

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

MODÉLISATION DES CRUES EN RÉGIME NIVAL À L'EST DU CANADA

THÈSE

PRÉSENTÉE

COMME EXIGENCE PARTIELLE DU

DOCTORAT EN SCIENCES DE LA TERRE

ET DE L'ATMOSPHÈRE

PAR

NAWRES YOUSFI

MARS 2024

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de cette thèse se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.12-2023). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

REMERCIEMENTS

Tout d'abord, je voudrai remercier mon directeur de recherche, monsieur Philippe Gachon qui m'a dirigé tout au long de ces années. Il était toujours disponible, à l'écoute de mes nombreuses questions, et s'est toujours intéressé à l'avancement de mes travaux. Les nombreuses discussions que nous avons eues ainsi que ses conseils ont bien amélioré les résultats finaux de ce travail. Je le remercie également pour le partage des connaissances et les corrections rapides.

Je remercie également mon co-directeur monsieur Salaheddine El Adlouni qui m'a codirigé tout au long de ma thèse de doctorat et celle de ma maîtrise. Grâce à son encadrement depuis la maîtrise, j'ai eu l'occasion d'améliorer mes connaissances en probabilité et en statistiques. Cette formation m'a donné un outil performant pour valoriser mes connaissances ainsi que mes décisions en ingénierie. Un remerciement tout spécial à la confiance qu'il m'a accordée en m'impliquant dans plusieurs projets de recherche, dans lesquelles j'ai acquis de l'expérience et j'ai amélioré mes compétences professionnelles.

Je tiens également à remercier monsieur Simon Michael Papalexiou : professeur assistant au département de génie civil à l'Université de Calgary pour sa contribution constructive dans la rédaction de notre deuxième article présenté sous forme du chapitre 3 dans cette thèse.

Je souhaite aussi remercier mon collègue Hussein Wazneh qui m'a aidé dans le traitement des données de télédétections que j'ai utilisées dans la réalisation de mon projet de doctorat. Je le remercie également pour les différentes discussions constructives que nous avons eues dans l'objectif d'améliorer l'analyse de mes résultats préliminaires.

Finalement, je remercie tous mes professeurs et mes collègues du département de sciences de la terre et de l'atmosphère, à l'Université du Québec à Montréal.

DÉDICACE

Je dédie cette thèse de doctorat :

À ma chère fille Janna qui est née durant ces années de thèse et qui m'a donné plus de motivation pour réaliser ce travail.

À mes chers parents qui m'ont soutenu et encouragé à poursuivre mes études doctorales, malgré toutes les contraintes que j'ai eues.

À mon époux qui a partagé avec moi tous les moments d'émotion lors de la réalisation de ce travail.

À mes deux frères qui m'ont chaleureusement supporté et encouragé tout au long de mon parcours.

À mes meilleures amies Yasmine et Azza qui n'ont pas arrêté de m'encourager et de me motiver durant toutes ces années de recherche.

À ma famille, mes proches et à ceux qui me donnent de l'amour et de l'encouragement.

À tous ceux que j'aime.

AVANT-PROPOS

La modélisation des événements extrêmes est une problématique d'actualité, vu les changements climatiques et les effets impliqués. J'ai commencé ce sujet depuis ma maîtrise, et j'ai développé des modèles probabilistes pénalisés pour estimer les crues en fonction d'une seule covariable. J'ai intégré les précipitations totales dans ce modèle et j'ai estimé les débits de toutes les stations hydrométriques jaugés au Nouveau-Brunswick sachant cette covariable (précipitations totales). C'était ma première expérience dans la recherche scientifique et j'étais très motivée par les développements que j'ai faits et les résultats que j'ai obtenus. Mon travail de maîtrise a abouti à la rédaction de deux articles scientifiques. J'avais l'avantage d'être co-dirigé par M. Salaheddine El Adlouni : professeur au département de statistique et mathématique à l'Université de Moncton, alors que j'étais étudiante au département de génie civil. Cette collaboration entre les deux disciplines a enrichi mes connaissances en probabilité et en statistique, ce qui m'a permis de faire le lien entre l'hydrologie et les précipitations en utilisant des liens statistiques et probabilistes. Ma motivation ne s'est pas arrêtée au niveau de la maîtrise, et je me suis trouvée poser plusieurs questions de recherche et ouvrir plusieurs pistes à développer. Cette motivation m'a poussé à appliquer dans un programme de doctorat lié à cette problématique et poursuivre ma recherche afin de vérifier toutes les hypothèses que j'avais en tête.

J'ai eu l'encouragement de mon co-directeur pour aller en avant et j'ai eu l'avantage d'être encadré par M. Philippe Gachon : professeur au département de Géographie et spécialiste en hydroclimatologie. Cette nouvelle collaboration n'a fait qu'enrichir mes connaissances surtout en météorologie et en climatologie : deux domaines dont je n'avais pas une grande connaissance alors qu'ils représentent les dynamos de mon projet. Avec les lectures approfondies lors des premières sessions, j'ai eu une base sur laquelle je me suis appuyée pour pouvoir faire le lien entre les différentes composantes du système atmosphérique et les inondations. De plus, les actualités sur les inondations, au cours des dernières années (inondations, 2017, 2019, 2020, 2021), m'ont encore motivé, vu que mon projet est une problématique actuelle et les résultats de mon projet vont être un centre d'intérêt pour plusieurs domaines comme l'infrastructure urbaine, la gestion des risques naturels et les assurances. Autrement dit, plusieurs domaines cherchent la disponibilité des outils d'estimation des crues en tenant compte du changement climatique. Vu mon domaine d'expertise en hydraulique et hydrologie, je peux mieux voir l'utilité de ce que je suis en train de développer dans la conception et la modélisation hydraulique. De plus, je trouve que le domaine de génie-conseil a besoin de proposer de nouveaux modèles d'estimation qui tiennent compte des conditions

climatiques existantes et projetées pour substituer la majoration utilisée actuellement (majoration pour le changement climatique) par un coefficient spécifique au site d'étude. Ceci va certainement améliorer la qualité d'estimation et mieux analyser le lien de cause à effet.

Pour toutes ces raisons, j'ai décidé de commencer une thèse de doctorat et la poursuivre jusqu'à la fin, malgré toutes les contraintes que j'ai eues durant ces années. Ces années n'étaient pas les plus simples et le défi de ne pas lâcher mon objectif a nécessité beaucoup de sacrifices. Grâce à l'encouragement et l'appui de mes directeurs, ainsi que ma famille, je me trouve finir ma thèse avec deux articles scientifiques qui développent deux modèles pertinents et qui offrent une innovation dans la modélisation des crues printanières en régime nival. Je finis également cette thèse avec une connaissance approfondie qui me permet de mieux analyser le système hydrologique dans un contexte du changement climatique. Ces connaissances enrichissent encore ma formation académique et mon expertise professionnelle et me permettent par conséquent d'avoir une meilleure analyse de la réalité, une meilleure vision et une meilleure conception.

TABLE DES MATIÈRES

REMERCIEMENTS	ii
DÉDICACE	iii
AVANT-PROPOS.....	iv
LISTE DES FIGURES.....	ix
LISTE DES TABLEAUX	xi
LISTE DES ABRÉVIATIONS, DES SIGLES ET DES ACRONYMES.....	xii
LISTE DES SYMBOLES ET DES UNITÉS	xv
RÉSUMÉ	xvi
ABSTRACT	xviii
CONTRIBUTION À LA SCIENCE	1
CONTRIBUTION DES AUTEURS	3
INTRODUCTION GÉNÉRALE	5
CHAPITRE 1 États des connaissances, problématique et objectifs	9
1.1 Mise en contexte.....	9
1.2 Mécanismes physiques impliqués : occurrence et caractéristiques des aléas hydrométéorologiques 11	
1.2.1 Facteurs de grande échelle.....	12
1.2.1.1 L'oscillation nord-atlantique (NAO).....	12
1.2.1.2 L'Atlantic Multidecadal Oscillation (AMO)	13
1.2.1.3 Le Baffin Island-West Atlantic (BWA) Index.....	13
1.2.2 Facteurs d'échelle locale.....	14
1.2.3 Autres facteurs.....	14
1.3 Outils de modélisation disponibles pour la simulation et la prévision des crues.....	15
1.3.1 Modèles numériques	16
1.3.2 Modèles Pluie-Débit	17
1.3.2.1 Modèles conceptuels :.....	17
1.3.2.2 Modèles physiques	17
1.3.2.3 Modèles empiriques	18
1.3.3 Modèles événementiels.....	18
1.4 Théorie des valeurs extrêmes (TVE).....	19
1.4.1 Principales distributions.....	19

1.4.1.1	Distribution GEV	19
1.4.1.2	Distribution GPD	20
1.4.1.3	Méthodes de sélection du seuil	20
1.5	Problématique liée à la région d'étude, objectifs et hypothèses	22
1.5.1	Problématique	22
1.5.2	Objectifs	23
1.5.3	Hypothèses	24

CHAPITRE 2 Multivariate spring floods estimation using the GEV-B-Splines model: The Saint-Jean River case study, New-Brunswick (Canada)25

2.1	Abstract	27
2.2	Abreviations	28
2.3	Introduction	29
2.4	Study area and hydrometric stations	32
2.5	Meteorology.....	36
2.5.1	Derived indices.....	36
2.5.1.1	Meteorological derived indices	36
2.5.1.2	Teleconnection indices	40
2.5.2	Research methods	42
2.5.2.1	The GEV distribution model.....	42
2.5.2.2	The non-stationary GEV-B-Splines model.....	43
2.5.2.3	The developed models.....	44
2.6	Results.....	46
2.6.1	Hydrographic characteristics upstream and downstream of the watershed.....	46
2.6.2	Single links between maximum flow and covariables	47
2.6.3	Combined effects of covariables on maximum flow from developed models.....	50
2.6.4	Results of the GEV-B-Splines with mixed covariates	52
2.7	Discussion and conclusion.....	59
2.8	Credit authorship contribution statement.....	60
2.9	Declaration of competing interest	60
2.10	Data availability statements.....	60
2.11	Acknowledgments.....	61

CHAPITRE 3 Mixture Probability Models with Covariates: Application in Estimating Risk of Hydroclimatic Extremes.....62

3.1	Abstract.....	64
3.2	Introduction	65
3.3	Material and methods.....	68
3.3.1	Study area	68
3.3.2	GGP mixture Model	69

3.4	Application of the proposed model	74
3.4.1	Choice of threshold and selection of extreme events	74
3.4.2	Simulation study	78
3.5	Results and discussion.....	79
3.5.1	Simulation study results.....	79
3.5.2	Application results	83
3.6	Discussion.....	87
3.7	Summary and conclusions.....	88
3.8	Credit authorship contribution statement.....	89
3.9	Declaration of competing interest	89
3.10	Data availability statements.....	89
3.11	Acknowledgments.....	89
	CHAPITRE 4 Discussion générale	91
4.1	Mise en contexte.....	91
4.2	Résultats obtenus et limites.....	92
4.3	Nouvelles connaissances.....	94
4.4	Synthèse et recommandations	95
	CONCLUSION GÉNÉRALE	97
	ANNEXE A The Gamma distribution	102
	ANNEXE B The Generalized Pareto Distribution (GPD)	103
	BIBLIOGRAPHIE.....	104

LISTE DES FIGURES

Figure 2.1: Localisation of the Saint John River watershed situated in Canadian provinces of New-Brunswick (NB), Quebec (QC), and the American State of Maine (ME, É.-U.). The main hydrometric (meteorological) stations, Fort-Kent (Edmundston) and Mactaquac (Fredericton) are in blue (red/yellow for temperature/precipitation) symbols. 34

Figure 2.2. The annual maximum flow observed over the period January-June at the stations Fort-Kent (01AD002, red line), for the period 1934-1994 and Mactaquac (01AK004, blue line), for the period 1967-1994..... 47

Figure 2.3. Estimation of the quantiles of extreme flows in nival regime using the *GEV3* with $X''' = \text{TXD}$, *PTOTB*, *BWAB*. The illustration is given for two extreme values (min and max) of the *BWA* index observed for the period 1934-2014, and two return periods $T= 20$ (left) and 100 years (right) for the Saint John River at Fort-Kent station..... 53

Figure 2.4. Probability density functions of annual maximum flow at Fort-Kent station (Saint John River) for the stationary model (without any covariate -black line-) and the model *GEV3* with $X''' = \text{TXD}$, *PTOTB*, *BWAB* with an illustration for two values, minimal (blue line) and maximal (red line) of the *BWAB* index. 54

Figure 2.5. Interannual standardized anomalies over the 1934-1994 period of annual maximum flows observed between January and June at Fort-Kent (blue line) and of seasonal mean *BWA* computed between December and March (pink curve). The maximum flow is selected using the time series of observed flow (measured every 15 minutes at the hydrometric station). All standardized values are computed using the mean and standard deviation over the 1934-1994 climatological period, both for the maximum flows and the seasonal mean *BWA*..... 55

Figure 2.6. Estimation of the quantiles of extreme flows in nival regime using the *GEV3* with $X''' = (\text{PTOTB}, \text{RAIND}, \text{and } \text{NAOB})$ for $T= 20$ (left) and 100 years (right) return period at Mactaquac (Saint John River). Illustration given for a minimal (390 mm) and maximal (910 mm) possible values of the covariate *PTOTB*. 56

Figure 2.7. Probability density functions of annual maximum flow at Mactaquac station for the stationary model (without any covariate -red line-) and the *GEV3* with $X''' = (\text{PTOTB}, \text{RAIND}, \text{and } \text{NAOB})$ model with an illustration for two values, minimal (blue line) and maximal (black line) of the *NAOB* index. 57

Figure 2.8. Interannual standardized anomalies over the 1967-1994 period of annual maximum flows observed between January and June at Mactaquac (blue line) and of seasonal mean *NAO* computed between December and March (pink curve). The maximum flow is selected using the time series of observed flow (measured every 15 minutes at the hydrometric station). All standardized values are computed using the mean and standard deviation over the 1967-1994 climatological period, both for the maximum flows and the seasonal mean *BWA*..... 58

Figure 3.1. The study area of Saint-John River (in green) with the location of the Fort-Kent hydrometric station located upstream of the watershed..... 69

Figure 3.2. Illustration of the difference between mixture model proposed in the present study (left) and the hybrid distributions 72

Figure 3.3. Polar plot of the daily flow at Fort-Kent station, located upstream of the Saint-John River, New Brunswick, Canada (see Figure 3.1). Hydrologic data are available from 1927 to 2014 and units are in $m^3.s^{-1}$. Annual maxima are represented by black dots, and maximum/minimum flow records are in red/green dots. Values are given for each Julian day (from 0 to 365). 75

Figure 3.4. Selected events using a fixed daily baseflow at Fort-Kent (Saint John River) using the observed values over the 1989-1993 period. Each daily maximum of flow is plotted in cyan dots, and baseflow is plotted in blue line. 77

Figure 3.5. Probability density functions of the simulated distributions (Table 3.2) with different shapes. 79

Figure 3.6. Mean relative bias (%) and its 95% confidence intervals for the quantile estimators with return period $T = 2, 10, 100$ and 1000 years given by AMS, POT and GGP mixture models..... 80

Figure 3.7. (a) Quantiles for given return periods, estimated by Mixture and Hybrid models with the same components Gamma and GPD. (b) The pdfs of the mixture and Hybrid models for the case study. 82

Figure 3.8. Probability density function of (a) the GGP mixture model for selected daily streamflow events, and (b) GEV for the series of maxima at Fort-Kent, Saint-John River. 84

Figure 3.9. (a) GGP mixture fit to all independent event at Fort-Kent (Saint-John River) and (b) the AMS-GEV model fit with quantiles of return periods for both GGP mixture and GEV model at Fort-Kent.85

Figure 3.10. Confidence intervals at 95% level for the quantiles of the GGP mixture model computed by bootstrap. The daily flow data series from the Fort-Kent station is used, from all data covering the 1927-2014 period. 86

LISTE DES TABLEAUX

Tableau 2.1. Hydrometric and meteorological stations and data used in the study. The location of each station is given in Figure 2.1.	35
Tableau 2.2. Meteorological derived indices and their definition with their respective acronyms.	37
Tableau 2.3. Teleconnection derived indices and aggregated definition at seasonal and monthly scales before and during the flood month.....	39
Tableau 2.4.. Teleconnection derived indices and aggregated definition at seasonal and monthly scales before and during the flood month.....	41
Tableau 2.5. Flow characteristics upstream at Fort Kent (1934-1994) and downstream at Mactaquac (1967-1994) of the Saint John watershed.....	46
Tableau 2.6. Spearman correlation between the maximum flow and all covariables or climate indices (presented in Tables 2 and 3) computed for Fort-Kent and Mactaquac hydrometric stations. Correlation is computed depending on hydrometric station and common period of data availability (1934-2014 for Fort-Kent and 1967-1994 for Mactaquac). The rho and p-values are provided for each pair of correlation and for each station.....	49
Tableau 2.7. Identification of the best significant covariate with the GEV-B-Splines model using one (Model1,1), two (Model1,2), three (Model1,3) covariates and the stationary model (<i>GEV0</i>) at Fort-Kent (located upstream of the Saint John River, see Figure 2.1). The best combination, i.e. lowest AIC values, is in bold (using three covariates)	50
Tableau 2.8. Identification of the best significant covariate with the GEV-B-Splines model using one (Model2,1), two (Model2,2), three (Model2,3) covariates and the stationary model (<i>GEV0</i>) at Mactaquac (located downstream of the Saint John River, see Figure 2.1). The best combination, i.e. lowest AIC values, is in bold (using three covariates)	51
Tableau 3.1. Integration of covariates in the mixture model	71
Tableau 3.2. Data generation scenarios with lower and heavy tailed distributions.....	78

LISTE DES ABRÉVIATIONS, DES SIGLES ET DES ACRONYMES

AF : Analyse Fréquentielle

AIC: Akaike's Information Criterion

AMS: Annual Maxima Series

AMO: Atlantic Multidecadal Oscillation

AMO_B: Monthly mean AMO values between December and March

AMO_D: Monthly mean AMO values of the flood month

BWA: Baffin Island-West Atlantic

BWA_B: Seasonal BWA values between December and March

BWA_D: Monthly mean BWA values of the flood month

C_g: Skewness coefficient

C_v: Variation coefficient

EVT: extreme value theory

FA: Frequency Analysis

FTC: Freeze-Thaw Cycles

FTC_B : The number of FTC before the flood month.

FTC_D: The number of FTC during the flood month.

FTCM_B: The mean average of the number of FTC between preceding November and the flood month.

FTCN_B: The minimum of FTC number between preceding November and the flood month.

FTCX_B : The maximum of FTC number between preceding November and the flood month.

GAM: Gamma distribution

GEV: Generalized Extreme Value

GGP: Gamma-Generalized Pareto mixture model (proposed model)

GP: Generalized Pareto

GPD: Generalized Pareto Distribution

iid: independent and identically distributed

INV-GAM: Inverse Gamma distribution

MC: Monte Carlo

NC-Chi2 : Non-centred χ^2

NAO: North Atlantic Oscillation

NAO_D: Monthly mean NAO values during the flood month

NAO_B: Seasonal mean NAO values between December and March

NB: Nouveau-Brunswick

POT : Peaks over threshold

PTOT_B: Total precipitation, calculated and accumulated between preceding November and the flood month (mm).

PSOL_B: The solid phase of precipitation of the month preceding the flood occurrence, as this affects the flood intensity (mm).

Q_{max}: Annual maximum flow

RB: Relative Bias

RAIN_D: Monthly accumulated of daily rainfall during the flood month in mm.

SNOW_B: Total accumulation of daily snowfall between preceding November and the flood month (mm).

SNOW_D: Monthly accumulated of snowfall during the flood month in mm.

TX_D: The monthly mean of the maximum daily temperature during the flood month (°C).

LISTE DES SYMBOLES ET DES UNITÉS

B: Betta

θ : Tetta

% : pourcentage

\mathbb{R} : Nombres réels

μ : parameter de position

κ : parameter de forme

°C : degré Celsius

cm : centimètre

mm : millimètre

m³/s: mètre cube par seconde

n : nombre

y: variable à estimer

Y_T : Quantile

RÉSUMÉ

Les inondations sont les catastrophes naturelles les plus fréquentes et les plus coûteuses au Canada. Ces événements sont responsables des dégâts majeurs touchant l'infrastructure urbaine, l'environnement, l'écologie, les assurances, l'économie, la santé physique et/ou mentale des citoyens. Ces inondations touchent particulièrement l'est canadien, notamment lors des épisodes de crues printanières. Au Nouveau-Brunswick les principales régions inondables se situent au sein du bassin du fleuve Saint-Jean. Dans ce contexte, il est important de bien comprendre les mécanismes impliqués dans la génération de ces événements hydrométéorologiques extrêmes, afin de mieux les modéliser et en améliorer la prévention.

L'objectif principal de cette thèse est de modéliser les crues printanières qui se produisent dans l'est du Canada, et mieux comprendre l'effet de certaines variables météorologiques et climatiques sur les débits printaniers historiques à l'aide des modèles statistiques. Les approches proposées sont illustrées pour une station hydrologique en amont et une autre en aval du bassin du fleuve Saint-Jean.

Le premier chapitre présente un état des lieux des connaissances, dans lequel les principaux outils numériques déterministes, utilisés dans la modélisation hydrologique et/ou hydraulique, sont brièvement présentés. De plus, les méthodes probabilistes utilisées comme alternatives aux méthodes déterministes sont également présentées, ainsi que les avantages et limites de chacune des méthodes ou outils. Le chapitre 2 représente l'utilisation du modèle probabiliste *Generalized Extreme Value* (GEV) dans l'estimation des débits maximums printaniers en fonction des indices météorologiques et climatiques proposés. Les indices sont développés en se basant sur les données historiques brutes et certaines analyses statistiques. La liaison de dépendance entre les variables est représentée par une fonction B-Splines. Le travail a abouti à la rédaction d'un premier article à soumettre dans le *Journal of Stochastic Environmental Research and Risk Assessment* (SERRA).

Dans le troisième chapitre de cette thèse, un modèle de mélange a été proposé pour améliorer l'estimation des extrêmes de crues à l'aide des données de la station hydrométrique de Fort-Kent. Le modèle développé propose une alternative aux techniques de choix du seuil et combine les distributions de Pareto Généralisée et Gamma. Cette méthode permet de prendre en considération plus qu'un extrême annuel en considérant tous les événements indépendants observés durant la saison à l'étude.

Ce travail a fait l'objet d'un article publié dans le *Journal of Hydrologic Engineering*. Les résultats ont été comparés avec le modèle GEV, utilisé et présenté au chapitre 2, et ont montré une meilleure flexibilité dans la modélisation des séries non stationnaires ou dépendantes. Le modèle de mélange proposé, décrit adéquatement aussi bien la partie centrale, que les deux queues de distribution des débits.

En conclusion, ces deux modèles proposés dans cette thèse de doctorat sont flexibles et peuvent être appliqués dans n'importe quel site d'étude, afin de modéliser les débits extrêmes en tenant compte des caractéristiques météorologiques et climatiques propres au site.

Le modèle GEV-B-Spline permet d'estimer les débits en fonction de plusieurs covariables explicatives et en se basant sur les débits maximums annuels historiques. Cependant, cette méthode néglige certains événements importants comme les crues en automne. Le modèle de mélange proposée, propose une combinaison de deux distributions pour considérer tous les événements indépendants et enrichir par conséquent la série de données.

Mots clés : Inondation printanière, aléa hydrométéorologique, téléconnexion, modélisation probabiliste, analyse fréquentielle, modèle de mélange, extrême hydrologique.

ABSTRACT

Floods are the most frequent and costly natural disasters in Canada. These events are responsible for major damage affecting urban infrastructure, the environment, ecology, insurance, the economy, and the physical and/or mental health of citizens. These floods particularly affect eastern Canada, particularly during episodes of spring flooding. In New Brunswick the main flood regions are located within the Saint John River basin. In this context, it is important to fully understand the mechanisms involved in the generation of these extreme hydrometeorological events, in order to better model them and improve their prevention.

The main objective of this thesis is to model the spring floods that occur in eastern Canada, and to better understand the effect of certain meteorological and climatic variables on historical spring flows using statistical models. The proposed approaches are illustrated for one hydrological station upstream and another downstream of the Saint John River basin.

The first chapter presents an inventory of knowledge, in which the main deterministic numerical tools, used in hydrological and/or hydraulic modeling, are briefly presented. In addition, probabilistic methods used as alternatives to deterministic methods are also presented, as well as the advantages and limitations of each method or tool.

Chapter 2 represents the use of the probabilistic Generalized Extreme Value (GEV) model in the estimation of maximum spring flows based on the proposed weather and climate indices. The indices are developed based on raw historical data and certain statistical analyses. The dependency link between variables is represented by a B-Splines function. The work resulted in the writing of a first article to be submitted to the journal of Stochastic Environmental Research and Risk Assessment (SERRA).

In the third chapter of this thesis, a mixing model was proposed to improve the estimation of flood extremes using data from the Fort-Kent hydrometric station. The developed model offers an alternative to threshold choice techniques and combines the Generalized Pareto and Gamma distributions. This method makes it possible to take into account more than one annual extreme by considering all the independent events observed during the season under study.

This work was the subject of an article published in the Journal of Hydrologic Engineering. The results were compared with the GEV model, used and presented in Chapter 2, and showed better flexibility in modeling non-stationary or dependent series. The proposed mixing model adequately describes both the central part and the two tails of flow distribution.

In conclusion, these two models proposed in this doctoral thesis are flexible and can be applied in any study site, in order to model extreme flows taking into account the meteorological and climatic characteristics specific to the site.

The GEV-B-Spline model makes it possible to estimate flows according to several explanatory covariates and based on historical maximum annual flows. However, this method neglects certain important events

such as floods in autumn. The proposed mixture model proposes a combination of two distributions to consider all independent events and therefore enrich the data series.

Keywords: Spring flood, hydrometeorological hazard, teleconnection, probabilistic modeling, frequency analysis, mixing model, hydrological extreme.

CONTRIBUTION À LA SCIENCE

Les nouvelles approches développées dans cette thèse de doctorat, pour la modélisation des crues en régime nival à l'est du Canada ainsi que les résultats de l'évaluation des modèles proposés en utilisant des combinaisons d'indices (météorologiques et climatiques), apportent les contributions techniques et scientifiques suivantes :

Contributions techniques :

- Développement d'un modèle de prévision des crues en utilisant un modèle probabiliste pour les valeurs extrêmes avec plus d'une seule covariable explicative. Il s'agit du modèle GEV-B-Splines (Generalized-Extrême-Values with B—Splines) qui permet d'intégrer des indices météorologiques et climatiques et tenir compte de la non-stationnarité dans les séries, et des liens non-linéaires entre les covariables utilisées dans la modélisation des débits.
- Modélisation et analyse des crues en amont et en aval du bassin versant de la rivière Saint-Jean (Nouveau-Brunswick, Canada) en utilisant une combinaison de facteurs météorologiques et climatiques.
- Développement d'une nouvelle approche d'estimation du risque hydrologique en combinant deux distributions probabilistes : la distribution Gamma (Gamma Distribution, GD) et la distribution Paréto Généralisée (Generalized Pareto Distribution, GPD) afin de représenter adéquatement les valeurs moyennes et les extrêmes de débits quotidiens observés.
- Appliquer le modèle probabiliste mixte avec des données hydrométriques observées pour illustrer les avantages du modèle de mélange pour augmenter la taille de la série étudiée et améliorer la précision des estimations.

Contributions scientifiques :

- L'intégration de l'interaction des variables climatiques et météorologiques sous forme de covariables dans la modélisation permet de mieux comprendre le cycle hydrologique dans le site

modélisé et analyser ses différentes composantes afin d'identifier les variables ayant des effets significatifs sur la réponse hydrologique.

- Le modèle développé offre un moyen d'estimation des débits de conception en tenant compte des variations climatiques et météorologiques spécifiques au site jaugé. Il améliore les approches non stationnaires existantes en intégrant plus d'une covariable explicative en lien avec l'événement extrême étudié.
- Identifier les limites du modèles GEV-B-Splines à plusieurs covariables, basé sur l'approche du Bloc maximum qui considère un seul maximum annuel et élimine parfois des événements d'intérêt jouant un rôle important dans la modélisation hydrologique des crues en régime nival.
- Proposer une alternative au modèle GEV-B-Splines limité à la base de données historiques et développer le modèle de mélange qui combine la distribution GPD et Gamma permettant de choisir tous les événements indépendants observés.
- Rédaction de deux articles scientifiques. Le premier article, sur le modèle GEV-B-Splines à plusieurs covariables, à soumettre dans le *Journal of Stochastic Environmental Research and Risk Assessment* (SERRA) le 18/03/2024. Le deuxième article scientifique, illustré dans le développement du modèle de mélange, est publié dans le *Journal of Hydrologic Engineering* le 14/02/2023.
- Contribution scientifique et transfert de connaissances via une conférence scientifique présentée lors du 86^e congrès de l'ACFAS qui a eu lieu le 10 mai 2018, intitulée Modélisation spatiale des crues en régime nival au Nouveau-Brunswick.

CONTRIBUTION DES AUTEURS

Cette thèse de doctorat portait sur la modélisation des crues en régime nival à l'est du Canada. Grâce à cette thèse, j'ai eu l'occasion d'enrichir mes connaissances scientifiques sur les différents modèles déterministes et statistiques qui peuvent être utilisés dans la modélisation des débits extrêmes et les avantages et les limites de chaque modèle étudié. Ces limites ont permis d'établir les bases du raffinement proposé dans le développement de nouveaux modèles probabilistes et statistiques dans la modélisation des débits extrêmes, surtout en hydraulique fluviale. Dans le cadre de cette recherche, j'ai développé un modèle GEV-B-Splines à plusieurs covariables sous le langage informatique Matlab et j'ai eu le support de mon codirecteur M. Salaheddine El Adlouni dans l'algorithme, la programmation ainsi que les tests préliminaires et l'amélioration de la présentation des résultats.

Le modèle GEV-B-Splines a permis l'intégration des variables climatiques et météorologiques pour estimer les débits extrêmes en régime nival. Pour ce faire, j'ai développé mes connaissances en climatologie et en météorologie grâce à des lectures approfondies et à l'encadrement de mon directeur Philippe Gachon. Les données climatiques étaient développées et analysées par le centre de recherche ESCER alors que les données météorologiques étaient téléchargées à partir d'Environnement Canada. Les indices météorologiques et climatiques étaient développés suite à des lectures, des tests préliminaires, des analyses des résultats et des discussions avec mes directeurs. Mon objectif n'était pas de réutiliser les indices déjà développés dans la littérature, mais de trouver d'autres indices et de nouvelles combinaisons qui expliquent les inondations en régime nival. Ce processus a pris beaucoup de temps parce que j'analysais les événements historiques événement par événement et les indices qui étaient développés étaient spécifiques à l'événement en question. Durant cette étape, l'analyse méthodique des données météorologiques et climatiques m'a permis de choisir les combinaisons de covariables. Les commentaires constructifs de mes directeurs ont encore amélioré le choix et la définition des covariables. Ce travail a amené à la rédaction d'un premier article à soumettre dans le *Journal of Stochastic Environmental Research and Risk Assessment* (SERRA), développant une nouvelle méthode pour l'estimation des crues en régime nival en intégrant plus qu'une covariable explicative afin d'analyser l'effet des conditions météorologiques et climatiques sur les inondations.

Dans une deuxième partie de cette thèse, j'ai contribué au développement d'une nouvelle méthode d'estimation du risque hydrologique en combinant deux distributions probabilistes : la distribution Gamma et la distribution Pareto généralisée. Cette dernière approche permet d'augmenter la taille de la série des données et par conséquent améliorer la précision des estimations. Deux co-auteurs sur les deux articles sont mes directeurs de recherche : M. Philippe Gachon et M. Salaheddine El Adlouni. M. Gachon a contribué dans les commentaires sur le contenu scientifique et la manière d'analyser les résultats alors que M. El Adlouni a contribué au développement de l'algorithme, les codes Matlab ainsi que les commentaires et les corrections. L'un des articles de la thèse a un autre co-auteur : M. Simon Michael Papalexiou : Professeur assistant au département de génie civil à l'Université de Calgary. Ce chercheur a contribué au développement du modèle de mélange, à l'amélioration de la qualité de l'article par les différentes révisions et commentaires.

Le travail de thèse est donc le fruit d'une collaboration scientifique entre trois chercheurs provenant de trois universités différentes et de trois disciplines complémentaires.

INTRODUCTION GÉNÉRALE

Les inondations sont les risques naturels les plus fréquents et répandus à l'échelle mondiale (WMO, 2021). Dans le contexte des changements climatiques, ces aléas hydrométéorologiques constituent une menace de plus en plus préoccupante pour le bien-être socioéconomique et la santé des populations à travers le monde, en particulier dans les pays en développement (Cassar et al., 2017; IPCC, 2022). Selon Barredo et al. (2012), l'effet des inondations peut être direct comme les préjudices à la santé et à la sécurité des populations (cf. WMO, 2021), la destruction des ponts, des systèmes de fourniture d'énergie en particulier électrique et des systèmes de drainage, ou indirect comme la perte de la productivité et des opportunités d'affaire ou les conséquences sur la santé mentale des sinistrés (Généreux et al., 2022). Ces effets combinés ne font qu'amplifier les conséquences des inondations, surtout dans les zones urbaines où la population et les infrastructures sont particulièrement exposées, voire vulnérables (Cea et Costabile, 2022).

Au Canada, les inondations sont les catastrophes naturelles les plus coûteuses et les plus fréquentes (Oubennaceur et al., 2022). Depuis les années 1970, les aides fédérales transférées aux provinces dans le cadre des Accords d'aide financière en cas de catastrophes (AAFC) pour les inondations couvrent un total de 4,9 milliards de dollars, soit environ deux tiers des dépenses totales des AAFC (BAC, 2019). Selon la même source, ces coûts ont presque quadruplé en 40 ans, passant de 300 millions de dollars dans les années 1970 à 1,2 milliard de dollars dans les années 2000 pour atteindre 3,7 milliards de dollars les quatre premières années de cette décennie. Ces statistiques confirment l'ampleur des dégâts associés aux inondations canadiennes touchant l'infrastructure, l'environnement, l'économie et les coûts de recouvrement engendré, selon l'étude de Buttle et al. (2016). Dans ce contexte, il est donc important de s'intéresser à ces événements de plus en plus fréquents et destructeurs en termes d'impacts économiques (WMO, 2021; IPCC, 2022). Oubennaceur et al. (2022) ont classé les inondations en quatre types à savoir : les inondations fluviales y compris les inondations par embâcles, les inondations pluviales (ruissellement de surface vers une rivière), les inondations côtières (ondes de tempête et provenant de marée côtière) et enfin les inondations induites par l'activité humaine.

Dans l'est du Canada, soit du Québec jusqu'à la Nouvelle-Écosse, incluant l'Île-du-Prince-Édouard, le régime nival ou nivo-pluvial des principaux cours d'eau (cf. Burn et Whitfield, 2017) est affecté par l'apport d'eau de fonte via les quantités (parfois importantes) de neige accumulées durant l'hiver et par le type, la

durée, l'intensité et l'occurrence des événements de pluie qui surviennent au moment du dégel, ainsi que par les conditions de surface et dans le sol (ex. processus de gel-dégel, eau initialement contenue dans le sol, état de la surface, etc.). La combinaison de tous ces facteurs météorologiques et physiographiques a un effet sur la variation des niveaux et des débits d'eau, tant au niveau de l'exutoire d'un bassin versant que plus en amont, avec l'effet majeur des embâcles de glace (Rokaya et al., 2018). Au Nouveau-Brunswick, les inondations majeures apparaissent le plus souvent lors de la période de crues printanières, soit lors de la phase de fonte de la neige ou de la phase de dégel, et l'arrivée de précipitations liquides, avec parfois des pluies abondantes et des embâcles de glace qui peuvent survenir aux printemps (Burrell et Beltos, 2022; Cui et Cui, 2019). Dans ce contexte, il est essentiel afin de bien les anticiper et réduire les conséquences associées à ces aléas, de bien comprendre et d'intégrer l'ensemble des facteurs physiques responsables de l'occurrence, de la durée et de l'intensité des inondations.

À la suite du réchauffement printanier, observé surtout ces dernières décennies et son effet sur la fonte rapide de la neige et les différents processus hydrologiques au sein de la forêt boréale de l'est du Canada, plusieurs recherches récentes ont permis d'établir les principales causes naturelles, reliées aux changements dans les régimes de températures et de précipitation, sur l'occurrence des inondations (cf. Benoit et al., 2022; Buffin-Bélanger et al., 2022; Bourgault et al., 2022). Par exemple, Assani (2022) a analysé les impacts des chutes de neige sur la variabilité spatio-temporelle des débits annuels moyens de 17 rivières dans le sud du Québec afin d'étudier le lien entre l'écoulement fluvial et la fonte de la neige. Cette étude a montré que la neige est fortement corrélée avec les débits annuels moyens, à l'aide de données observées qui couvrent la période 1930-2019. En définitive, les inondations printanières dépendent donc de facteurs multiples et complexes, nommément (ceux d'origine: météorologique, physiographique et/ou géographique, environnementale et humaine). La combinaison de certains facteurs complexes comme les ouragans, les ondes de tempête, les précipitations et la fonte rapide des neiges sont également responsables des inondations (Nasr et al., 2021). Les changements et les variations des caractéristiques des inondations sont donc complexes à appréhender, et nécessitent une approche multidimensionnelle (Burn and Whitfield, 2017), et l'identification et l'évaluation des liens d'interdépendance entre différents facteurs naturels et anthropiques, comme l'effet des terres cultivées sur les caractéristiques de la réponse hydrologique (Zaffaroni, 2023), liens qui sont encore mal compris et/ou modélisés.

Alors que les inondations pourraient augmenter en fréquence ou seront de plus en plus sévères et dommageables en raison des changements climatiques et socioéconomiques à venir (IPCC, 2022), il devient de plus en plus important d'améliorer les modèles existants afin de favoriser une meilleure prévention et développer les stratégies d'adaptation appropriées vis-à-vis des risques d'inondations, en raison des dommages et des conséquences croissantes sur le plan humain, économique et social. Pour cela, il est impératif de bien identifier les facteurs responsables des inondations et d'améliorer la modélisation des crues en régime nival, afin de mieux évaluer les changements potentiels futurs dans les risques d'inondations printanières au pays.

Dans ce contexte, cette thèse de doctorat a pour objectif de proposer de nouvelles méthodes probabilistes pour améliorer l'estimation des crues en régime nival à l'est du Canada. Cette amélioration est basée sur une meilleure compréhension des divers facteurs qui affectent un régime nival, et qui peuvent avoir un effet direct ou indirect sur la variation des débits extrêmes.

Le chapitre I présente un état des lieux des connaissances sur les mécanismes connus impliqués dans l'occurrence des inondations printanières, notamment dans l'est du Canada, soit le bassin versant du fleuve Saint-Jean (situé pour l'essentiel du Nouveau-Brunswick) qui constituera notre région d'étude. La problématique du risque d'inondation dans ce bassin versant est en particulier décrite, ainsi que les différentes méthodes existantes de modélisation des crues, incluant les avantages et les limites de chacune d'entre elles et les besoins d'amélioration des outils probabilistes ou statistiques disponibles. Ce chapitre se termine par l'identification des objectifs de cette étude, et des hypothèses utilisées pour réaliser ce travail.

Dans le chapitre II, un modèle de prévision des crues printanières est présenté, en tenant compte des effets de différentes covariables météorologiques et indices climatiques. Il est présenté sous forme d'article à soumettre à la revue *Stochastic Environmental Research and Risk Assessment (SERRA)*. Ce modèle a été appliqué dans le bassin versant du fleuve Saint-Jean.

Dans le chapitre III, un autre modèle de mélange est présenté, celui-ci combine deux approches en comblant les limites du premier modèle proposé afin d'améliorer l'estimation des débits extrêmes en régime nival. Ce chapitre est également sous forme d'un article qui a été publié en 2022 dans la revue *Journal of Hydrologic Engineering*.

Dans le chapitre IV, une discussion générale est présentée en mettant l'accent sur les développements scientifiques réalisés dans le cadre de cette thèse de doctorat et les problèmes résolus ainsi que les limites de chaque approche.

Une conclusion générale est présentée à la fin de cette thèse et synthétise les principaux résultats de notre étude, ainsi qu'une mise en perspective des principaux résultats et des travaux à poursuivre.

CHAPITRE 1

États des connaissances, problématique et objectifs

Ce chapitre présente l'état actuel des connaissances sur les inondations, notamment dans l'est du Canada, incluant les outils de modélisation disponibles et les avantages et limites de ceux-ci ainsi que la problématique de recherche, en particulier dans le bassin versant du fleuve Saint-Jean qui constitue notre région d'étude. Il termine par la présentation des objectifs et des hypothèses de recherche.

1.1 Mise en contexte

À l'échelle mondiale et selon la base internationale de données sur les désastres (EM-DAT) du Centre de recherche sur l'épidémiologie des catastrophes (CRED) enregistrée durant la période 1970-2019, 50% de toutes les catastrophes naturelles (y compris celles d'origine technologique), 45% des décès déclarés et 74% des pertes économiques signalées (soit 2,06 millions de morts et 3,6 mille milliards de dollars US, respectivement), sont liés aux risques d'origine météorologique, climatique et hydrologique (WMO, 2021).

Parmi les catastrophes naturelles enregistrées durant cette période (1970-2019), les inondations représentent 44% des 11 072 événements signalés (24 % sont des inondations fluviales et 14% sont des inondations générales) et 31 % des 3,6 mille milliards de dollars US de pertes économiques enregistrées durant cette période (20% causées par les crues fluviales, 8% par les crues générales et 3% par les crues soudaines; WMO, 2021).

Au Canada, les inondations sont devenues une priorité importante, aux échelles fédérale, provinciale et municipale (Golnaraghi et al., 2020-a). En effet, selon les statistiques récentes, les coûts annuels de recouvrement fédéral dans le cadre des AAFC, entre 2017 et 2022, devraient atteindre en moyenne 902 millions \$CAD, dont 673 millions \$CAD sont attribuables aux inondations (Golnaraghi et al., 2020-a), soit environ 75% du total. Parmi les différents événements historiques d'inondations, on distingue celles printanières enregistrées en avril 2019 touchant l'est du Canada, à la suite de fortes précipitations liquides durant la période de fonte de la neige (Sécurité Publique Canada, 2022). Les inondations les plus graves, quant au nombre de personnes affectées, ont eu lieu au Québec, où plus de 10 000 personnes ont été évacuées où un seul décès a été enregistré à Pontiac (QC). Ce type d'inondations majeures se produit parfois au printemps au Québec (plus ou moins fréquemment selon le cas, cf. Burn et Whitfield, 2016; Zadeh et al., 2020; Burn et Whitfield, 2023), notamment dans les municipalités de l'Estrie, à savoir les villes

de Sherbrooke, Bromptonville, Richmond et Windsor avec leur lot d'inconvénients et de dommages associés (Saint-Laurent and Hähni, 2008). Ce fut également le cas dans la vallée du Richelieu au printemps 2011 où une inondation majeure a duré plus de six semaines et a engendré des dommages importants en terme de destruction de propriétés et d'impacts négatifs sur l'aquaculture dans son bassin versant coûtant environ 90 millions de dollars (International Joint Commission (IJC), 2013; Saad et al., 2016). De plus, en avril et mai 2017, le sud du Québec a connu des inondations printanières majeures, notamment pour la plupart des tributaires du fleuve Saint-Laurent près de Montréal et la rivière des Outaouais qui ont nécessité l'évacuation obligatoire dans les municipalités québécoises de : Gatineau, Pontiac, Rigaud et Montréal et les secteurs environnants (Sécurité publique Canada, 2022). Ces inondations de l'Est canadien ont également touché la province du Nouveau-Brunswick. En 2018 et 2019, les inondations, dans cette province, ont causé des dégâts estimés à 110 millions de dollars canadiens, correspondant à plus de 12 000 propriétés endommagées, sans tenir compte des conséquences psychosociales des sinistrés (Duhamel et al., 2022). Ces inondations ont également causé la fermeture de la route transcanadienne entre Fredericton et Moncton pendant plusieurs jours, en particulier lors de l'inondation de mai 2018 (Lau, 2020; Henry et al., 2020). Cet événement est associé aux débordements des rivières du bassin versant du fleuve Saint-Jean et a entraîné l'évacuation de 1 400 personnes (Sécurité publique Canada, 2022).

Au sein du bassin du fleuve Saint-Jean, le régime des principaux cours d'eau est de type nival ou nivopluvial (Golnaraghi et al., 2020-b). Ce régime hydrologique est caractéristique de la plupart des rivières de l'est du Canada, avec des niveaux d'eau et des débits maximums au printemps alimentés, par la fonte des neiges, l'arrivée de précipitations liquides durant la phase du dégel, ainsi que de bas niveaux d'eau en été et en hiver, notamment durant la période froide ou de gel de l'eau. Dans le contexte du réchauffement climatique, plusieurs aspects du cycle hydrologique seront affectés notamment via l'augmentation de l'évaporation, la modification des régimes (type de précipitation) et des intensités de précipitations, les processus impliqués dans le stockage de surface de l'eau, y compris la génération et la fonte de la neige, la débâcle des glaces et la fonte des glaciers (Whitfield, 2012).

Dans les années à venir, les catastrophes naturelles liées au climat et leurs pertes économiques associées devraient augmenter (IBC, 2018). De plus, les inondations extrêmes, en raison des effets des changements climatiques, seront à la hausse au cours du 21^e siècle (Guo et al., 2023). Toutefois, la complexité de ce phénomène d'inondation printanière, affecté par la combinaison des différentes variables, nécessite de bien identifier et comprendre les mécanismes impliqués à la fois via des facteurs de grande échelle (ex.,

téléconnexion et variabilité de la circulation atmosphérique et océanique) et d'échelle régionale, notamment selon les conditions de surface du bassin versant considéré, en particulier dans l'est du Canada (Whitfield, 2012). Par exemple, dans cette région (l'est du Canada), la fréquence et l'intensité des dépressions extratropicales, qui affectent l'occurrence, la durée, l'intensité et le type de précipitations, sont les plus importantes au pays (ex., Poan et al., 2018). Lawford et al. (1995) ont étudié les conditions météorologiques responsables des trois principaux types d'inondations au Canada à savoir : les inondations causées par les pluies abondantes et le ruissellement des eaux pluviales, les inondations urbaines qui constituent un cas particulier d'inondations reliées le plus souvent à des systèmes orageux ou des pluies soudaines et intenses (ponctuelles) et peuvent être exacerbées par l'extrême imperméabilité de la surface de la ville le plus souvent asphaltée ou bétonnée, et enfin les inondations causées par la fonte de la neige, comme le cas du bassin versant du fleuve Saint-Jean. Les inondations dues à la fonte de la neige sont communes à travers le Canada, et la modélisation de celles-ci nécessite donc la prise en compte explicite des caractéristiques du couvert nival. Par exemple, la fonte des neiges tend à être plus graduelle dans les zones boisées, centrales et orientales du Canada (Lawford et al., 1995), ce qui modifie l'apport en eau issu du couvert de neige lors de la phase de dégel printanier. En général, le régime hydrologique nival de l'est du Canada est marqué par une saisonnalité particulière, fonction de la quantité de neige accumulée durant la saison automnale et hivernale (Octobre-Mai). Selon l'ampleur de la fonte de neige et des précipitations liquides, certaines crues printanières peuvent entraîner des inondations majeures. Dans ce contexte, la conception des aménagements des cours d'eau, le dimensionnement des ouvrages de rétention et d'évacuation ou de drainage des eaux de ruissellement en zones urbaines (Aubert, 2012) nécessite de bien évaluer ou anticiper les inondations printanières à l'échelle régionale, voire locale. Les différents types d'inondations sont en général prévisibles selon la durée et l'intensité des événements de précipitations (Lawford et al., 1995; St-Hilaire et al., 2003).

1.2 Mécanismes physiques impliqués : occurrence et caractéristiques des aléas hydrométéorologiques

Bien que la circulation continentale de l'eau soit en premier lieu influencée par le climat, les caractéristiques physiographiques et l'aménagement du territoire ont aussi un effet significatif sur le régime hydrologique (Anctil et al., 2012; Bourgault et al., 2022; Yu et Bourque, 2022). De plus, plusieurs mécanismes de grande échelle, qui affectent les conditions atmosphériques et océaniques, sont également à l'origine de fluctuations dans les régimes de température et de précipitations (cf. IPCC, 2022), et des caractéristiques des inondations (Kundzewicz et al., 2019).

Dans cette section, une étude des mécanismes et des facteurs impliqués dans l'occurrence, la durée ou l'intensité des inondations printanières sera réalisée, aussi bien via des processus de grande échelle (au niveau de l'atmosphère) que ceux d'échelle locale (au niveau du bassin versant et ses caractéristiques hydrométéorologiques). Bien que, l'intensité, la durée, la fréquence et l'occurrence de l'événement d'inondation ne feront pas l'objet d'une étude détaillée dans cette thèse, ni seront intégrées dans le modèle, certains facteurs précurseurs des inondations printanières seront identifiés et évalués dans le détail.

1.2.1 Facteurs de grande échelle

Selon Esfandiari et al. (2020), Mazouz et al. (2012), Villarini et al. (2011) et Turcotte et al. (2010), l'intensité des crues printanières au nord-est de l'Amérique du Nord est principalement causée par l'accumulation de la neige en hiver, le taux de fonte du couvert nival ainsi que l'intensité des précipitations liquides au printemps. Comme mentionné précédemment, Burn et Whitfield (2017) ont montré l'importance du rôle combiné de la neige (quantité et fonte) et de la pluie importante dans les crues en régime nival ou nivopluvial au Canada et aux États-Unis. Cette étude a également permis d'explorer les relations entre les régimes d'inondations et quelques indices de téléconnexions, à savoir : l'oscillation décennale du Pacifique (Pacific Decadal Oscillation, PDO), l'oscillation multidécaennale de l'Atlantique (Atlantic Multidecadal Oscillation, AMO) et l'indice d'oscillation australe (Southern Oscillation Index, SOI) à partir des anomalies annuelles de ces indices (cf. Maurer et al., 2004). Shabbar et al. (1997) a développé le Baffin-Ouest Atlantique (Baffin Island-West Atlantic, BWA) qui explique, mieux que l'indice (North Atlantic Oscillation, NAO), la variabilité de la température dans le nord-est de l'Amérique du Nord où se situe la zone d'étude considérée dans cette thèse. Dans cette section, plusieurs indices de téléconnexion sont présentés ainsi que les liens connus avec les caractéristiques des inondations ou des régimes de températures et de précipitations. Parmi les indices étudiés, on a retenu ceux ayant une corrélation significative avec les événements extrêmes enregistrés au bassin de l'étude, à savoir la NAO, l'AMO et le BWA.

1.2.1.1 L'oscillation nord-atlantique (NAO)

L'oscillation nord-atlantique (North Atlantic Oscillation, NAO, Hurrell, 1995) est un mode de variabilité climatique à grande échelle qui affecte la circulation atmosphérique dans l'Atlantique Nord. Les phases positives et négatives de l'indice NAO sont associées à des changements dans l'emplacement et l'intensité du courant-jet et de la trajectoire des tempêtes dans l'Atlantique Nord (NCEI, 2018; Greene, 2012). Elles sont également associées à des fluctuations à grande échelle des trajectoires de transport zonal et

méridien de chaleur et d'humidité (Greene, 2012). Par conséquent, ces phases de la NAO sont également associées à des anomalies de température et de précipitations à l'est de l'Amérique du Nord, y compris l'est du Canada, et en Europe occidentale et centrale (Nolin et al., 2021; Cassou, 2004). Pour cette raison, l'effet de la NAO sur l'occurrence et les caractéristiques des crues printanières peut être non négligeable ou doit être considéré, notamment sur la sévérité des inondations (ex., Santos et al., 2018), compte tenu de l'effet potentiel de celle-ci sur le régime des précipitations responsable des inondations à l'est du Canada (Bourgault et al., 2022; Benoit et al., 2022). L'indice NAO est donc un des facteurs intégrés dans le modèle d'estimation des débits printaniers développé dans cette étude (cf. chapitre 2).

1.2.1.2 L'Atlantic Multidecadal Oscillation (AMO)

L'Atlantic Multidecadal Oscillation (AMO) est une mesure de la variation de la température de surface océanique (TSO) dans l'océan Atlantique Nord, qui influence le climat des masses continentales adjacentes (O'Reilly et al., 2016). D'après Bonsal et Shabbar (2011), depuis la fin des années 90, l'AMO est dans une phase positive, correspondant donc à un réchauffement prononcé de l'Atlantique Nord, potentiellement relié à des événements de blocages dans cette région, ce qui peut engendrer des anomalies significatives de températures et de précipitations dans l'est du Canada (Wazneh et al., 2021), potentiellement responsables des inondations (Häkkinen et al., 2011).

L'étude de Burn et Whitfield (2017) a d'ailleurs montré un lien significatif entre les différentes phases de l'AMO et les quantiles d'inondation dans l'est de l'Amérique du Nord. Celle-ci a été confirmée par Hodgkins et al. (2017) qui ont établi des liens significatifs entre l'occurrence des inondations majeures et l'AMO, dans divers bassins versants de l'Amérique du Nord et de l'Europe. L'effet de l'AMO sur les quantiles de débits printaniers sera donc inclus dans le modèle de prévision détaillé dans le chapitre 2.

1.2.1.3 Le Baffin Island-West Atlantic (BWA) Index

L'indice de l'île de Baffin-Ouest Atlantique (Baffin Island-West Atlantic, BWA; Shabbar et al., 1997) est un indice qui mesure la différence d'anomalie de géopotential à 500 hPa entre l'île de Baffin (65°N - 60°W) et l'est de l'Atlantique Nord (30°N - 60°W; Shabbar et al., 1997). Shabbar et al., (1997) ont suggéré que le BWA explique mieux la forte variabilité de température dans le nord-est de l'Amérique du Nord que la NAO. Coulibaly et al. (2000) ont également conclu qu'une meilleure estimation du ruissellement annuel dans le nord-est du Canada est obtenue en intégrant le BWA.

Selon Wazneh et al. (2021), une relation significative de cause à effet est présente entre les indices BWA, NAO et certaines caractéristiques des blocages atmosphériques dans l'Atlantique Nord. Dans ce contexte, l'indice BWA sera également inclus dans notre modèle de prévision des inondations (cf. chapitre 2).

1.2.2 Facteurs d'échelle locale

Une bonne estimation des inondations repose sur la connaissance des conditions hydrologiques du bassin versant qui précèdent l'événement, et les types ou les conditions physiographiques de surface au sein de ce bassin (Nied et al., 2017). Par exemple, Assani (2023) a identifié une corrélation positive entre les débits, le drainage et les chutes de neige, et une corrélation négative avec la superficie des milieux humides et la température printanière maximale quotidienne. Yousfi et El-Adlouni (2016) ont également démontré l'effet des précipitations totales sur les maximums des débits annuels dans la province du Nouveau-Brunswick. D'autres paramètres météorologiques, à l'échelle locale, doivent être pris en considération de façon combinée soit la température, la fonte de la neige et le contenu en eau de la neige (Benoit et al., 2022).

Par ailleurs, d'autres paramètres physiographiques à l'échelle du bassin versant doivent également être intégrés dans l'estimation des crues extrêmes, et particulièrement des crues printanières. Par exemple, la forme géométrique du bassin versant joue un rôle important dans la réponse hydrologique suite à une inondation : un bassin circulaire n'a pas la même réponse qu'un bassin allongé ou carré (Douvinet et al., 2008). De plus, les caractéristiques physiographiques du bassin versant avant les inondations sont des facteurs très importants à considérer (un bassin urbain ou rural, couvert ou non en neige, pente faible ou forte).

Dans notre étude, quelques paramètres météorologiques seront étudiés dans le modèle de crues, à savoir : la température, les précipitations (liquide et solide) et quelques indices dérivés (comme le degré de gel-dégel basé sur les températures maximales et minimales quotidiennes). La forme géométrique ne sera pas intégrée dans le modèle, mais sa pertinence sera évaluée lors de l'interprétation des résultats.

1.2.3 Autres facteurs

Certaines activités humaines comme les prélèvements en eau, la navigation, l'occupation des sols par la pratique agricole et certains travaux d'ingénieries reliés aux ressources en eau, comme l'irrigation et la construction des barrages, peuvent affecter le débit des cours d'eau (Blöschl et al., 2007; Di Baldassarre et al., 2013; Jiang et al., 2017; McMillan et al., 2016; Merz et al., 2012; Montanari et al., 2017 Vogel et al.,

2015; Yan et al., 2016). Tout ceci affecte non seulement le système hydrographique au niveau d'un bassin versant, tout comme la forme du cours d'eau principal et ses affluents et les caractéristiques du ruissellement, mais peut aussi avoir des effets majeurs sur les écosystèmes forestiers, la composition et le fonctionnement écologique (Huntington et al., 2009), voire même sur les inondations dans le bassin du fleuve Saint-Jean selon le niveau de coupe forestière (Yu et Bourque, 2022).

Même si ces activités humaines affectent le débit des cours d'eau, elles sont difficilement quantifiables à l'échelle locale (Oubennaceur et al., 2022). La prédominance des autres facteurs (climatiques et météorologiques) plus facilement quantifiable est en effet avérée dans de nombreuses études, et ce sont ces derniers qui seront donc privilégiés dans notre évaluation des crues. L'intervention humaine ne sera pas prise en compte dans notre étude.

1.3 Outils de modélisation disponibles pour la simulation et la prévision des crues

Dans la littérature, plusieurs auteurs ont proposé une classification des modèles hydrologiques (Ambroise, 1999; Padmanabhan et al., 1995; Refsgaard et al., 2010). Par exemple, Storm and Refsgaard (1990) ont classé les modèles hydrologiques en trois catégories à savoir : les modèles empiriques, les modèles conceptuels et les modèles distribués à base physique. Refsgaard (1997) a classé les modèles hydrologiques, selon la description du processus physique, en deux classes à savoir : les modèles conceptuels et les modèles physiques. Il a également classé les modèles selon la distribution spatiale des processus au niveau d'un bassin versant et il a distingué, dans ce cas, les modèles globaux et les modèles distribués.

Dans cette revue, et en se basant sur l'étude de Kumar et al. (2023), les modèles disponibles étaient classés en trois classes, selon l'utilisation en hydrologie. On distingue, par conséquent, : 1) les modèles numériques (Kumar et al., 2023; Aneed et al., 2016), 2) les modèles Pluie-Débit qui peuvent également être classés en sous-modèles conceptuels, physiques ou empiriques (cf. Kumar et al., 2023), et enfin 3) les modèles événementiels y compris les modèles probabilistes dont certains seront développés dans cette recherche doctorale.

1.3.1 Modèles numériques

Les processus hydrologiques varient à l'échelle spatio-temporelle en fonction des conditions météorologiques et physiographiques du bassin versant (Fortin et al., 1995). Selon Fortin et Turcotte (2006), la modélisation hydrologique numérique s'intéresse donc au cheminement des précipitations liquides et solides une fois qu'elles touchent le sol et jusqu'à ce que le volume d'eau ruisselé quitte la zone modélisée. Par conséquent, ces modèles simulent le comportement de l'eau au sein d'un bassin versant (Anees et al., 2016). Cette modélisation numérique permet ainsi de déterminer la réponse d'un bassin versant, soit une crue, à la suite d'une pluie (Anctil et al., 2012).

Selon la revue de Kumar et al. (2023), ces modèles décrivent le flux hydraulique d'un cours d'eau, en considérant les caractéristiques hydrométéorologiques (précipitations et ruissellement) et géophysiques (géométrie du lit du cours d'eau et sa rugosité). Parmi ces modèles, le modèle HEC-RAS (1D et 2D) a été créé par le corps des ingénieurs de l'armée américaine (Khattak et al., 2016). Les avantages et les inconvénients des modèles numériques varient d'un modèle à un autre. Par exemple, HEC-RAS est l'un des modèles les plus utilisés et reconnus en hydraulique fluviale, notamment en ingénierie communautaire. Il a la capacité de simuler aussi bien les écoulements stationnaires que non-stationnaires (Kumar et al., 2023). Malgré les avantages de HEC-RAS, ce modèle reste limité dans la représentation des géométries et des limites de bassins complexes. De plus, sa capacité est limitée dans la modélisation des interactions entre l'eau et l'environnement, comme par exemple le transport des sédiments (Kumar et al., 2023).

La modélisation hydrologique consiste donc à simuler et analyser le débit d'eau dans un cours d'eau ou dans une rivière à l'aide d'outils mathématiques et informatiques (Herrera et al., 2022). Ces modèles nécessitent des données physiographiques (pente, occupation du sol, infiltration, coefficient de ruissellement) et météorologiques (pluies historiques) du bassin versant. Les extrants des modèles hydrologiques, nécessaires à la simulation des débits de crue, sont utilisés dans la modélisation hydraulique (comportement de chaque section du cours d'eau suite à une crue) pour simuler par exemple, le comportement de l'eau lors d'une inondation (Munir et al., 2019), incluant les vitesses et les niveaux d'eau dans les différentes sections d'un cours d'eau. Ces modèles (hydrologiques et hydrauliques) sont complexes et nécessitent des connaissances et des compétences spécialisées pour être développés et bien interprétés (Maskrey et al., 2022). Ceci présente un obstacle d'utilisation surtout pour les non spécialistes (Kumar et al., 2023). Bien que ces modèles numériques ont la capacité de simuler les écoulements

stationnaires et non stationnaires et qu'ils sont reconnus en hydraulique fluviale et en ingénierie (surtout HECRAS 1D et 2D; Kumar et al., 2023), ces modèles présentent toujours des limites dans leur utilisation. Par exemple, l'ensemble des caractéristiques du bassin versant et les processus physiques complexes ne sont pas intégrés dans ce genre de modèles (Anees et al., 2016).

1.3.2 Modèles Pluie-Débit

Autres que les modèles numériques, on distingue les modèles Pluie-Débit qui sont des modèles hydrologiques pour simuler la transformation de pluie en ruissellement dans un bassin versant (Moradkhani et Sorooshian, 2008). Ces modèles peuvent être classés en trois catégories, à savoir : les modèles conceptuels, les modèles physiques et les modèles empiriques (d'après Kumar et al., 2023).

1.3.2.1 Modèles conceptuels :

Les modèles conceptuels sont des modèles utilisés pour représenter un cycle hydrologique simplifié (Kumar et al., 2023). Ils sont faciles à utiliser, ne demandent pas beaucoup de paramètres et permettent de comprendre le comportement hydrologique au sein d'un petit ou moyen bassin versant. Cependant, ces modèles, ne permettent pas de représenter avec précision le processus physique impliqué dans la génération du ruissellement, ainsi que l'effet des changements dans l'occupation du sol et du climat. Par contre, ces modèles peuvent ne pas représenter avec précision les processus physiques impliqués dans la génération du ruissellement, et ont une capacité limitée à simuler les effets des changements socio-environnementaux ou dans l'aménagement du territoire (Bai et al., 2015; McGrath et al., 2018).

1.3.2.2 Modèles physiques

Les modèles physiques sont également des modèles Pluie-Débit comme le modèle HEC-HMS (Anderson et al., 2002) qui représente le processus physique dans le calcul du ruissellement. Les limites de ce modèle résident dans la nécessité d'obtenir des données détaillées pour représenter des processus complexes, comme la saturation initiale du sol et le taux d'infiltration. Par conséquent, le temps de simulation est très élevé et l'utilisation de ces méthodes reste limitée (Kumar et al., 2023). Malgré sa capacité à représenter le processus physique dans la génération du ruissellement dans les grands bassins, ce type de modèle nécessite des données détaillées et des ressources informatiques qui peuvent être complexes et nécessitent un temps de simulation élevé (Bai et al., 2019; Montanari et Koutsoyiannis., 2019).

1.3.2.3 Modèles empiriques

Les modèles empiriques sont des modèles qui représentent la relation statistique entre les apports en précipitations et le débit de ruissellement observés (Kumar et al., 2023). Ces modèles sont simples et efficaces et ne nécessitent que des données historiques de précipitations et de débits afin de simuler les hydrogrammes de crues. Cependant, ces modèles ne permettent pas de représenter le processus physique responsable de la génération du ruissellement et d'intégrer l'effet du changement climatique ou de l'occupation du sol en dehors des données historiques utilisées pour calibrer le modèle (Kumar et al., 2023; Triguero et al., 2015).

1.3.3 Modèles événementiels

La troisième catégorie de modèles hydrologiques disponibles correspond aux modèles événementiels. Ces modèles n'ont pas besoin de décrire le cheminement de l'événement pour le modéliser, comme dans le cas des modèles physiques (Kumar et al., 2023). Ces modèles sont probabilistes ou statistiques et visent à définir la distribution de probabilité des crues extrêmes sur les sites étudiés, et donc à associer une période de retour aux valeurs de débits extrêmes (Brigode et al., 2015). En général, les approches probabilistes utilisées dans la modélisation hydrologique exigent moins de temps de calcul et moins de données physiographiques que les modèles déterministes.

L'Analyse Fréquentielle Hydrologique (AFH) est l'un des outils statistiques les plus utilisés pour déterminer la probabilité d'occurrence d'un événement d'intérêt dans le futur, en se basant sur la distribution des événements passés (Li et al., 2019). Elle repose sur la détermination de la loi de probabilité des événements extrêmes (Knebl et al., 2005; Pappenberger et al., 2005). Pour décrire le comportement des extrêmes, différentes distributions ou lois de probabilités disponibles sont utilisées (Bracken et al., 2018).

Dans cette section, la théorie des valeurs extrêmes sera présentée brièvement ainsi que ses deux principales distributions probabilistes : la distribution GEV (*Generalized Extreme Value*) et la distribution GPD (*Generalized Pareto Distribution*). Ces distributions sont basées sur l'analyse de la fréquence potentielle (probabiliste) des événements, contrairement aux modèles déterministes qui sont basés sur la reproduction du processus physique dans le temps à l'aide d'une simulation explicite des séries temporelles de débits ou de niveaux d'eau.

1.4 Théorie des valeurs extrêmes (TVE)

La théorie des valeurs extrêmes est une branche de la statistique qui étudie le comportement asymptotique d'une distribution de probabilité d'un échantillon de données, pour estimer l'occurrence des minimums ou des maximums (Roberts, 1999). Selon cette théorie, développée par Fisher et Tippet (1928) et formalisée par Gnedenko (1943), l'approche du *Block Maxima* (BM) converge asymptotiquement vers la distribution des valeurs extrêmes généralisées (GEV; Drissia et al., 2019; AghaKouchak et Nasrollahi, 2010; Coles, 2001; Shinyie et al., 2013; Vasiliades et al., 2015) alors que l'approche du *Peaks Over Threshold* (POT) est basée sur la convergence asymptotique des dépassements du seuil vers la distribution GPD (Ammar et al., 2020). Les deux approches sont brièvement présentées dans les sous-sections suivantes afin d'identifier la différence, l'utilité et les limites de chacune.

1.4.1 Principales distributions

1.4.1.1 Distribution GEV

La distribution GEV a été proposée par Jenkinson (1955), en se basant sur la TVE (Fisher et Tippet, 1928). Cette distribution combine trois distributions connues, à savoir : Gumbel (I), Fréchet (II) et Weibull (III). Le développement théorique de cette distribution est détaillé dans le chapitre 2, et le chapitre 3. La distribution GEV est la distribution limite des BM obtenus à partir d'une série temporelle quotidienne d'une variable. Si la valeur maximale est sélectionnée dans un bloc de temps prédéterminé (chaque année, par exemple), la série temporelle résultante suivra une distribution GEV à trois paramètres, à savoir : les paramètres de position, d'échelle et de forme (Westra et al., 2013), ce qui permet de décrire un large éventail de comportements de la queue.

Une distribution stationnaire GEV est une distribution dans laquelle les paramètres de cette distribution sont constants dans le temps (Bracken et al., 2018). Par contre, afin de tenir compte de la non-stationnarité et des effets du changement climatique sur les extrêmes hydroclimatiques (Katz, 2010; Milly et al., 2008), l'inclusion explicite de l'évolution des paramètres dans le temps et des effets des interactions entre la variable à estimer (ex., le débit extrême) et la variable explicative ou covariable (ex., les paramètres météorologiques ou physiographiques d'un bassin versant) est essentielle. Ceci permet de modéliser l'ensemble des caractéristiques des crues ou des débits en général. L'utilisation des modèles probabilistes non stationnaires, basés sur les différentes approches de l'AFH, permet d'incorporer des covariables variant dans le temps aux estimations des paramètres de GEV (Aryal et al., 2009; Cheng et al., 2014; Machado et al., 2015, 2011; Tan and Gan, 2015). En se basant sur ces avantages, cette distribution a été

utilisée dans cette thèse de doctorat. La limite de cette méthode est dans l'élimination de certains événements, potentiellement importants à considérer, mais qui ne représentent pas des maxima annuels de débits (comme les extrêmes de débits enregistré durant l'été ou l'automne). Dans ce contexte, la distribution GPD présentée dans la suite peut comporter certains avantages.

1.4.1.2 Distribution GPD

La distribution GPD est un modèle de distribution probabiliste à deux paramètres, introduits par le théorème de Pickands (1975). Ce dernier considère tous les événements qui dépassent un certain seuil, contrairement à l'approche du BM (ayant la GEV comme distribution) qui considère un seul maximum par an. La section suivante détaille les différentes méthodes de choix du seuil, selon la littérature, afin de mettre en valeur ce modèle (GPD) et les défis dans ce choix.

La distribution GPD est la distribution limite de l'approche des dépassements au seuil (POT), et elle est utilisée en hydrologie pour modéliser les seuils de contrôle des inondations (Lawrence, 2020; Lyamuremye et al., 2019; Choulakian and Stephens, 2001). Dans ce cas, on établit une valeur extrême fixe à partir de la série des seuils supérieure ou égale à cette valeur sélectionnée. La limite de cette approche est dans le choix du seuil (Jonathan et al., 2021). Le seuil doit être suffisamment grand pour que les excès (valeurs supérieures au seuil) suivent une distribution GPD, mais un seuil surestimé conduit à une taille d'échantillon réduite et augmente la variance des estimations (Curceac et al., 2020). Par ailleurs, un seuil trop faible augmente la taille de l'échantillon et le biais des estimations, car la distribution empirique s'écarte d'un modèle GPD parfait (Scarrott et MacDonald, 2012). Différentes études ont utilisé ce modèle dans l'estimation des événements extrêmes en hydrologie, comme l'étude de Zhao et al. (2019). Cette distribution a été proposée dans le chapitre 3, et comparée avec la méthode GEV présentée au chapitre 2.

1.4.1.3 Méthodes de sélection du seuil

Le choix du seuil optimal dans la modélisation des extrêmes hydrologiques avec la méthode du POT reste un défi toujours d'actualité (Bücher et Zhou, 2021). Plusieurs problèmes affectent la performance des méthodes de sélection du Seuil (TS), selon Kiran et Srinivas (2020). Plusieurs études comme celles de Caires (2009), Cunnane (1973), Madsen et al. (1997), et Wang (1991) ont montré, à partir des simulations Monte-Carlo et des applications hydrologiques, l'efficacité du choix de données dans l'approche du seuil, comparée à celle du BM. Cette comparaison concernait l'utilisation de l'information enregistrée afin

notamment de réduire la variabilité dans l'estimation du quantile. Malgré les avantages de cette méthode, elle est moins utilisée que la méthode du BM, en raison de l'absence d'une méthode systématique pour tester la structure de dépendance de la série chronologique (série de données). Au cours des dernières années, différentes méthodes ont été développées pour sélectionner le seuil optimal en respectant les hypothèses de chaque approche (Scarrott et MacDonald, 2012; Caeiro et Gomes, 2015; et Langousis et al., 2016). Ces méthodes peuvent être classées en quatre catégories, à savoir : les méthodes non paramétriques, les méthodes graphiques, les méthodes bayésiennes et les modèles de mélange présentés comme alternative à la sélection du seuil. Celles-ci sont présentées dans la suite.

a. Méthodes non paramétriques

Les méthodes non paramétriques identifient le seuil, en localisant le point de changement entre les extrêmes et les régions non extrêmes présentes dans les séries de données. Gerstengarbe et Werner (1989) utilisent une version séquentielle du test de Mann-Kendall (Mann, 1945; Kendall, 1975) pour localiser le point de rupture. Cette méthode est théoriquement incorrecte, et elle a conduit à des résultats erronés selon Langousis et al. (2016).

b. Méthodes graphiques

Les méthodes graphiques impliquent l'extraction des seuils correspondant à différentes valeurs de seuil, et visent à analyser graphiquement les paramètres adaptés au seuil du GPD. Le seuil optimal est identifié en localisant un point du graphique où un comportement linéaire est observé dans les paramètres GPD avec des seuils croissants. Ces méthodes sont les plus utilisées parce qu'elles sont simples à utiliser. Le choix du seuil, cependant, nécessite un jugement et une expertise subjective et les tracés graphiques peuvent être parfois non interprétables. Ainsi, le processus ne peut pas être automatisé lorsqu'un grand nombre d'échantillons doit être analysé (Kiran et Srinivas, 2021).

c. Méthodes bayésiennes

Les méthodes bayésiennes (par exemple Bermudez et al., 2001; Northrop et al., 2017) traitent le seuil comme paramètre dans un modèle hiérarchique, avec une distribution a priori appropriée. L'estimation du seuil est une moyenne pondérée des quantiles sur plusieurs valeurs possibles du seuil afin de réduire la sensibilité des estimateurs obtenus à l'aide d'un seul seuil. La limite de cette méthode bayésienne est la nécessité de spécifier la forme de la distribution a priori, et les valeurs des hyperparamètres (paramètres dont la valeur est utilisée pour contrôler le processus d'apprentissage automatique; Bermudez et al., 2001).

d. Modèles de mélange

Les modèles de mélange considèrent le seuil comme le point de jonction de la partie queue de la distribution GPD, et la partie centrale est représentée par une autre distribution (e.g., Behrens et al., 2004; MacDonald, 2011; Naveau et al., 2016). Ces modèles considèrent le seuil comme étant un paramètre à estimer, au lieu de le sélectionner comme un a priori. Ainsi, ils peuvent tenir compte de l'incertitude découlant de la sélection du seuil qui sépare les deux distributions du modèle de mélange. Dans ces modèles, il est nécessaire de s'assurer que le modèle de la partie centrale et le modèle de queue sont robustes pour une estimation efficace des paramètres.

Bien que plusieurs modèles sont disponibles et peuvent être utilisés dans la modélisation des crues en régime nival, chaque approche présente des avantages et des limites. Il devient donc pertinent, dans le cadre de cette thèse de doctorat, d'améliorer ces approches de modélisation et proposer des nouveaux outils, pratiques, simples et peu coûteux en temps de calcul, basées sur des données historiques disponibles et permettant d'intégrer les variables météorologiques et les indices climatiques, représentatifs de certains processus physiques au sein d'un bassin versant.

1.5 Problématique liée à la région d'étude, objectifs et hypothèses

1.5.1 Problématique

Comme suggéré précédemment, la prévision des inondations est primordiale pour renforcer la résilience face à ces catastrophes naturelles (Cui and Cui, 2019). Au Nouveau-Brunswick, les crues et les inondations sont affectées par les caractéristiques complexes reliées au régime nival. De plus, les effets du climat et des conditions météorologiques à l'échelle globale et régionale (respectivement) doivent être inclus en tenant compte des covariables météorologiques clés, telles que les températures, l'équivalence en eau de la neige, et les précipitations, ainsi que les covariables de grande échelle, telles que l'AMO, la NAO (Bracken et al., 2018; Renard, 2011) et le BWA (Shabbar et al., 1997; Wazneh et al., 2021).

Dans le contexte des types d'inondations printanières dans l'Est canadien, et notamment celles qui caractérisent le fleuve Saint-Jean, il est primordial de proposer des outils de modélisations des crues plus efficaces et plus pertinents, voire d'améliorer les approches probabilistes existantes présentées précédemment afin de combler les limites des modèles disponibles. Le bassin du fleuve Saint-Jean est l'un des bassins versants potentiellement inondables chaque printemps (Esfandiari et al., 2020). Ceci est

également confirmé par les statistiques disponibles dans la base de données de Sécurité Publique Canada (2022), qui démontrent que les inondations enregistrées au Nouveau-Brunswick sont majoritairement causées par le débordement du fleuve Saint-Jean, comme ce fut le cas de 2017 à 2019. Ce bassin versant constituera la région d'étude de cette thèse de doctorat. Le modèle GEV-B-Splines a été sélectionné pour évaluer les effets de l'interaction entre plusieurs variables météorologiques et climatiques sur les inondations printanières, vu sa flexibilité et sa structure permettant d'intégrer des covariables dans son paramètre de position. De plus, selon la théorie des valeurs extrêmes, la méthode des BM, qui sera utilisée dans cette thèse, converge vers le modèle GEV (cf. Fisher, 1928).

Malgré les avantages présentés précédemment de la méthode du BM, la taille de la série reste un défi. En effet, cette méthode demande une grande base de données et elle ne tient pas compte de tous les événements enregistrés par année. Par exemple, si le maximum de débit a été enregistré au printemps, la série de données n'inclura que cette valeur, malgré l'enregistrement d'autres événements durant le reste de l'année, en particulier en automne ou en hiver. Donc, cette méthode ne tient compte que d'un seul événement par année, ce qui représente une simplification de la réalité ou une limite afin de tenir compte de tous les événements d'inondations à travers l'année. Dans ce contexte, la méthode du seuil provenant du théorème de Pickands sera proposée dans cette thèse afin de ne pas éliminer arbitrairement des événements importants en sélectionnant un seul événement par année (dans l'approche du BM). Cette approche nécessite à son tour que les données soient indépendantes et identiquement distribuées (IID). Pour valider cette hypothèse, seulement les événements indépendants observés chaque année seront choisis. Pour garantir cette indépendance, la méthode de séparation des événements sera appliquée (Kiran et Srinivas, 2020). Tous les événements retenus seront ajustés à un mélange de distribution: une distribution GPD pour la queue droite de la distribution et une distribution Gamma pour les événements les plus fréquents afin de sélectionner le maximum d'événements hydrologiques pouvant être responsables d'une inondation majeure.

1.5.2 Objectifs

L'objectif principal de cette étude est de proposer de nouveaux modèles pour l'estimation des crues printanières dans le contexte du changement climatique (en mode non stationnaire), et de les évaluer dans le bassin versant le plus inondable au Nouveau-Brunswick, soit celui du fleuve Saint-Jean.

Les objectifs spécifiques sont les suivants:

- Identifier les variables et leur combinaison à l'échelle des bassins versants responsables des crues printanières observées, voire des risques d'inondations.
- Développer des modèles probabilistes pour l'estimation des événements de crues en intégrant plus qu'une covariable météorologique et/ou climatique.
- Développer un modèle mixte qui vient combler les limites des méthodes classiques (POT et BM) et ainsi proposer un modèle de mélange qui améliore l'estimation des extrêmes de débits et capable de réduire les erreurs associées.

1.5.3 Hypothèses

Plusieurs hypothèses seront utilisées dans ce travail :

- Les variables les plus significatives à considérer sont d'origines climatiques et météorologiques principalement.
- Les variables identifiées comme étant responsables des crues historiques sont les mêmes dans les conditions futures et/ou représentatives des changements en cours et à venir.
- Certains effets en cascades menant à des changements futurs jamais observés dans les régimes de crues passés ne pourront être explorés.

Dans le deuxième chapitre qui suit, rédigé sous forme d'article à soumettre à la revue *Stochastic Environmental Research and Risk Assessment* (SERRA), un modèle GEV-B-Splines est proposé en intégrant des covariables responsables des crues printanières dans le bassin versant du fleuve Saint-Jean. Il répond aux objectifs spécifiques 1 et 2, présentés ci-dessus.

Le troisième chapitre, publié dans la revue *Journal of Hydrologic Engineering*, présente une nouvelle approche de modélisation des crues avec une comparaison entre le GEV et le GPD, ainsi qu'une méthode de sélection des seuils. Cet article répond à l'objectif spécifique 3, présenté ci-dessus.

CHAPITRE 2

Multivariate spring floods estimation using the GEV-B-Splines model: The Saint-Jean River case study, New-Brunswick (Canada)

Ce chapitre est présenté sous forme d'un article scientifique à soumettre à la revue *Stochastic Environmental Research and Risk Assessment (SERRA)*. La référence détaillée de cet article est :

Yousfi, N., El Adlouni, S.E., and Gachon, P. (2024). Multivariate spring floods estimation using the GEV-B-Splines model: The Saint-John River case study, New-Brunswick (Canada). *Stochastic Environmental Research and Risk Assessment (SERRA)*. To be submitted on March, 2024.

Dans ce chapitre, la méthode du bloc maxima est développée pour estimer les débits printaniers maximums en utilisant le modèle GEV-B-Splines, en mode non stationnaire. Afin de tenir compte des changements climatiques, discutés dans le chapitre précédent, des variables météorologiques et des indices climatiques de téléconnexion ont été intégrés dans le modèle sous forme de covariables. Le modèle est appliqué dans le bassin versant du fleuve Saint-Jean situé au Nouveau-Brunswick (Canada), et plus précisément aux stations hydrométriques Fort-Kent et Mactaquac (en amont et en aval de ce bassin, respectivement). L'originalité de ce travail est dans le choix des covariables et leur combinaison dans la modélisation améliorée ou plus robuste des débits extrêmes, en tenant compte des caractéristiques de ceux-ci en amont et en aval du bassin versant.

Résumé

Les outils de prévision et de prévention liés aux crues printanières dans un contexte de changements globaux sont devenus une préoccupation importante dans la gestion des risques en raison des conséquences humaines et économiques de ces aléas hydrométéorologiques, notamment dans l'est du Canada. Ceci est particulièrement intéressant pour le bassin versant de la rivière Saint-Jean qui est le plus grand bassin du Nouveau-Brunswick (NB), situé dans l'est du Canada, où des inondations importantes peuvent survenir. L'objectif principal de cette recherche est de mieux évaluer la dépendance entre des variables climatiques et météorologiques importantes et les débits printaniers dans ce bassin versant spécifique, afin de proposer une méthode d'estimation des crues printanières multivariée. Deux stations hydrométriques (Fort-Kent et Mactaquac) ont été utilisées avec les stations météorologiques les plus proches (Edmundson et Fredericton) pour évaluer le comportement hydrologique en amont et en aval du bassin versant. La crue printanière est représentée par le débit maximal annuel observé entre janvier et juin.

L'approche GEV-B-Splines (*Generalized Extreme Value with B-Splines Functions*) a été développée pour intégrer différentes covariables sélectionnées. Le critère d'information Akaike (AIC) a été utilisé pour sélectionner les meilleurs modèles proposés.

Pour les deux stations, les modèles retenus sont ceux à trois variables démontrant l'importance des précipitations totales accumulées entre l'automne et le mois de crue, ainsi que les indices de téléconnexion. Cette étude a permis de proposer de nouveaux indices météorologiques et de téléconnexion, spécifiques aux sites d'étude et basés sur l'analyse du cycle hydrologique. Grâce à ce modèle, l'estimation des crues printanières peut être améliorée en intégrant davantage d'informations caractéristiques du régime neigeux.

Mots clés : GEV-B-Splines, NAO, BWA, Fort-Kent, Mactaquac.

2.1 Abstract

The forecasting and prevention tools related to spring floods in the context of global change have become an important concern in risk management due to the human and economic consequences of these hydrometeorological hazards, especially in the eastern Canada. This is of particular interest for the Saint John River watershed which is the largest basin in New Brunswick (NB), located in eastern Canada, where major floods can occur. The main objective of this research is to better evaluate the dependence between important climatic and meteorological variables and spring flows in this specific watershed, to propose a multivariate spring flood estimation method. Two hydrometric stations (Fort-Kent and Mactaquac) were used with the nearest meteorological stations (Edmundson and Fredericton) to evaluate the hydrologic behavior in upstream and downstream of the watershed. The spring flood is represented by the annual maximum flow observed between January and June.

The GEV-B-Splines approach (Generalized Extreme Value with B-Splines functions) was developed to integrate different selected covariates. The Akaike Information Criterion (AIC) was used to select the best proposed models.

For both stations, the selected models are those with three variables showing the importance of total precipitation accumulated between the fall and the flood month, as well as teleconnexion indices. This study made it possible to propose new meteorological and teleconnexion indices, specific to the study sites and based on the analysis of the hydrological cycle. Thanks to this model, the estimate of spring floods can be improved by integrating more information characteristic of the snow regime.

Keywords: GEV-B-Splines, NAO, BWA, Fort-Kent, Mactaquac.

2.2 Abbreviations

AIC: Akaike's Information Criterion

AMO: Atlantic Multidecadal Oscillation

AMO_B: Monthly mean AMO values between December and March

AMO_D: Monthly mean AMO values of the flood month

BWA: Baffin Island-West Atlantic

BWA_B: Seasonal BWA values between December and March

BWA_D: Monthly mean BWA values of the flood month

FA: Frequency Analysis

FTC: Freeze-Thaw Cycles

FTC_B : The number of FTC before the flood month.

FTC_D: The number of FTC during the flood month.

FTC_M: The mean average of the number of FTC between preceding November and the flood month.

FTC_N: The minimum of FTC number between preceding November and the flood month.

FTC_X : The maximum of FTC number between preceding November and the flood month.

GEV: Generalized Extreme Value

NAO: North Atlantic Oscillation

NAO_D: Monthly mean NAO values during the flood month

NAO_B: Seasonal mean NAO values between December and March

NB: New-Brunswick

PTOT_B: Total precipitation, calculated and accumulated between preceding November and the flood month (mm).

PSOL_B: The solid phase of precipitation of the month preceding the flood occurrence, as this affects the flood intensity (mm).

Q_{max} : Annual maximum flow

RAIN_D: Monthly accumulated of daily rainfall during the flood month in mm.

SNOW_B: Total accumulation of daily snowfall between preceding November and the flood month (mm).

SNOW_D: Monthly accumulated of snowfall during the flood month in mm.

TX_D: The monthly mean of the maximum daily temperature during the flood month (°C).

2.3 Introduction

Globally, extreme floods and droughts are becoming more frequent and even more severe because of global warming (Berghuijs et al. 2017; Kundzewicz et al. 2019), through complex or compound effects from changes in the water cycle (IPCC 2023). Over most of Canada, the flood regime reveals increased prominence of rainfall floods and decreased prevalence of snowmelt contributions to flood responses (Burn and Whitfield 2016). Among historic floods that have affected Canada, recent events were particularly severe, as the Richelieu River flood in 2011 (Saad et al. 2016), the 2013 Calgary and Toronto floods, the 2017 flood in southern Quebec (Lin et al. 2018), the 2019 flood in the Ottawa River basin (Benoit et al. 2022) and the 2018 and 2019 floods in southern New-Brunswick (Turcotte et al. 2019).

According to Humes and Dublin (1988), most of the flood damage in the province of New-Brunswick occur due to ice-jam flooding within the Saint John River Basin. It can be explained, according to Faghfour et al. (2023), by the configuration of river system in eastern Canada and the climatic conditions characterized by the cold winter with significant snow cover create favorable conditions for flooding caused by snowmelt and favorite the ice-jam formation (Rokaya et al. 2018). Thus, it's important to more study floods of NB in the context of climate change and the interaction between the hydrological cycle and its physical components (Faghfour et al. 2023).

Recently, the study of Yu and Bourque (2022) has allowed to characterize the key discharge variables responsible for the daily dynamics of water flow in four sub-catchments of the upper Saint John River in

New Brunswick (Canada). The authors demonstrated that, for four sub-basins, the daily water flow is mainly controlled by seasonal surges in air temperature and cumulative snow degree-days that both influence the production and timing of snowmelt. Another recent study over eastern Canada from Nolin et al. (2022) reveals also that warmer late spring mean temperatures on average may contribute, among other factors, to advance the spring break-up and to likely shift the contribution of snow to rain in spring flooding processes.

The occurrence and duration of spring floods are also linked with both large-scale influences from atmosphere-ocean interactions or teleconnections, namely El Niño Southern Oscillation (ENSO) or the North Atlantic Oscillation (NAO) (see Guimarães Nobre et al. 2020, 2017), Atlantic Multidecadal Oscillation (AMO) (See Burn and Whitfield 2017), Baffin Island-West Atlantic (BWA) (See Coulibaly et al., 2000), and regional-scale feedbacks from cryosphere conditions (Das et al. 2022; Turcotte et al. 2019) or basin-scale physiographic features (ex. From the presence or absence of forest cover; Yu and Bourque 2022).

For example, Gado and Nguyen (2016), investigated the nonstationary behavior of flood peaks in 32 watersheds in Quebec and they found that all significant trends are negative by reporting a general impression on the possible impacts of climate change on Quebec flood estimation without study this dependence. Yousfi and El Adlouni (2016) was developed the penalized GEV-B-Splines and they used it to estimate the annual maximum flows in gauged hydrometric stations in all the Province of NB. They used the total precipitation as covariate without a solid study of the meteorological effect in annual floods. Also, authors didn't study the climatic effect on the hydrometric data. In 2020, Xu et al., studied the flood dependences on climatic and meteorological variables using copula-based modeling and they applied this method on the longest river in China (Yangtze River Basin). Recently, Faghfour et al. (2023), studied the effect of climate change on flooding conditions in Eel River basin (NB) to estimate the future floods using the modified Artificial Neural Network (ANN) in Matlab. On the other hand, this method is a black box which allows dependences to be identified based on historical data, without giving the calculation details.

In this context, the originality of this work consists of analyzing the factors responsible for a flood during snowfall in the example of the Saint John watershed and integrating them into flood modeling. Several studies have investigated on a global scale successful approaches to establish an efficient and accurate quantification of floods and their associated risks (Mangini et al. 2018; Zischg et al. 2018). Classical Flood Frequency Analysis (FFA) assumes stationarity (i.e., constant parameters) of the probability distribution

functions (pdfs) of water flow. However, with the hydrological impacts from non-stationary climatic conditions, low frequency climate variability (e.g. for example from the ENSO phenomenon) and other human related factors (e.g. land use changes), frequency analysis have to incorporate the non-stationarity or the time-variation of the underlying distribution parameters (Ngongondo et al. 2019). To include these non-stationarity factors, the relationship between model parameters and covariates needs to be developed, as linear and quadratic polynomial functions (El Adlouni and Ouarda 2008; Jiang and Kang 2019; Ouarda and El Adlouni 2011). Other research proposed the use of semi-parametric functions like smooth models (Chavez-Demoulin and Davison 2005). Nasri et al. (2013) developed the GEV-B-Splines with covariates in a Bayesian framework. Recently, Calderón-Vega et al. (2019) expressed parameters in terms of harmonic functions whose amplitudes are regression parameters to be estimated. This diversity in the expression of relationships among model parameters and covariates (dependence) gives rise to diversity in the non-stationary models.

In this context, non-stationary Frequency Analysis (FA) of extreme hydrologic data has been widely used, especially, for problems related to flood risk and water quality managements (Salas et al. 2018). This statistical method aims to estimate occurrence probabilities of rare hydrologic events by fitting a probability distribution to observed extremes (Bracken et al. 2018; Knebl et al. 2005; Pappenberger et al. 2005). The GEV model was frequently used in Canada because of its ability to capture a wide range of tail behaviors and multivariate causal factors, and has been used in numerous past hydrologic studies (Davison et al. 2012; Ghosh and Mallick 2011; Hodgkins et al. 2017; Katz 2010; Najafi and Moradkhani 2014; Renard 2011; Yousfi and El Adlouni 2016). For these reasons, the use of the GEV distribution in the modeling of the Saint John River floods, characterized by a maritime climate and a mixed nivo-pluvial flow regime, is often recommended (Dam et al. 2022). Also, according to the literature, no study has been done with the GEV-B-Spline and by combining both meteorological and climatic variables especially in the estimation of spring flow quantiles in the Saint John River basin.

The main objective of this study is then to propose a non-stationary flood estimation model using the GEV-B-Spline method which considers both meteorological and teleconnection indices as covariates to estimate future spring floods. A multivariate GEV-B-Splines is proposed using one, two and three covariates, and tested over the Saint John River basin in eastern Canada. In our case, the effect of the NAO, BWA and AMO are identified and evaluated on the characteristics of floods over the recent decades, as these teleconnection indices are then used as covariates in the GEV model.

In the next section, a description of study area is presented. Next, the methodology including data, covariables and methods are provided, including the proposed model and its development. Results are shown in the 4th section. Finally, a concluding and discussion are presented in the last sections.

2.4 Study area and hydrometric stations

The study region is the Saint John River (see Figure 1), an international watershed connecting Canada (Quebec and New Brunswick) and the United States (Maine). It is characterized by an area of approximately 55,200 km², of which over half is situated in New Brunswick (Andrews et al. 2020; Newton and Burrell 2016), and corresponds to the major waterway which runs throughout the province (McGrath et al. 2015). It's approximately 673 km long, and most of it is situated in the province of New-Brunswick (Olthof 2017). Average monthly temperature in this province is between -10°C in January and 19°C in July. Annual precipitation is approximately 1,100 mm (El-Jabi et al. 2016).

According to Caissie and Robichaud (2009), total precipitation tends to be highest in southern parts of the New-Brunswick province towards the North Atlantic seaboard, while the northern part receives relatively higher amounts of snow due to colder winter than in southern or eastern areas.

Two hydrometric stations were chosen along the Saint John River. One station, Fort-Kent hydrometric station (01AD002), is located upstream of the basin and another one, Mactaquac hydrometric station (01AK004), is situated in downstream of the basin (see Figure 1). These two available stations have been selected to allow the comparison of the hydrological response upstream and downstream of the basin according to various climatic conditions across the watershed. These hydrometric stations data are from the National Water Data Archive, HYDAT (Environnement et Changement climatique Canada 2019), and provide annual instantaneous peaks/ maximum flows (in m³/s) from 1934 to 2014 for the Fort-Kent station, and from 1967 to 1994 for the Mactaquac one (see details in Table 1). The objective of this study is to model the floods in nival regime, therefore the hydrometric data which represents the annual maximum were sorted on the March-June period, so that only the maximum recorded between March 1 and June 30 have been kept. Based on this selection, a single maximum value, recorded on October 2013, was excluded for Fort-Kent station and another one recorded also on October 1990 was eliminated for Mactaquac station. The rest of the annual maximums are recorded in the period defined above as the spring period. This means that the maximum flows are really recorded in the spring.

The drainage area corresponding to the flow measured at Fort-Kent is around 14,700 km², and the one measured at Mactaquac is around 39,900 km². Note that the upper Saint John River before Fredericton (see Figure 1) is dammed at several places (e.g., at Edmundston, Tobique Narrows, Grand Falls, Beechwood, and Mactaquac), affecting the normal flow regime of the river (Kidd et al. 2011). These dams affect the Saint John River's hydrology, ice regime and ecology (Burrell and Beltaos 2022). Mactaquac dam is located at 80km downstream of Woodstock's dam and it has contributed to eliminate the devastating ice-jam floods which is occurred along the downstream tidal portion of the Saint John River (Burrell and Beltaos 2022).

Near each hydrometric station, used in this study, two meteorological stations have been selected to provide covariates characterizing the regional or local climate conditions. The meteorological stations are part of the meteorological observed network of Environment and Climate Change Canada (ECCC 2021), namely stations located in Edmundston and Fredericton (see Figure 1) closed to Fort-Kent and Mactaquac hydrometric stations, respectively. The Table 1 presents the hydrological and meteorological variables obtained from these stations. For the two hydrologic stations, the period of available data covers the 1934-2014 and 1967-1994 periods for Fort-Kent and Mactaquac, respectively, with annual maximum flows obtained from measured values at each station every 15 minutes. For the meteorological stations, variables include total, solid and liquid precipitations and maximum/minimum temperature (Tmax/Tmin) both at daily scale, and available from 1871 or 1916/1933 and 2022 or 2017/2029, respectively (according to the considered station). The common period (1967-1994) will be used for the hydrometeorological analysis and the training of the GEV model.

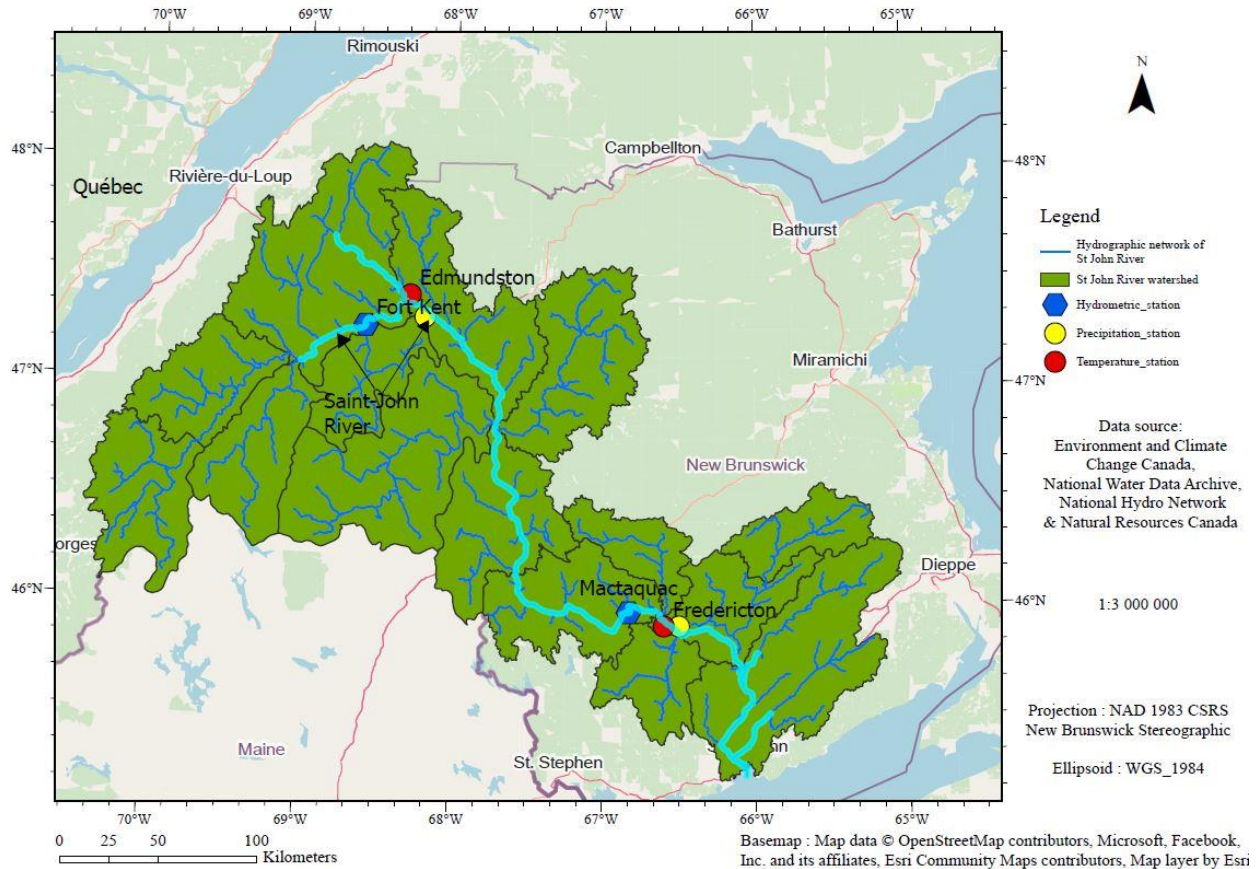


Figure 2.1: Localisation of the Saint John River watershed situated in Canadian provinces of New-Brunswick (NB), Quebec (QC), and the American State of Maine (ME, É.-U.). The main hydrometric (meteorological) stations, Fort-Kent (Edmundston) and Mactaquac (Fredericton) are in blue (red/yellow for temperature/precipitation) symbols.

From the raw meteorological variables (i.e. daily precipitation and daily temperatures time series) presented in Table 2.1, some indices are derived, as presented in the following section, and will be used as covariables to reconstruct daily and seasonal flood events at the two hydrometric stations.

Tableau 2.1. Hydrometric and meteorological stations and data used in the study. The location of each station is given in Figure 2.1.

Station name (Identification number)	Latitude (°N) Longitude (°W)	Variable (unit)	Frequency	Period of available data
Hydrologic stations				
Fort-Kent (01AD002)	47.25°N - 68.59°W	Annual maximum flows (m ³ /s)	≈ Every 15 minutes	1934-2014
Mactaquac (01AK004)	45.96°N - 66.83°W			1967-1994
Meteorological stations				
Edmundston (810AL00* & 8101303**)	47.3°N - 68.2°W & 47.4N - 68.3°W	Precipitation* (mm) &	Daily	1933-2002* & 1916-2017**
Fredericton (8101500* & 8101605**)	45.9°N - 66.5°W & 45.9°N - 66.6°W	Maximum & Minimum temperature** (°C)		1871-2012* & 1871-2019**

2.5 Meteorology

The methodology consists of analyzing the data to develop climatic and meteorological indices derived from the raw data and subsequently choosing combinations to integrate into the proposed model. For this reason, the methodology will be subdivided into two major parts: Derived indices (section 3.1) and Research method (section 3.2).

2.5.1 Derived indices

2.5.1.1 Meteorological derived indices

The occurrence, duration, and intensity of imprinted floods in the nival regime are essentially determined by the temperature regime in winter and spring which directly and indirectly influences the types of precipitation, as well as its duration and occurrence (solids versus liquid), and periods of water retention or storage (through snow and ice at the surface and within the soil) and runoff and infiltration (rapidity of melting of the land cryosphere, etc.). In this context, thirteen derived meteorological indices have been developed, based on the date of flood events identified from the annual instantaneous peaks data (during the spring) from both climatological values and interannual hydrologic hazards. Those indices are defined in Table 2.2.

Tableau 2.2. Meteorological derived indices and their definition with their respective acronyms.

Acronyms	Definition (unit)	
TX_D	Monthly mean of daily maximum temperature during the flood month (°C).	The flood month (1934-2014)
$RAIN_D$	Monthly accumulated of daily rainfall during the flood month (mm)	
$SNOW_D$	Monthly accumulated of daily snowfall during the flood month (mm)	
SWE_D	Snow Water Equivalent during the flood month (mm)	
FTC_D	Number of freeze-thaw cycles (FTC) during the flood month	
$PTOT_B$	Monthly accumulated of daily total precipitation before the flood month (mm)	The period between 1 st November and the flood month (1934-2014)
$SNOW_B$	Total accumulation of daily snowfall before the flood month (mm)	
$FTCC_B$	Cumulative freeze-thaw cycle number before the flood month	
$FTCM_B$	Mean average of the number of freeze-thaw cycles before the flood month	
$FTCX_B$	Maximum of FTC number before the flood month	
$FTCN_B$	Minimum of FTC number before the flood month	
$PSOL_B$	Solid precipitation of the month before the flood month (mm)	The month preceding the flood event (1934-2014)
FTC_B	Number of freeze-thaw cycle (FTC) of the month before the flood month (mm)	

The development of the meteorological indices was based on a physical analysis of the flood events and the available explanatory variables. Qualitative analysis was done based on the daily flow and flood dates,

and before (indices or subscript B for each meteorological index presented in Table 2) those events since the fall from early November to the flood month, as well as during the flood month (indices or subscript D for each meteorological index presented in Table 2.2).

A first analysis of nonparametric statistical dependence between the extreme flow and each meteorological index presented in Table 2 was done using the Spearman rank dependence (Spearman 1904) to identify some potential monotonic relationship. The following definition and explanation for each selected index are briefly presented in Table 2.3.

Tableau 2.3. Teleconnection derived indices and aggregated definition at seasonal and monthly scales before and during the flood month.

Acronyms	Definition
TX _D	TX _D corresponds to the monthly mean of the maximum daily temperature during the flood month that may affect the occurrence and rapidity of thawing period (ex., snowmelt) and runoff.
PTOT _B :	Total precipitation is calculated and accumulated between preceding November and the flood month, as this affects the total amount of available water (solid and liquid) before the flood event.
SNOW _B :	Total accumulation of daily snowfall between preceding November and the flood month, as this affects the total amount of available water on the ground before the flood and the runoff during the thawing period.
PSOL _B :	The solid phase of precipitation of the month preceding the flood occurrence, as this affects the flood intensity.
RAIN _D :	Monthly accumulated of daily rainfall during the flood month (liquid phase of available water), as this affects the flood duration and intensity.
SNOW _D :	Monthly accumulated of snowfall during the flood month (solid phase of available water), as this affects the flood intensity.
SWE _D	The snow water equivalent during the flood month, which affects the runoff, is reconstructed from the total accumulated snowfall or solid precipitation (mm). The assumption is made based on the 1/10 rule (1 cm of snow equal 10 mm of water; Chen et al., 2022). In this approximation, we neglect the quantity of snow lost by runoff and evaporation or sublimation, so this can be a serious source of errors in the estimation of available water from the snow cover.
FTCC _B :	Cumulative freeze-thaw cycle (FTC) number or event between preceding November and the flood date is used as covariable that influences the available water from the snow cover available on the ground just before and during the thawing period (see Fortin, 2003; Saad et al., 2016). This FTC index is computed from daily Tmax and Tmin using the following joint occurrence the same day, i.e. Tmax ≥ 0°C and Tmin < 0°C. (Fortin, 2003).
FTCM _B :	The mean average of the number of FTC between preceding November and the flood month is also used as covariable to reconstruct or infer the available water at the start of the thawing period that influences the intensity of flood.
FTCX _B and FTCN _B :	The maximum and minimum of FTC number (respectively) between preceding November and the flood month are also used.
FTC _B and FTC _D :	The number of FTC before and during the flood month (respectively) are used as complementary covariables to infer the available water during the thawing period, useful for both duration and intensity of flood (see Saad et al., 2016).

2.5.1.2 Teleconnection indices

In our study, three main teleconnection indices have been used, based on previous studies of Burn and Whitfield (2017), Coulibaly et al. (2000), Hodgkins et al. (2017) and Shabbar et al. (1997), namely:

- AMO index: This index has been computed using the approach of Trenberth and Shea (2006), where the North Atlantic sea surface temperature (SST) anomalies covering the area between 0 and 60°N in latitudes, and 0 and 80°W in longitude. The global warming of all oceanic basins (from 60°S to 60°N) was subtracted from the AMO index of 1934-2019, and AMO monthly average relative to 1901-1970 as in Trenberth and Shea (2006). The AMO index was recently used in Burn and Whitfield (2017) to evaluate the links with increasing/decreasing trends in quantile flood regimes of various Canadian streamflow.
- NAO index: This index is that of Hurrell (1995) and corresponds to the differences of Mean Sea Level Pressure (MSLP) across the North Atlantic, calculated from (or around) the meteorological station data time series between Lisbon (Portugal) and Stykkishólmur (Iceland). Data used in this paper, computed over the 1901-2019 period at monthly scale. According to the different phases of the NAO index, Guimarãez Nobre et al. (2017 and 2020) have found a strong link between NAO and extreme rainfall, and flood occurrence and damages across Europe.
- BWA index: The BWA is an index developed by Shabbar et al. (1997), defined as difference of 500-hPa geopotential height between Baffin Island (52°N, 60°W) and the western North Atlantic (30°N, 60°W). Coulibaly et al. (2000) found that the BWA provides a good forecast for annual runoff over northeastern Canada. That's why this index was included to evaluate its links as covariate for flood occurrence modeling.

All the three teleconnection indices have been computed over the 1901-2019 at a monthly scale using the CERA-20C reanalysis datasets (Laloyaux et al. 2018), as in the recent study of Wazneh et al. (2021). Those will be used as covariates to derive and integrate into the probabilistic model, to evaluate the influence of climate indices on the hydrological response over our study area.

The teleconnection data series have been selected to cover the period between 1934 and 2014 for the Fort-Kent station and 1967-1994 for the Mactaquac station (see available data in Table 4). For each teleconnection index, the Table 4 presents the six aggregated indices at seasonal, over the preceding winter season (December-March, indices with subscript B in Table 4) as recommended by Coulibaly et al.

(2000), and monthly scales during the flood month (indices with subscript D in Table 2.4). This recommendation was statistically tested and compared with the effect of monthly indices on throughput. The dependence was less important with the monthly values contrary to the seasonal data. For this reason, in the presented analysis, the recommendation of Coulibaly et al. (2000) has been followed.

The proposed aggregated teleconnection indices presented in Table 2.4 are the result of dependence tests between the raw data and the historical flood flows. The objective of the development of these indices is to propose indices which integrate the remote climatic influences, and which have the strongest dependence with the flow. These six teleconnection indices were proposed to find the best explanatory combination for the spring flood event recorded in each year. Akaike's Information Criterion (AIC; Akaike 1998) has been used next to choose the best combination (See Eq. (9) and table 7 and table 8 of the result subsection).

Tableau 2.4.. Teleconnection derived indices and aggregated definition at seasonal and monthly scales before and during the flood month.

Acronyms	Definition	Period of defined indices
NAO_D	Monthly mean NAO values during the flood month	The flood month (1934-2014)
AMO_D	Monthly mean AMO values of the flood month	
BWA_D	Monthly mean BWA values of the flood month	
NAO_B	Seasonal mean NAO values between December and March	The period between December 1 st and March 30 st (1934-2014)
AMO_B	Seasonal mean AMO values between December and March	
BWA_B	Seasonal mean BWA values between December and March	

2.5.2 Research methods

In the following, the GEV distribution will be first presented before to introduce the GEV-B-Splines model in detail. Proposed non-stationary GEV-B-Splines models will be also presented in the following section. The developed models using identified best covariates will be discussed in the last sub-section.

2.5.2.1 The GEV distribution model

The GEV distribution was proposed first by Jenkinson (1955) following the original distribution of Fisher and Tippett (1928). The flows estimation was based on the approach of the blocks of the seasonal maxima (Bloc of Maxima, BM). According to Extreme Value Theory (EVT; Coles, 2001), the BM of independent and identically distributed data (iid) converges asymptotically to the GEV distribution (Aghakouchak and Nasrollahi, 2010; Shinyie et al., 2013; Vasiliades et al., 2015). This probabilistic distribution is widely applied to model hydrologic process extremes (De Paola et al., 2018), such as floods (Martins and Stedinger, 2000), rainfall (Coles and Tawn, 1996), runoff (Jiang et al., 2023) and sea waves (De Hann and De Ronde, 1998). It combines three distribution families: Gumbel (I), Fréchet (II) and Weibull (III) distributions (Jenkinson, 1955; Jiang and Kang, 2019). Its distribution function is given by Eq. (1):

$$F_{GEV}(y, \mu, \sigma, \kappa) = \begin{cases} \exp \left[- \left(1 + \kappa \left(\frac{y-\mu}{\sigma} \right) \right)^{\frac{1}{\kappa}} \right], & \kappa \neq 0 \\ \exp \left[- \exp \left(- \frac{y-\mu}{\sigma} \right) \right], & \kappa = 0 \end{cases} \quad (1)$$

where y is the variable to estimate, μ is the position parameter, σ is the scale parameter and κ is the shape parameter.

The family I is characterized by two parameters defined on \mathbb{R} and correspond to the case which $\kappa = 0$.

The family II is characterized by three parameters defined on $\left] -\infty, \mu + \frac{\sigma}{\kappa} \right[$, for $\kappa < 0$

The family III is characterized by three parameters defined on the interval $\left] -\frac{\sigma}{\kappa} + \mu, +\infty \right[$, for $\kappa > 0$.

The estimation of the flood rate of an extreme event amount is done by the determination of the three unknown parameters (μ, σ, κ) . Various methods have been developed to estimate the GEV distribution's parameters, such as the Method of Moments (MM; Fisher, 1929), the Maximum Likelihood Method (ML;

Smith, 1985) and the Probability Weighted Moments (PWM; Hosking, 1986; Hosking, 1985). For the GEV model, the most used method is the ML. Coles (2001) has proposed an univariate approach of the GEV distribution and an estimation of its parameters by the ML method. Then, Katz et al. (2002) have shown that this method can be erratic when the samples are small ($n \leq 25$) and especially in the estimation of extreme quantiles by the GEV. This explains the development of Generalized Maximum Likelihood (GML) estimators, introduced by Martins and Stedinger (2000), for classical GEV with constant parameters. This method considers the Beta distribution of parameters ($u = 6, v = 9$) as a prior distribution for the shape parameter. This Beta distribution is centered at the value of -0.1 and set to the interval $[-0.5; 0.5]$. This shows that the GML method is a special case of the Bayesian approach. This method has been generalized in the non-stationary case by El Adlouni et al., (2007) considering the same prior distribution for the shape parameter. Yousfi and El Adlouni (2016) developed regularized estimators in the case of non-stationary GEV-B-Splines model with different penalty functions.

2.5.2.2 The non-stationary GEV-B-Splines model

In hydrology, the flood probability distribution can change with time, given the occurrence and intensity of hydrometeorological phenomena and climate change (Masson-Delmotte et al., 2022). This implies the non-stationarity of the probabilistic event. In this case, stationary Frequency Analysis (FA) is no longer applicable, and it is essential to use approaches that consider non-stationarity. In this case, the GEV parameters are function of time or other covariates. The quantile Y_T depends on these parameters (μ_X, σ, κ) , where μ_X is the located parameter which depends on the covariate vector $X = (X_1, X_2, \dots, X_l)$.

The model considered in this study is a generalization of the non-stationary GEV-B-Splines model (Yousfi and El Adlouni, 2016) for the case of multivariate covariates. This model assumes that σ is constant, $\kappa \sim \text{Beta}(u = 6, v = 9)$ and μ depends on the covariate vector X . In the present study, we will develop a flood estimate in a general context of nivo-pluvial regime using a vector of d covariates $X = (X_1, \dots, X_d)$.

Since hydrologic non-stationarity is associated with on-going climate change at various timescales and with low frequency climate oscillation (Milly et al., 2008), we will combine meteorological and climate-teleconnection indices as covariates, as presented in the section 2.2.

The interaction between the position's parameter and covariates is represented by B-Splines functions given by Eq. (2), whatever the number of covariates, such as:

$$\mu_{\underline{X}} = \sum_{i=1}^d g_i(X_i) \quad (2)$$

where $g_i(X_i)$ is written in the case of each covariate $i = (1, \dots, d)$ by Eq. (3):

$$g(x_i) = \beta_0 + \sum_{j=1}^m \beta_j B_{j,l}(x_i) \quad (3)$$

where $B_{j,l}(x)$ represents the basic function splines, m is the number of nodes and l is polynomials' degree.

B-Splines are defined by recurrence relations given by the following Eq. (4) and (5):

For $j = 0, \dots, m - l - 2$

$$B_{j,l}(x) = \frac{x-x_j}{x_{j+l}-x_j} B_{j,l-1}(x) + \frac{x_{j+l+1}-x}{x_{j+l+1}-x_{j+1}} B_{j+1,l-1}(x) \quad (4)$$

For $j = 0, \dots, m - 2$

$$B_{j,0}(x) = \begin{cases} 1, & \text{Si } x_j < x < x_{j+1} \\ 0 & \text{else.} \end{cases} \quad (5)$$

In the this model, σ is assumed to be a constant, $\kappa \sim \text{Beta}(u = 6, v = 9)$ defined on the interval

$] - 0.5, 0.5[$ and $\mu_{\underline{X}} = \sum_j (1 - B) \times \left(\frac{\beta_0}{\beta} \right)$ where B is the B-Spline function: a function of covariates and 1 is the unit vector of size d .

2.5.2.3 The developed models

To evaluate different combination of covariates on the extreme spring flow using a non-stationary GEV model, three non-stationary models (GEV_1 , GEV_2 and GEV_3) will be compared with the classic stationary model GEV_0 , namely:

Model₁: Non-stationary GEV-B-Splines integrating one covariate:

$$GEV_1 \text{ with } X' = (Cov1) \quad (6)$$

Model₂: Non-stationary GEV-B-Splines integrating two covariates:

$$GEV_2 \text{ with } X'' = (Cov1, Cov2) \quad (7)$$

Model₃: Non-Stationary GEV-B-Splines integrating three covariates:

$$GEV_3 \text{ with } X''' = (Cov1, Cov2, Cov3) \quad (8)$$

Each model was tested with one or more covariates (see section 2.2) to find the best combinations. The AIC will also be used to select the best representative probabilistic model.

As previously suggested, different weather and climate covariates will be exploited in the GEV model with 2 and 3 covariates. In order to choose the most efficient model, the calculation of a decision support criterion, like the AIC, is necessary. The AIC equation is a combination relationship between the Kullback-Leibler information (Kullback-Leibler, 1951) and Fisher's log-likelihood function (Kim et al., 2017). The AIC's Eq. (9) is given by:

$$AIC = -2 \log(ML) + 2k \quad (9)$$

Where $\log(ML)$ is the maximized log likelihood function under the proposed model, and k is the number of parameters in each model.

The choice of the most efficient model will be made in comparing the values of the AIC between models. The model with the lowest AIC value will be considered as the most efficient model. As a result, this criterion will be considered for choosing either a two or three covariate model, to select the best combination of covariates. For each model, the extreme flood flow will be estimated by using one covariate and combining two and three covariates among the covariates discussed in the previous section. The best combination of covariates will be selected by calculating the AIC criterion. Results will be also compared with the AIC of the stationary-GEV-B-Splines. The model will be applied to the two hydrometric data series (Fort-Kent and Mactaquac stations) to compare the hydrological behavior upstream and downstream of the Saint John River watershed.

2.6 Results

2.6.1 Hydrographic characteristics upstream and downstream of the watershed

From the two hydrometric stations used in this study, Table 2.5 presents comparative statistics of maximum, minimum and median flows computed over the common time series (1967-1994). As suggested from the flow statistics, the flood maximum is mainly observed during the spring, and the record was recorded on April 30, 1979, in both upstream and downstream of the watershed (4,280 and 10,700 m³/s, respectively). The minimum flow occurs in general by the end of winter, early spring (April or May) before the thawing period. The maximum/median/minimum flows are 2.5/2.47/2.85 higher at Mactaquac than at Fort-Kent. This means that the Fort-Kent station has such important flows variability (high/low flow records) on the upstream location in higher elevation, and its hydrological response is also strongly affected by cryospheric processes during the cold season that influence more strongly the low and high flow (see the recent work of Das et al. 2022).

Tableau 2.5. Flow characteristics upstream at Fort Kent (1934-1994) and downstream at Mactaquac (1967-1994) of the Saint John watershed.

	Fort-Kent (01AD002)	Date	Mactaquac (01AK004)	Date
Annual maximum flow (m ³ /s)	4280	30-04-1979	10700	30-04-1979
Annual minimum flow (m ³ /s)	1350	06-04-1988	3850	06-05-1967
Median of annual maximum flow (m ³ /s)	2410	N/A	5960	N/A

The chronological series of maximum of annual flows of the two stations available over two different periods having a common period (1967-1994) are illustrated in Figure 2.2 to compare the hydrological behavior of the two stations during a same period. Over the common period between 1967 and 1994, we note that the monotony of the maximum flows' variation is similar among the two stations (synchronicity in the maximum/minimum flow values). However, the flows from Mactaquac station are much higher than those from Fort-Kent, not only because it is downstream within the watershed, but also, given the presence of the dams closed to this station, which influence the observed flow during the period of high-water levels' management.

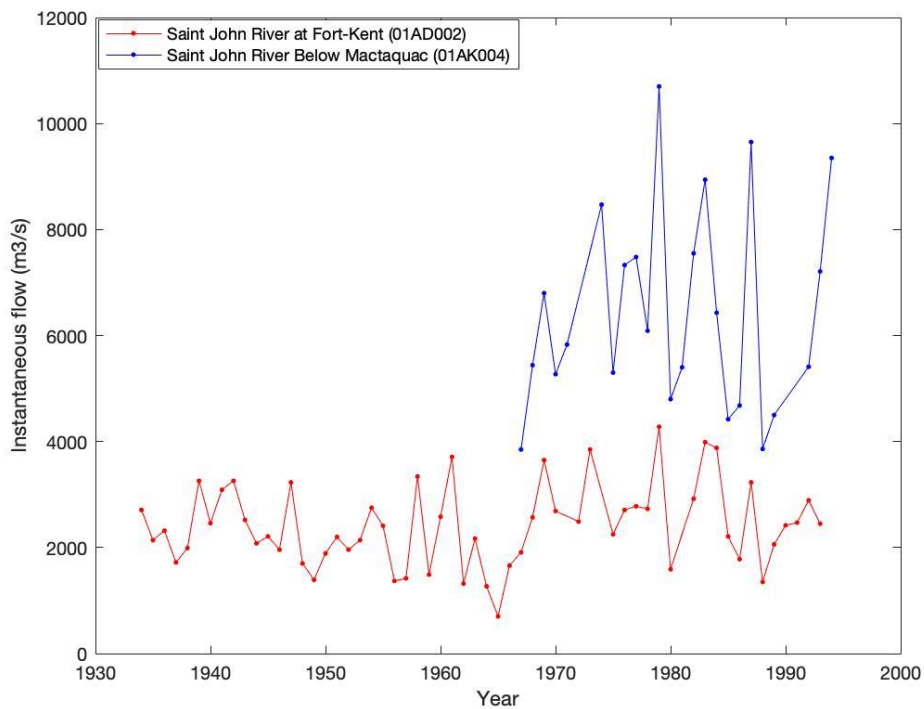


Figure 2.2. The annual maximum flow observed over the period January-June at the stations Fort-Kent (01AD002, red line), for the period 1934-1994 and Mactaquac (01AK004, blue line), for the period 1967-1994.

2.6.2 Single links between maximum flow and covariables

An analysis of Spearman rank correlations between covariables and the maximum of spring maximum flow (between March 1 and June 30) is presented in the Table 2.5 for both hydrometric stations, with the rho

and p-values. In summary, this correlation analysis reveals that At Fort-Kent, a positive correlation (at the 5% significance level, i.e. p-value < 0.05, bold values in Table 2.6) between the maximum flows and $PTOT_B$, $SNOW_B$, and $PSOL_B$ is observed. This means an increasing trend in flow (Q_{max}) when total precipitation, snow and/or solid precipitation are more important before the flood month.

A negative correlation (at the 2% significance level, i.e. p-value < 0.02, bold values in Table 2.6) is recorded between Q_{max} and the AMO index before the flood month. This means that when the AMO is negative/positive during the winter month the maximum flow increases/decreases, signifying that colder/warmer SSTs in the North Atlantic before the thawing period affect the risk of flooding in the upstream part of the Saint John watershed.

At Mactaquac: No significant correlation is found (at the 2-29% significance level) between maximum flow and meteorological covariables and climate indices. Only, weak correlation is found (around 0.22-0.24 in absolute values) with the NAO and BWA in winter (before the flood month), as well as $RAIN_D$ (corresponding to the monthly accumulated daily rainfall during the flood month) at the 30% significance level (p-value < 0.3, see Table 2.6).

Tableau 2.6. Spearman correlation between the maximum flow and all covariables or climate indices (presented in Tables 2 and 3) computed for Fort-Kent and Mactaquac hydrometric stations. Correlation is computed depending on hydrometric station and common period of data availability (1934-2014 for Fort-Kent and 1967-1994 for Mactaquac). The rho and p-values are provided for each pair of correlation and for each station.

Variable of interest	Tested indices	Fort-Kent (01AD002)		Mactaquac (01AK004)	
		rho	p-value	rho	p-value
Annual maximum flows Q_{\max} (m^3/s)	TX_D	-0.0858	0.5259	0.0831	0.6995
	$RAIN_B$	0.3153	0.0169	0.1755	0.4121
	$SNOW_B$	0.3209	0.0149	0.1755	0.4121
	$RAIN_D$	0.1744	0.1946	-0.2287	0.2825
	$SNOW_D$	0.0736	0.5863	0.1611	0.4521
	$PSOL_B$	0.2644	0.0469	0.1611	0.4521
	SWE_D	0.0736	0.5863	0.1611	0.4521
	FTC_D	0.0141	0.9168	0.0864	0.6880
	FTC_B	-0.0287	0.8319	0.0187	0.9307
	$FTCC_D$	-0.0642	0.6354	-0.0267	0.9013
	$FTCM_B$	0.1236	0.3596	-0.0450	0.8345
	$FTCX_B$	-0.2112	0.1149	0.0462	0.8302
	$FTCN_B$	-0.1640	0.2227	0.0087	0.9677
	NAO_B	0.0367	0.7864	0.2186	0.3047
	AMO_B	-0.3082	0.0197	-0.1012	0.6381
	BWA_B	0.0213	0.8750	-0.2436	0.2514
	NAO_D	-0.1789	0.1830	-0.1207	0.5744
	AMO_D	-0.0550	0.6843	0.0188	0.9305
BWA_D	0.0583	0.6666	-0.0708	0.7423	

According to these preliminary results, the linear and single interaction of climatic and meteorological variables with the maximum flow is more evident in the upper Saint John River basin (near Fort-Kent), than in the lower part of the basin (near Mactaquac) which is also more potentially affected by dams and regulation of water flows. Based on the evaluated correlations, it seems potentially pertinent to alternatively test more than one covariate that may have a synergistic effect (on the maximum flow) when they occur together. In this context, the following section details the combinations between the covariates and the maximum flow, with an emphasis on the most interesting results of various combinations.

2.6.3 Combined effects of covariables on maximum flow from developed models

Fifty-seven models were tested with different covariables, and the respective AIC were calculated with the non-stationary (GEV_3) and stationary (GEV_0) models, for the Fort-Kent and Mactaquac stations (see Tables 6 and 7, respectively). The best models obtained for the Fort-Kent station with the lowest value of AIC are non-stationary models, using 1, 2 and 3 covariables, presented in Table 2.7. The best model and combination is the GEV-B-Splines including three covariates, i.e. the model that combines TX_D , $PTOT_B$ and BWA_B with an AIC value of 898.94. This is the lowest value in all combinations and is much lower than the stationary model (AIC = 923.31). Hence, in all cases whatever the number of covariables included in the GEV-B-Spline model, the non-stationary model performs better than the stationary ones.

Tableau 2.7. Identification of the best significant covariate with the GEV-B-Splines model using one (Model1,1), two (Model1,2), three (Model1,3) covariates and the stationary model (GEV_0) at Fort-Kent (located upstream of the Saint John River, see Figure 2.1). The best combination, i.e. lowest AIC values, is in bold (using three covariates)

Models	Covariate(s) vector	AIC
$Model_{1.1}$	$X'=(PTOT_B)$	913.36
$Model_{1.2}$	$X''=(PTOT_B, RAIN_D)$	906.62
$Model_{1.3}$	$X'''=(TX_D, PTOT_B, BWA_B)$	898.94
GEV_0	923.31	

Tableau 2.8 presents the results using the GEV-B-Splines model with one, two and three covariates and using hydrometeorological data and climatic indices from the Mactaquac station. As for the Fort-Kent station, fifty-seven models were tested, and their AIC was calculated and compared with the stationary AIC (stationary model GEV_0). As for the upstream station, the best combinations have been obtained with the non-stationary model with three covariates (based on the lowest AIC). However, the stationary model appears to give slightly better results than the single or two covariates GEV no-stationary model. However, the use of the two covariates model $Model_{2,2}$ is still more relevant than that of $Model_{2,1}$ (one single covariable) as for the upstream station. Again, by integrating three covariates, the $Model_{2,3}$ gives the lowest AIC results. These covariables are $PTOT_B$, $RAIN_D$, and NAO_B . These combinations are quite like the ones used for the Fort-Kent station, as total precipitation is used in both cases, maximum temperature and total amount of rainfall are linked in spring (see Wazneh et al., 2020), and the (anti)correlation between NAO and BWA indices are strong in winter months (see Wazneh et al. 2021). In summary, the $Model_{2,3}$ is chosen as the best model for the estimation of spring floods downstream of the watershed in the hydrometric station of Mactaquac.

Tableau 2.8. Identification of the best significant covariate with the GEV-B-Splines model using one (Model2,1), two (Model2,2), three (Model2,3) covariates and the stationary model (GEV_0) at Mactaquac (located downstream of the Saint John River, see Figure 2.1). The best combination, i.e. lowest AIC values, is in bold (using three covariates)

Models	Covariate(s) vector	AIC
$Model_{2,1}$	$X'=(TX_D)$	430.38
$Model_{2,2}$	$X''=(TX_D, NAO_B)$	425.97
$Model_{2,3}$	$X'''=(PTOT_B, RAIN_D, NAO_B)$	415.62
GEV_0		425.90

2.6.4 Results of the GEV-B-Splines with mixed covariates

To illustrate the effect of integration of the three covariates on the estimation of spring flooding in the Saint John River basin, the $Model_{1,3}$ and $Model_{2,3}$ were used and calibrated with the historical data of three variables and the hydrometric data of Fort-Kent and Mactaquac stations, respectively. For both stations, the results are obtained to project for 20 and 100 years return periods, and compared with the conventional stationary model, which assumes that the data are identically distributed.

Figures 2.3, 2.4 and 2.5 present the results upstream of the river (Fort-Kent) while Figures 2.6, 2.7 and 2.8 show those downstream (Mactaquac). According to the estimated quantiles for each return periods, quantiles present a quite similar asymptotic component for the two (20 and 100 years) return periods. For the Fort-Kent station, the quantile estimate considers the historical data of the three integrated covariates (TX_D , $PTOT_B$ and BWA_B), and the projection of the results is a function of the BWA_B anomaly values (see Figure 3). For both return periods, the spring flow tends to more often decrease/increase for BWA values less/above than 0.5, whatever the two confidence intervals tested (95% and 99%). This suggests the presence of a threshold effect for the climate index BWA_B on the monotony of the spring flow. In addition, the spring flow tends to increase over time and the hydrological risk tends to increase over time, as presented in Figure 3. Indeed, for the same value of BWA equal to 0, the quantile over 20 years is 4000 m^3/s while for the same value of BWA, the quantile of 100 years is estimated over 5000 m^3/s , while for the same value of BWA the quantile of 100 years is estimated over 5000 m^3/s , for the same value of BWA.

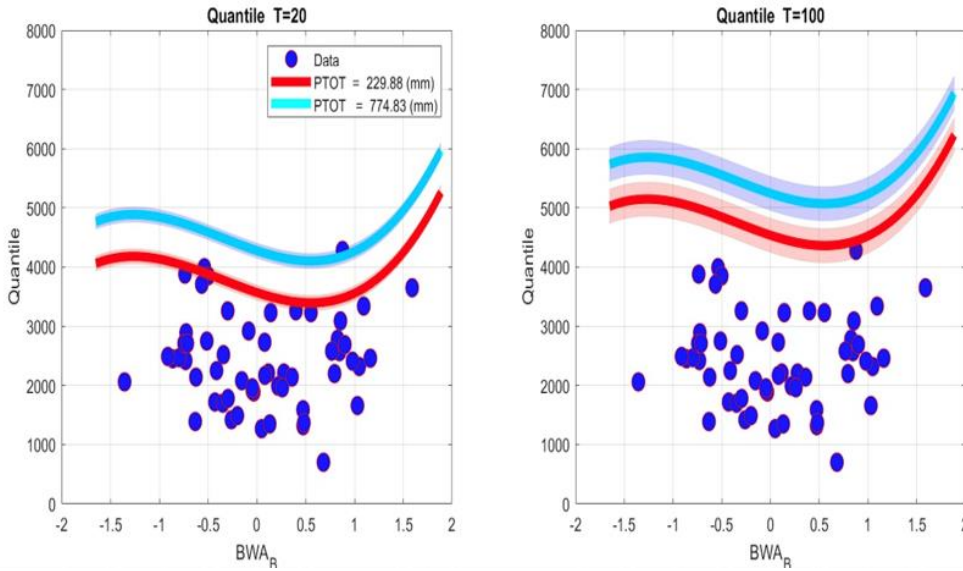


Figure 2.3. Estimation of the quantiles of extreme flows in nival regime using the GEV_3 with $X''' = (TX_D, PTOT_B, BWA_B)$. The illustration is given for two extreme values (min and max) of the BWA index observed for the period 1934-2014, and two return periods $T= 20$ (left) and 100 years (right) for the Saint John River at Fort-Kent station.

Figure 2.3 allows visualizing the PDF of the extreme flows as a function of the BWA_B index. In this Figure, the display function of the BWA_B index has been projected according to the historical extreme flows from the minimum value of BWA_B (-1.3554) to the maximum ones (1.591). Moreover, the PDF of the stationary model is also included to facilitate the comparison with the non-stationary GEV-B-Splines with three selected covariates. As shown in Figure 4, the probability of having high flow rates is higher with positive than negative BWA_B values.

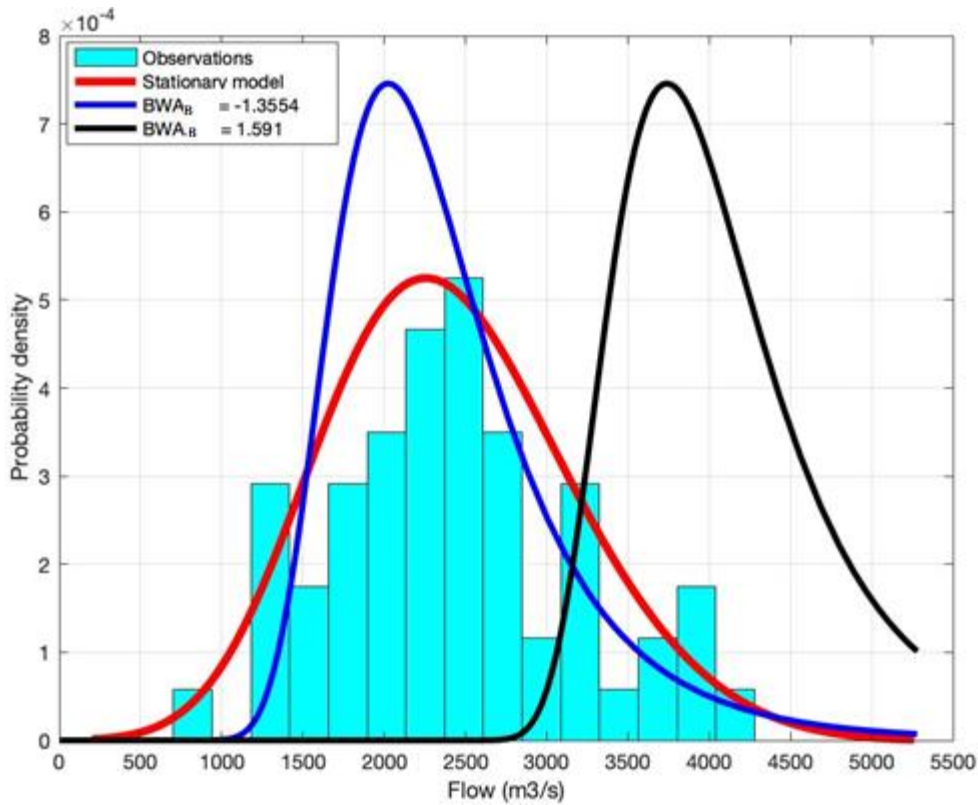


Figure 2.4. Probability density functions of annual maximum flow at Fort-Kent station (Saint John River) for the stationary model (without any covariate -black line-) and the model GEV_3 with $X''' = (TX_D, PTOT_B, BWA_B)$ with an illustration for two values, minimal (blue line) and maximal (red line) of the BWA_B index.

Indeed, according to the projection of historical observations, with the maximum historical value of BWA_B which is a positive value (1.591), the approximate flow is 3700 m³/s. On the other hand, with the minimum BWA_B value, which is in turn a negative value (-1.3554), the observed flow rate is 1800 m³/s. In Figure 5, the comparison of the standardized anomalies of annual maximum flow and of the seasonal BWA_B over 1934-1994 period confirm that periods of higher flows are observed during positive phases of the BWA. However, this is not systematic with 10/14 occurrences of higher/lower flows observed during these positive BWA phases.

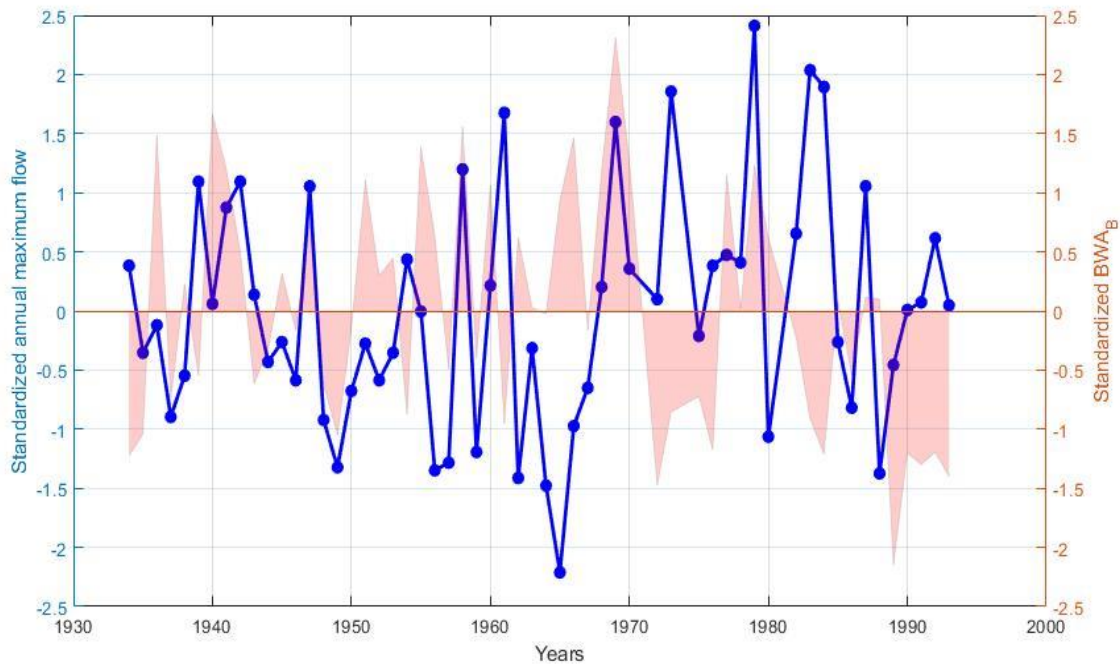


Figure 2.5. Interannual standardized anomalies over the 1934-1994 period of annual maximum flows observed between January and June at Fort-Kent (blue line) and of seasonal mean BWA computed between December and March (pink curve). The maximum flow is selected using the time series of observed flow (measured every 15 minutes at the hydrometric station). All standardized values are computed using the mean and standard deviation over the 1934-1994 climatological period, both for the maximum flows and the seasonal mean BWA.

As mentioned previously and shown in Table 2.5, there is absence of a cause-and-effect relationship between BWA and annual maximum flow. Figure 5 reveals also that BWA is also strongly affected by high interannual and low frequency variability with positive phases more often observed before 1970s, and negative phases in 1980s and 1990s. These correspond to high records of positive/negative flow anomalies in 1960s/1980s. The highest annual maximum flow was recorded in 1979 during a positive BWA phase.

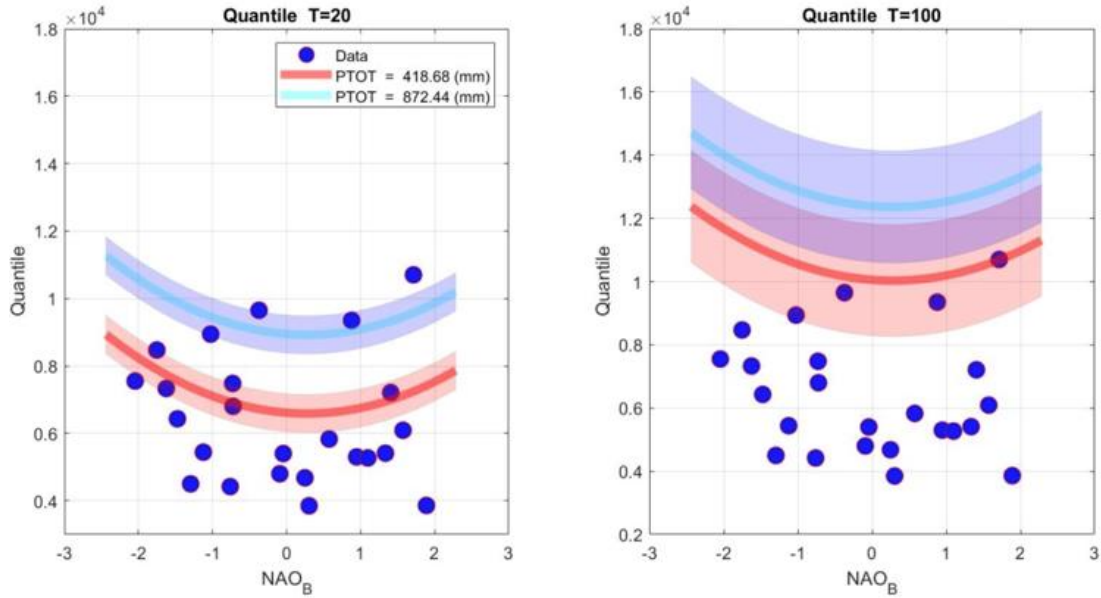


Figure 2.6. Estimation of the quantiles of extreme flows in nival regime using the GEV_3 with $X''' = (PTOT_B, RAIN_D, \text{ and } NAO_B)$ for $T = 20$ (left) and 100 years (right) return period at Mactaquac (Saint John River). Illustration given for a minimal (390 mm) and maximal (910 mm) possible values of the covariate $PTOT_B$.

In this context, Figure 2.6 illustrates the estimation results with the selected model for the Fort-Kent station ($Model_{1,3}$ in Table 2.8). The integration of the total precipitation with the maximum temperatures of the peak month (between March and May) and the winter average of the BWAB indices yielded the best estimating model of the snowfall floods (according to AIC index). Figure 6 presents also the quantile estimate from the $Model_{2,3}$ using the three chosen covariates (POT_B , $RAIN_D$, and NAO_B) and the historical data of the Mactaquac station. According to the results, the estimated quantiles have the same shape for the two return periods tested with a difference in the estimated maximum flows, which is normal because of the common distribution. We notice a proportionality the total precipitations from November to the peak's month and the quantile up to a maximum constant between 418.68 mm and 872.44 mm.

According to fig 2.7, two NAO_B phases are detected (the positive and the negative phasis). The estimate quantile is symmetric to the zero. The further the NAO value is from zero, the higher the flow rate is. This can be explained by temperature and extreme snowmelt accompanied by heavy rains which can lead to a significant rise in the water level of the river. Also, the Mactaquac Dam, located in the downstream portion

of the Saint John watershed, often cannot hold all of the melted snow and ice in the upstream portion, so water can be released downstream (Esfandiari et al. 2022). So this stable value of precipitation despite the increasing flow can be increased by water management at the level of the dams which consists in emptying the dam when it reaches a certain saturation threshold.

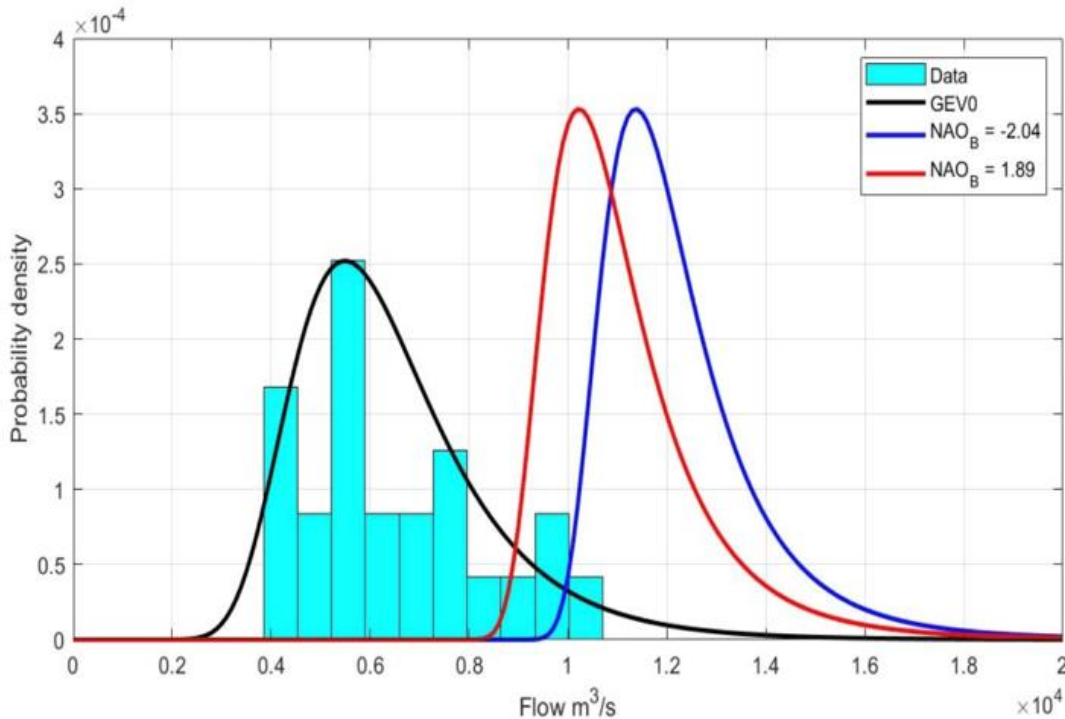


Figure 2.7. Probability density functions of annual maximum flow at Mactaquac station for the stationary model (without any covariate -red line-) and the GEV_3 with $X''' = (PTOT_B, RAIN_D, \text{and } NAO_B)$ model with an illustration for two values, minimal (blue line) and maximal (black line) of the NAO_B index.

Figure 2.7 presents the PDF of the NAO_B as a function of the spring flow characteristics, using the historical extreme flows according to the minimum and maximum values of NAO_B (-2.04 and 1.89, respectively). The flows are more often lower with the minimum NAO_B than those with the maximum NAO_B . This means that the positive phase of the NAO_B has an amplifying effect on the spring flow estimate. Also, the stationary model which does not integrate any covariate tends to generate lower hydrological risk of higher flows. The duration, intensity and type of precipitation once the flood has started are also factors to consider which influence the intensity and duration of the highest flows (see the recent study in

southern Québec in Benoit et al. 2022). Land use planning also plays an important role (see Yu and Bourque, 2022).

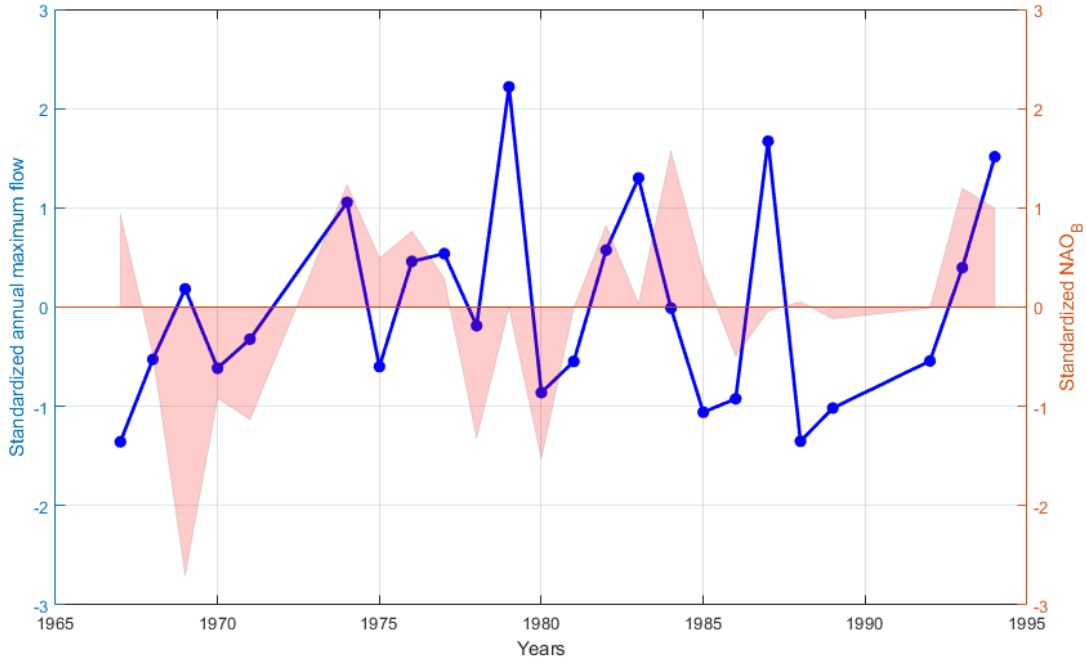


Figure 2.8. Interannual standardized anomalies over the 1967-1994 period of annual maximum flows observed between January and June at Mactaquac (blue line) and of seasonal mean NAO computed between December and March (pink curve). The maximum flow is selected using the time series of observed flow (measured every 15 minutes at the hydrometric station). All standardized values are computed using the mean and standard deviation over the 1967-1994 climatological period, both for the maximum flows and the seasonal mean BWA.

Figure 2.8 presents the interannual anomalies of the standardized annual maximum flows and NAO_B over the 1967-1994 period. This confirms that higher flow values are more often observed during positive phase of the NAO. However, the most extreme flow anomalies are recorded when NAO is in neutral phase, such as in 1979, 1983, and 1987. The lowest annual maximum flows in 1970s and 1980s do not necessarily occur during negative phases of the NAO, but also during neutral NAO phases as for higher flow events.

In summary, the proposed non-stationary models with three covariables are the most relevant qualitatively and statistically compared to the stationary models. Although results in the two stations show

different hydrological behavior upstream and downstream of the Saint John River watershed, the explanatory covariates or combination of covariables are quite similar or at least they are issued from the same physical processes (from temperature and precipitation regimes) or teleconnection indices (BWA or NAO). This further argues not to use the stationary model in the estimation of floods whatever the region, to reduce the risk of underestimation of higher flows.

2.7 Discussion and conclusion

The objective of this study was to model the risk of high flow events within the Saint John River system in New Brunswick. This study was realized using two selected hydrometric stations, one at Fort-Kent (Maine) and Mactaquac (NB) located respectively in the upper and lower part of the watershed, and the nearest meteorological stations in Edmundson (NB) and Fredericton (NB). Several combinations of covariables from meteorological and teleconnection indices have been developed, allowing the identification and evaluation of several GEV-non-stationary models with one, two and three covariates. A quantitative analysis was carried out to evaluate non-linear dependence between the selected indices and the seasonal flow records to have an optimal criterion to identify covariates.

The best GEV-non-stationary model selected for the Fort-Kent station combines the three variables:

- TX_D corresponds to the monthly mean of the maximum daily temperature during the flood month that may affect the occurrence and rapidity of thawing period (ex., snowmelt) and runoff.
- $PTOT_B$: Total precipitation accumulated between preceding November and the flood month, as this affects the total amount of available water (solid and liquid) before the flood event.
- BWA_B : Seasonal mean BWA values between December and March.

For the Mactaquac station, the best GEV-non-stationary model combines the three covariates:

- $PTOT_B$: as for Fort-Kent.
- $RAIN_D$: Monthly accumulated of daily rainfall during the flood month (liquid phase of available water), as this affects the flood duration and intensity.
- NAO_B : Seasonal mean NAO values between December and March.

The results obtained for the two stations confirm the need to study the hydrological behavior of the specific study site to better predict the hydrological risk. Then, it is more appropriate to use probabilistic models with several covariates such as GEV-B-Splines to integrate as much information as possible, and reduce the uncertainty in the estimation of hydrological risk. This will in turn improve flood protection facilities on the Saint John River, which records at least one major spring flooding each year (Canadian Disaster Database 2022). However, further analysis needs to incorporate other components of the hydrological cycle, as management aspects or forest or land cover modifications (as in Yu and Bourque 2022), all those link to human intervention (modification of the territory, presence of dams, etc.). Further application needs also to consider the development of plausible climate change information on flood risks over the whole Saint John River system, and to compare those with the recent study of Budhathoki et al. (2022).

Finally, the integration and combination of explanatory variables in flood estimation models improves scientific knowledge in causal factors associated to floods, and especially the decisions taken to reduce their consequences. Therefore, it is recommended to test other hydrological behaviors in vulnerable watersheds exposed to spring flooding, to define a hydrological risk estimation model compatible with local scale hydrometeorological hazards.

2.8 Credit authorship contribution statement

Nawres Yousfi: Methodology, Investigation, Data curation, Formal Analysis, Writing - Original Draft. Salah-Eddine El Adlouni: Conceptualization, Methodology, Writing- Review & Editing. Philippe Gachon: Conceptualization, Methodology, Writing - Review & Editing.

2.9 Declaration of competing interest

No potential conflict of interest was reported by the authors.

2.10 Data availability statements

Hydrometric data used in this study are available on the official site of environment Canada (Environnement et Changement climatique Canada, 2022).

Matlab codes were developed by the authors. Direct requests for these codes can be made to co-author Salaheddine El Adlouni (E-mail: salah-eddine.el.adlouni@umoncton.ca).

2.11 Acknowledgments

We acknowledge the financial support from the Natural Sciences and Engineering Research Council of Canada (NSERC), individual Grants of Pr. Philippe Gachon (NSERC-RGPIN-2016-06436 and NSERC-RGPIN-2022-05032) and of Pr. Salah El Adlouni (NSERC- RGPIN-2019-05746). We also acknowledge other financial supports from the Strategic research chair of UQAM hold by Pr. Philippe Gachon, by the UQAM under the scholarship for the exemption of additional tuition fees for foreign students, and by the UQAM's Faculty of Sciences under the programme "faculty financial support".

CHAPITRE 3

Mixture Probability Models with Covariates: Application in Estimating Risk of Hydroclimatic Extremes

Ce chapitre est présenté sous forme d'un article scientifique publié dans le *Journal of Hydrologic Engineering*. La référence détaillée de cet article est :

Yousfi, N., El Adlouni, S.E., Papalexiou, S.M., and Gachon, P. (2022). "Mixture Probability Models with Covariates : Application in Estimating Risk of Hydroclimatic Extrêmes". *Journal of Hydrologic Engineering*. DOI : 10.1061/JHYEFF/HEENG-5831.

Dans ce chapitre, une nouvelle approche de modélisation des extrêmes est proposée pour combler les limites des deux principales méthodes utilisées (POT et BM présentées au Chapitre 1). Les résultats de ce modèle sont comparés avec le GEV (la distribution de la série BM) et GPD (la distribution de la série POT) ainsi que la distribution Gamma, afin de démontrer la flexibilité de cette approche aussi bien au centre de la distribution qu'aux deux différentes queues de celle-ci.

Résumé

La modélisation des événements extrêmes est importante dans de nombreux domaines scientifiques, incluant l'environnement, le génie civil et l'évaluation des impacts et des risques. Parmi les méthodes de modélisation disponibles, on distingue les modèles statistiques permettant d'estimer la fréquence et l'intensité des extrêmes. Ce type de modèle est régulièrement utilisé dans les procédures permettant d'anticiper l'évolution potentielle des événements extrêmes. La théorie des valeurs extrêmes (TVE) fournit une base théorique pour l'estimation statistique des événements extrêmes à l'aide de l'analyse fréquentielle (AF). Le défi de ce type de modélisation est de savoir quand utiliser la méthode du Bloc Maxima (BM) ou la méthode de dépassement de seuil (*Peaks-Over-Threshold*, POT). Chacune a ses propres limites. La POT permet de décrire les principales caractéristiques des séries extrêmes observées. La sélection du seuil dans cette méthode est toujours difficile et peut affecter la précision des résultats simulés, et la crédibilité des changements dans les valeurs extrêmes. Pour relever ce défi, les modèles de mélange offrent plus de flexibilité pour représenter des échantillons avec des données non homogènes. Cette étude présente le modèle de mélange de Gamma-Pareto-Généralisé (GGP) pour estimer l'occurrence du risque d'extrêmes hydroclimatiques. Le modèle est développé dans sa forme générale et les événements hydrométéorologiques extrêmes dépendent de covariables multidimensionnelles. Un algorithme de maximum de vraisemblance (ML) est proposé pour estimer les paramètres avec une contrainte sur le paramètre de forme de la distribution Pareto généralisée (*Generalized Pareto Distribution*, GPD). Une simulation de Monte Carlo (MC) compare le modèle proposé à l'approche POT classique, à seuil fixe, et à la série de maximums annuels de débit. L'approche proposée est appliquée sur un ensemble de données hydrologiques quotidiennes provenant d'une station observée située dans le fleuve de Saint-Jean à Fort-Kent (01AD002), au Nouveau-Brunswick, Canada. Les résultats montrent une flexibilité dans la modélisation des extrêmes pour les séries chronologiques dépendantes ou non stationnaires, et décrivent adéquatement la partie centrale des fréquences observées, ainsi que les queues de distribution.

Mots clés : Seuil, Distribution Gamma, GPD, Maximum de vraisemblance, MC

3.1 Abstract

Modeling of extreme events is important in many scientific fields, including environmental and civil engineering, and impacts and risk assessments. Among available methods, statistical models that allow estimating extremes' frequency and intensity are regularly used in procedures to anticipate potential changes in extreme events. Extreme value theory provides a theoretical basis for statistical estimation of extreme events using frequency analysis. The challenge in modeling is knowing when to use the block maxima method or the Peaks-Over-Threshold (POT) method. Each has its drawbacks. POT describes the main characteristics of the observed extreme series; the threshold selection is always challenging and might affect the accuracy of the simulated results and the credibility of changes in extreme values. To encompass this challenge, mixture models offer more flexibility to represent samples with nonhomogeneous data. This study presents the gamma generalized Pareto (GGP) mixture model for estimating risk occurrence of hydroclimatic extremes. The model was developed in its general form, whereas the observed hydrometeorological extreme events depend on multidimensional covariates. A maximum likelihood algorithm is proposed to estimate the parameters with a constraint on the shape parameter of the generalized Pareto (GP) distribution. A Monte Carlo (MC) simulation compared the proposed model with the classical POT approach, with a fixed threshold, and the annual maximum series of streamflow. The approach was applied using a daily hydrological data set from an observed station located in the Saint John River at Fort Kent (01AD002), New Brunswick, Canada. The results show a flexibility to model extremes for dependent or nonstationary time series and adequately describes the central part of the observed frequencies, as well as the tails.

Keywords: Mixture model; Generalized Pareto distribution; Gamma distribution; Generalized Extreme Value distribution; hydrological extremes, river flow, Saint John River.

3.2 Introduction

Modeling of extreme events is critical for risk assessment, especially considering the ongoing climate change, which will strongly affect the hydrometeorological hazards and flood events across the world (IPCC 2021). Statistical techniques have been developed previously to estimate these flooding risks (Salas and Obeysekera 2019). Annual maximum series (AMS) and peaks over threshold (POT) are typically used to estimate the intensity and frequency of hydrometeorological extremes. Since its introduction, extreme value theory (EVT) (Fisher and Tippett 1928; Jenkinson 1955; Pickands 1975) has become a standard tool to characterize rare events, and it aims at estimating the extremal behavior of observed phenomena by asymptotic probability distributions. According to EVT, the AMS approach converges asymptotically toward the generalized extreme value (GEV) distribution, while the POT approach converges asymptotically toward the generalized Pareto distribution (GPD) (e.g., AghaKouchak and Nasrollahi 2010; Coles 2001; Shinyie et al. 2013; Vasiliades et al. 2015). These probabilistic models have been widely applied in the field of civil engineering. For example, Nyaupane et al. (2018) used GEV to estimate flooding and map floodplains for Hurricane Harvey in Buffalo Bayou. Typically, a parametric extreme value model describes the upper or lower tails of a probabilistic distribution, which is fitted to available extreme value data. Different modeling techniques are available depending on the structure of the available data and the data generation process of interest (El Adlouni et al. 2007). Model performance is evaluated by how well it describes the observed tail behavior of the sample data. If the model provides a good fit, it is used for the extrapolation of quantities of interest (typically high quantiles), with estimation of the associated extrapolation uncertainty. EVT assumes that observations should be independent and identically distributed (IID). In addition, the sample size should be large depending on the considered characteristics and the desired level of precision. According to Volpi et al. (2019), the data selection approach results in the elimination of many observations that may be significant in estimating return periods. The limited sample size in hydrological records significantly affects the reliability of the return period estimates from rare extreme events. Therefore, the convergence to the asymptotic distribution can be slow and may require a large number of events (Papalexiou et al. 2013). Actually, there are cases where convergence is not achieved even for very large samples; for example, stretched exponential or lognormal tails should converge to the Gumbel distribution, although this convergence is unrealistically slow. In this context, and to solve this problem, modeling extreme events has experienced significant developments for parameter estimation, threshold selection, and asymptotic behavior (Papalexiou et al. 2013; Yang et al. 2019). Thus, extreme value analysis has become an area of increased interest, with major developments being made in univariate, multivariate, dependent, and nonstationary extremal modeling (Behrens et al. 2004; de Melo

Mendes and Lopes 2004; Tancredi et al. 2006; Carreau and Bengio 2009). Theoretically, the two approaches should converge asymptotically to the same results, but the POT approach is more adapted to reality because the AMS approach rejects a large part of the information by selecting the annual maximums (Kiran and Srinivas 2021). Other research has taken into consideration covariate dependence (both parametric and nonparametric) for nonstationary sequences (Smith 1987; Davison and Ramesh 2000; Hall and Tajvidi 2000; Pauli and Coles 2001; Chavez-Demoulin and Davison 2005; Yee and Stephenson 2007; Eastoe and Tawn 2009). Davison and Smith (1990), McNeil and Frey (2000), and Ferro and Segers (2003) studied the dependence among extremes, while Coles and Tawn (1991, 1994) and Heffernan and Tawn (2004) studied multivariate extremes. However, apart from classical approaches in parameter estimation, many recent studies have shown the importance of integrating covariates into the parameters of the two main approaches, that is, block maxima and POT (Yousfi and El Adlouni 2016; Thiombiano et al. 2017). The POT approach describes the main characteristics of the extreme time series. Both intensity and occurrence models are considered to estimate extreme events risks. The challenge with the POT model is the threshold selection. It is often set with statistical, physical, or a priori considerations (Lang et al. 1999; Davison and Smith 1990). To overcome this subjective selection, mixture models consider two probability distributions, the first for the lower and upper parts to define the threshold and the second for the exceedances. For the mixture models, the threshold is one of the parameters to be estimated (Frigessi et al. 2002; Behrens et al. 2004; Scarrott and MacDonald 2012). However, the combined distribution creates an artificial discontinuity at the threshold, which contradicts the reality and seems inappropriate in the simulation of hydrometeorological time series. A solution put forward in the literature is to impose a continuity constraint, of the entire distribution, at the threshold (Carreau and Bengio 2009; Nascimento et al. 2012). This constraint has a significant effect on the tail of the distribution and therefore on the estimation of extreme events. It leads automatically to an underestimation of the extreme event intensity (Mo et al. 2019). The classical POT approach might also underestimate the extreme events (Lang et al. 1999). This is due to fitting more data on the left-hand side of the distribution and the effects are more pronounced for large or low-frequency return periods (e.g., greater than 10 years). To avoid prior determination of the threshold, several studies investigated the possibility to include the determination of threshold values, of the POT method, in the estimation method of the parameters (Solari and Losada 2012). This would provide another way to estimate threshold values that is complementary to the original approach, which unfortunately cannot be automated and requires user intervention. Such models apply hybrid distributions to parametrically model the entire range of values and use the estimated upper threshold values to define the series of peaks in the POT method (Solari and Losada 2012). This is very

similar to the original approach, where the threshold is a parameter of the probability distribution (Figure. 3.2). The proposed model is a real mixture of the two components and affects both tails of the selected data. The hybrid model in the last reference was fitted to daily precipitations. Other studies used mixture models for AMS, such two gamma distributions or lognormal with GPD (Evin et al. 2011). The context is different for the proposed model, which makes sure the selected daily flows are independent for annual picks. The use of the separation step is important to make sure the selected events are independent. Here we explore the gamma generalized Pareto (GGP) mixture model for risk estimation, where the risk can be applied in different fields (health, finance, or hydrology). The models proposed in the literature are hybrid distributions. In these models, the observations below the threshold follow a distribution (central part) and the GPD distribution is considered for the part above the threshold (Kiran and Srinivas 2021). In these hybrid models, the threshold is estimated with the other parameters of the two distributions [e.g., lognormal-GPD in Solari and Losada (2012)]. The originality of this work is in (1) proposing a new mixing model for modeling hydrological extremes, different from the hybrid models often cited in the literature; (2) clarifying the difference between mixture models and hybrid models by developing the equations for a mixture model in case of covariates (climatic indices or seasonal indexes); (3) evaluating the bias and the error of estimation of the quantiles of the mixing model compared with the classical approaches of distribution of annual maxima (GEV) and overshoots at a threshold with a GPD; and (4) providing the baseflow separation approach developed by Fischer et al. (2021) as a solution to the retained event dependence problem of the hydrograph. The model was developed in its general form, where the observed events depend on multidimensional covariates and knowing that different mixture models could be used. The proposed model was compared with the classical POT approach with fixed threshold and the AMS through simulated data. The approach was applied for daily flow data set from the observed hydrometric station located at Fort Kent in the upstream part of the Saint John River watershed, in New Brunswick, Canada. In that case, the hydrologic risk was estimated from the proposed GGP mixture model. Mathematical modeling is subject to uncertainties from multiple sources, such as measurement and input errors, as well as model structure. Each of these sources of uncertainty contributes to the generation of uncertainty in the modeling results. An uncertainty analysis for input data and model parameters can be performed by specifying the ranges of their variations and identifying the most likely values from these ranges. For example, in hydrological modeling, the uncertainties associated with the amount and timing of flow can be presented by incorporating the full range of simulation results from a hydrological model (Abdi et al. 2021). This study aims to (1) advance the statistical development of the GGP mixture model in its general form, (2) integrate covariates to account for climate fluctuations, and (3) propose an efficient

algorithm to compute confidence intervals of the estimated quantiles. The following section presents material and methods, the study area, and the main characteristics of the GGP mixture model. Theoretical development is then presented, with hydrological data preparation and extreme events selection, and the simulation experiment. Simulation and case study results are presented in the next section, which includes a discussion of the proposed approach. A discussion is then presented. Finally, a summary and conclusions are presented along with future research recommendations.

3.3 Material and methods

The proposed model, based on the combination of two distributions (GPD and gamma) was applied to hydrological data from the Fort Kent station, located upstream of the Saint John watershed (presented in the following), characterized by major annual floods. The proposed model is detailed in “GGP Mixture Model.” This approach is based on the separation of events. The theoretical development of this approach is detailed in “Results and Discussion.”

3.3.1 Study area

The study area corresponds to the Saint John River’s watershed located in New Brunswick, Canada. Maximum daily flow of the time series covering the period between 1927 and 2014 from the National Water Data Archive, HYDAT, from the Fort Kent hydrological observed station (01AD002), located upstream of this watershed, was used (Environnement et Changement climatique Canada 2018). In this area, most flood disasters are the result of snowmelt runoff, storm rainfall, or combinations of snowmelt runoff and storm rainfall in the spring season (El Jabi and Caissie 2018). The Saint John River runs for 673 km, starting in the northern part of Maine, northeast into the province of New Brunswick, Canada (Figure 3.1). Steep slopes and fast-running streams characterize many parts of this watershed. It then flows southeast through the province into the Maritime Plain (Bay of Fundy), which is relatively flat with many lakes and bogs (Newton and Burell 2016). The climate of this study area is cold without a dry season and with a warm summer (Peel et al. 2007; Beck et al. 2018) with a mean annual air temperature between 5.8°C in the south and 3.2°C in the north (El Jabi and Caissie 2018). Most of this basin has a humid continental climate, except for the regions along the Bay of Fundy coasts, which have a maritime climate (Boisvert et al. 2017). In general, the climate gets colder and drier toward the north. The peak discharge usually occurs in the spring (April or May), corresponding with the periods of snowmelt (Newton and Burell 2016).

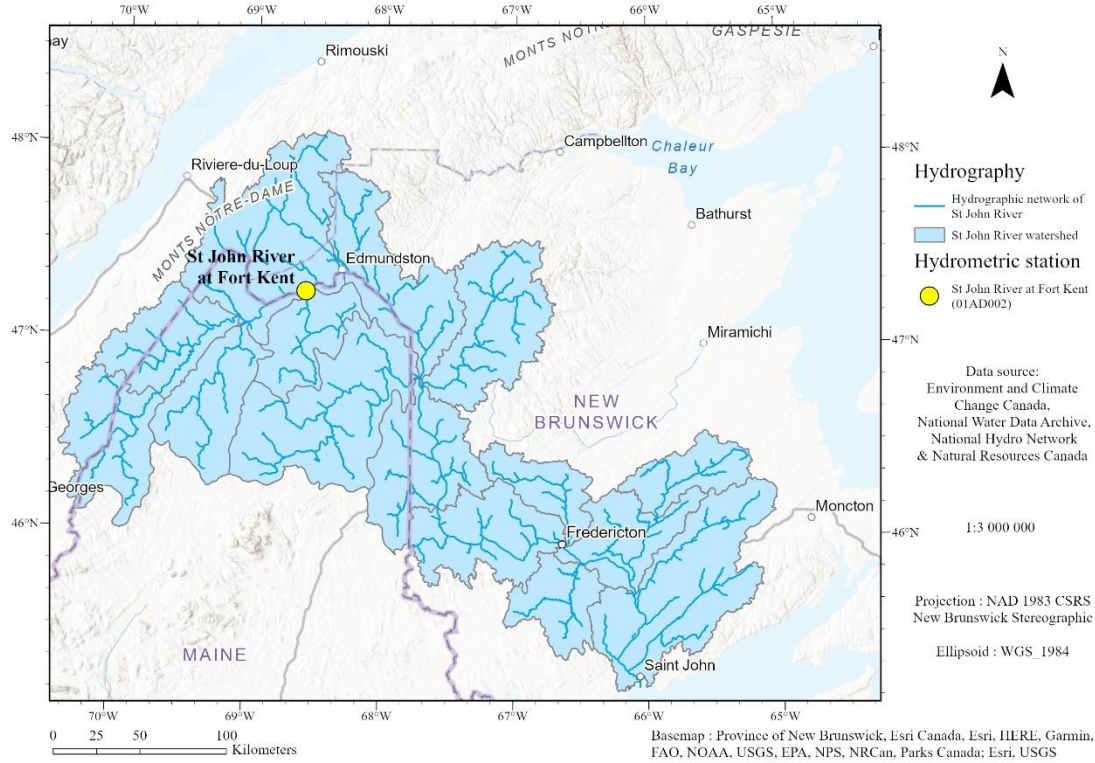


Figure 3.1. The study area of Saint-John River (in green) with the location of the Fort-Kent hydrometric station located upstream of the watershed.

3.3.2 GGP mixture Model

For a sequence of independent and identically distributed variables $\{Y_i, i = 1, \dots, n\}$, and corresponding observed data $\{y_i, i = 1, \dots, n\}$ the probability density function (pdf) of mixture model with k components is given by:

$$f(y_i) = \sum_{j=1}^k c_j g_j(y_i | \theta_j) , \quad i = 1, \dots, n \quad (10)$$

where $g_j(\cdot | \theta_j)$, $j = 1, \dots, k$; are the pdf's of the mixture models with parameter vectors θ_j and weights c_j such that, $c_j \geq 0$ and $\sum_{j=1}^k c_j = 1$.

Here we fix $k = 2$ using the Gamma and Generalized Pareto distributions as components. The Gamma density is:

$$g_1(y|\theta_1) = \frac{1}{b^{a\Gamma(a)}} y^{a-1} \exp^{-y/b}; \quad y > 0, \quad a > 0, \quad b > 0; \quad (11)$$

with parameters $\theta_1 = (a, b)'$ and skewness coefficient $Cs = 2/\sqrt{a}$. Details of the gamma's development law are in appendix A (Annexe A).

The second component, the pdf of the exceedances $(y_i - u)$ above a given threshold u , is:

$$g_2(y_i|\theta_2) = \begin{cases} \frac{1}{\beta} \left(1 + \kappa \left(\frac{y_i - u}{\beta} \right) \right)^{-1-1/\kappa} & \kappa \neq 0 \\ \frac{1}{\beta} \exp(-(y_i - u)/\beta) & \kappa = 0 \end{cases} \quad (12)$$

where $\theta_2 = (u, \kappa, \beta)'$ with $\beta > 0$ and $\kappa \in \mathbb{R}$ being the scale and the shape parameters, respectively. Details of the GPD development law are also in appendix A.

In the presence of temporal non-stationarity and/or dependence on other variables (Yilmaz et al., 2017; El Adlouni et al. 2007), the models with covariates, allow to estimate conditional quantile curves. Let $\mathbf{V} = (\mathbf{v}_1, \dots, \mathbf{v}_m)'$ and $\mathbf{W} = (\mathbf{w}_1, \dots, \mathbf{w}_m)'$ be the covariate matrices for the mixture components and the weights coefficients, respectively.

We assume that the scale parameters of the Gamma and the GPD distributions depend on the covariate \mathbf{V} through the relation:

$$b(\mathbf{v}_r) = \exp(\mathbf{v}_r' \boldsymbol{\gamma}) \quad \text{and} \quad \beta(\mathbf{v}_r) = \exp(\mathbf{v}_r' \boldsymbol{\delta}) \quad (13)$$

The second assumption concerns the weight coefficients that follow Logistic regression model on with \mathbf{W} as the covariate

$$c_1(\mathbf{w}_r, \boldsymbol{\tau}) = \frac{1}{1 + \exp(\mathbf{w}_r' \boldsymbol{\tau})} \quad \text{and} \quad c_2(\mathbf{w}_r, \boldsymbol{\tau}) = 1 - c_1(\mathbf{w}_r, \boldsymbol{\tau}) \quad (14)$$

where $\boldsymbol{\gamma} = (\gamma_0, \gamma_1)'$, $\boldsymbol{\delta} = (\delta_0, \delta_1)'$ and $\boldsymbol{\tau} = (\tau_0, \tau_1)'$. The table1 simplifies the parameters associated with each relation in this proposed mixed model. For $k = 2$ (see equation (1)), the proposed model can be written as:

$$f(y_i) = \sum_{j=1}^k c_j g_j(y_i|\boldsymbol{\theta}_j) = c_1 g_1(y_i|\boldsymbol{\theta}_1) + c_2 g_2(y_i|\boldsymbol{\theta}_2) , i = 1, \dots, n \quad (15)$$

To simplify the understanding of the model which combines two probabilistic distributions: the gamma law and the GPD law. Table 3.1 makes it possible to distinguish the parameters associated with each distribution as well as the unknown variables. This makes it possible to better visualize the vector of covariates to be determined in the proposed model.

Tableau 3.1. Integration of covariates in the mixture model

$f(y_i) = c_1 g_1(y_i \boldsymbol{\theta}_1) + c_2 g_2(y_i \boldsymbol{\theta}_2) , \quad i = 1, \dots, n$		
Equation	variables	unknown variables
$c_1 g_1(y_i \boldsymbol{\theta}_1)$	$\boldsymbol{\theta}_1 = (a, b)'$ $b(\mathbf{v}_r) = \exp(\mathbf{v}_r' \boldsymbol{\gamma})$ $c_1(\mathbf{w}_r, \boldsymbol{\tau}) = \frac{1}{1 + \exp(\mathbf{w}_r' \boldsymbol{\tau})}$	$a \text{ and } b$ $\boldsymbol{\gamma} = (\gamma_0, \gamma_1)'$ $\boldsymbol{\tau} = (\tau_0, \tau_1)'$
$c_2 g_2(y_i \boldsymbol{\theta}_2)$	$\boldsymbol{\theta}_2 = (u, \kappa, \beta)'$ $\beta(\mathbf{v}_r) = \exp(\mathbf{v}_r' \boldsymbol{\delta})$ $c_2(\mathbf{w}_r, \boldsymbol{\tau}) = 1 - c_1(\mathbf{w}_r, \boldsymbol{\tau})$	$u, \kappa \text{ and } \beta$ $\boldsymbol{\delta} = (\delta_0, \delta_1)'$ $\boldsymbol{\tau} = (\tau_0, \tau_1)'$

The parameters to be estimated are $\boldsymbol{\theta}_1 = \{a, \boldsymbol{\gamma}\}$ for the Gamma and $\boldsymbol{\theta}_2 = \{u, \kappa, \boldsymbol{\delta}\}$ for the GPD components. Let $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\tau}\}$ the set of all the parameters for the mixture model with covariates.

Figure 3.2-a illustrates the pdf of a GGP mixture model in the absence of the covariates effects and special case of the parameters: $\boldsymbol{\theta}_1 = \{a = 20, \boldsymbol{\gamma} = (1, 0)\}$, $\boldsymbol{\theta}_2 = \{u = 0, \kappa = 0.1, \boldsymbol{\delta} = (80, 0)\}$ and $c_1 = 0.6$. This figure also compares the three distributions: the Gamma, the GPD and the proposed model (mixture model) which combines the two distributions. As shown in Figure 3.2.a, the GGP mixture model exhibits

some interesting behaviors such as bimodal distributions or heavy-tailed models with high frequencies for values close to zero.

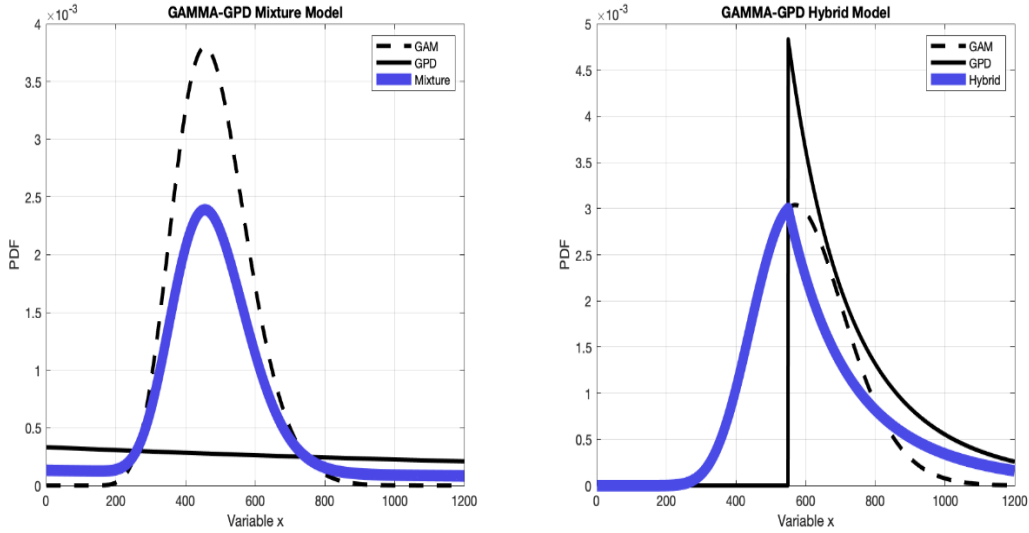


Figure 3.2. Illustration of the difference between mixture model proposed in the present study (left) and the hybrid distributions

Such shapes, as shown in figure3.2, are more likely to be encountered in observed situation for flood and precipitations' extremes (El Adlouni et al., 2007). The hazard function of the mixture, as shown in, has different shape distribution when compared to common distributions, which are monotonic increasing, decreasing, or constant (Figure 3.2-b).

To realize a statistical inference for the GGP mixture model, latent variables $Z_i, i = 1, \dots, n;$ with $Z_i \in \{0,1\}$ are introduced (Cassella et al., 2002). The latent variable allows an explicit expression of the likelihood by coding $z_i = 1$ when the observation y_i belongs to the Gamma component and $z_i = 0$ for the GPD component dataset. In the rest of the manuscript, the observations that belong to the Gamma component will be noted y_{iG} and those of the GPD by y_{iP} . Thus, the set of observed data is related to the latent variable by:

$$Z_i = y_{iGAM} z_i + y_{iGPD} (1 - z_i) \quad ; \quad i = 1, \dots, n \quad (16)$$

The conditional probability distribution of the latent variable satisfies:

$$P(Z_i = j | \mathbf{w}_i, \boldsymbol{\theta}) = \begin{cases} c_1(\mathbf{w}_i, \boldsymbol{\tau}) & \text{if } j = 0 \\ c_2(\mathbf{w}_i, \boldsymbol{\tau}) & \text{if } j = 1 \end{cases} \quad (17)$$

Then the conditional density function of Y_i is given by:

$$f(y_i | Z_i = j, \mathbf{v}_i, \mathbf{w}_i, \boldsymbol{\theta}) = g_{j+1}(y_i | \mathbf{v}_i, \mathbf{w}_i, \boldsymbol{\theta}_j) \quad \text{when } j = 0, 1 \quad (18)$$

and

$$h_i = P(Z_i = 1 | y_i, \mathbf{v}_i, \mathbf{w}_i, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \frac{c_2(\mathbf{w}_i, \boldsymbol{\tau}) g_2(y_i | \mathbf{v}_i, \mathbf{w}_i, \boldsymbol{\theta}_2)}{c_1(\mathbf{w}_i, \boldsymbol{\tau}) g_1(y_i | \mathbf{v}_i, \mathbf{w}_i, \boldsymbol{\theta}_1) + c_2(\mathbf{w}_i, \boldsymbol{\tau}) g_2(y_i | \mathbf{v}_i, \mathbf{w}_i, \boldsymbol{\theta}_2)} \quad (19)$$

Then, the conditional distribution of the latent variable Z_i , with respect to $(y_i, \mathbf{v}_i, \mathbf{w}_i, \boldsymbol{\theta})$, is a Bernoulli random variable (Garrido and Cepeda-Cuervo, 2012) with success h_i (9), and its probability mass function is:

$$\begin{aligned} P(z_i | y_i, \mathbf{v}_i, \mathbf{w}_i, \boldsymbol{\theta}) &= h_i^{z_i} (1 - h_i)^{1-z_i} \quad (20) \\ &= \frac{(c_2(\mathbf{w}_i, \boldsymbol{\tau}) g_2(y_i | \mathbf{v}_i, \mathbf{w}_i, \boldsymbol{\theta}_2))^{z_i} (c_1(\mathbf{w}_i, \boldsymbol{\tau}) g_1(y_i | \mathbf{v}_i, \mathbf{w}_i, \boldsymbol{\theta}_1))^{1-z_i}}{c_1(\mathbf{w}_i, \boldsymbol{\tau}) g_1(y_i | \mathbf{v}_i, \mathbf{w}_i, \boldsymbol{\theta}_1) + c_2(\mathbf{w}_i, \boldsymbol{\tau}) g_2(y_i | \mathbf{v}_i, \mathbf{w}_i, \boldsymbol{\theta}_2)} \\ &= \frac{(c_2(\mathbf{w}_i, \boldsymbol{\tau}) g_2(y_i | \mathbf{v}_i, \mathbf{w}_i, \boldsymbol{\theta}_2))^{z_i} (c_1(\mathbf{w}_i, \boldsymbol{\tau}) g_1(y_i | \mathbf{v}_i, \mathbf{w}_i, \boldsymbol{\theta}_1))^{1-z_i}}{f(y_i | \mathbf{v}_i, \mathbf{w}_i, \boldsymbol{\theta})} \end{aligned}$$

The likelihood function of the couple (Y, Z) with the sample $(\mathbf{y}, \mathbf{z}) = ((y_1, z_1), \dots, (y_n, z_n))$ and the vector of the parameters $\boldsymbol{\theta}$, is:

$$\begin{aligned} L(\mathbf{y}, \mathbf{z}; \boldsymbol{\theta} | \mathbf{V}, \mathbf{W}) &= \prod_{i=1}^n f(y_i | \mathbf{v}_i, \mathbf{w}_i, \boldsymbol{\theta}) \cdot P(z_i | y_i, \mathbf{v}_i, \mathbf{w}_i, \boldsymbol{\theta}) \quad (21) \\ &= \prod_{i=1}^n (c_2(\mathbf{w}_i, \boldsymbol{\tau}) g_2(y_i | \mathbf{v}_i, \boldsymbol{\theta}_2))^{z_i} (c_1(\mathbf{w}_i, \boldsymbol{\tau}) g_1(y_i | \mathbf{v}_i, \boldsymbol{\theta}_1))^{1-z_i} \\ &= \prod_{i=1}^n \left(\frac{e^{\mathbf{w}_i' \boldsymbol{\tau}}}{1 + e^{\mathbf{w}_i' \boldsymbol{\tau}}} g_2(y_i | \mathbf{v}_i, \boldsymbol{\theta}_2) \right)^{z_i} \left(\frac{1}{1 + e^{\mathbf{w}_i' \boldsymbol{\tau}}} g_1(y_i | \mathbf{v}_i, \boldsymbol{\theta}_1) \right)^{1-z_i} \end{aligned}$$

where V and W are the observed covariate vectors (Equations (4) and (5)). This function could be maximised using numerical algorithms to determine the maximum likelihood estimates (MLE). In the present study, we propose the mixture model to fit daily flow to estimate extremes for given return periods, in the Saint John River basin.

3.4 Application of the proposed model

The model was applied to the maximum annual flows using daily time series from the Fort Kent station to estimate the hydrological risk associated with different return periods. The model was compared with the GEV distribution, based on the AMS approach, to show the usefulness of the proposed mixture model fitted to separate independent events.

3.4.1 Choice of threshold and selection of extreme events

Daily flow time series at the Fort Kent station (Saint John River, eastern Canada) are available from 1927 to 2014 (Environnement et Changement climatique Canada 2018). As shown in Figure 3.3, which presents the polar distribution of annual maximum flows, the majority of observed annual maximums occur between March and May, with some exception in September and during the fall season. The extreme records for minimum and maximum annual flow values were observed in 1965 and 2008, with 691 and 4,630 m^3s^{-1} , respectively. Hence, the use of a selection procedure for each annual maximum flow event that provides IID of data is needed to be able to apply the EVT approach.

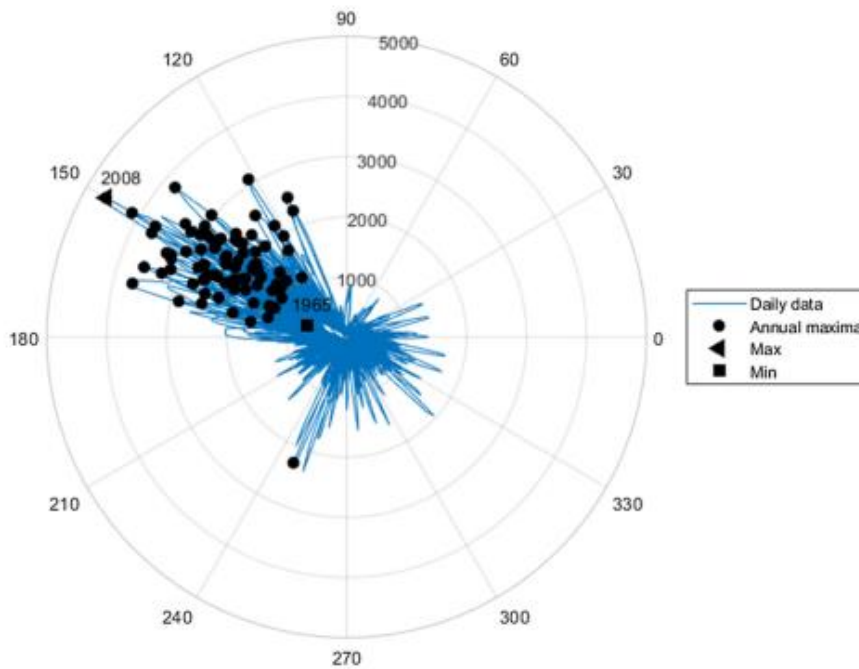


Figure 3.3. Polar plot of the daily flow at Fort-Kent station, located upstream of the Saint-John River, New Brunswick, Canada (see Figure 3.1). Hydrologic data are available from 1927 to 2014 and units are in $\text{m}^3.\text{s}^{-1}$. Annual maxima are represented by black dots, and maximum/minimum flow records are in red/green dots. Values area given for each Julian day (from 0 to 365).

Although the annual maximum flow plays a crucial role in the flood risk assessment, it does not fully characterize a flood event. On the other hand, the approach of annual maximums is more often considered for rainfall-runoff modeling or frequency analysis inference. However, for diagnostic evaluation purposes, it is imperative to analyze all important events observed during the year to evaluate the shape of the floods' distribution (Blöschl et al. 2013). Therefore, a detailed knowledge of exhaustive characteristics of floods is essential to apply a flood typology (Tarasova et al. 2019; Sikorska et al. 2015; Merz and Blöschl 2003; Fischer et al. 2019) or to select criteria for evaluating flood regimes (Viglione et al. 2009). In this study, the objective is to consider almost all information produced by the observed daily flow data set to make a complete statistical inference, so as not to limit the analysis to annual maximums. In such a case, the mixture models will capture the behavior of the daily flow for both distribution tails in addition to the central parts. The analysis of all these events is essential for understanding the nature of the processes leading to extreme floods, and therefore leads to a robust and credible assessment of the flood risk. To

select independent events, a baseflow approach was applied and only their peaks were selected to fit the GGP mixture distribution. There are few methods designed specifically for the selection of floods from daily time series. The most widely used method is based on the threshold approach, where events that are above the chosen threshold are isolated (Kundzewicz 2019; Mangini et al. 2018). Thresholds can be chosen based on average runoff or the smallest annual maximum (Fischer and Schumann 2016). This method is often used in the context of flood statistics because it easily provides the flood peak. This approach was adopted for our case study to illustrate the use of the proposed model. Fischer et al. (2021) give more details on separating floods and baseflow estimation. They propose a new method of statistical flood separation to automate the analysis of a long series of daily flows to obtain flood events. In addition, rainfall inducing floods is received with this method, which allows the estimation of rainfall inducing floods as well as the runoff coefficient. In our case, the first quartile was considered as the threshold to estimate the baseflow for the Saint John River station. The algorithm used is based on the methodology described in Fisher et al. (2021). Extremes correspond to the maximums of independent events with respect to this baseline flow. Figure. 3.4 presents some examples of independent maximum flow events using a time window over the period 1989–1993 from the daily flows observed at the Fort Kent station. All independent events were selected with a large range of intensities. Such variability could not be represented by the classical distributions usually used in hydrological frequency analysis.

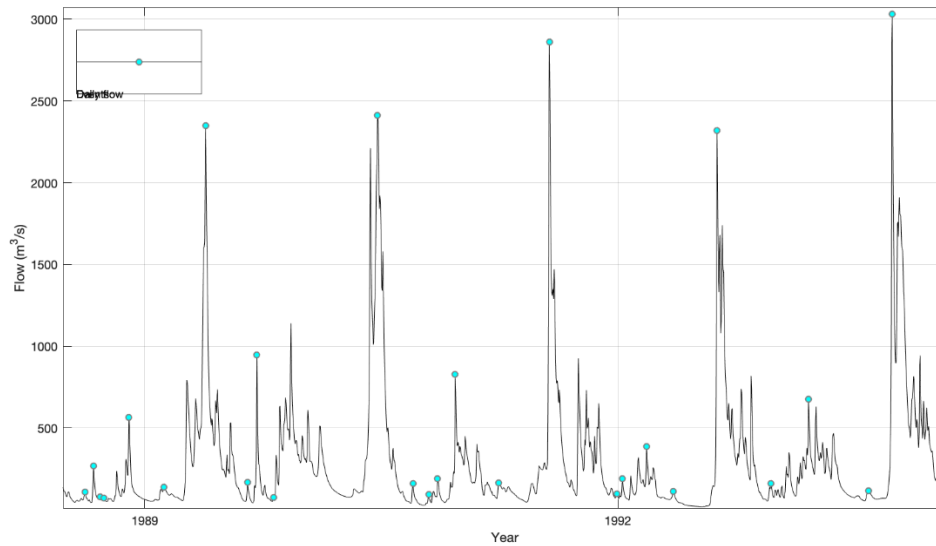


Figure 3.4. Selected events using a fixed daily baseflow at Fort-Kent (Saint John River) using the observed values over the 1989-1993 period. Each daily maximum of flow is plotted in cyan dots, and baseflow is plotted in blue line.

Because the Saint John River is one of the largest rivers in eastern Canada, the response of daily flow from any specific meteorological event can be slow, and therefore this watershed has a long recession system. For other rivers, shorter flow durations could be sufficient to ensure the independence of the selected events. Over the 88 years (1927–2014) at Fort Kent station, 362 events were selected, with a mean of $\lambda \approx 4.11$ events per year. This expected number of annual events is used to compute the probability of non-exceedance p as a function of the return period T . Then, for a given return period T the probability of non-exceedance is $p = 1 - (\lambda T)^{-1}$.

Figure 3.4 illustrates the POT approach in a general way, with a very low threshold to include almost all hydrograph events. However, for each hydrological event, the start of the flood as well as the end were determined by the approach detailed in Fischer et al. (2021). Indeed, the maximums of the selected events would be independent and will be considered to fit the mixture distribution.

3.4.2 Simulation study

A simulation study is given in this section to assess the performance of the model and present the estimation procedure. Random data sets were generated from a GGP mixture model with different characteristics of variation and skewness. A simulation study was considered to compare the performance of the GGP mixture model to the classical extreme models, i.e., AMS and POT. The comparison was realized considering the whole process by estimating the distributions of the quantiles with given return periods for the three models: AMS, POT, and GGP mixture. For this study, six parametric distributions were considered (MacDonald 2012). These distributions possess varying behaviors in terms of the shape (no mode, modal) and the lower tail with different combination of the skewness (Cs) and variation coefficients (Cv), as presented in Table 2. The simulated cases represent distributions with small, moderate, and high skewness illustrated by noncentered χ^2 (NC-Chi2) distribution (Cases 1 and 2 in Table 2), gamma (GAM, Cases 3 and 4 in Table 2) and inverse gamma distributions (INV-GAM, Cases 5 and 6 in Table 2), respectively.

Tableau 3.2. Data generation scenarios with lower and heavy tailed distributions.

	Mean	STD	Cv	Cs
Case 1: NC-Chi2 (6,6)	12	6	0.5	0.88
Case 2: NC-Chi2 (1,12)	13	7.02	0.54	0.83
Case 3: GAM (2,6)	12	8.5	0.71	1.41
Case 4: GAM (1,12)	12	12	1	2
Case 5: INV-GAM (7,70)	12	5.2	0.44	2.23
Case 6: INV-GAM (5,50)	12	7.21	0.57	3.46

3.5 Results and discussion

3.5.1 Simulation study results

The shapes of the simulated distributions are presented in Figure 3.5. Simulated values were independent and identically distributed. However, as observed from the station data sets at Fort Kent, daily flow was highly autocorrelated. This autocorrelation structure should be removed by considering a lag between two consecutive events. For the case study, separating floods and baseflow estimation are considered to ensure the selection of independent events (Fischer et al. 2021).

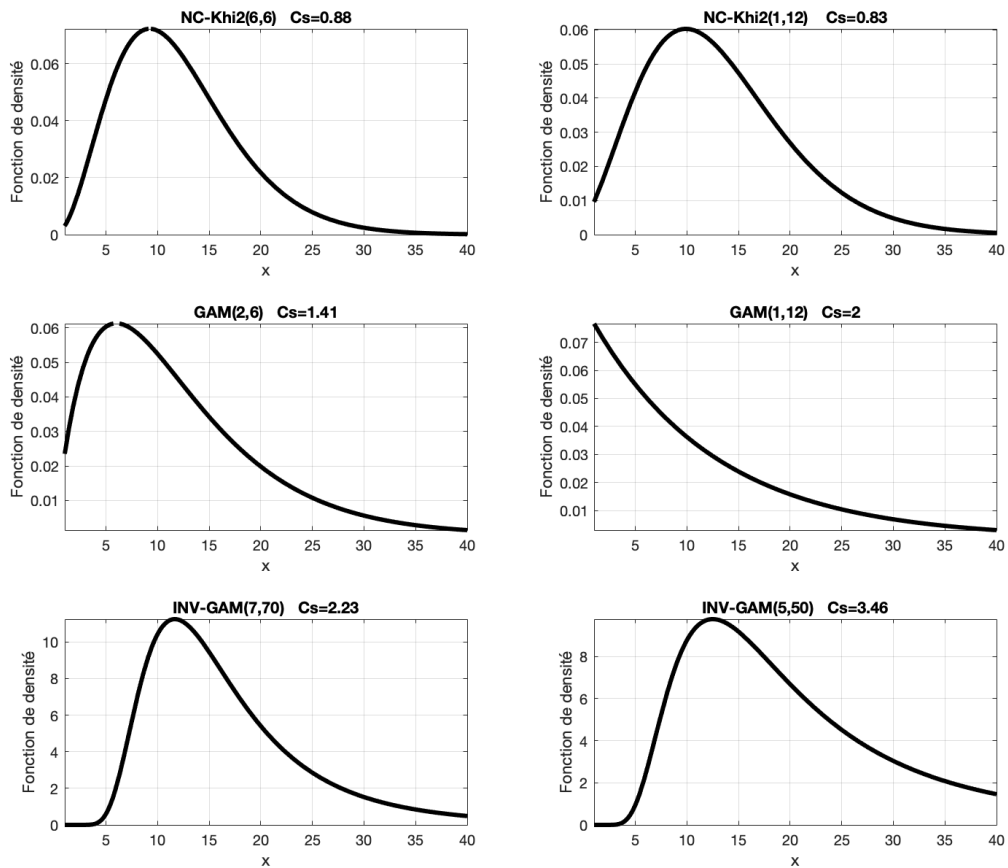


Figure 3.5. Probability density functions of the simulated distributions (Table 3.2) with different shapes.

For comparison purposes, three models were fitted to simulate daily flow data. Both AMS and POT approaches were compared with the proposed mixture model. For the POT model, the thresholds were fixed to the 90th percentile of the generated data sets. The assessment of the approaches was based on the quantile relative bias (RB) : $(RB) \left(\widehat{Q}_{T,t}/Q_T - 1 \right)_{t=1,\dots,N_s}$ (22)

and its standard error of distributions with return period $T = 100$ years. $\widehat{Q}_{T,t}$ is the estimated quantile for the t^{th} sample; $N_s = 1000$ the number of simulated samples and Q_T is the true value of the quantile.

In Figure 3.6, the boxplot shows the quantile estimators' errors for the three models: AMS, POT approach and GGP mixture model. The Generalized maximum likelihood approach is used to estimate the parameters of the GEV distribution for the AMS data, and GPD for the POT approach (Martins and Stedinger, 2001). We used the 90th percentile as threshold for the POT approach in all simulated data (presented in Figure 3.6), using the 6 cases scenarios presented in Table 3.2.

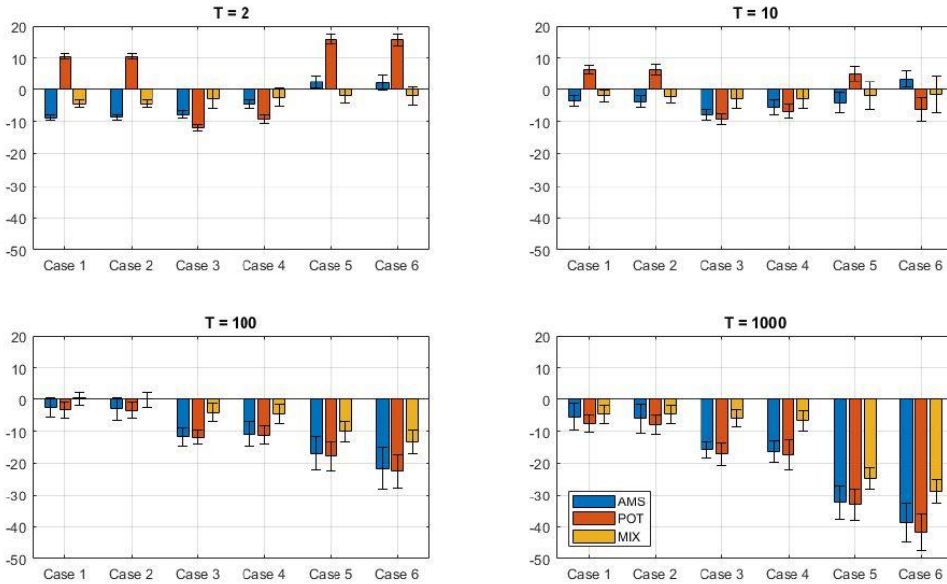


Figure 3.6. Mean relative bias (%) and its 95% confidence intervals for the quantile estimators with return period $T = 2, 10, 100$ g and 1000 years given by AMS, POT and GGP mixture models.

The quantile of return period T is computed for the non-exceedance probability, $p = 1 - 1/365 T$, of the daily dataset. The relative bias presented in Figure 3.6 as percentage of the real value of the N_s quantiles

generated with the six models with different tail behaviour. The first two models with light tail, the 3rd and 4th have moderate tails and heavy tails for the last two Inverse Gamma with high skewness.

Results show that for $T = 2$ or 10 years, the relative bias is less than 10%, for almost all cases and both AMS and mixture models. The POT approach may lead to high bias, especially for $T = 2$. This is due mainly to threshold selection, which affects the lower tail of the distribution. For the quantiles of return period $T = 100$ years, the mixture model presents very small relative bias (less than 5%) and standard error, for the four first generating distributions. When the samples are generated by one of the two last models, with heavy tail behaviour, the relative bias of the mixture model associated to $T = 100$, is less than 12%, while it can exceed 21% and 22% for the AMS and POT approaches, respectively (as shown in Figure 3.6).

Quantiles of high return periods are usually used for the design of large hydraulic dams, and their estimations may have large differences among potential models. When samples are generated from any of the four first models (Cases 1, 2, 3 and 4 with small and moderate skewness), the relative bias of the quantile with $T = 1,000$ years, estimated by mixture distribution, is less than 8%. However, the estimators given by the AMS and POT approaches may have relative biases of 16% and 18%, respectively. Higher values of the relative bias are reached for Cases 3 and 4 (presented in Table 2, i.e., gamma distribution with skewness equal to 1.41 and 2). In the case of high return period and for generating distribution with very heavy tails (Cases 5 and 6 in Table 3.2), the quantile estimators of the quantiles may have large negative relative biases. This underestimation is more important for AMS and POT approaches, than for the mixture model (see Figure 3.6). For $T = 100$, the absolute value of the relative bias exceeds 15% and 20% for both AMS and POT, with samples generated by distributions of cases 5 and 6 (Table 3.2), respectively. The proposed mixture model leads to an important improvement of the quantile estimation. The relative bias is less than 10%, in absolute value, for case 5 and 13% for case 6. For $T = 1000$, the absolute value of the relative bias exceeds 30% and 38% for both AMS and POT, with samples generated by distributions of cases 5 and 6 (Table 3.2), respectively. The relative bias of this quantile estimates by mixture model is equal to -22% and to illustrate the difference between the proposed model and the hybrid model, the quantile of each distribution is presented in figure 3.7.

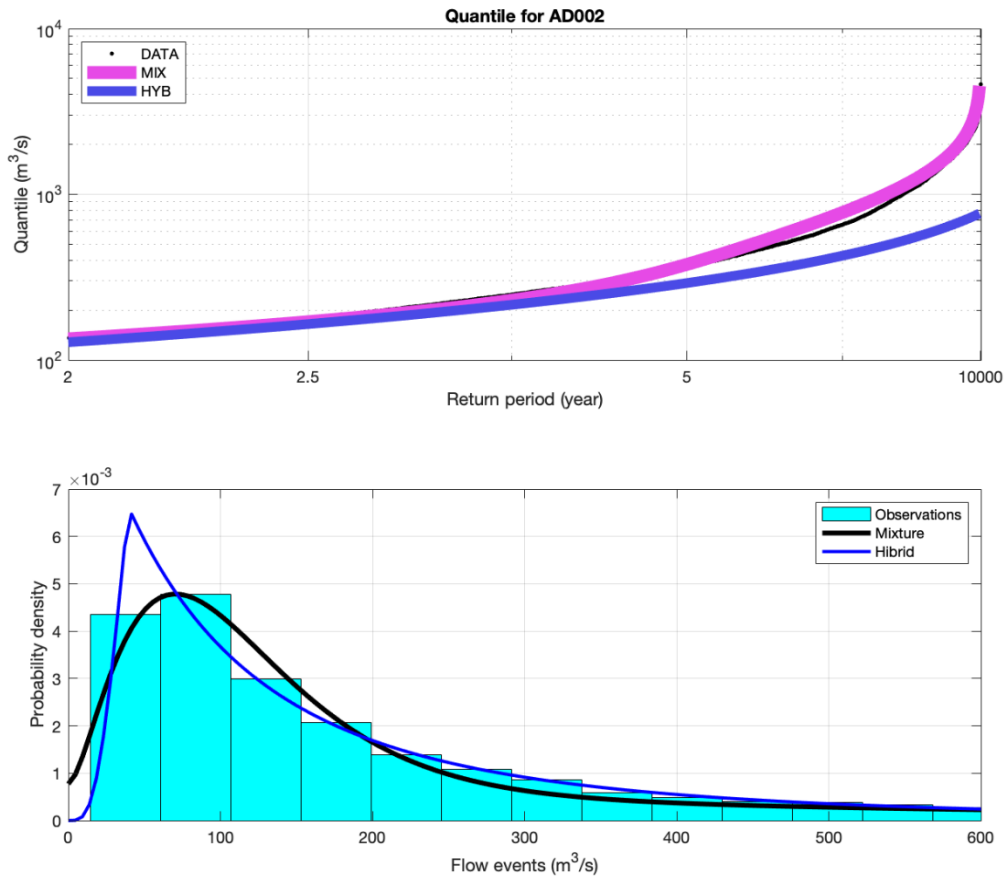


Figure 3.7. (a) Quantiles for given return periods, estimated by Mixture and Hybrid models with the same components Gamma and GPD. (b) The pdfs of the mixture and Hybrid models for the case study.

The comparison of the mixture model with hybrid model clearly shows the problem of abrupt change between the two components for the hybrid model. Indeed, this model is like the classical POT model, where the threshold is estimated with the parameters of the gamma distribution and GPD. However, the mixture model allows the construction of a new smooth distribution with the advantages of the two distributions: gamma for the central part and GPD for extremes. Results show the capacity of the mixture model to give a smooth fit of the data histogram. The estimated return period events given by the mixture model are also higher than the hybrid model estimates for a T value larger than 5 years.

3.5.2 Application results

In this section, we present the fit of the studied models to flow data records at the Fort Kent station on Saint John River. The PDF of the mixture model with the two components is flexible to fit many data shapes (Figure 3. 8). The advantage of the mixture model is mainly due to the length of the data set, which increases the collected information on all support of the fitted distribution. This is also important to reduce the parameters' estimation uncertainty. The PDF of the GEV fitted to the annual maximums data set, shown in Figure 3. 7(b), is presented for comparative purposes to estimate events with given return periods. The comparison of the two results [Figures. 3.8(a and b)] clearly shows that the GEV approach eliminates events below 500 m³=s in the lower tail, whereas the mixed model based on the separation of events and the choice of threshold takes these events into consideration in the modeling. Therefore, the mixed model does not neglect extreme events that can have a destructive effect on the infrastructure and the population in general.

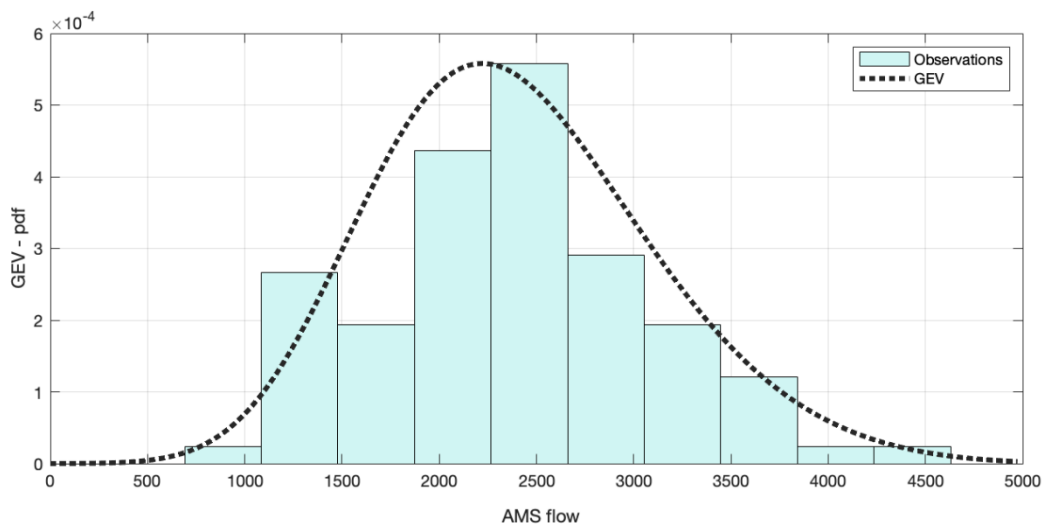
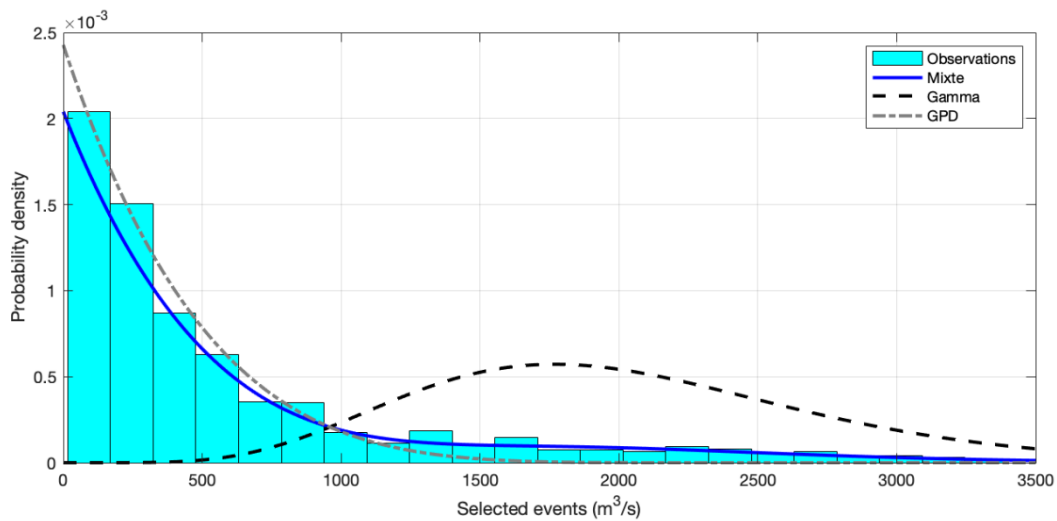


Figure 3.8. Probability density function of (a) the GGP mixture model for selected daily streamflow events, and (b) GEV for the series of maxima at Fort-Kent, Saint-John River.

The quantiles of the mixture model were estimated by numerical approximation of the PDF integrals. For the case study, we used a Monte Carlo simulation to estimate the quantiles of the mixture model with their confidence intervals deduced by bootstrap of $N \approx 1,000$ realizations (Bickel and Freedman 1981). In Figure 3. 9, the quantiles corresponding to return periods T from 1 to 1,000 years, for the GGP mixture model fitted to all independent events and the GEV distribution fitted to AMS (solid circles) are presented.

The confidence intervals of the AMS-GEV model were deduced from asymptotic distribution of the maximum likelihood estimates.

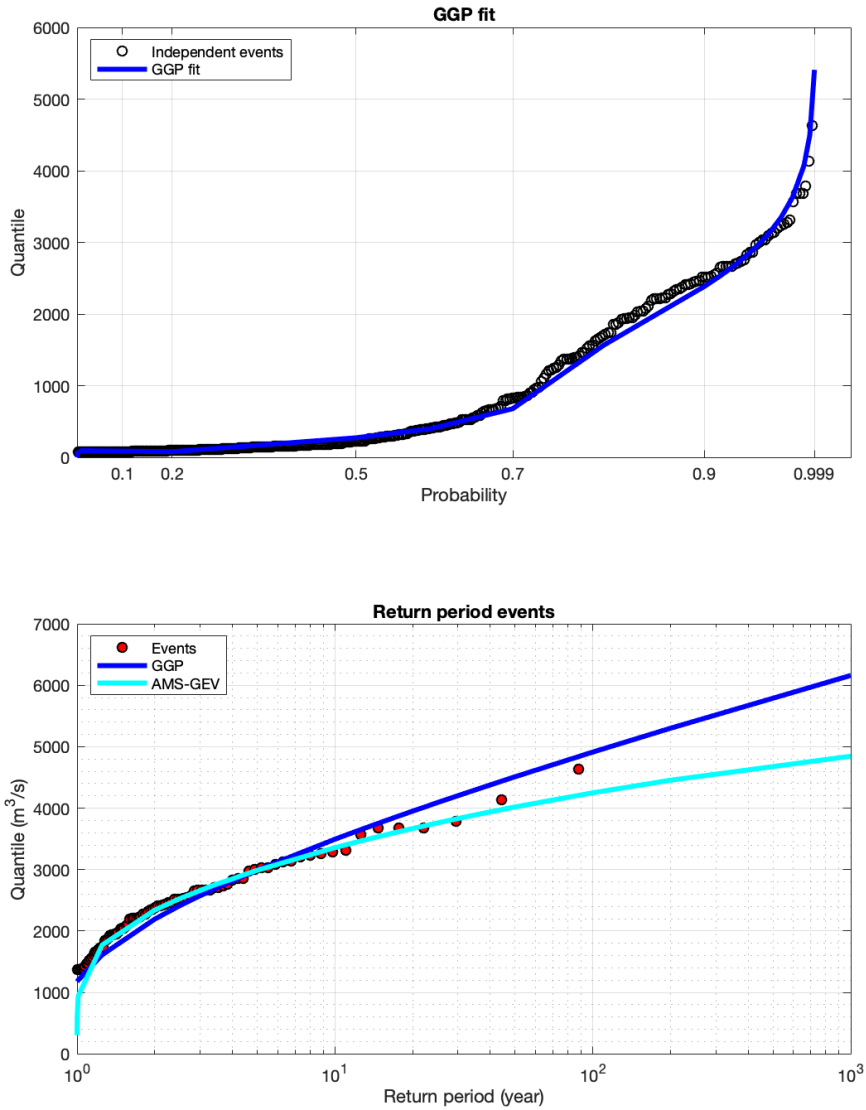


Figure 3.9. (a) GGP mixture fit to all independent event at Fort-Kent (Saint-John River) and (b) the AMS-GEV model fit with quantiles of return periods for both GGP mixture and GEV model at Fort-Kent.

As revealed in Figure 3. 9, the main advantage of the mixture model is its capacity to fit the lower tail as well as the upper tail of the daily distribution data (as also noted in Figure 3.7). Results illustrate such

characteristic of the proposed model. For moderate return period $5 \leq T \leq 20$ years, representing the central part of the distribution, the mixture model estimated quantiles are close to the AMS fitted by GEV distribution (Figure 3.10).

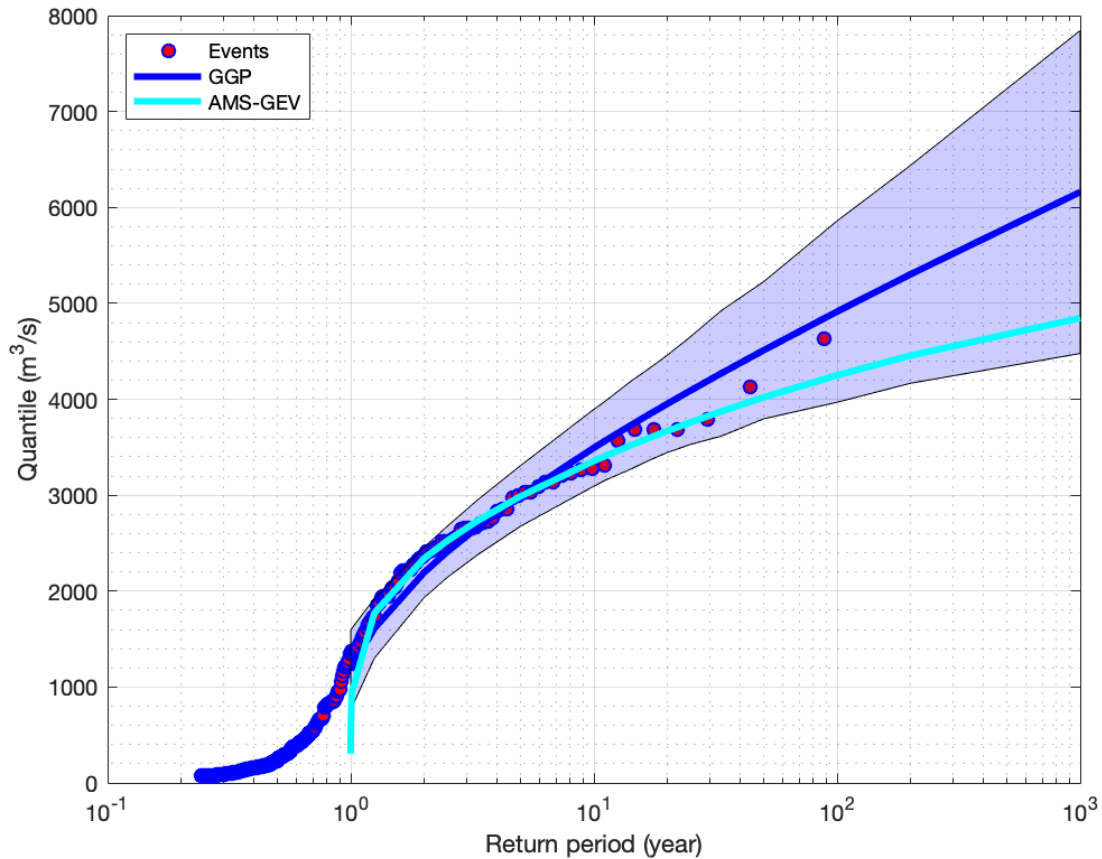


Figure 3.10. Confidence intervals at 95% level for the quantiles of the GMP mixture model computed by bootstrap. The daily flow data series from the Fort-Kent station is used, from all data covering the 1927-2014 period.

For small return periods, the quantiles of the mixture model are significantly higher than that given by the GEV model. Indeed, the existence of one small maximum affects the GEV fit and leads to a small quantile for the lower tail. We also note the capacity of the mixture model to represent a very heavy right tail. An adequate fit of both tails is of a great importance for environmental studies, such as the estimation of in-stream flow in winter, as well as dams' conception to estimate extreme events

In Figure 3.10, the confidence level for the quantiles of the proposed model computed by bootstrap represents the range of values within which we are 95% certain to find the true value we are looking for. According to the results obtained, the GGP mixture model gives an adequate fit of both tails for different return periods.

The flow quantile is well centered in the confidence interval for the different return periods (Fig. 10). This approach based on the separation of events allows keeping as many events as possible while ensuring independence between these events. With this, even complex forms of flood hydrographs, such as multiple peaks or prefloods, can be separated

The proposed approach offers flexibility to model extremes for dependent or nonstationary time series and adequately describes the central part of the observed frequencies, as well as the tails. Such models could be used for modeling extremes in the presence of inhomogeneity such as seasonality or mixture rainfall and snowmelt flood regimes (Filipova et al. 2019), variation between ebb and flood conditions for tidal velocity (Novo and Kyojuka 2019), and climate change effects on precipitations and floods (Yilmaz et al. 2017; Yang et al. 2019).

3.6 Discussion

An important hypothesis that should be studied is that of the homogeneity of the series observed. Indeed, in terms of seasonal variability, the meteorological phenomena at the origin of floods are multiple. In a region such as the one considered in the case study, there are spring floods caused by snowmelt combined with liquid precipitation, or unforeseen events that form torrential flows. These are often caused by severe thunderstorms. For the case study, these events, flash floods, occur in summer or early fall. This diversity requires the use of probabilistic models considering the heterogeneity of observed events. Mixture models, by their construction, make it possible to represent unusual forms of the probability densities that could be observed for certain regions. The distribution of the mixture makes it possible to have high probabilities for low flows and a heavy tail for the right tail corresponding to extreme events. The usual distributions (GEV, gamma, lognormal, Pearson) do not make it possible to represent all these events. The overshoot model, based on the GPD distribution alone, faces the problem of threshold choice and high uncertainty due to the reduced number of events retained above the threshold.

3.7 Summary and conclusions

In this study, we introduced a general form of the GGP mixture model to combine the properties of the central part and the lower and upper tail of extreme values. We examined the mixture weight parameter and defined a new latent variable accordingly. We detailed the use of the likelihood expression for model fitting, especially in the case of a model with a high number of parameters. Simulation studies were considered to compare the mixture model with the AMS-GEV model. Results show that the AMS-GEV model leads to an underestimate of the quantiles with high probabilities of non-exceedance. A case study was performed for an event with return period 100 years and shows that the classical models (GP) lead to up to 10% underestimation of the quantile generated data. Also, the range of the quantile estimators given by AMS-GEV model was larger than that estimated by the GGP mixture model, especially for low values of the gamma component's skewness (5% for the mixture model and 20% for the AMS-GEV model).

Finally, we illustrated the use of the proposed mixture on the observed daily flow at Fort Kent station, located upstream on the Saint John River watershed (New Brunswick, Canada). Results illustrate the flexibility of the GGP mixture model to capture the shape of the data distribution for both tails. This property is very important when dealing with hydrological extremes, especially in climate zones like Canadian conditions. Indeed, Saint John River floods are mainly due to a combination of several meteorological and environmental (surface) conditions. The spring extreme flood events are caused by snowmelt combined with rain on snow occurrences from March to April. Such a complex system or processes combined with seasonal and intra-annual variability leads to unusual distributions with a heavy tail. The first three moments are not sufficient to describe the shape of these distributions. The mixture model with gamma (subexponential class of distribution) and generalized Pareto (heavy-tailed class) offers more flexibility to represent these observed hydrometeorological phenomena. Comparison of the quantiles estimated by the mixture model and the classical models shows the capacity of the proposed model to adequately fit both sides of the daily data set in terms of bias and variance. The case study illustrates the flexibility of the proposed model to fit a variety of empirical distributions generated by inhomogeneous hydrological events. However, because the case study used only a data set from the upstream sector of the Saint John River flow (i.e., from the Fort Kent station), to illustrate the importance to use the mixture model, further studies need to be done over the downstream sector of the river. Indeed, other sectors have different climatic conditions with respect to the upstream conditions of the basin, along with different physiographic characteristics. The hydrological events in the downstream are more complex

and are usually caused by other compound effects from various meteorological conditions. Nevertheless, the form of the mixture model adapts well to the data set and can be easily used from other case studies and hydrometeorological conditions. In fact, the comparison, with the GEV model with the annual maximum series, indicates the importance to take into account all available information on the variance of the observed events, and not restrict it to a single observation per year. Statistical implementation of the mixture model is simple and user friendly. Only five parameters need to be estimated, and with the size of the series is becoming large (on average 10 events per year), an efficient estimation is ensured for the parameters. The confidence intervals show the estimators' efficiency, with a very narrow width even for large return periods.

3.8 Credit authorship contribution statement

Nawres Yousfi: Methodology, Investigation, Data curation, Formal Analysis, Writing - Original Draft. Salah-Eddine El Adlouni: Conceptualization, Methodology, Writing- Original draft. Simon Michael Papalexiou: Conceptualization, Methodology, Writing - Review & Editing. Philippe Gachon: Conceptualization, Methodology, Writing - Review & Editing.

3.9 Declaration of competing interest

No potential conflict of interest was reported by the authors.

3.10 Data availability statements

Hydrometric data used in this study are available on the official site of environment Canada (Environnement et Changement climatique Canada, 2022).

Matlab codes were developed by the authors. Direct requests for these codes can be made to co-author Salaheddine El Adlouni (E-mail: salah-eddine.el.adlouni@umoncton.ca).

3.11 Acknowledgments

We acknowledge the financial support from the Natural Sciences and Engineering Research Council of Canada (NSERC), individual Grants of Pr. Philippe Gachon (NSERC-RGPIN-2016-06436 and NSERC-RGPIN-2022-05032) and of Pr. Salah El Adlouni (NSERC- RGPIN-2019-05746). We also acknowledge other financial supports from the Strategic research chair of UQAM hold by Pr. Philippe Gachon, by the UQAM under the

scholarship for the exemption of additional tuition fees for foreign students, and by the UQAM's Faculty of Sciences under the programme "faculty financial support".

CHAPITRE 4

Discussion générale

4.1 Mise en contexte

La compréhension, l'analyse et l'amélioration de la modélisation du risque d'inondation sont devenues un aspect important des politiques de gestion et de réduction des risques d'inondation (Oubennaceur et al., 2022), surtout après les inondations printanières qui ont touché l'Est canadien en 2017, 2018 et 2019 (Sécurité Publique Canada, 2023). Dans ce contexte, il est important de mieux analyser ces phénomènes et d'améliorer les outils disponibles dans le cadre de la gestion du risque hydrologique, soit de permettre une meilleure préparation et prévention des inondations.

La fréquence et le niveau de risque associé à ce genre d'événement peuvent être expliqués par la variabilité saisonnière des cycles locaux de précipitations, des caractéristiques d'évaporation et de la fonte de la neige influençant la plupart des régimes d'écoulement fluvial au Canada (Dettinger et Diaz, 2000). Dans ce contexte, et afin d'améliorer l'estimation des crues printanières, responsables des inondations et des dégâts associés, il est important d'améliorer les outils de modélisation disponibles ainsi que les variables utilisés dans les estimations.

Compte des enjeux des inondations du bassin versant Saint-Jean, notamment reliés aux caractéristiques, à la fréquence et à l'ampleur de celles-ci (cf. Duhamel et al., 2022; Das et al., 2022; Yevdokimov, 2022), de nouveaux modèles probabilistes ont donc été proposés dans ce travail pour tenir compte des aléas météorologiques dans l'estimation des crues printanières à l'est du Canada. Une analyse des données hydrométéorologiques et climatiques historiques a permis de proposer des nouvelles variables dérivées des données brutes, et permettant de mieux évaluer les inondations printanières enregistrées aux stations hydrologiques utilisées (en amont et en aval du bassin versant).

Les modèles GEV ont été utilisés dans un cadre non-stationnaire, et en intégrant les covariables développées dans des équations B-Splines afin d'exprimer la dépendance entre le débit estimé et les variables développées. C'est une nouvelle approche pour intégrer le maximum d'information explicative dans le modèle afin d'améliorer la qualité de l'estimation. Ces modèles offrent un moyen pour comprendre et analyser les variables météorologiques et climatiques afin de développer des modèles adéquats et spécifiques au site d'étude.

Dans un deuxième temps, un modèle GGP a été proposé afin d'améliorer la base de données historiques des événements à tester. En effet, le fait de prendre un seul événement par année (maximum annuel) néglige certains événements moins importants que le maximum annuel (crues de l'automne capable d'inonder le site). Dans ce contexte, la deuxième méthode fusionne deux modèles probabilistes (GP) et Gamma et profite de la distribution de ces lois probabilistes. La manière de choisir les événements garantit l'indépendance de ces événements.

4.2 Résultats obtenus et limites

L'application du modèle GEV-B-Splines à plusieurs covariables a permis de réaliser plusieurs scénarios spécifiques aux deux stations hydrométriques testées, à savoir : Fort-Kent et Mactaquac, situées dans la province du Nouveau-Brunswick en amont et en aval du bassin versant du fleuve Saint-Jean, respectivement. En se basant sur la théorie des valeurs extrêmes (Fisher et Tippett, 1928), pour appliquer le modèle GEV, il faut respecter la nature de la série de données qui doit être une série temporelle d'une variable ayant une valeur maximale sélectionnée sur un bloc de temps prédéfini (Bloc Maxima) en utilisant le maximum de données disponibles. L'utilisation d'une seule observation par année ne permet pas d'analyser le comportement hydrologique et l'estimation du risque. Il est donc important d'utiliser des modèles qui intègrent le maximum d'information comme le modèle GEV-B-Splines utilisé dans le cadre de cette thèse de doctorat. Ce modèle permet d'intégrer plusieurs covariables liées aux crues printanières, et par conséquent améliorer la qualité de l'information dans l'estimation du débit.

En se basant sur les données hydrologiques, météorologiques et climatiques des stations jaugées (voir tableau 2.1.), de nouveaux indices ont été développés pour le bassin versant du fleuve Saint-Jean afin d'être intégrés dans le modèle GEV-B-Splines à plusieurs covariables. Ces indices étaient développés d'une manière spécifique au site d'étude en analysant les événements de crues avec la date d'enregistrement et en développant des indices climatiques et météorologiques basés sur les enregistrements qui précèdent la date de l'événement hydrologique. L'intégration de ces indices dans le modèle GEV-B-Splines a permis d'estimer les quantiles de débits extrêmes pour différentes périodes de retour et sachant les conditions climatiques et météorologiques du site. Grâce aux analyses de présélection des covariables, basées sur le test de corrélation de Spearman, certaines covariables ont montré une corrélation significative positive avec les débits printaniers surtout pour la station Fort-Kent (comme le $RAIN_B$, le $SNOW_B$), alors que l' AMO_B a montré une corrélation significative négative pour la même station, et une corrélation négative significative pour la station Mactaquac. Les résultats de cette corrélation négative obtenue avec

l'AMO viennent confirmer l'étude de McCabe et al. (2008) qui a montré que l'AMO est le principal facteur climatique affectant la variabilité interannuelle du débit fluvial en Amérique du Nord. Les autres covariables présentaient une corrélation non-significative avec le débit historique, peu importe les deux stations testées. L'intégration des covariables dans les modèles GEV-B-Splines a permis de réaliser vingt-neuf scénarios pour chaque site jaugé, et d'identifier l'effet synergique ou antagoniste de certaines covariables climatiques ou météorologiques sur le débit de crue printanier.

Cet effet synergétique a été démontré dans les simulations des modèles sélectionnés, soit ceux qui intègrent ou pas une ou plusieurs covariables fortement liées avec le débit (selon les résultats du tableau 2.5). En effet, le modèle sélectionné pour la station de Fort-Kent intégrant le BWA_B , a présenté une faible corrélation avec la série des débits de la station, soit une valeur de 0,05 (Tableau 2.1). Autrement dit, la corrélation entre le débit et cet indice climatique était non significative. Par contre, en intégrant le BWA avec d'autres indices météorologiques, le modèle était sélectionné, ce qui témoigne de l'effet synergique de cet indice avec les autres covariables sélectionnées dans la station Fort-Kent. Ces résultats viennent confirmer ceux de Coulibaly et al. (2000) qui ont constaté que l'intégration de l'indice BWA dans la prévision du ruissellement annuel au nord-est du Canada donne de meilleurs résultats que l'inclusion de l'indice NAO.

Finalement, les résultats pour la station de Fort-Kent ont démontré que l'intégration du vecteur ou de trois covariables (TX_D , $PTOT_B$ and BWA_B) a donné de meilleurs résultats dans l'estimation des crues, selon le critère AIC. Par contre, pour la station de Mactaquac les résultats ont révélé un comportement hydrologique dépendant d'une autre combinaison de facteurs, à savoir: ($PTOT_B$, $RAIN_D$, and NAO_B). Bien que les variables sélectionnées dans les deux stations ne sont pas les mêmes, le BWA et le NAO sont fortement corrélés, et la température maximale est liée aux événements de pluie sur neige au printemps et aux extrêmes de précipitation (cf. Wazneh et al., 2021 et 2022). Ceci confirme la dépendance indirecte entre toutes les variables sélectionnées dans les deux stations.

La sélection des covariables $PTOT_B$ et $RAIN_D$ dans les deux modèles confirme l'étude récente de Burn et Whitfield (2023) concernant l'influence croissante des précipitations sur le changement des régimes d'inondations au Canada. Budhathoki et al., (2022) ont également démontré, suite à une analyse du bilan hydrique du bassin versant du fleuve Saint-Jean, que le ruissellement et l'évapotranspiration représentent environ l'équivalent de 99% des précipitations totales qui tombent sur ce bassin versant, précipitations qui exercent un rôle crucial sur le régime hydrologique et sa modélisation au sein de ce bassin. De plus,

selon la même étude, les simulations des débits futurs prévoient des débits hivernaux plus élevés au cours des périodes 2041-2070 et 2071-2100, par rapport à la période 1991-2020. Ceci met plus en valeur l'importance de développer des outils de modélisation du risque hydrologique, spécifique à ce bassin problématique ayant un comportement hydrologique très variable dans le temps et dans l'espace. Plusieurs facteurs peuvent expliquer cette forte variabilité spatio-temporelle, notamment selon les conditions météorologiques et climatiques variables qui affectent l'amont (où se situe Fort-Kent) et l'aval (où se situe Mactaquac). Par ailleurs, la station de Mactaquac est localisée en aval des barrages et son écoulement est donc contrôlé contrairement à la station de Fort-Kent dont l'écoulement en amont est naturel et non contrôlé. D'autres facteurs qui n'ont pas été investigués dans notre étude, qui comporte donc plusieurs limites, pourraient expliquer cette variabilité dans les débits simulés ou estimés, notamment l'occupation du sol, l'infiltration, la profondeur de la nappe, le type du sol, l'aménagement du terrain et l'imperméabilisation. La limite de l'application du modèle proposé dans cette thèse est donc dans la non-intégration des covariables d'origine physiographique qui influencent le cycle hydrologique et le ruissellement, et par conséquent le déclenchement des inondations printanières. De plus, les facteurs d'exposition et de vulnérabilité affectés par les aménagements du territoire peuvent aussi être importants à inclure, car ils peuvent affecter non seulement le risque d'inondations mais aussi ces conséquences.

4.3 Nouvelles connaissances

L'originalité du développement fait dans cette thèse réside dans les deux modèles probabilistes dont les applications sont spécifiques au site d'intérêt. À notre connaissance, ces indices sont développés pour la première fois, de cette manière, quelque soit pour Fort-Kent ou pour Mactaquac.

Le premier modèle offre une approche pour estimer les crues en intégrant le maximum d'information. Bien que ce modèle offre un moyen d'améliorer la qualité de l'estimation, il nécessite une analyse approfondie et une connaissance du cycle hydrologique et ses différentes interactions avec le climat afin de développer les indices climatiques les plus significatifs.

Le modèle de mélange offre une approche combinant deux distributions probabilistes qui permettent de chercher plus qu'un seul événement par année. Le modèle permet également de sélectionner les événements indépendants et par conséquent combler la limite de l'approche du Block Maxima sans être obligé de définir un seuil.

4.4 Synthèse et recommandations

Cette thèse de doctorat propose des liens à prendre en considération entre les indices climatiques et météorologiques et les débits locaux observés. Plusieurs indices ont été développés et testés à l'aide des scénarios afin de choisir la meilleure combinaison à appliquer selon le lieu. Au préalable, plusieurs variables ont été éliminées et d'autres variables n'ont pas été considérées comme l'intensité, la durée et la fréquence de l'événement de crue, ainsi que les indices de blocage atmosphérique qui ont un effet important sur les régimes de précipitation et de températures dans l'ouest de l'Atlantique Nord (cf. Wazneh et al., 2021). Il est donc recommandé dans de futurs travaux d'approfondir l'analyse de ces variables afin de développer des indices et des combinaisons de covariables qui reflètent les changements en cours et à venir du cycle hydrologique du bassin versant du fleuve Saint-Jean.

De plus, parmi les hypothèses utilisées, l'effet de l'occupation du sol, de l'urbanisation et de l'intervention humaine n'a pas été considéré, alors que des études récentes ont montré l'importance de ces facteurs sur l'occurrence des inondations dans le bassin versant du fleuve Saint-Jean (Yu and Bourque, 2022). Il est donc recommandé de poursuivre ce travail en intégrant ces covariables dans le modèle statistique afin d'évaluer l'effet de ces facteurs sur les inondations.

Par ailleurs, il serait pertinent de tester les modèles proposés pour d'autres comportements hydrologiques dans d'autres bassins versants vulnérables et/ou exposé aux crues printanières, afin de définir un modèle d'estimation du risque hydrologique compatible avec les aléas hydrométéorologiques propres aux différentes échelles locales concernées. Par exemple, l'utilisation de données de grille de variables météorologiques, comme les réanalyses (ex., ERA5-Land; Muñoz-Sabater, 2021) ou les produits comme Daymet (Daily Surface Weather Data) disponible à 1 km de résolution (ORNL DACC, 2023) pourraient permettre de développer des informations spatiales à l'aide de toutes les données de stations hydrologiques présentes au sein du bassin. Le développement des covariables basées sur les données de grille permettraient également de tester d'autres combinaisons de covariables.

Par ailleurs, le nouveau modèle de mélange proposé offre des pistes innovantes de recherche. En effet celui-ci permet l'intégration de différentes covariables testées comme celles utilisées dans le modèle GEV-B-Splines proposé. Il serait pertinent de poursuivre son développement pour d'autres bassins versants, et ainsi évaluer son potentiel dans l'estimation des crues printanières dans d'autres régions. Le modèle a été développé dans sa forme générale et peut donc être utilisé dans n'importe quel site là où les données

hydrométéorologiques sont disponibles. Les résultats peuvent être comparés avec ceux du modèle GEV-B-Splines afin de choisir le meilleur modèle d'estimation du risque hydrologique à l'échelle locale.

CONCLUSION GÉNÉRALE

1. Synthèse des approches et des résultats

L'objectif général de cette thèse de doctorat était de modéliser les crues en régime nival à l'est du Canada en proposant des améliorations de certains modèles probabilistes disponibles, afin d'intégrer les conditions météorologiques et climatiques (en mode non stationnaire) dans la modélisation. L'application des modèles développés était sur le bassin versant Saint-Jean particulièrement vulnérable et exposé à l'inondation printanière.

Les objectifs spécifiques étaient de : 1) Identifier les variables et leur combinaison à l'échelle du bassin versant responsable des crues printanières observées, voire des risques d'inondations; 2) Développer des modèles probabilistes pour l'estimation des événements de crues en intégrant plus qu'une covariable météorologique et/ou climatique; et 3) Développer un modèle mixte qui vient combler les limites des méthodes classiques (*Peaks-Over-Threshold*, POT et *Block Maxima*, BM) et ainsi proposer un modèle de mélange qui améliore l'estimation des extrêmes de débits et réduit les erreurs associées.

Les variables météorologiques et climatiques ont été identifiées événement par événement de crue, afin d'analyser leurs influences sur le débit avant la date de ladite crue. La combinaison des indices développés a été évaluée avec des tests préliminaires (tests de corrélations linéaires, non-linéaires et de tendances de Man-Kendall). Plusieurs combinaisons ont été testées afin d'identifier les indices qui ont un effet significatif sur la valeur de débit printanier au niveau des stations testées. Cette évaluation a permis de choisir des combinaisons de covariables à intégrer dans le modèle GEV-B-Splines avec une, deux et trois covariables afin de choisir le modèle le plus adéquat dans l'estimation des quantiles de débits pour chaque station testée, et pour différentes périodes de retour. Le critère AIC a été utilisé pour identifier le meilleur modèle non stationnaire qui a été comparé avec le modèle stationnaire classique qui n'intègre pas de covariables. Le modèle choisi pour la station de Fort-Kent, située en amont du bassin versant du fleuve Saint-Jean, combinait la moyenne mensuelle de la température quotidienne maximale pendant le mois de la crue, les précipitations totales accumulées entre le mois de novembre précédent et le mois de la crue, et les valeurs de l'indice BWA moyennées entre décembre et mars. Pour la station de Mactaquac, située en aval du bassin, la meilleure combinaison dans ce cas incluait les précipitations totales avant l'événement, comme dans le cas de Fort-Kent, le cumul mensuel de précipitations quotidiennes pendant

le mois de l'inondation, et les valeurs de NAO moyennées entre le mois de décembre et le mois de mars. Ces résultats ont confirmé que l'intégration de plus de paramètres ou de covariables améliore l'estimation des quantiles de crue. De plus, des indices de téléconnexion ont montré un effet de seuil (0.5 pour le BWA_B et 0 pour la NAO_B). La monotonie du quantile estimé change en atteignant le seuil. À noter que comme la distribution GEV utilisée dans l'article présenté au Chapitre 2 et basée sur l'approche BM qui consiste à identifier un seul maximum par année, l'élimination de plusieurs événements répétitifs au cours d'une année a été au préalable réalisée.

Un modèle de mélange a été développé afin d'améliorer l'estimation des extrêmes de débits et réduire l'erreur d'estimation en lien avec la taille de l'échantillon disponible. Ce modèle de mélange est une combinaison pondérée des distributions Gamma et GPD. En appliquant ce modèle sur les données de la station de Fort-Kent, les résultats ont montré une amélioration de la précision des estimations par rapport aux méthodes classiques GEV, GPD et Gamma. L'application avec les séries de débits de Fort-Kent était une illustration de la flexibilité du modèle au niveau de la forme de la distribution des événements retenus. Le développement du modèle de mélange et les résultats obtenus ont été présentés dans l'article 2 présenté au chapitre 3 répondant ainsi au troisième objectif spécifique. Ce modèle offre une nouvelle piste de modélisation hydrologique, surtout en hydraulique fluviale où la conception des grands ouvrages hydrauliques est basée sur la modélisation du débit maximum selon la durée de service ou la période de retour d'intérêt.

2. Nouvelles perspectives

Développer des modèles spécifiques à la station hydrométrique

Généralement, l'analyse fréquentielle est appliquée à une série de données annuelles (Block maxima) et le modèle GEV (stationnaire ou non-stationnaire) donne des estimations de quantiles pour différentes périodes de retour. Les valeurs des débits extrêmes estimés, dans le cas d'un bassin versant de 25 km² et plus, sont souvent utilisées comme intrants dans les modèles hydrauliques comme HEC-RAS afin d'estimer les hauteurs d'eau maximales et concevoir par conséquent les ouvrages hydrauliques (par exemple les ponceaux et les ponts). Cependant, l'est du Canada est caractérisé par un régime nival et des crues printanières exceptionnelles. L'analyse des facteurs physiques (variables météorologiques et indices climatiques) responsables des crues printanières, permet de mieux choisir les variables responsables de la réponse hydrologique d'un bassin versant suite à une crue. Cette méthode d'analyse a abouti à développer

des covariables spécifiques aux événements de crues enregistrés, et par conséquent le modèle GEV-B-Splines à plusieurs covariables qui était spécifique aux stations étudiées (Fort-Kent et Mactaquac). Dans ce cas, l'analyse fréquentielle appliquée converge vers les statistiques de débits de la station testée, selon le critère statistique utilisé (AIC). Ce genre d'analyse permet donc d'améliorer la qualité et la quantité d'information à considérer, et par conséquent de fournir des estimations plus spécifiques au site jaugé.

La combinaison des paramètres physiques les plus significatifs impliqués dans le cycle hydrologique avec l'analyse des données historiques des débits permet d'inclure les avantages de deux approches de modélisation, et de proposer un GEV-B-Splines à plusieurs covariables significatives et spécifiques à un site prédéfini. Cette nouvelle approche permet de mieux modéliser le risque hydrologique et potentiellement de contribuer à réduire les désastres associés en améliorant les outils de prévention.

Proposer un modèle de mélange avec covariables

La deuxième approche du modèle de mélange ouvre la porte à une nouvelle méthode d'estimation probabiliste améliorée, et qui peut être également jumelée avec les avantages du modèle déterministe (en intégrant des covariables physiques basées sur le même raisonnement). Ce modèle pourra être intégré dans un logiciel comme HYFRAN PLUS (El Adlouni et al., 2008) afin d'aider les gestionnaires à mieux évaluer le risque hydrologique associé en tenant compte des changements climatiques.

Modélisation spatiale des crues à l'aide des copules

Cette thèse de doctorat a montré une variabilité spatiale dans l'effet des conditions météorologiques sur les crues printanières dans le bassin versant de Saint-Jean à l'aide de deux modèles probabilistes. Ceci met en évidence la nécessité de régionaliser les modèles afin de prendre en considération le climat local et au sein du bassin versant de la rivière à modéliser. Il serait donc intéressant de poursuivre ce travail, en utilisant ou en développant par exemple un modèle spatial basé sur les copules (ex., comme dans Saad et al., 2015) qui permet d'estimer les crues dans un site non jaugé, à l'aide des sites jaugés, et en intégrant l'information sur le changement climatique.

3. Transfert à un nouvel environnement

Passer des modèles probabilistes généralisés à des modèles spécifiques à la station hydrométrique

Cette thèse de doctorat ouvre la porte vers de nouvelles méthodes qui intègrent les connaissances scientifiques et pratiques. En effet, les approches spécifiques proposées permettent de combiner l'approche stochastique avec les séries de débits observés, et ainsi de tenir compte des facteurs physiques responsables des caractéristiques de débits extrêmes d'un site en particulier. Grâce à cette approche, le modélisateur se trouve devant l'obligation de comprendre son système (qui est le bassin versant) et de bien l'analyser afin de choisir les meilleures covariables ayant un effet sur la réponse hydrologique et par conséquent avoir une meilleure estimation du risque hydrologique. Avec les modèles proposés et surtout la manière de développer les covariables, un transfert de connaissance est rassuré en comprenant la méthodologie derrière le développement des indices météorologiques et climatiques proposés.

4. Limitations et nouvelles pistes de recherche.

Une des limites du travail réalisé est l'aspect spatial qui n'a pas été étudié. En effet, les modèles développés intègrent le paramètre de position de la densité de probabilité à estimer (avec l'échelle et la forme) en se basant sur les données historiques des débits. De plus, même si les GEV-B-Splines proposés sont des modèles flexibles et peuvent être appliqués à n'importe quelle station, la contrainte spatiale ou de dépendance régionale à l'échelle du bassin versant n'est pas en réalité prise en compte. Dans ce contexte, il serait intéressant de considérer d'autres approches pour représenter les structures de dépendance spatiale afin de fournir des estimations des événements extrêmes pour des sites non jaugés. Le développement basé sur des copules, qui sont très utilisés en hydrologie, pourrait être considéré pour répondre à ce besoin d'informations spatialisées.

5. Recommandation pour des études ultérieures

Le modèle de mélange proposé dans ce travail permet d'intégrer l'approche POT et de choisir un seuil afin d'intégrer tous les événements indépendants, et par conséquent enrichir la base de données. Ce modèle est jumelé avec la distribution Gamma pour profiter des avantages des deux distributions au niveau des deux queues de la distribution ainsi que la partie centrale. Les modèles de mélange sont bien étudiés en statistique théorique. Il serait intéressant d'inclure les effets des covariables au niveau des paramètres du modèle pour mieux représenter les fluctuations temporelles. Le développement serait similaire à celui présenté pour le modèle GEV-B-Splines.

ANNEXE A

The Gamma distribution

A random variable Y is a Gamma distributed with shape parameter a and scale parameter b (strictly positive), we note $Y \sim \Gamma(a, b)$, if its probability density function is given by:

$$f(y; a, b) = f(y; \boldsymbol{\theta}) = \frac{1}{b^a \Gamma(a)} y^{a-1} \exp(-y/b) \quad ; \quad y > 0, \quad a > 0, \quad b > 0; \quad \text{Eq. (A.1)}$$

with parameters $\boldsymbol{\theta} = (a, b)'$.

ANNEXE B

The Generalized Pareto Distribution (GPD)

The probability density function (pdf) of a Generalized Pareto Distribution (GPD) with shape, scale and location parameters κ ; β and u , respectively, noted $Y \sim \text{GPD}(u, \kappa, \beta)$, if its probability density function has the form:

$$g(y|\boldsymbol{\theta}) = \begin{cases} \frac{1}{\beta} \left(1 + \kappa \left(\frac{y-u}{\beta} \right) \right)^{-1-1/\kappa} & \kappa \neq 0 \\ \frac{1}{\beta} \exp(-(y-u)/\beta) & \kappa = 0 \end{cases} \quad \text{Eq. (A.2)}$$

Where $Y = y - u$, are the exceedances and $\boldsymbol{\theta} = (u, \kappa, \beta)'$ the vector of the parameters.

BIBLIOGRAPHIE

- Abdi, B., O. Bozorg-Haddad, et X. Chu. (2021). Uncertainty analysis of model inputs in riverine water temperature simulations. *Sci. Rep.* 11 (1):19908. <https://doi.org/10.1038/s41598-021-99371-0>
- AghaKouchak, A., et Nasrollahi, N. (2010). Semi-parametric and Parametric Inference of Extreme Value Models for Rainfall Data. *Water Resour. Manag.* 24, 1229–1249. <https://doi.org/10.1007/s11269-009-9493-3>
- Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle, in: *Selected Papers of Hirotugu Akaike*. Springer, pp. 199–213. DOI: 10.1007/978-1-4612-1694-0_15
- Ambroise, B. (1999). La dynamique du cycle de l'eau dans un bassin versant. *Process. Facteurs Model.* Ed. H G Bucar. ISBN 973-98954-2-5.
- Ammar, M. E., Gharib, A., Islam, Z., Davies, E. G., Seneka, M., et Faramarzi, M. (2020). Future floods using hydroclimatic simulations and peaks over threshold: An alternative to nonstationary analysis inferred from trend tests. *Advances in Water Resources*, 136, 103463. <https://doi.org/10.1016/j.advwatres.2019.103463>
- Ancil, F., Rousselle, J., et Lauzon, N. (2012). *Hydrologie: cheminements de l'eau* (Presses inter Polytechnique).
- Anderson, M.L., Chen, Z.-Q., Kavvas, M.L., et Feldman, A. (2002). Coupling HEC-HMS with atmospheric models for prediction of watershed runoff. *J. Hydrol. Eng.* 7, 312–318.
- Andrews, S.N., Linnansaari, T., Curry, R.A., Leblanc, N.M., et Pavey, S.A. (2020). Winter ecology of striped bass (*Morone saxatilis*) near its northern limit of distribution in the Saint John River, New Brunswick. *Environ. Biol. Fishes* 103, 1343–1358. <https://doi.org/10.1007/s10641-020-01027-x>
- Anees, M.T., Abdullah, K., Nawawi, M.N.M., Ab Rahman, N.N.N., Piah, A.R.M., Zakaria, N.A., Syakir, M.I., Omar, M. (2016). Numerical modeling techniques for flood analysis. *J. African Earth Sci.* 124, 478–486.
- AQUAVEO. (2021). Surface Water Modeling System. Disponible sur: <https://www.aquaveo.com/about-us>
- Arsenault, R., Poulin, A., Côté, P., et Brissette, F. (2013). Comparison of stochastic optimization algorithms in hydrological model calibration. *J. Hydrol. Eng.* 19, 1374–1384.
- Aryal, S.K., Bates, B.C., Campbell, E.P., Li, Y., Palmer, M.J., et Viney, N.R. (2009). Characterizing and modeling temporal and spatial trends in rainfall extremes. *J. Hydrometeorol.* 10, 241–253.

- Assani, A. A. (2022). Variability of Mean Annual Flows in Southern Quebec (Canada). *Water*, 14(9), 1370. <https://doi.org/10.3390/w14091370>
- Assani, A. A. (2023). Comparison of Spatio-Temporal Variability of Daily Maximum Flows in Cold-Season (Winter and Spring) in Southern Quebec (Canada). *Hydrology*, 10(2), 44.
- Aubert, Y. (2012). Estimation des valeurs extrêmes de débit par la méthode Shyreg : réflexions sur l'équifinalité dans la modélisation de la transformation pluie en débit (Paris 6).
- Ayari, K., Djebbi, M., et Chkroun, H. (2016). Cartographie du risque d'inondation de la ville de Mdjez El Bab par débordement de la Medjerda. *Larhyss J.* 285–307.
- BAC: Bureau d'Assurance du Canada. (2019). Options for managing flood costs of Canada's highest risk residential properties. A report of the National Working Group on Financial Risk of Flooding. <http://assets.ibc.ca/Documents/Studies/IBC-Flood-Options-Paper-EN.pdf>
- Bai, P., Liu, X., Liang, K., et Liu, C. (2015). Comparison of performance of twelve monthly water balance models in different climatic catchments of China. *J. Hydrol.* 529, 1030–1040.
- Bai, Y., Zhang, Z., et Zhao, W. (2019). Assessing the Impact of Climate Change on Flood Events Using HEC-HMS and CMIP5. *Water Air Soil Pollut.* 230, 119.
- Barredo, J.I., Saurí, D. et Llasat, M.C. (2012). Assessing trends in insured losses from floods in Spain 1971–2008. *Nat. Hazards Earth Syst. Sci.* , 12, 1723–1729. <https://doi.org/10.5194/nhess-12-1723-2012>
- Beck, H., Zimmermann, N., McVicar, T., Vergopolan, N., Berg, A., et Wood, E. F. (2018). Present and future Köppen-Geiger climate classification maps at 1-km resolution." *Sci. Data* 5: 180214. <https://doi.org/10.1038/sdata.2018.214>.
- Behrens, C. N., Lopes, H. F., et Gamerman, D. (2004). Bayesian analysis of extreme events with threshold estimation. *Statistical Modelling*, 4(3), 227–244. <https://doi.org/10.1191/1471082X04st075oa>.
- Benoit, C., Demers, I., Roberge, F., Gachon, P. et Laprise, R. (2022). Inondations des printemps 2017 et 2019 dans le bassin versant de la rivière des Outaouais (Québec, Canada) : analyse des facteurs physiographiques et météorologiques en cause. Dans T. Buffin-Bélanger, D. Maltais et M. Mario Gauthier (dir.). *Les inondations au Québec : risques, aménagement du territoire, impacts socioéconomiques et transformation des vulnérabilités* (p. 29–58). Presse de l'Université du Québec.
- Berghuijs, W.R., Aalbers, E.E., Larsen, J.R., Trancoso R, et Woods, R.A. (2017). Recent changes in extreme floods across multiple continents. *Environ. Res. Lett.* 12 (11) <https://doi.org/10.1088/1748-9326/aa8847>.
- Bermudez, P. Z., Turkman, M. A., et Turkman, K. F. (2001). A predictive approach to tail probability estimation. *Extremes*, 4(4), 295–314.

- Bickel, P. J., D. A. et Freedman. (1981). Some asymptotic theory for the bootstrap. *Ann. Stat.* 9 (6): 1196–1217. <https://doi.org/10.1214/aos/1176345637>.
- Blöschl, G., Ardoin-Bardin, S., Bonell, M., Dorninger, M., Goodrich, D., Gutknecht, D., Matamoros, D., Merz, B., Shand, P., Szolgay, J. (2007). At what scales do climate variability and land cover change impact on flooding and low flows? *Hydrol. Process.* 21, 1241–1247. <https://doi.org/10.5194/hess-17-5197-2013>.
- Boisvert, J., N. El-Jabi, S. E. El Adlouni, D. Caissie, A. N. Thiombiano. (2017). “New Brunswick hydrometric network analysis and rationalization.” *Can. J. Civ. Eng.* 44 (10): 829–837. <https://doi.org/10.1139/cjce-2016-0487>.
- Bonsal, B., et Shabbar, A. (2011). Oscillations climatiques à grande échelle ayant une incidence sur le Canada, de 1900 à 2008 . (Conseils canadiens des ministres des ressources).
- Bourgault, M. A., Boivin, M., Roy, R., Desrochers, G., et Anctil, F. (2022). Regards sur les mécanismes et les facteurs contrôlant les inondations des bassins versants du Québec méridional. Les inondations au Québec: Risques, aménagement du territoire, impacts socioéconomiques et transformation des vulnérabilités. https://books.google.ca/books?hl=fr&lr=&id=OY-ZEAAAQBAJ&oi=fnd&pg=PT37&dq=Bourgault+et+al.,+2022+flood&ots=vwC2pDaO2x&sig=7ct8jn_ehUrD7sErY5bGn2TDDyA#v=onepage&q=Bourgault%20et%20al.%2C%202022%20flood&f=false
- Bracken, C., Holman, K. D., Rajagopalan, B., et Moradkhani, H. (2018). A Bayesian hierarchical approach to multivariate nonstationary hydrologic frequency analysis. *Water Resources Research*, 54(1), 243-255. <https://doi.org/10.1002/2017WR020403>
- Brigode, P., Mićović, Z., Bernardara, P., Paquet, E., Gailhard, J., Garavaglia, F., et Ribstein, P. (2015). Méthodes probabilistes et déterministes d'estimation des débits extrêmes: comparaison de résultats sur deux bassins versants de Colombie-Britannique (Canada). *La Houille Blanche*, (3), 72-78. <https://doi.org/10.1051/lhb/20150036>
- Bücher, A., et Zhou, C. (2021). A horse race between the block maxima method and the peak-over-threshold approach. *Statistical Science*, 36(3), 360-378.
- Budhathoki S, Rokaya P, et Lindenschmidt K-E. (2022). Impacts of future climate on the hydrology of a transboundary river basin in northeastern North America. *J Hydrol* 605:127317. <https://doi.org/10.1016/j.jhydrol.2021.127317>
- Budhathoki, S., Rokaya, P., et Lindenschmidt, K. E. (2022). Impacts of future climate on the hydrology of a transboundary river basin in northeastern North America. *Journal of Hydrology*, 605, 127317.
- Buffin-Bélanger, T., Maltais, D., et Gauthier, M. (2022). *Les inondations au Québec: Risques, aménagement du territoire, impacts socioéconomiques et transformation des vulnérabilités*. PUQ. <https://www.puq.ca/catalogue/livres/les-inondations-quebec-4082.html>

- Burn, D. H., et Whitfield, P. H. (2023). Climate Related Changes to Flood Regimes Show an Increasing Rainfall Influence. *Journal of Hydrology*, 129075.
- Burn, D.H., et Whitfield, P.H. (2017). Changes in cold region flood regimes inferred from long-record reference gauging stations. *Water Resour. Res.* 53, 2643–2658. <https://doi.org/10.1002/2016WR020108>
- Burn, D.H., et Whitfield P.H. (2016). Changes in floods and flood regimes in Canada, *Canadian Water Resources Journal / Revue canadienne des ressources hydriques*, 41:1-2, 139-150, DOI: [10.1080/07011784.2015.1026844](https://doi.org/10.1080/07011784.2015.1026844)
- Burrell, B. C., et Beltaos, S. (2022). Ice and Hydroclimatic Regimes of the Saint John River, NB, Canada. Conference Paper: 26 th IAHR International Symposium on Ice. <https://doi.org/10.1139/cjce-2014-0372>
- Buttle, J.M., Allen, D.M., Caissie, D., Davison, B., Hayashi, M., Peters, D.L., Pomeroy, J.W., Simonovic, S., St-Hilaire, A., et Whitfield, P.H. (2016). Flood processes in Canada: Regional and special aspects. *Can. Water Resour. J.*, 41, 7–30. <https://doi.org/10.1080/07011784.2015.1131629>
- Cairo, F., et Gomes, M. I. (2015). Threshold selection in extreme value analysis. In D. K. Dey, & J. Yan (Eds.), *Extreme value modeling and risk analysis: Methods and applications* (pp. 69–82). Boca Raton, FL: CRC Press.
- Caires, S. (2009). A comparative simulation study of the annual maxima and the peaks-over-threshold methods (Deltares Report 1200264-002 for Rijkswaterstaat, Waterdienst).
- Caissie, D., et Robichaud, S. (2009). Towards a better understanding of the natural flow regimes and streamflow characteristics of rivers of the Maritime Provinces. *Canadian technical Report of Fisheries and Aquatic Sciences* 2843: Viii+53p.
- Calderón-Vega, F., Mössö, C., García-Soto, A. D., et Delgadillo-Ruiz, E. (2019). Single site extreme wave analysis in the Pacific Ocean comparing stationary and non-stationary GEV models. *Curr. J. Appl. Sci. Technol*, 32(6), 1-12.
- Canadian Disaster database. (2022). Available on: <https://cdd.publicsafety.gc.ca/rs/lts-eng.aspx?cultureCode=frCa&provinces=4&eventTypes=%27FL%27&normalizedCostYear=1>
- Carreau, J., et Y. Bengio. (2009). A hybrid Pareto model for asymmetric fat-tailed data: The univariate case. *Extremes* 12 (5): 53–76. <https://doi.org/10.1007/s10687-008-0068-0>
- Cassar, A., Healy, A., et von Kessler. (2017). Trust, Risk, and Time Preferences After a Natural Disaster: Experimental Evidence from Thailand. *World Dev.* 94, 90–105. <https://doi.org/10.1016/j.worlddev.2016.12.042>

- Cassella, J. P., Cassella, S., Smith, I. (2002). Synergistic antifungal activity of tea tree (*Melaleuca alternifolia*) and lavender (*Lavandula angustifolia*) essential oils against dermatophyte infection. *Int. J. Aromather.* 12 (1): 2–15. <https://doi.org/10.1054/ijar.2001.0127>.
- Cassou, C. (2004). Du changement climatique aux régimes de temps: l'oscillation nord-atlantique. *La Météorologie*, 45, 21–32.
- Cea, L., et Costabile, P. (2022). Flood Risk in Urban Areas: Modelling, Management and Adaptation to Climate Change. A Review. *Hydrology*, 9(3), 50. <https://doi.org/10.3390/hydrology9030050>
- Chavez-Demoulin, V., et Davison, A. C. (2005). Generalized additive modelling of sample extremes. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(1), 207-222. <https://doi.org/10.1111/j.1467-9876.2005.00479.x>
- Chen, Y., Tan, Y., She, H., Zhang, M., Jiang, X., Guo, P., et Zhang, Y. (2022). Characteristics of slow-release ice and snow melting micro-surfacing materials. *International Journal of Pavement Engineering*, 1-16. <https://doi.org/10.1080/10298436.2022.2120986>
- Cheng, L., AghaKouchak, A., Gilleland, E., et Katz, R.W. (2014). Non-stationary extreme value analysis in a changing climate. *Clim. Change* 127, 353–369.
- Choulakian, V., et Stephens, M.A. (2001). Goodness-of-Fit Tests for the Generalized Pareto Distribution. *Technometrics* 43, 478–484.
- Coles, S. (2001). Introduction, in: Coles, S. (Ed.), *An Introduction to Statistical Modeling of Extreme Values*, Springer Series in Statistics. Springer London, London, pp. 1–17. https://doi.org/10.1007/978-1-4471-3675-0_1
- Coles, S. G., et Tawn, J. A. (1991). Modelling extreme multivariate events. *J. R. Stat. Soc. B* 53 (2): 377–392. <https://doi.org/10.1111/j.2517-6161.1991.tb01830.x>.
- Coles, S. G., et Tawn, J. A. (1994). Statistical methods for multivariate extremes: An application to structural design. *J. R. Stat. Soc. C* 43 (1): 1–31. <https://doi.org/10.2307/>
- Coles, S.G., et Tawn, J.A. (1996). A Bayesian analysis of extreme rainfall data. *J. R. Stat. Soc. Ser. C Appl. Stat.* 45, 463–478. <https://doi.org/10.2307/2986068>
- Coulibaly, P., Anctil, F., Rasmussen, P., et Bobée, B. (2000). A recurrent neural networks approach using indices of low-frequency climatic variability to forecast regional annual runoff. *Hydrological Processes*, 14(15), 2755-2777.
- Cui, C., et Cui, L. (2020). An Innovative Flood Prediction System Using Improved Machine Learning Approach. *The Canadian Science Fair Journal*, 2(2). <https://creativecommons.org/licenses/by-nc-sa/4.0>

- Cunnane, C. (1973). A particular comparison of annual maxima and partial duration series methods of flood frequency prediction. *Journal of Hydrology*, 18(3–4), 257–271.
- Curceac, S., Atkinson, P. M., Milne, A., Wu, L., et Harris, P. (2020). An evaluation of automated GPD threshold selection methods for hydrological extremes across different scales. *Journal of Hydrology*, 585, 124845. <https://doi.org/10.1016/j.jhydrol.2020.124845>.
- Dam, T. T., Ta, T. A., et Mai, T. (2022). Submodularity and local search approaches for maximum capture problems under generalized extreme value models. *European Journal of Operational Research*, 300(3), 953-965. <https://doi.org/10.1016/j.ejor.2021.09.006>
- Das, A., Budhathoki, S., et Lindenschmidt, K. E. (2022). A stochastic modelling approach to forecast real-time ice jam flood severity along the transborder (New Brunswick/Maine) Saint John River of North America. *Stochastic Environmental Research and Risk Assessment*, 36(7), 1903-1915. <https://doi.org/10.1007/s00477-022-02234-x>
- Davison, A. C., Ramesh, N. I. (2000). Local likelihood smoothing of sample extremes. *J. R. Stat. Soc. B* 62 (1): 191–208. <https://doi.org/10.1111/1467-9868.00228>.
- Davison, A. C., Smith, R. L. (1990). Models for exceedances over high thresholds. *J. R. Stat. Soc. B* 52 (3): 393–442. <https://doi.org/10.1111/j.2517-6161.1990.tb01796.x>.
- Davison, A.C., Padoan, S.A., et Ribatet, M. (2012). Statistical Modeling of Spatial Extremes. *Statist. Sci.* 27 (2) 161 - 186. <https://doi.org/10.1214/11-STS376>
- De Haan, L., et De Ronde, J. (1998). Sea and wind: multivariate extremes at work. *Extremes* 7–45. <https://doi.org/10.1023/A:1009909800311>.
- de Melo Mendes, B. V., et Lopes, H. F. (2004). Data driven estimates for mixtures *Comput. Stat. Data Anal.* 47 (3): 583–598. <https://doi.org/10.1016/j.csda.2003.12.006>.
- De Paola, F., Giugni, M., Pugliese, F., Annis, A., et Nardi, F. (2018). GEV Parameter Estimation and Stationary vs. Non-Stationary Analysis of Extreme Rainfall in African Test Cities. *Hydrology* 5, 28. <https://doi.org/10.3390/hydrology5020028>
- Dettinger, M.D., et Diaz, H.F. (2000). Global characteristics of stream flow seasonality and variability. *J Hydrometeorol* 1:289–310. doi: [https://doi.org/10.1175/1525-7541\(2000\)001<0289:GCOSFS>2.0.CO;2](https://doi.org/10.1175/1525-7541(2000)001<0289:GCOSFS>2.0.CO;2)
- Di Baldassarre, G., Kooy, M., Kemerink, J.S., et Brandimarte, L. (2013). Towards understanding the dynamic behaviour of floodplains as human-water systems. *Hydrol. Earth Syst. Sci.* 17, 3235–3244.
- Douvinet, J., Delahaye, D., et Langlois, P. (2008). Modélisation de la dynamique potentielle d'un bassin versant et mesure de son efficacité structurelle. *Cybergeo: European Journal of Geography*.

- Drissia, T. K., Jothiprakash, V., et Anitha, A. B. (2019). Flood frequency analysis using L moments: a comparison between at-site and regional approach. *Water Resources Management*, 33(3), 1013-1037.
- Duhamel, F., Germain, D., et Fortin, G. (2022). Flood risk quantification and mapping: An integrated vulnerability-based approach at the local scale in New Brunswick, Canada. *International Journal of Disaster Risk Reduction*, 83, 103403.
- Eastoe, E. F., et Tawn, J. A. (2009). Modelling non-stationary extremes with application to surface level ozone. *J. R. Stat. Soc. C* 58 (1): 25–45. <https://doi.org/10.1111/j.1467-9876.2008.00638.x>.
- ECCC: Environment and Climate change Canada. (2021). https://climat.meteo.gc.ca/historical_data/search_historic_data_f.html, Accessed April, 2021.
- El Adlouni, S., Bobée, B., et Samoud, O. (2013). GUIDE POUR LES MÉTHODES DU SYSTÈME D'AIDE À LA DÉCISION (SAD) DE HYFRAN-PLUS.
- El Adlouni, S., et Ouarda, T. B. (2008). Étude de la loi conjointe débit-niveau par les copules: Cas de la rivière Châteauguay. *Canadian Journal of Civil Engineering*, 35(10), 1128-1137. <https://doi.org/10.1139/L08-054>
- El Adlouni, S., Ouarda, T., Zhang, X., Roy, R., et Bobée, B. (2007). Generalized maximum likelihood estimators for the nonstationary generalized extreme value model. *Water Resour. Res.* 43. <https://doi.org/10.1029/2005WR004545>
- El Jabi, N., et Caissie, D. (2018). Characterization of natural and environmental flows in New Brunswick, Canada. *River Res. Appl.* 35 (1):14–24. <https://doi.org/10.1002/rra.3387>.
- El-Jabi, N., Caissie, D., et Turkkan, N. (2016). Flood analysis and flood projections under climate change in New Brunswick. *Canadian Water Resources Journal/Revue canadienne des ressources hydriques*, 41(1-2), 319-330. <https://doi.org/10.1080/07011784.2015.1071205>
- Environnement et Changement Climatique Canada. (2018). Archives nationales des données hydrologiques: HYDAT. Accessed March 7, 2018. <https://www.canada.ca/fr/environnement-changement-climatique/services/eau-aperçu/volume/surveillance/relevés/produits-données-services/archives-nationales-hydat.html>.
- Environnement et Changement climatique Canada. (2022). Archives nationales des données hydrologiques : HYDAT [WWW Document] . aem. URL <https://www.canada.ca/fr/environnement-changement-climatique/services/eau-aperçu/volume/surveillance/relevés/produits-données-services/archives-nationales-hydat.html>, Accessed Mai, 2021.
- Esfandiari, M., Abdi, G., Jabari, S., McGrath, H., et Coleman, D. (2020). Flood hazard risk mapping using a pseudo supervised random forest. *Remote Sensing*, 12(19), 3206. <https://doi.org/10.3390/rs12193206>

- Evin, G., Merleau, J., et Perreault, L. (2011). Two-component mixtures of normal gamma and Gumbel distributions for hydrological applications. *Water Resour. Res.* 47 (5): W08525. <https://doi.org/10.1029/2010WR010266>.
- Faghfour A, Hentati A, Guillaume F, et Germain D. (2023) A novel statistical model for flood prediction in the Eel River watershed, New Brunswick, Canada. *Water Science* 37.1: 251-268.
- Ferro, C. A., et Segers, J. (2003). Inference for clusters of extreme values. *J. R. Stat. Soc. B* 65 (2): 545–556. <https://doi.org/10.1111/1467-9868.00401>.
- Filipova, V., Lawrence, D., et Skaugen, T. (2019). A stochastic event-based approach for flood estimation in catchments with mixture rainfall and snowmelt flood regimes. *Nat. Hazards Earth Syst. Sci. J.* 19 (Apr):1–18. <https://doi.org/10.5194/nhess-19-1-2019>.
- Fischer, S., et Schumann, A. (2016). Robust flood statistics: Comparison of peak over threshold approaches based on monthly maxima and TL-moments. *Hydrol. Sci. J.* 61 (3): 457–470. <https://doi.org/10.1080/02626667.2015.1054391>.
- Fischer, S., Schumann, A., et Bühler, P. (2019). Timescale-based flood typing to estimate temporal changes in flood frequencies. *Hydrol. Sci. J.* 64 (15): 1867–1892. <https://doi.org/10.1080/02626667.2019.1679376>.
- Fischer, S., Schumann, A., et Bühler, P. (2021). A statistics-based automated flood event separation. *J. Hydrol. X* 10 (2021): 100070. <https://doi.org/10.1016/j.hydroa.2020.100070>.
- Fisher RA, et Tippett LHC. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24(2):180-190. <https://doi.org/10.1017/S0305004100015681>
- Fisher, R.A. (1929). Tests of significance in harmonic analysis. *Proc R Soc Lond A* 125, 54–59. <https://doi.org/10.1098/rspa.1929.05105>
- Fortin, G. (2003). Conceptualisation des processus de formation, d'évolution et de désagrégation des couches de glace dans un couvert nival saisonnier. (PhD Thesis). Université du Québec, Institut national de la recherche scientifique.
- Fortin, J., Moussa, R., Bocquillon, C., et Villeneuve, J. (1995). Hydrotel, un modèle hydrologique distribué pouvant bénéficier des données fournies par la télédétection et les systèmes d'information géographique. *Rev. Sci. Eau J. Water Sci.* 8, 97–124.
- Fortin, V., et Turcotte, R. (2006). Le modèle hydrologique MOHYSE. Note Cours Pour SCA7420 Dép. Sci. Terre L'Atmosphère Univ. Qué. À Montr. 23.

- Frigessi, A., Haug, O., et Rue, H. (2002). A dynamic mixture model for unsupervised tail estimation without threshold selection. *Extreme* 5 (3): 219–235. <https://doi.org/10.1023/A:1024072610684>.
- Gado, T.A, et Nguyen, V.T.V . (2016). An at-site flood estimation method in the context of nonstationarity II. Statistical analysis of floods in Quebec. *Journal of Hydrology*, 535, 722-736. <https://doi.org/10.1016/j.jhydrol.2015.12.064>
- Garrido, L. L., et Cepeda Cuervo, E. (2012). Mixture of distributions in the biparametric exponential family: A Bayesian approach. *Commun.Stat.- Simul. Comput.* 41 (3): 355–375. <https://doi.org/10.1080/03610918.2011.592245>.
- Généreux, M., Landaverde, E. et Maltais, D. (2022). Conséquences sociosanitaires des inondations printanières de 2019 : comprendre afin de mieux agir. Dans T. Buffin-Belanger, D. Maltais et M. Gauthier, *Les inondations au Québec : risques, aménagement du territoire, impacts socioéconomiques et transformation des vulnérabilités* (p. 365-384)
- Gerstengarbe, F. W., et Werner, P. C. (1989). A method for the statistical definition of extreme-value regions and their application to mete-orological time series. *Zeitschrift für Meteorologie*, 39(4), 224–226.
- Ghosh, S., et Mallick, B. K. (2011). A hierarchical Bayesian spatio-temporal model for extreme precipitation events. *Environmetrics*, 22(2), 192-204. <https://doi.org/10.1002/env.1043>
- Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Annals of mathematics*, 423-453.
- Golnaraghi, M., Surminski, S., Mehryar, S., Kousky, C., et Roezer, V. (2020-b). Building flood resilience in a changing climate. <https://rnrf.org/wp-content/uploads/2021/09/RRJV36N4.pdf>
- Golnaraghi, M., Thistlethwaite, J., Henstra, D., et Stewart, C. (2020-a). Flood Risk Management in Canada: Building flood resilience in a changing climate. *The Geneva Association: Geneva, Switzerland*, 2021.
- Graham, F. S., et Brown, J. N. (2020). El Niño, La Niña, and the Southern Oscillation. In *Atmosphere and Climate* (pp. 221-230). CRC Press.
- Greene C. (2012). The winters of our discontent. Loss of Arctic sea ice is stacking the deck in favour of harsh winter weather in the U.S. and Europe. *Sci. Am.* 307: 50–55.
- Guimarães Nobre, G., de Moel, H., Giuliani, M., Bischiniotis, K., Aerts, J. C., et Ward, P. J. (2020). What will the weather do? forecasting flood losses based on oscillation indices. *Earth's Future*, 8(3), e2019EF001450. <https://doi.org/10.1029/2019EF001450>

- Guimarães Nobre, G., Jongman, B., Aerts, J. C. J. H., et Ward, P. J. (2017). The role of climate variability in extreme floods in Europe. *Environmental Research Letters*, 12(8), 084012. DOI 10.1088/1748-9326/aa7c22
- Guo, X., Cheng, J., Yin, C., Li, Q., Chen, R., et Fang, J. (2023). The extraordinary Zhengzhou flood of 7/20, 2021: How extreme weather and human response compounding to the disaster. *Cities*, 134, 104168.
- Häkkinen, S., Rhines, P.B., et Worthen, D.L. (2011). Atmospheric blocking and Atlantic multidecadal ocean variability. *Science* 334, 655–659.
- Hall, P., et Tajvidi, N. (2000). Nonparametric analysis of temporal trend when fitting parametric models to extreme-value data. *Stat. Sci.* 15 (2):153–167. <https://doi.org/10.1214/ss/1009212755>.
- He, Y., Peng, L., Zhang, D., et Zhao, Z. (2022). Risk Analysis via Generalized Pareto Distributions. *Journal of Business & Economic Statistics*, 40(2), 852-867.
- Heffernan, J. E., et Tawn, J. A. (2004). A conditional approach for multivariate extreme values (with discussion). *J. R. Stat. Soc. B* 66 (3): 497–546. <https://doi.org/10.1111/j.1467-9868.2004.02050.x>.
- Henry, S., Laroche, A. M., Hentati, A., et Boisvert, J. (2020). Prioritizing Flood-Prone Areas Using Spatial Data in the Province of New Brunswick, Canada. *Geosciences*, 10(12), 478.
- Herrera, P.A., Marazuela, M.A., et Hofmann, T. (2022). Parameter estimation and uncertainty analysis in hydrological modeling. *WIREs Water*, 9, e1569.
- Hodgkins, G.A., Whitfield, P.H., Burn, D.H., Hannaford, J., Renard, B., Stahl, K., Fleig, A.K., Madsen, H., Mediero, L., Korhonen, J., Murphy, C., et Wilson, D. (2017). Climate-driven variability in the occurrence of major floods across North America and Europe. *J. Hydrol.* 552, 704–717. <https://doi.org/10.1016/j.jhydrol.2017.07.027>
- Hosking, J.R.M. (1986). The theory of probability weighted moments: Yorktown Heights, NY, IBM Research Division, TJ Watson Research Center. Research Report RC–12210.
- Hosking, J.R.M., Wallis, J.R., et Wood, E.F. (1985). Estimation of the Generalized Extreme-Value Distribution by the Method of Probability-Weighted Moments. *Technometrics* 27, 251–261. <https://doi.org/10.1080/00401706.1985.10488049>
- Humes TM, et Dublin J. (1988). A comparison of the 1976 and the 1987 St. John River ice jam flooding with emphasis on antecedent conditions. In *Proceedings of the Workshop on Hydraulics of River Ice/Ice Jams* (pp. 43-62).
- Huntington, T.G., Richardson, A.D., McGuire, K.J., et Hayhoe, K. (2009). Climate and hydrological changes in the northeastern United States: recent trends and implications for forested and aquatic

ecosystems This article is one of a selection of papers from NE Forests 2100: A Synthesis of Climate Change Impacts on Forests of the Northeastern US and Eastern Canada. *Can. J. For. Res.* 39, 199–212.

Hurrell, J.W. (1995). Decadal trends in the North Atlantic Oscillation: regional temperatures and precipitation. *Science* 269, 676–679. <https://doi.org/10.1126/science.269.5224.676>

International Building Code (IBC). (2018). International Building Code [a compilation of flood resistant provisions, prepared by FEMA]. https://www.fema.gov/sites/default/files/2020-08/fema_2018-ibc-codes-flood-provisions.pdf

International Joint Commission (IJC). (2013). Plan d'étude pour la détermination des mesures visant à atténuer les inondations et leurs répercussions dans le bassin versant du lac Champlain et de la rivière Richelieu. Ott. Wash. Groupe Trav. Int. Plan D'étude Lac Champlain Rivière Richelieu 151 Pp.

IPCC (Intergovernmental Panel on Climate Change). (2021). Summary for policymakers. In *Climate change 2021: The physical science basis. Contribution of Working Group I to the sixth assessment report of the intergovernmental panel on climate change*. Cambridge, UK: Cambridge University Press.

IPCC (Intergovernmental Panel on Climate Change). (2022). *Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* [H.-O. Pörtner, D.C. Roberts, M. Tignor, E.S. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, B. Rama (eds.)]. Cambridge University Press. Cambridge University Press, Cambridge, UK and New York, NY, USA, 3056 pp., doi:10.1017/9781009325844.

Jalili Pirani, F., et Najafi, M. R. (2022). Multivariate analysis of compound flood hazard across Canada's Atlantic, Pacific and Great Lakes coastal areas. *Earth's Future*, 10(8), e2022EF002655.

Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) of meteorological elements. *Q. J. R. Meteorol. Soc.* 81 (5): 158–171. <https://doi.org/10.1002/qj.49708134804>.

Jiang, C., Xiong, L., et Xu, W. (2023). Nonstationary Hydrological Distribution Estimation Using Hierarchical Model with Stochastic Covariates. *Journal of Hydrologic Engineering*, 28(4), 04023009.

Jiang, C., Xiong, L., Guo, S., Xia, J., et Xu, C.-Y. (2017). A process-based insight into nonstationarity of the probability distribution of annual runoff. *Water Resour. Res.* 53, 4214–4235.

Jiang, S., et Kang, L. (2019). Flood frequency analysis for annual maximum streamflow using a non-stationary GEV model. In *E3S Web of Conferences* (Vol. 79, p. 03022). EDP Sciences. <https://doi.org/10.1051/e3sconf/20197903022>

- Jonathan, P., Randell, D., Wadsworth, J., et Tawn, J. (2021). Uncertainties in return values from extreme value analysis of peaks over threshold using the generalised Pareto distribution. *Ocean Engineering*, 220, 107725.
- Katz, R. W. (2010). Statistics of extremes in climate change. *Climatic change*, 100(1), 71-76. <https://doi.org/10.1007/s10584-010-9834-5>
- Katz, R.W., Parlange, M.B., et Naveau, P. (2002). Statistics of extremes in hydrology. *Adv. Water Resour.* 25, 1287–1304. [https://doi.org/10.1016/S0309-1708\(02\)00056-8](https://doi.org/10.1016/S0309-1708(02)00056-8)
- Kendall, M.G. (1975). Rank Correlation Methods, Oxford Univ. Press, New York., No 98
- Khattak, M.S., Anwar, F., Saeed, T.U., Sharif, M., Sheraz, K., et Ahmed, A. (2016). Floodplain Mapping Using HEC-RAS and ArcGIS: A Case Study of Kabul River. *Arab. J. Sci. Eng.* 41, 1375–1390.
- Kidd, S.D., Curry, A.C., et Munkittrick, K.R. (eds.). (2011). The Saint John River: A state of environment report. Canadian Research Institute, UNB, 175 p.
- Kim, Y., et Bang, H. (2017). Monte-Carlo calculation of cramer-rao bound for non-Gaussian recursive filtering. In *Proc. Asia–Pacific Int. Symp. Aerosp. Technol.*
- Kiran, K. G., et Srinivas, V. V. (2021). A Mahalanobis distance-based automatic threshold selection method for peaks over threshold model. *Water Resour. Res.* 57 (1): e2020WR027534. <https://doi.org/10.1029/2020WR027534>.
- Knebl, M. R., Yang, Z. L., Hutchison, K., et Maidment, D. R. (2005). Regional scale flood modeling using NEXRAD rainfall, GIS, and HEC-HMS/RAS: a case study for the San Antonio River Basin Summer 2002 storm event. *Journal of Environmental Management*, 75(4), 325-336. <https://doi.org/10.1016/j.jenvman.2004.11.024>
- Kullback, S., et Leibler, R.A. (1951). On information and sufficiency. *Ann. Math. Stat.* 22, 79–86.
- Kumar, V., Sharma, K. V., Caloiero, T., Mehta, D. J., et Singh, K. (2023). Comprehensive overview of flood modeling approaches: A review of recent advances. *Hydrology*, 10(7), 141. <https://doi.org/10.3390/hydrology10070141>
- Kundzewicz, Z. W. (2019). Changes in flood risk in Europe. London: CRC Press.
- Kundzewicz, Z.W., Szwed, M., et Pińskwar, I. (2019). Climate Variability and Floods—A Global Review. *Water* 2019, 11, 1399. <https://doi.org/10.3390/w11071399>

- Laloyaux P, de Boisseson E, Balmaseda M, Bidlot J-R, Broennimann S, et Buizza R. (2018). CERA-20C: a coupled reanalysis of the Twentieth Century. *J Adv Model Earth Syst* 10(5):1172–1195. <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018M. S001273>
- Lang, M., Ouarda, T. B. M. J., et Bobée, B. (1999). "Towards operational guidelines for over-threshold modeling." *J. Hydrol.* 225 (Jun): 103–117. [https://doi.org/10.1016/S0022-1694\(99\)00167-5](https://doi.org/10.1016/S0022-1694(99)00167-5).
- Langousis, A., Mamalakis, A., Puliga, M., et Deidda, R. (2016). Threshold detection for the generalized Pareto distribution: Review of representative methods and application to the NOAA NCDC daily rainfall database. *Water Resources Research*, 52(4), 2659–2681
- Latif, S., et Simonovic, S. P. (2023). Compounding joint impact of rainfall, storm surge and river discharge on coastal flood risk: An approach based on 3D Fully Nested Archimedean Copulas. *Environmental Earth Sciences*, 82(2), 63.
- Lau, R. (2020). Flooding Forces Highway Closure between Moncton and Fredericton. Available online: <https://globalnews.ca/news/4184512/flooding-forces-highway-closure-between-moncton-and-fredericton/> (accessed on 4 August 2020).
- Lawford, R.G., Prowse, T.D., Hogg, W.D., Warkentin, A.A., et Pilon, P.J. (1995). Hydrometeorological aspects of flood hazards in Canada. *Atmosphere-Ocean* 33, 303–328.
- Lawrence, D. (2020). Uncertainty introduced by flood frequency analysis in projections for changes in flood magnitudes under a future climate in Norway. *Journal of Hydrology: Regional Studies*, 28, 100675.
- Li, H., Wang, D., Singh, V. P., Wang, Y., Wu, J., Wu, J., et Zhang, J. (2019). Non-stationary frequency analysis of annual extreme rainfall volume and intensity using Archimedean copulas: A case study in eastern China. *Journal of hydrology*, 571, 114-131.
- Lin, H., Mo, R., Vitart, F., et Stan, C. (2018). Eastern Canada flooding 2017 and its subseasonal predictions. *Atmosphere-Ocean*, 57(3),195-207. <https://doi.org/10.1080/07055900.2018.1547679>
- Lyamuremye, E., Mung'atu, J., et Mwita, P. (2019). Extreme value modelling of rainfall using poisson-generalized pareto distribution: A case study tanzania. *nt. J. Stat. Distrib. Appl*, 5, 67-75.
- MacDonald, A. (2011). Extreme value mixture modelling with medical and industrial applications, Christchurch, New Zealand: (Doctoral dissertation, PhD thesis): University of Canterbury.
- Machado, M.J., Botero, B.A., López, J., Francés, F., Díez-Herrero, A., et Benito, G. (2015). Flood frequency analysis of historical flood data under stationary and non-stationary modelling. *Hydrol. Earth Syst. Sci.* 19, 2561–2576.

- Madsen, H., Rasmussen, P. F., et Rosbjerg, D. (1997). Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events: 1. At-site modeling. *Water Resources Research*, 33(4), 747–757.
- Mangini, W., Viglione, A., Hall, J., Hundscha, Y., Ceola, S., Montanari, A., Rogger, M., Salinas, J.L., Borzi, I et Parajka, J. (2018). Detection of trends in magnitude and frequency of flood peaks across Europe. *Hydrological Sciences Journal*, 63(4), 493-512. <https://doi.org/10.1080/02626667.2018.1444766>
- Mann, H. B. (1945). Nonparametric tests against trend, *Econometrica*, 13, 245-259.
- Martins, E. S., et Stedinger, J. R. (2001). Historical information in a generalized maximum likelihood framework with partial duration and annual maximum series. *Water Resour. Res.* 37 (10): 2559–2567. <https://doi.org/10.1029/2000WR000009>.
- Martins, E.S., et Stedinger, J.R. (2000). Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. *Water Resour. Res.* 36, 737–744. <https://doi.org/10.1029/1999WR900330>
- Maskrey, S.A., Mount, N.J., et Thorne, C.R. (2022). Doing flood risk modelling differently: Evaluating the potential for participatory techniques to broaden flood risk management decision-making. *J. Flood Risk Manag.*, 15, e12757.
- Masson-Delmotte, V., Zhai, P., Pörtner, H. O., Roberts, D., Skea, J., et Shukla, P. R. (2022). Global Warming of 1.5° C: IPCC Special Report on Impacts of Global Warming of 1.5° C above Pre-industrial Levels in Context of Strengthening Response to Climate Change, Sustainable Development, and Efforts to Eradicate Poverty. Cambridge University Press.
- Maurer, E. P., Lettenmaier, D. P., et Mantua, N. (2004). Variability and potential sources of predictability of North American runoff, *Water Resour. Res.* 40, W09306, doi: [10.1029/2003WR002789](https://doi.org/10.1029/2003WR002789).
- Mazouz, R., Assani, A.A., Quessy, J.-F., et Légaré, G. (2012). Comparison of the interannual variability of spring heavy floods characteristics of tributaries of the St. Lawrence River in Quebec (Canada). *Adv. Water Resour.* 35, 110–120.
- McCabe, G.I, Betancourt, IL., Gray, S.T., Palecki, M.A., et Hidalgo, H.G. (2008). Associations of multi-decadal sea-surface temperature variability with US drought. *Quaternary International*, 188, 31-40.
- McGrath, H., Bourgon, J.F., Proulx-Bourque, J.S., Nastev, M., et Abo El Ezz, A. (2018). A comparison of simplified conceptual models for rapid web-based flood inundation mapping. *Nat. Hazards*, 93, 905–920.

- McGrath, H., Stefanakis, E., et Nastev, M. (2015). Sensitivity analysis of flood damage estimates: A case study in Fredericton, New Brunswick. *International Journal of Disaster Risk Reduction*, 14, 379–387. <https://doi.org/10.1016/j.ijdr.2015.09.003>
- McMillan, H., Montanari, A., Cudennec, C., Savenije, H., Kreibich, H., Krueger, T., Liu, J., Mejia, A., Loon, A.V., et Aksoy, H. (2016). Panta Rhei 2013–2015: global perspectives on hydrology, society and change. *Hydrol. Sci. J.* 61, 1174–1191.
- McNeil, A. J., et Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. *J. Empirical Finance* 7 (3–4): 271–300. [https://doi.org/10.1016/S0927-5398\(00\)00012-8](https://doi.org/10.1016/S0927-5398(00)00012-8).
- Merz, B., Vorogushyn, S., Uhlemann, S., Delgado, J., et Hündebach, Y. (2012). HESS Opinions More efforts and scientific rigour are needed to attribute trends in flood time series. *Hydrol. Earth Syst. Sci.* 16, 1379–1387.
- Merz, R., et Blöschl, G. (2003). A process typology of regional floods. *Water Resour. Res.* 39 (Dec): 12. <https://doi.org/10.1029/2002WR001952>.
- Milly, P.C.D., Betancourt, J., Falkenmark, M., Hirsch, R.M., Kundzewicz, Z.W., Lettenmaier, D.P., et Stouffer, R.J. (2008). Stationarity is dead: Whither water management? *Science* 319, 573–574. <https://doi.org/10.1126/science.1151915>
- Mo, C., Ruan, Y, He, J., Jin, J. L. , Liu, P., et G. Sun. (2019). Frequency analysis of precipitation extremes under climate change. *Int. J. Climatol.* 39 (3): 1373–1387. <https://doi.org/10.1002/joc.5887>.
- Montanari, A., et Koutsoyiannis, D. (2012). A blueprint for process-based modeling of uncertain hydrological systems. *Water Resour. Res.*, 48, 2011WR011412.
- Montanari, A., Young, G., Savenije, H.H.G., Hughes, D., Wagener, T., Ren, L.L., et Munich RE. (2017). Natural catastrophes 2015: Analyses, Assessments, Positions. TOPICS GEO.
- Moradkhani, H., et Sorooshian, S. (2008). General Review of Rainfall–runoff Modeling: Model Calibration, Data Assimilation, and Uncertainty Analysis. In *Hydrological Modelling and the Water Cycle: Coupling the Atmospheric and Hydrological Models*; Sorooshian, S., Hsu, K.L., Coppola, E., Tomassetti, B., Verdecchia, M., Visconti, G., Eds.; Springer: Berlin/Heidelberg, Germany; pp. 1–24.
- Munir, B.A., Ahmad, S.R., et Hafeez, S. (2019). Integrated Hazard Modeling for Simulating Torrential Stream Response to Flash Flood Events. *ISPRS Int. J. Geo-Inf.*, 9, 1.
- Muñoz-Sabater, J., Dutra, E., Agustí-panareda, A., Albergel, C., Hersbach, H., Martens, B., Miralles, D. G., Piles, M., et Rodríguez-fernández, N. J. (2021). ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data Discussions*, March, 1–50. <https://doi.org/https://doi.org/10.5194/essd-2021-82>

- Najafi, M. R., et Moradkhani, H. (2014). A hierarchical Bayesian approach for the analysis of climate change impact on runoff extremes. *Hydrological processes*, 28(26), 6292-6308. <https://doi.org/10.1002/hyp.10113>
- Nascimento, F. X., Brígido, C., Glick, B. R., Oliveira, S., et Alho, L. (2012). Mesorhizobium ciceri LMS-1 expressing an exogenous 1-aminocyclopropane-1-carboxylate (ACC) deaminase increases its nodulation abilities and chickpea plant resistance to soil constraints. *Lett. Appl. Microbiol.* 55 (1): 15–21. <https://doi.org/10.1111/j.1472-765X.2012.03251.x>.
- Nasr, A.A., Wahl, T., Rashid, M.M., Camus, P., et Haigh, I.D. (2021). Assessing the dependence structure between oceanographic, fluvial, and pluvial flooding drivers along the United States coastline. *Hydrol. Earth Syst. Sci.* 25, 6203–6222.
- Nasri, B., Adlouni, S.E., et Ouarda, T.B.M.J. (2013). Bayesian Estimation for GEV-B-Spline Model. *Open J. Stat.* 03, 118–128. <https://doi.org/10.4236/ojs.2013.32013>
- Natural Resources Canada (NRCan). (2019). Federal Hydrologic and Hydraulic Procedures for Flood Hazard Delineation- Version 1.0. Federal Flood Mapping Guidelines Series. https://publications.gc.ca/collections/collection_2020/rncan-nrcan/m45/M45-113-2019-eng.pdf
- Naveau, P., Huser, R., Ribereau, P., et Hannart, A. (2016). Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection. *Water Resources Research*, 52(4), 2753–2769
- NCEI. (2018). North Atlantic Oscillation (NAO) | Teleconnections | National Centers for Environmental Information (NCEI).
- Newton, B., et Burrell, B.C. (2016). The April–May 2008 flood event in the Saint John River Basin: Causes, assessment and damages. *Canadian Water Resour. J. /Revue Canadienne des Ressources Hydriques.* 41 (1–2): 118–128. <https://doi.org/10.1080/07011784.2015.1009950>.
- Ngongondo, C., Zhou, Y., et Xu, C. Y. (2019). Non-stationary frequency analysis of extreme Lake Malawi levels. In *Geophysical Research Abstracts*, 21, (1-1).
- Nied, M., Schröter, K., Lüdtkke, S., Nguyen, V. D., et Merz, B. (2017). What are the hydro-meteorological controls on flood characteristics?. *Journal of Hydrology*, 545, 310-326.
- Nikoo, M., Ramezani, F., Hadzima-Nyarko, M., Nyarko, E.K., et Nikoo, M. (2016). Flood-routing modeling with neural network optimized by social-based algorithm. *Nat. Hazards*, 82, 1–24.
- Nolin, A. F., Girardin, M. P., Tardif, J. C., Guo, X. J., Conciatori, F., et Bergeron, Y. (2022). A 247-year tree-ring reconstruction of spring temperature and relation to spring flooding in eastern boreal Canada. *International Journal of Climatology*, 42(12), 6479-6498. <https://doi.org/10.1002/joc.7608>

- Nolin, A. F., Tardif, J. C., Conciatori, F., Kames, S., Meko, D. M., et Bergeron, Y. (2021). Multi-century tree-ring anatomical evidence reveals increasing frequency and magnitude of spring discharge and floods in eastern boreal Canada. *Global and Planetary Change*, 199, 103444.
- Northrop, P. J., Attalides, N., et Jonathan, P. (2017). Cross-validators extreme value threshold selection and uncertainty with application to ocean storm severity. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 66(1), 93–120.
- Novo, P. G., et Kyozyuka, Y. (2019). Analysis of turbulence and extreme current velocity values in a tidal channel. *J. Mar. Sci. Technol.* 24 (5): 659–672. <https://doi.org/10.1007/s00773-018-0601-z>.
- Nyaupane, N., Bahandari, S., Rahaman, M. M., Wagner, v., Kalra, A., Ahmed, S., et Gupta, R. (2018). Flood frequency analysis using generalized extreme value distribution and floodplain mapping for Hurricane Harvey in Buffalo Bayou. In *Proc., World Environmental and Water Resources Congress 2018: Watershed Management, Irrigation and Drainage, and Water Resources Planning and Management*, 364–375. Reston, VA: ASCE.
- Olthof, I. (2017). Mapping Seasonal Inundation Frequency (1985–2016) along the St-John River, New Brunswick, Canada using the Landsat Archive [WWW Document]. URL <https://www.mdpi.com/2072-4292/9/2/143> (accessed 2.4.19).
- O'Reilly, C. H., Huber, M., Woollings, T., et Zanna, L. (2016). The signature of low-frequency oceanic forcing in the Atlantic Multidecadal Oscillation, *Geophys. Res. Lett.*, 43, 2810–2818, doi:[10.1002/2016GL067925](https://doi.org/10.1002/2016GL067925).
- ORNL DAAC. (2021). (Page consultée le 18 janvier 2021). Daymet: Daily Surface Weather Data on a 1-km Grid for North America, Version 2, [En ligne]. (https://daac.ornl.gov/DAYMET/guides/Daymet_mosaics.html).
- Osborn, T.J. (2004). Simulating the winter North Atlantic Oscillation: the roles of internal variability and greenhouse gas forcing. *Clim. Dyn.* 22, 605–623.
- Ouarda, T. B., et El-Adlouni, S. (2011). Bayesian nonstationary frequency analysis of hydrological variables 1. *JAWRA Journal of the American Water Resources Association*, 47(3), 496-505. <https://doi.org/10.1111/j.1752-1688.2011.00544.x>
- Oubennaceur, K., Chokmani, K., Lessard, F., Gauthier, Y., Baltazar, C., et Toussaint, J. P. (2022). Understanding Flood Risk Perception: A Case Study from Canada. *Sustainability*, 14(5), 3087. <https://doi.org/10.3390/su14053087>
- Padmanabhan, C., Rook, T.E., et Singh, R. (1995). Modeling of automotive gear rattle phenomenon: state of the art (SAE Technical Paper).

- Papalexiou, S. M., Koutsoyiannis, D., et Makropoulos, C. (2013). How extreme is extreme? An assessment of daily rainfall distribution tails. *Hydrol. Earth Syst. Sci.* 17 (Jun): 851–862. <https://doi.org/10.5194/hess-17-851-2013>.
- Pappenberger, F., Beven, K., Horritt, M., et Blazkova, S. J. J. O. H. (2005). Uncertainty in the calibration of effective roughness parameters in HEC-RAS using inundation and downstream level observations. *Journal of hydrology*, 302(1-4), 46-69. <https://doi.org/10.1016/j.jhydrol.2004.06.036>
- Pauli, F., et S. Coles. (2001). Penalized likelihood inference in extreme value analyses. *J. Appl. Stat.* 28 (5): 547–560. <https://doi.org/10.1080/02664760120047889>.
- Peel, M. C., Finlayson, B. L., et McMahon, T.A. (2007). Updated world map of the Köppen-Geiger climate classification. *Hydrol. Earth Syst. Sci. Discuss.* 11 (5): 1633–1644. <https://doi.org/10.5194/hess-11-1633-2007>.
- Pickands, J. (1975). Statistical inference using extreme order statistics. *Ann. Stat.* 3 (1): 119–131. <https://doi.org/10.1214/aos/1176343003>.
- Poan, E.D., Gachon, P., Laprise, R., Aider, R., et Dueymes, G. (2018). Investigating added value of regional climate modeling in North American winter storm track simulations. *Clim. Dyn.* 50, 1799–1818.
- Qian, M., Jones, C., Laprise, R, et Caya, D. (2008). The influence of NAO and the Hudson Bay sea-ice on the climate of eastern Canada. *Climate Dynamics*, 31, 169-182.
- Refsgaard, J.C. (1997). Parameterisation, calibration and validation of distributed hydrological models. *J. Hydrol.* 198, 69–97.
- Refsgaard, J.C., Storm, B., et Clausen, T. (2010). Système Hydrologique Européen (SHE): review and perspectives after 30 years development in distributed physically-based hydrological modelling. *Hydrol. Res.* 41, 355.
- Renard, B. (2011). A Bayesian hierarchical approach to regional frequency analysis. *Water Resources Research*, 47(11). <https://doi.org/10.1029/2010WR010089>
- Rincón, D., Velandia, J. F., Tsanis, I., et Khan, U. T. (2022). Stochastic Flood Risk Assessment under Climate Change Scenarios for Toronto, Canada Using CAPRA. *Water*, 14(2), 227. <https://doi.org/10.3390/w14020227>
- Roberts, S. J. (1999). Novelty detection using extreme value statistics. *IEE Proceedings-Vision, Image and Signal Processing*, 146(3), 124-129.
- Rokaya, P., Budhathoki, S., et Lindenschmidt, K.-E. (2018). Trends in the Timing and Magnitude of Ice-Jam Floods in Canada. *Sci. Rep.* 8, 5834. <https://doi.org/10.1038/s41598-018-24057-z>

- Rousseau, A.N., Savary, S., et Konan, B. (2008). Implantation du modèle HYDROTEL sur le bassin de la rivière Montmorency afin de simuler les débits observés et de produire des scénarios de crues du printemps pour l'année 2008. (INRS, Centre Eau, Terre et Environnement).
- Saad, C., St-Hilaire, A., El-Adlouni, S. et Gachon, P. (2015). A nested multivariate copula approach to hydrometeorological simulations of spring floods : the case of the Richelieu River (Québec, Canada) record flood. *Stochastic Environmental Research and Risk Assessment*, 29(1), 275–294. <http://dx.doi.org/10.1007/s00477-014-0971-7>
- Saad, C., St-Hilaire, A., Gachon, P., et El Adlouni, S. (2016). The 2011 flood event in the Richelieu River basin: Causes, assessment and damages. *Canadian Water Resources Journal/Revue canadienne des ressources hydriques*, 41(1-2), 129-138. <https://doi.org/10.1080/07011784.2014.999825>
- Saint-Laurent, D., et Hähni, M. (2008). Crues et inondations majeures des villes de l'Estrie : variations climatiques et modifications anthropiques (Québec, Canada). *Environnement Urbain / Urban Environment*, 2, 50–72. <https://doi.org/10.7202/019221ar>
- Salas, J. D., et Obeysekera, J. (2019). Probability distribution and risk of the first occurrence of k extreme hydrologic events. *J. Hydrol. Eng.* 24 (10):04019032. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001809](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001809).
- Salas, J. D., Obeysekera, J., et Vogel, R. M. (2018). Techniques for assessing water infrastructure for nonstationary extreme events: a review. *Hydrological Sciences Journal*, 63(3), 325-352. <https://doi.org/10.1080/02626667.2018.1426858>
- Santos, M., Fragoso, M., et Santos, J.A. (2018). Damaging flood severity assessment in Northern Portugal over more than 150 years (1865–2016). *Nat Hazards* 91, 983–1002. <https://doi.org/10.1007/s11069-017-3166-y>
- Scarrott, C., et MacDonald, A. (2012). A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT-Stat. J.* 10 (1):33–60. <https://doi.org/10.57805/revstat.v10i1.110>.
- Sécurité publique Canada. (2022). Mise à jour de Sécurité Publique : Inondations récentes au Canada. Disponible sur le site : [Base de données canadienne sur les catastrophes \(securitepublique.gc.ca\)](https://www.securitepublique.gc.ca/fr/actualites/2022/05/2022-05-10-inondations-recentes-au-canada)
- Sécurité Publique Canada. (2023). Base de données canadienne sur les catastrophes causées par les inondations dans les provinces des Nouveau-Brunswick, Terre-Neuve-et-Labrador, Nouvelle-Écosse, Île-du-Prince-Édouard et Québec: événements enregistrés entre 2017 et 2023. site: <https://bdc.securitepublique.gc.ca/rs/lts-fra.aspx?cultureCode=fr-Ca&boundingBox=&provinces=4,5,7,10,11&eventTypes=%27FL%27&eventStartDate=%2720160101%27,%2720231231%27&injured=&evacuated=&totalCost=&dead=&normalizedCostYear=1&dynamic=false>. Site visité le 27-02-2023.
- Serago, J. M., et Vogel, R. M. (2018). Parsimonious nonstationary flood frequency analysis. *Advances in Water Resources*, 112, 1-16.

- Shabbar, A., Bonsal, B., et Khandekar, M. (1997). Canadian Precipitation Patterns Associated with the Southern Oscillation. *J. Clim.* 10, 3016–3027.
- Shabbar, A., Higuchi, K., Skinner, W., et Knox, J.L. (1997). The association between the BWA index and winter surface temperature variability over eastern Canada and west Greenland. *Int. J. Climatol.* 17, 1195–1210. [https://doi.org/10.1002/\(SICI\)1097-0088\(199709\)17:11<1195::AID-JOC190>3.0.CO;2-U](https://doi.org/10.1002/(SICI)1097-0088(199709)17:11<1195::AID-JOC190>3.0.CO;2-U)
- Shinyie, W.L., Ismail, N., et Jemain, A.A. (2013). Semi-parametric estimation for selecting optimal threshold of extreme rainfall events. *Water Resour. Manag.* 27, 2325–2352. <https://doi.org/10.1007/s11269-013-0290-7>
- Sikorska, A. E., Viviroli, D., et Seibert, J. (2015). Flood-type classification in mountainous catchments using crisp and fuzzy decision trees. *Water Resour. Res.* 51 (10): 7959–7976. <https://doi.org/10.1002/2015WR017326>.
- Smith, R. L. (1987). Estimating tails of probability distributions. *Ann. Stat.*15 (3): 1174–1207. <https://doi.org/10.1214/aos/1176350499>.
- Smith, R.L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika* 72, 67–90. <https://doi.org/10.1093/biomet/72.1.67>
- Solari, S., et Losada, M. A. (2012). A unified statistical model for hydrological variables including the selection of threshold for the peak over threshold method. *Water Resour. Res.* 48 (5): W10541. <https://doi.org/10.1029/2011WR011475>.
- St-Hilaire, A., Ouarda, T. B., Lachance, M., Bobée, B., Barbet, M., et Bruneau, P. (2003). La régionalisation des précipitations: une revue bibliographique des développements récents. *Revue des sciences de l'eau*, 16(1), 27-54. doi : <https://doi.org/10.7202/705497ar>
- Storm, B., et Refsgaard, A. (1990). Distributed physically-based modelling of the entire land phase of the hydrological cycle. In *Distributed Hydrological Modelling*, (Springer), pp. 55–69.
- Tan, X., et Gan, T.Y. (2015). Nonstationary analysis of annual maximum streamflow of Canada. *J. Clim.* 28, 1788–1805.
- Tancredi, A., Anderson, C., et O’Hagan, A. (2006). Accounting for threshold uncertainty in extreme value estimation. *Extremes* 9 (2): 87–106. <https://doi.org/10.1007/s10687-006-0009-8>.
- Tarasova, L., Merz, R., Kiss, A., Basso, S., Blöschl, G., Merz, B., et Wietzke, L. (2019). Causative classification of river flood events. *Wiley Interdiscip. Rev.: Water* 6 (4): e1353. <https://doi.org/10.1002/wat2.1353>.
- Thiombiano, A. N., El Adlouni, S., St-Hilaire, A., Ouarda, T. B., et El-Jabi, N. (2017). Nonstationary frequency analysis of extreme daily precipitation amounts in southeastern Canada using a peaks-overthreshold approach.” *Theor. Appl. Climatol.* 129 (1): 413–426. <https://doi.org/10.1007/s00704-016-1789-7>.

- Trenberth, K.E., et Shea, D.J. (2006). Atlantic hurricanes and natural variability in 2005. *Geophys. Res. Lett.* 33. <https://doi.org/10.1029/2006GL026894>
- Triguero, I., García, S., et Herrera, F., (2015). Self-labeled techniques for semi-supervised learning: Taxonomy, software and empirical study. *Knowl. Inf. Syst.*, 42, 245–284.
- Turcotte, B., Burrell, B. C., Beltaos, S., et She, Y. (2019). The impact of climate change on breakup ice jams in Canada: State of knowledge and research approaches. In *Proceedings of the 20th Workshop on the Hydraulics of Ice Covered Rivers*, Ottawa, Canada, available at: <http://cripe.ca/publications/proceedings/20> (last access: 20 July 2020).
- Turcotte, R., Fortier Filion, T.-C., Lacombe, P., Fortin, V., Roy, A., et Royer, A. (2010). Simulation hydrologique des derniers jours de la crue de printemps: le problème de la neige manquante. *Hydrol. Sci. Journal–Journal Sci. Hydrol.* 55, 872–882.
- US Army Corps of Engineering Hydrologic Engineering Center. 1964. About HEC. <https://www.hec.usace.army.mil/software/hec-ras/>. Visité le: 18-02-2023.
- Vasiliades, L., Galiatsatou, P., et Loukas, A. (2015). Nonstationary frequency analysis of annual maximum rainfall using climate covariates. *Water Resour. Manag.* 29, 339–358. <https://doi.org/10.1007/s11269-014-0761-5>
- Viglione, A., Merz, R et Blöschl, G. (2009). On the role of the runoff coefficient in the mapping of rainfall to flood return periods. *Hydrol. Earth Syst. Sci.* 13 (5): 577–593. <https://doi.org/10.5194/hess-13-577-2009>.
- Villarini, G., Smith, J.A., Baeck, M.L., et Krajewski, W.F. (2011). Examining flood frequency distributions in the Midwest US. *JAWRA J. Am. Water Resour. Assoc.* 47, 447–463.
- Vogel, R.M., Lall, U., Cai, X., Rajagopalan, B., Weiskel, P.K., Hooper, R.P., et Matalas, N.C. (2015). Hydrology: The interdisciplinary science of water. *Water Resour. Res.* 51, 4409– 4430.
- Volpi, E., Fiori, A., Grimaldi, S., Lombardo, F., et Koutsoyiannis, D. (2019). Save hydrological observations! Return period estimation without data decimation. *J. Hydrol.* 571 (Apr): 782–792. <https://doi.org/10.1016/j.jhydrol.2019.02.017>.
- Wang, B., Lee, M. Y., Xie, Z., Lu, M., et Pan, M. (2022). A new Asian/North American teleconnection linking clustered extreme precipitation from Indian to Canada. *npj Climate and Atmospheric Science*, 5(1), 1-8.
- Wang, Q. J. (1991). The POT model described by the generalized Pareto distribution with Poisson arrival rate. *Journal of Hydrology*, 129(1–4), 263–280.

- Wazneh, H., Arain, M.A., Coulibaly, P., et Gachon, P. (2020). Evaluating the Dependence between Temperature and Precipitation to Better Estimate the Risks of Concurrent Extreme Weather Events. *Advances in Meteorology*, 2020, Article ID 8763631. <http://dx.doi.org/10.1155/2020/8763631>.
- Wazneh, H., Gachon, P., Laprise, R., de Vernal, A., et Tremblay, B. (2021). Atmospheric blocking events in the North Atlantic: trends and links to climate anomalies and teleconnections. *Clim. Dyn.* 56, 2199–2221. <https://doi.org/10.1007/s00382-020-05583-x>.
- Westra, S., Alexander, L.V., et Zwiers, F.W. (2013). Global increasing trends in annual maximum daily precipitation. *J. Clim.* 26, 3904–3918.
- Whitfield, P.H. (2012). Floods in future climates: a review. *J. Flood Risk Manag.* 5, 336–365.
- World Meteorological Organization (WMO). (2021). Atlas of mortality and economic losses from weather, climate and water extremes (1970-2019), 1123WMO-No. 1267, p. 89, available at: https://library.wmo.int/doc_num.php?explnum_id=10989
- Yan, R., Huang, J., Wang, Y., Gao, J., et Qi, L. (2016). Modeling the combined impact of future climate and land use changes on streamflow of Xinjiang Basin, China. *Hydrol. Res.* 47, 356–372.
- Yang, L., Franzke, C. L. E., et Fu, Z. (2019). Power-law behaviour of hourly precipitation intensity and dry spell duration over the United States. *Int. J. Climatol.* 39 (5): 1–16. <https://doi.org/10.1002/joc.6343>.
- Yee, T. W., et Stephenson, A. G. (2007). Vector generalized linear and additive extreme value models. *Extremes* 10 (1): 1–19. <https://doi.org/10.1007/s10687-007-0032-4>.
- Yevdokimov, Y. (2022). Extreme Weather Events and Climate Change: the Case of the Province of New Brunswick in Atlantic Canada. *Економіка розвитку систем*, 2(1), 38-43.
- Yilmaz, A. G., Imteaz, M. A., et Perera, B. J. C. (2017). Investigation of non-stationarity of extreme rainfalls and spatial variability of rainfall intensity–frequency–duration relationships: A case study of Victoria, Australia. *Int. J. Climatol.* 37 (5): 430–442. <https://doi.org/10.1002/joc.4716>.
- Yousfi, N. (2016). Modélisation bayésienne non stationnaire des événements hydrométéorologiques extrêmes au Nouveau-Brunswick. Département de génie civil, Faculté d'ingénierie, Université de Moncton.
- Yousfi, N., et Adlouni, S. E. (2016). Regularized Bayesian estimation for GEV-B-splines model. *Stochastic Environmental Research and Risk Assessment*, 31, 535-550. <https://doi.org/10.1007/s00477-016-1295-6>
- Yu, X., et Bourque, C. P. A. (2022). Controls of contemporary (2001–2018) springtime waterflow dynamics in a Large, snowmelt-dominated basin in northeastern North America. *Journal of Hydrology X*, 14, 100117. <https://doi.org/10.1016/j.hydroa.2021.100117>

- Zadeh, S. M., Burn, D. H., et O'Brien, N. (2020). Detection of trends in flood magnitude and frequency in Canada. *Journal of Hydrology: Regional Studies*, 28, 100673. <https://doi.org/10.1016/j.ejrh.2020.100673>
- Zaffaroni, P. T., Houspanossian, J., Di Bella, C. M., et Jobbagy, E. G. (2023). Quantifying water cover shifts across the globe: following the steps of walking floods. *Authorea Preprints*. DOI: [10.22541/essoar.169008297.72074081/v1](https://doi.org/10.22541/essoar.169008297.72074081/v1)
- Zhang, Z., Stadnyk, T. A., et Burn, D. H. (2020). Identification of a preferred statistical distribution for at-site flood frequency analysis in Canada. *Canadian Water Resources Journal/Revue canadienne des ressources hydriques*, 45(1), 43-58. <https://doi.org/10.1080/07011784.2019.1691942>
- Zhao, X., Zhang, Z., Cheng, W., et Zhang, P. (2019). A new parameter estimator for the generalized Pareto distribution under the peaks over threshold framework. *Mathematics*, 7(5), 406.
- Zischg, A.P., Felder, G., Weingartner, R., Quinn, N., Coxon, G., Neal, J., Freer, J., et Bates, P. (2018). Effects of variability in probable maximum precipitation patterns on flood losses. *Hydrol. Earth Syst. Sci.* 22, 2759–2773. <https://doi.org/10.5194/hess-22-2759-2018>.