091	14,24	10,25	3,6	59	46	35
7042840	GRANDES BERGERONNES	32	60	0,0113	14	42	63	93	0,0084	6,36	15,9	12,44	47	67	61
7051120	CAPLAN	30	47	0,0105	12	39	58	86	0,009	6,47	12,23	18,59	45	58	72
7054095	LA POCATIERE CDA	29	44	0,0115	9	36	51	72	0,01	10,09	8,19	13,62	52	48	60
7057650	STE PERPETUE	32	33	0,0111	12	40	58	84	0,0081	2,99	12,5	18,12	39	61	73
7058560	TROIS PISTOLES	27	46	0,0116	9	34	49	70	0,0102	14,64	8,92	6,57	60	48	42
7060400	BAGOTVILLE A	25	60	0,0108	10	32	47	68	0,0063	13,81	8,92	6,23	56	46	39
7080600	BELLETERRE	26	46	0,0108	12	35	55	82	0,0094	7,33	10,62	23,28	43	55	79
7090120	AMOS	24	47	0,0128	11	32	49	74	0,0099	7,71	10,44	16,61	42	50	62

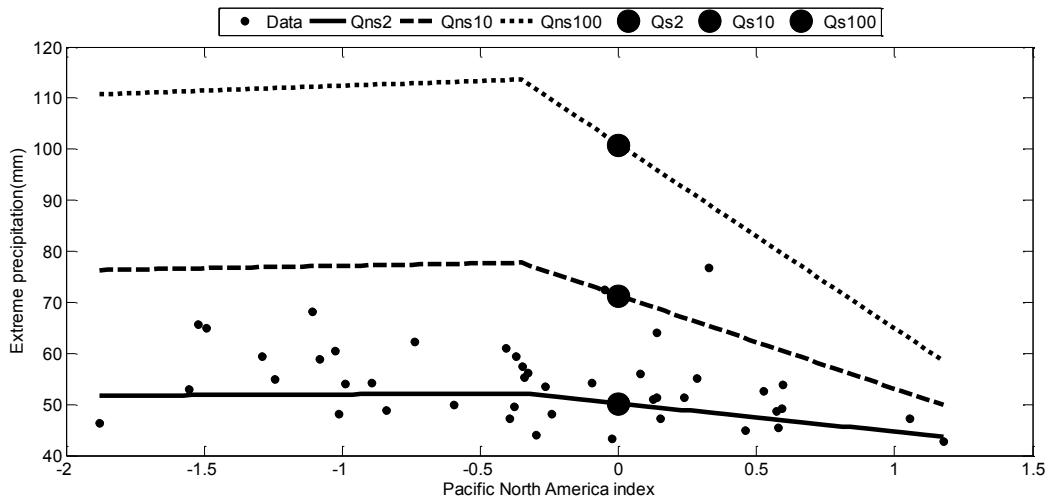
Annexe 26: Changement dans l'intensité des précipitations extrêmes durant les deux phases de la covariable climatique : Exemple 1 avec la Station 8300500



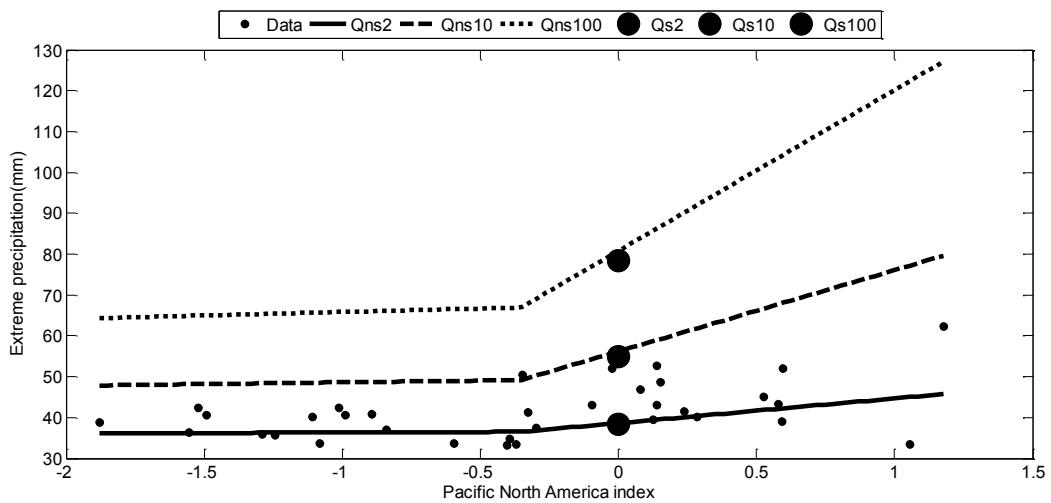
Annexe 27: Changement dans l'intensité des précipitations extrêmes durant les deux phases de la covariable climatique : Exemple 2 avec la Station 6057592



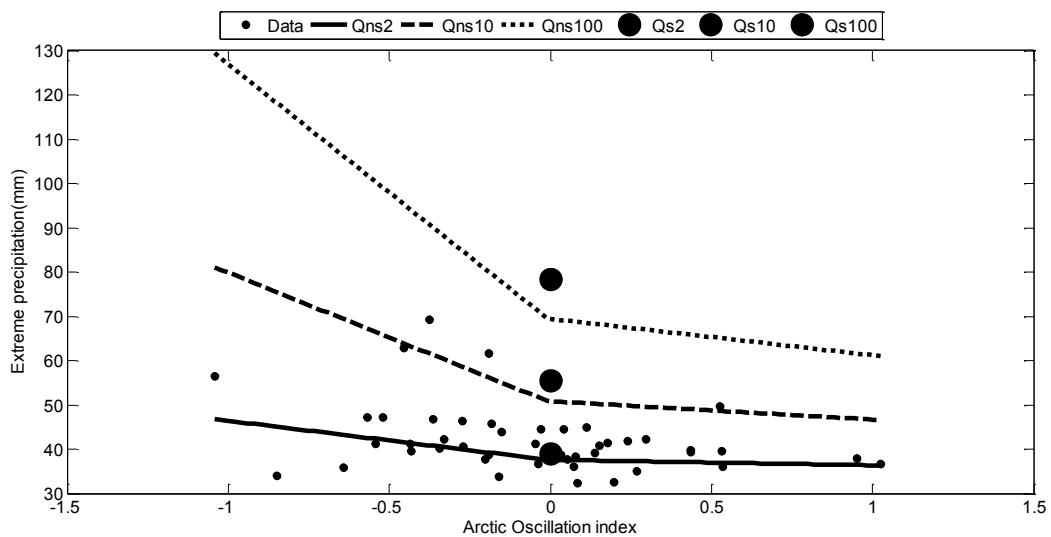
Annexe 28: Changement dans l'intensité des précipitations extrêmes durant la phase positive de la covariable climatique : Exemple 1 avec la Station 8403600



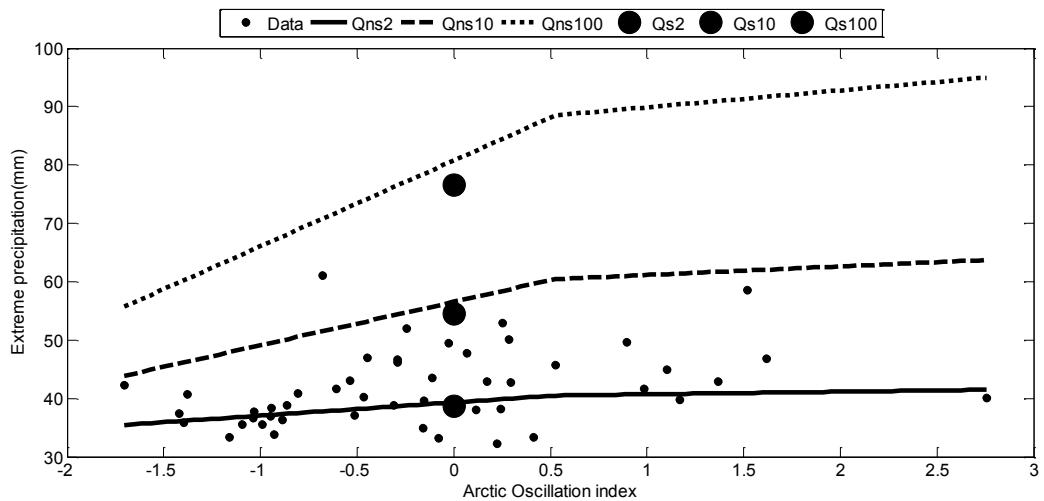
Annexe 29: Changement dans l'intensité des précipitations extrêmes durant la phase positive de la covariable climatique : Exemple 2 avec la Station 7020040



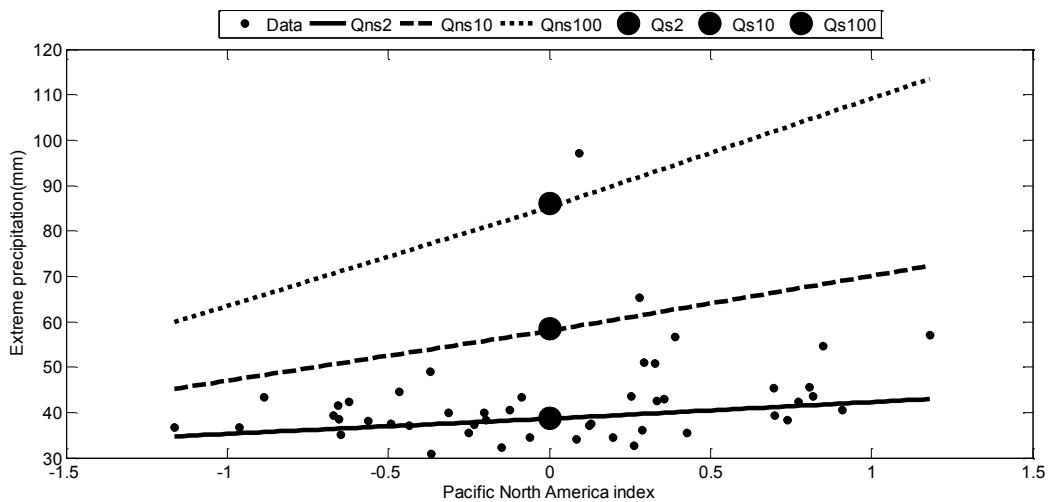
Annexe 30: Changement dans l'intensité des précipitations extrêmes durant la phase négative de la covariable climatique : Exemple 1 avec la Station 7022375



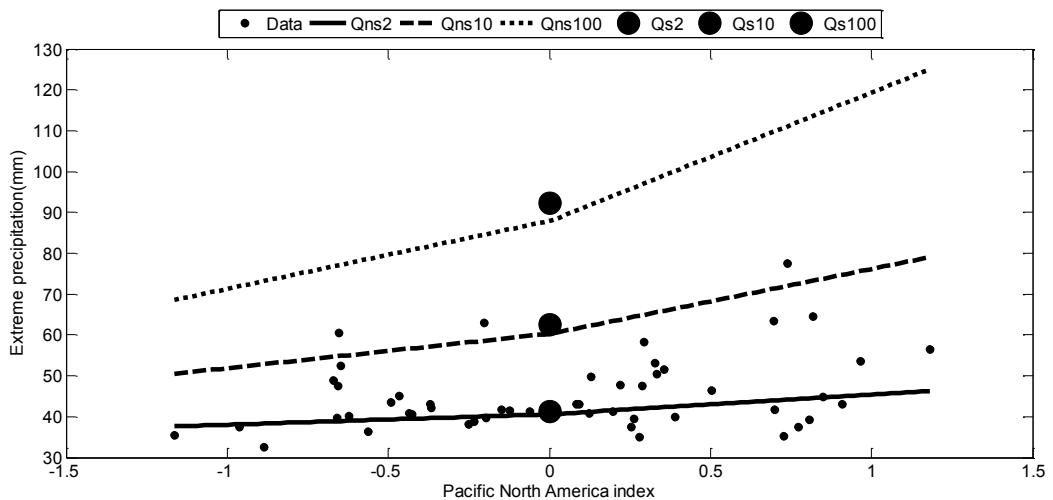
Annexe 31: Changement dans l'intensité des précipitations extrêmes durant la phase négative de la covariable climatique : Exemple 2 avec la Station 8300525



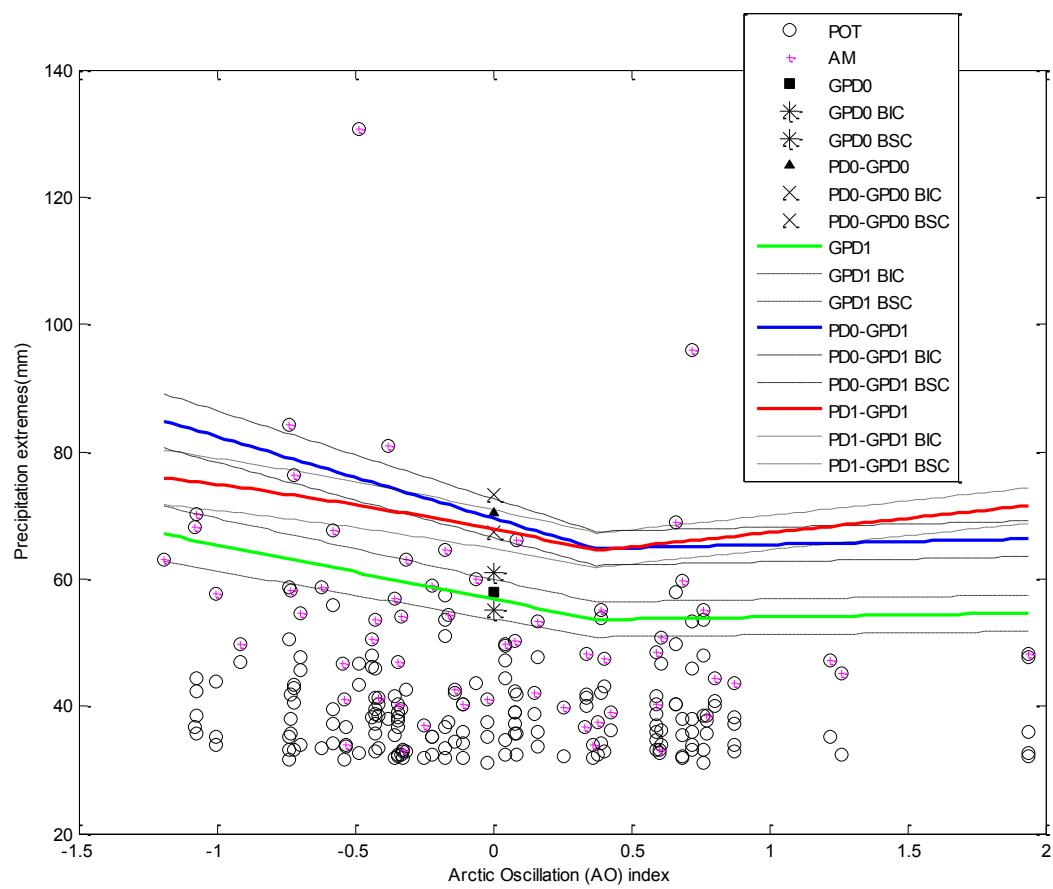
Annexe 32: Dépendance linéaire statistiquement significative : Exemple avec la Station 7051120



Annexe 33: Dépendance linéaire statistiquement non-significative : Exemple avec la Station 8203700



Annexe 34: Quantiles conditionnels à AO avec le modèle P-GPD à la station 8403800



Annexe 35: Quantiles conditionnels à AO avec le modèle P-GPD à la station 8203535

