Investigating the Modelling Uncertainties Associated with the Generation of Flood Projections

by

Mariana CASTANEDA-GONZALEZ

MANUSCRIPT-BASED THESIS PRESENTED TO ÉCOLE DE
TECHNOLOGIE SUPÉRIEURE IN PARTIAL FULFILLMENT FOR THE
DEGREE OF DOCTOR OF PHILOSOPHY
Ph.D.

MONTREAL, AUGUST 12TH, 2022

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC

ACKNOWLEDGMENT

Étude des incertitudes de modélisation associées à la génération des projections de crues

Mariana CASTANEDA-GONZALEZ

RÉSUMÉ

Les crues extrêmes continuent d'être l'une des catastrophes naturelles les plus menaçantes dans le monde en raison de leurs grands impacts sociaux, environnementaux et économiques. Des changements dans l'ampleur et la fréquence des crues ont été documentés au cours des dernières années, et il est prévu que les changements climatiques continuent d'affecter leur occurrence. Par conséquent, comprendre les impacts de changements climatiques à travers des simulations hydroclimatiques est devenu essentiel pour préparer des stratégies d'adaptation pour l'avenir. Présentement, la confiance dans les projections des crues est encore faible en raison des incertitudes liées à leur simulation, et de la complexité des caractéristiques locales influençant ces événements. L'objectif principal de cette thèse de doctorat est donc d'améliorer notre compréhension des incertitudes associées à la génération de projections des crues ainsi que d'évaluer des stratégies pour réduire ces incertitudes afin d'augmenter notre confiance sur les crues simulées. Pour répondre à l'objectif principal, ce projet visait à (1) quantifier les contributions à l'incertitude des différents éléments impliqués dans la chaîne de modélisation utilisée pour produire des projections de crues, et (2) évaluer les effets de différentes stratégies pour réduire les incertitudes associées aux modèles climatiques et hydrologiques dans des régions aux conditions hydroclimatiques diversifiées. Un total de 96 bassins situés au Québec (bassins dominés par les processus liés à la neige) et au Mexique (bassins dominés par les processus liés à la pluie), couvrant une grande variété de régimes climatiques et hydrologiques, ont été inclus dans l'étude.

La première étape a consisté à décomposer les contributions à l'incertitude de quatre sources impliquées dans la génération des projections de crues: (1) les modèles climatiques, (2) les méthodes de post-traitement, (3) les modèles hydrologiques, et (4) les distributions de probabilité utilisées dans les analyses fréquentielles. Une méthode de décomposition de la variance a permis de quantifier et de classer l'influence de chaque source d'incertitude sur les crues dans les deux régions étudiées et ce, par saison. Les résultats ont montré que les contributions à l'incertitude de chaque source varient selon les régions et les saisons. Les régions et les saisons dominées par la pluie ont montré que les modèles climatiques étaient la principale source d'incertitude, tandis que celles dominées par la fonte des neiges ont montré que les modèles hydrologiques étaient leur principale source d'incertitude. Ces résultats montrent les dangers d'utiliser des modèles climatiques et hydrologiques uniques, et soulignent l'importance des analyses d'incertitudes régionales.

La deuxième étape de ce projet de recherche s'est concentrée sur l'évaluation de stratégies visant à réduire l'enveloppe d'incertitude associé aux modèles hydrologiques sur les projections de crues. Cette étape comporte deux volets: (1) l'analyse de la fiabilité de la calibration et la validation des modèles hydrologiques sous un climat changeant et (2)

l'évaluation des effets des stratégies de pondération des simulations hydrologiques sur les projections de crues. Pour répondre à la première partie, différentes stratégies de calibration ont été testées et évaluées à l'aide de cinq modèles hydrologiques conceptuels globaux dans des conditions climatiques contrastées dont les longueurs ont varié de 2 jusqu'à 21 ans. Les résultats ont révélé que les conditions climatiques des données utilisées pour la calibration ont un impact plus important sur les performances des modèles hydrologiques que la longueur des séries temporelles climatiques. Les changements sur les précipitations ont généralement montré des impacts plus importants que les changements sur la température sur tous les bassins. Ces résultats suggèrent que des périodes de calibration et validation courtes qui sont plus représentatives des changements possibles des conditions climatiques pourraient être plus adéquates pour les études des impacts du changement climatique. Suite à l'obtention de ces résultats, les effets de différentes stratégies de pondération basées sur la robustesse des modèles hydrologiques (en conditions climatiques contrastées) ont été évalués sur les projections de crues des différents bassins. La pondération des simulation issues des mêmes cinq modèles hydrologiques globaux conceptuels, en fonction de leur robustesse, a montré certaines améliorations par rapport à l'approche traditionnelle de pondération égale, en particulier dans des conditions plus chaudes et plus sèches. De plus, les résultats montrent que la différence entre ces approches est plus prononcée sur les projections de crues, car différentes ampleurs de crues et signaux de changement climatique ont été produits. Des analyses supplémentaires effectuées à l'aide d'un modèle hydrologique semi-distribué et de base plus physique sur quatre bassins sélectionnés ont suggéré que ce type de modèle pouvait avoir une valeur ajoutée lors de la simulation d'étiages, et de hauts débits sur de petits bassins (dont la superficie est de l'ordre de 500 km^2). Ces résultats soulignent à nouveau l'importance de travailler avec des ensembles de modèles hydrologiques et présentent les impacts potentiels de la pondération des modèles hydrologiques sur les études d'impacts des changements climatiques.

La dernière étape de cette étude a porté sur l'évaluation des impacts de la pondération des simulations climatiques sur les projections des crues. Les différentes stratégies de pondération testées ont montré que la pondération des simulations climatiques peut améliorer la représentation des hydrogrammes annuels moyens par rapport à l'approche traditionnelle basée sur la « démocratie » des modèles. Cette amélioration a été principalement observée avec l'approche de pondération proposée dans cette thèse qui évalue la performance des débits saisonniers simulés par rapport aux observations. Les résultats ont également révélé que la pondération des simulations climatiques en fonction de leurs performances pouvait: (1) modifier l'ampleur de crues projetées, (2) modifier les signaux de changement climatique et (3) réduire les enveloppes d''incertitude des projections de crues résultantes. Ces effets étaient particulièrement clairs sur les bassins dominés par la pluie, où l'incertitude de la modélisation climatique joue un rôle principal. Ces résultats soulignent la nécessité de reconsidérer l'approche traditionnelle de la démocratie des modèles climatiques, en particulier lors de l'étude de processus avec des niveaux d'incertitude climatique plus élevés.

Enfin, les implications des résultats obtenus ont été discutées. Cette section met les principaux résultats en perspective et identifie différentes voies à suivre pour continuer à améliorer la compréhension des impacts du changement climatique sur l'hydrologie et accroître notre confiance dans les projections de crues qui guident les stratégies d'adaptation pour l'avenir.

Investigating the Modelling Uncertainties Associated with the Generation of Flood Projections

Mariana CASTANEDA-GONZALEZ

ABSTRACT

Extreme flood events continue to be one of the most threatening natural disasters around the world due to their pronounced social, environmental and economic impacts. Changes in the magnitude and frequency of floods have been documented during the last years, and it is expected that a changing climate will continue to affect their occurrence. Therefore, understanding the impacts of climate change through hydroclimatic simulations has become essential to prepare adaptation strategies for the future. However, the confidence in flood projections is still low due to the considerable uncertainties associated with their simulations, and the complexity of local features influencing these events. The main objective of this doctoral thesis is thus to improve our understanding of the modelling uncertainties associated with the generation of flood projections as well as evaluating strategies to reduce these uncertainties to increase our confidence in flood simulations. To address the main objective, this project aimed at (1) quantifying the uncertainty contributions of different elements involved in the modelling chain used to produce flood projections and, (2) evaluating the effects of different strategies to reduce the uncertainties associated with climate and hydrological models in regions with diverse hydroclimatic conditions. A total of 96 basins located in Quebec (basins dominated by snow-related processes) and Mexico (basins dominated by rain-related processes), covering a wide range of climatic and hydrological regimes were included in the study.

The first stage consisted in decomposing the uncertainty contributions of four main uncertainty sources involved in the generation of flood projections: (1) climate models, (2) post-processing methods, (3) hydrological models, and (4) probability distributions used in flood frequency analyses. A variance decomposition method allowed quantifying and ranking the influence of each uncertainty source on floods over the two regions studied and by seasons. The results showed that the uncertainty contributions of each source vary over the different regions and seasons. Regions and seasons dominated by rain showed climate models as the main uncertainty source, while those dominated by snowmelt showed hydrological models as the main uncertainty contributor. These findings not only show the dangers of relying on single climate and hydrological models, but also underline the importance of regional uncertainty analyses.

The second stage of this research project focused in evaluating strategies to reduce the uncertainties arising from hydrological models on flood projections. This stage includes two steps: (1) the analysis of the reliability of hydrological model's calibration under a changing climate and (2) the evaluation of the effects of weighting hydrological simulations on flood projections. To address the first part, different calibration strategies were tested and evaluated using five conceptual lumped hydrological models under contrasting climate conditions with

datasets lengths varying from 2 up to 21 years. The results revealed that the climatic conditions of the calibration data have larger impacts on hydrological model's performance than the lengths of the climate time series. Moreover, changes on precipitation generally showed greater impacts than changes in temperature across all the different basins. These results suggest that shorter calibration and validation periods that are more representative of possible changes in climatic conditions could be more appropriate for climate change impact studies. Following these findings, the effects of different weighting strategies based on the robustness of hydrological models (in contrasting climatic conditions) were assessed on flood projections of the different studied basins. Weighting the five hydrological models based on their robustness showed some improvements over the traditional equal-weighting approach, particularly over warmer and drier conditions. Moreover, the results showed that the difference between these approaches was more pronounced over flood projections, as contrasting flood magnitudes and climate change signals were observed between both approaches. Additional analyses performed over four selected basins using a semi-distributed and more physically-based hydrological model suggested that this type of models might have an added value when simulating low-flows, and high-flows on small basins (of about 500 km^2). These results highlight once again the importance of working with ensembles of hydrological models and presents the potential impacts of weighting hydrological models on climate change impact studies.

The final stage of this study focused on evaluating the impacts of weighting climate simulations on flood projections. The different weighting strategies tested showed that weighting climate simulations can improve the mean hydrograph representation compared to the traditional model "democracy" approach. This improvement was mainly observed with a weighting approach proposed in this thesis that evaluates the skill of the seasonal simulated streamflow against observations. The results also revealed that weighting climate simulations based on their performance can: (1) impact the floods magnitudes, (2) impact the climate change signals, and (3) reduce the uncertainty spreads of the resulting flood projection. These effects were particularly clear over rain-dominated basins, where climate modelling uncertainty plays a main role. These finding emphasize the need to reconsider the traditional climate model democracy approach, especially when studying processes with higher levels of climatic uncertainty.

Finally, the implications of the obtained results were discussed. This section puts the main findings into perspective and identifies different ways forward to keep improving the understanding of climate change impacts in hydrology and increasing our confidence on flood projections that are essential to guide adaptation strategies for the future.

Keywords: uncertainty, climate modelling, hydrological modelling, weighting models, floods.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABREVIATIONS

AOGCMs   Atmosphere-Ocean General Circulation Models

AR5   Fifth Assessment Report of the Intergovernmental Panel on Climate Change

AR6   Sixth Assessment Report of the Intergovernmental Panel on Climate Change

BDH   Banque de Données Hydriques

BIC   Bayesian Information Criterion

CAM   Central American domain

CanESM2   Canadian Earth System Model version 2

CEHQ   Centre d'Expertise Hydrique du Québec

CGCM   Canadian General Circulation Model

CMIP   Coordinated Modelling Intercomparison Project

CORDEX   Coordinated Regional Downscaling Experiment

CRCM5   Canadian Regional Climate Model version 5

CRED   Centre for Research on the Epidemiology of Disasters

DBC   Daily Bias Correction

DJF   December January February

DT   Daily translation

ED   Euclidean Distance

ESMs   Earth System Models

ET   Evapotranspiration

GCMs   General Circulation Models

GEV   Generalized Extreme Value distribution

GR4J   A lumped conceptual rainfall-runoff model with 6 and/or 9 parameters

XXX

HMETS	Hydrological Model École de Technologie Supérieure
JJA	June July August
KGE	Kling-Gupta Efficiency
LAMs	Limited Area Models
LOCI	Local intensity scaling
MAM	March April May
MME	Multi-Model Ensemble
MOHYSE	MOdèle HYdrologique Simplifié à l'Extrême
NSE	Nash-Sutcliffe Efficiency
QC	Quebec domain
RCM	Regional Climate Model
RCP	Representative Concentration Pathways
SON	September October November
SSP	Socio-Economic Pathways
UNDRR	United Nations Disaster Risk Reduction

LIST OF SYMBOLS

%	percent
°C	degree Celsius
m	meters
km	kilometer
s	second
W	watt

INTRODUCTION

Floods continue to be one of the most threatening natural disasters to our societies (Blöschl, 2022; Merz et al., 2021). Since the Fifth Assessment Report (AR5), the Intergovernmental Panel on Climate Change (IPCC) underlined the need to focus on changes on extreme climate and hydrological events (Hartmann et al., 2013; IPCC, 2013b).Currently, on their most recent Sixth Assessment Report (AR6), the IPCC shows with additional and strengthened evidence that the global water cycle has been impacted, and will continue to intensify with the rising global temperatures (IPCC, 2021). This non-stationary behaviour in hydrological processes has gained importance to the hydrological community as it directly impacts different fields that have traditionally relied in stationary conditions such as infrastructure design, water management, risk analysis, ecosystem assessments, and flood assessments (Byun & Hamlet, 2020).

During the last years, the impacts and costs of extreme flooding have shown great increases in different regions of the world (Allaire, 2018). According to the Centre for Research on the Epidemiology of Disasters (CRED) and the United Nations Disaster Risk Reduction (UNDRR), floods that occurred globally between 2000 and 2019 caused about $651 billion USD in damages alone (CRED, 2020). Just during the following years, important flooding events have occurred around the world. For instance, the major flood in Germany in 2021, the flood in the Henan province of China, the 2020 flood in Bangladesh, and the major flood in the Canadian province of British Columbia in 2021 (Blöschl, 2022). The historical great impacts of these events along with the observed increments on the severity and frequency of floods around the world (Tellman et al., 2021), makes more evident that the traditional approaches of flood management that rely on extrapolations from historical records are becoming insufficient.

Different strategies have been proposed and applied in flood-prone areas as a way to mitigate and adapt to extreme flooding events. A common approach for assessing and developing mitigation and adaptation plans consists of using hydrological projections obtained through the

combination of climate simulations, hydrological models and statistical analyses. However, there are no perfect models and methods, as stated by George EP Box "essentially, all models are wrong, but some are useful" (Box & Draper, 1987). Thus, numerous and varied climate, hydrological and statistical models with varying levels of complexity have been developed throughout the years. However, each of these tools, with their own advantages and disadvantages, add uncertainty to the resulting climate change impact assessments. Due to the high levels of uncertainty associated with the different stages of floods simulation and the complexity of interacting processes that influence their occurrence, the confidence on flooding projections remains low (Giuntoli, Prosdocimi, & Hannah, 2021; IPCC, 2021; Kundzewicz et al., 2014; Kundzewicz et al., 2018). Thus, improving our understanding of these extreme hydrological events, and evaluating the potential impacts of climate change on their occurrence has become of great importance for the research community (IPCC, 2018).

Presently, the use of multi-model ensembles (MMEs) has become the standard approach in climate change impact studies as it allows obtaining ranges of projections and quantifying the uncertainties arising from the different modelling stages(Kundzewicz et al., 2018). From these multi-model hydroclimatic ensembles, the mean and spread of the MMEs have been commonly used to describe and assess the potential effects of climate change on hydrology. This approach assumes that all simulations are equally plausible, meaning that all hydroclimatic simulations are considered equally. However, this "democracy" approach has been questioned as it overlooks the regional performance and adequacy of the models and methods included in the ensemble. Thus, different studies have underlined the need to reconsider this approach, and have proposed different strategies that might reduce the overall modelling uncertainties and help increasing our confidence on climate change impact studies. For this reason, the main objective of this project is to improve the understanding on the modelling uncertainties associated with the generation of flood projections and analyse different strategies to reduce them in order to increase our confidence in the simulated impacts of climate change on floods.

To investigate the main objective, this research project is divided in two main research lines: (1) the first research line aims at quantifying the uncertainty contributions of different elements

involved in the modelling chain used to generate flood projections; and (2) the second research line focuses on evaluating the effects of different strategies to reduce uncertainties associated with hydrological and climate models on flooding projections. To address these research lines, this project uses 96 North American basins located in the Canadian province of Quebec and the country of Mexico, to cover a diversity of hydroclimatic regimes. These two regions were selected as part of a long-term international collaboration between Quebec and Mexico, where this project contributes. More particularly, this doctoral thesis contributes through the project « *Complémentarité et diversité Québec-Mexique en matière de gestion de l'eau dans le contexte des changements climatiques: impacts sur les régimes hydriques et prise de décision en adaptation* » that investigates the impacts of climate change on hydrological extremes.

This thesis is composed of seven chapters. First, a detailed literature review describing and discussing state-of-the-art studies that are relevant to the research question as well as the main objectives of the present thesis are presented in Chapter 1. The following Chapter 2, presents the overall methodology of the present project and describes the present thesis organization in more detail. Chapters 3 to 6, present the four articles derived from the two research lines addressing this project's main objective. To put the results and findings into perspective, the last Chapter 7 provides a general discussion. Following this chapter, the main conclusions and recommendations are presented. A list describing other scientific contributions derived from the present thesis, such as collaboration articles and oral presentations are presented in the appendices.

CHAPTER 1

LITERATURE REVIEW

This chapter is dedicated to presenting a thorough review of relevant literature on flood simulations under climate change and their associated modelling uncertainties. First, a description of the different processes and interacting factors that influence and generate flooding events is presented. Then, relevant studies showing the impacts of climate change on floods are described followed by a detailed description of the different methodological stages needed to produce flood projections. Finally, the uncertainties associated to the different modelling stages are reviewed.

1.1 Flooding

As clearly defined in the AR6 of the IPCC, "*floods are the inundation of normally dry land*" (IPCC, 2021). In natural conditions, the occurrence of this common phenomenon can be driven by different climatic and hydrological factors, such as changes on rainfall, snow and ice melt, and land use variations. In addition to these hydroclimatic and human-driven factors, hydrological processes also vary in space and time. As presented in Figure 1.1, the spatial and temporal interrelationships of hydrological and meteorological processes cover wide ranges of variation going from one to thousands of meters in space, and from minutes to millennia in time. The various hydroclimatic factors that can influence the occurrence of a flood event and their spatiotemporal variations makes their detection and attribution a complex task (Wehner, Arnold, Knutson, Kunkel, & LeGrande, 2017).

Depending on the main hydroclimatic process generating flood events and their spatiotemporal scales, they can be classified into different types. For instance, (1) pluvial floods, (2) coastal floods, (3) river (fluvial) floods, (4) groundwater floods, (5) urban floods, among others (Blöschl, 2022; IPCC, 2021). Thus, the appropriate approach to study this complex hydrometeorological event, highly depends on the objective (Sivakumar, 2017). Most inland floods occur when river banks are surpassed (i.e., fluvial floods) or due to heavy precipitation

(pluvial flooding). Thus, studies evaluating flooding events, as performed in this research project, are often carried out at the river basin scale.
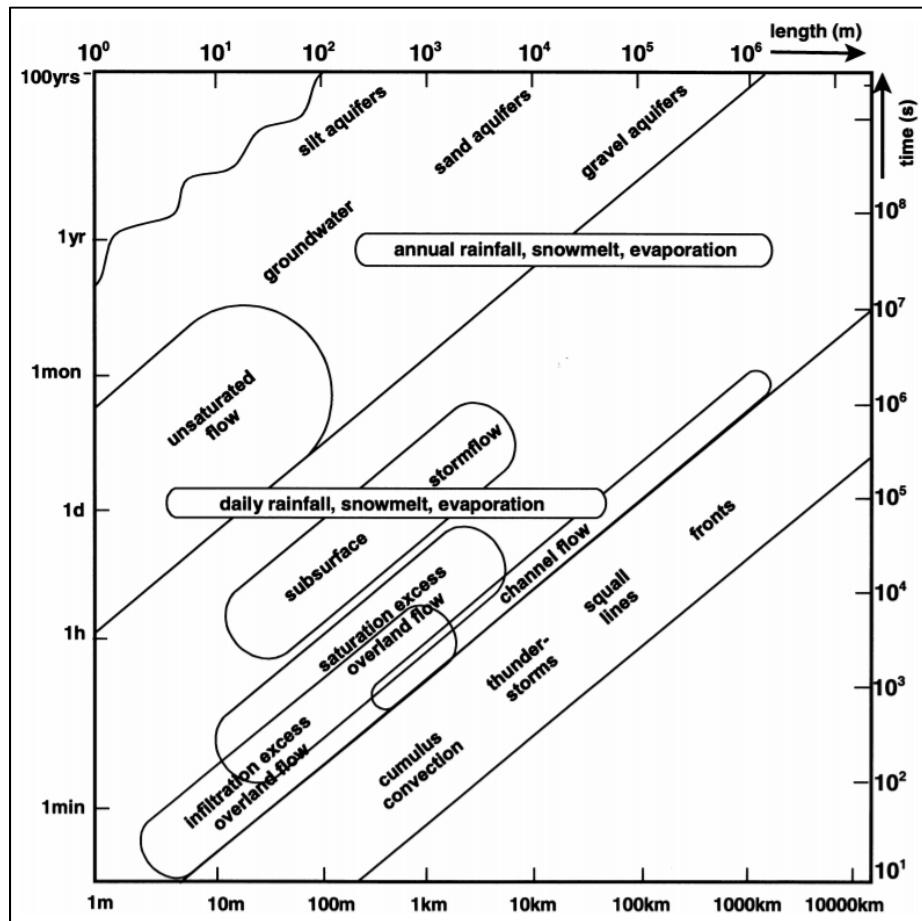
Figure 1.1 Space and time interrelationships in hydrometeorological processes (Blöschl & Sivapalan, 1995; Schulze, 2000) after Blöschl & Sivapalan, (1995)

1.2   Floods in a changing climate

This hydrological event not only has historically influenced societies and environments around the world, but it also continues to impact our modern societies (Camargo & Cortesi, 2019; Gaines, 2016; IPCC, 2013b, 2021). A recent study evaluating satellite images over all continents revealed that the proportion of population exposed to floods has increased by 20 to 24% between 2000 and 2015 (Tellman et al., 2021). At the same time, evidence of increases

on the severity, duration and frequency of floods has been documented in different regions of the world (IPCC, 2021; Merz et al., 2021; Tanoue, Hirabayashi, & Ikeuchi, 2016). These increments on population exposure along with the observed changes on flood patterns have raised concerns for the future, especially knowing that climate change has been associated with changes in different hydrological processes (e.g., precipitation trends, alteration on the melting of snow and ice, and runoff variations) that can influence the occurrence of floods (Bates, Kundzewicz, Wu, & Palutikof, 2008; Coppola, Raffaele, & Giorgi, 2016; Vormoor, Lawrence, Schlichting, Wilson, & Wong, 2016; Wehner et al., 2017).

Among the different hydroclimatic processes that influence the occurrence of flooding events, changes in precipitation have largely been studied during the last years (Martel, Brissette, Lucas-Picher, Troin, & Arsenault, 2021). Impacts on precipitation patterns have been anticipated as larger amounts of water vapor can be held in a warmer atmosphere, which can potentially lead to increases in rainfall in the future (Trenberth, 2011; Westra, Alexander, & Zwiers, 2013). During the last years, strengthened and increasing evidence based on observations and climate model projections have shown that extreme rainfall events are already intensifying and are expected to increase at both global and regional scales (IPCC, 2021; Martel, Mailhot, & Brissette, 2020), an issue that can directly impact future flooding events. The clear impacts on the water cycle have highlighted the importance of better understanding the effects of climate change on floods (IPCC, 2018, 2021). However, the complex interactions between the different hydroclimatic factors that influence floods' occurrence, makes their analysis a more complex task (Bates et al., 2008; Field, 2012; Kundzewicz et al., 2014).

Currently, mixed effects of climate change on floods were reported in the literature (IPCC, 2021). For instance, over snow-dominated regions, it has been observed that a warming climate can lead to an earlier and more extensive snowmelt season, resulting in generally smaller snowmelt-generated flood peaks (Peng et al., 2013; Rhoades, Jones, & Ullrich, 2018). However, over regions dominated by rainfall events, contrasting conclusions have been observed showing increments and decreases on flood peaks (Blöschl et al., 2019; Burn & Whitfield, 2016). Among them, a recent large-scale study by Blöschl et al. (2019) showed

general increases in winter and autumn rainfall-driven floods in northwestern European regions, while southern parts of the continent showed a decreasing behaviour. Similarly, varying flood projections depending on the basins' characteristics and their main flood-generating process have been observed with flood decreases in some regions and increasing floods in others (Burn & Whitfield, 2016; Giuntoli et al., 2021). These contrasting results show the strong regional dependency of flooding projections around the world (Blöschl et al., 2019; IPCC, 2021), and underline the importance of regional and/or local climate change impact studies to better understand the potential local impacts of a warming world.

1.3 Developing flood projections

Generating hydrological and flooding projections generally comprises a sequence of different modelling stages. In general, as represented in Figure 1.2, this chain of processes consists of a series of hydroclimatic modelling steps that generally include: (1) climate modelling, (2) post-processing methods, (3) hydrological modelling, and in the case of floods (4) flood frequency analyses.
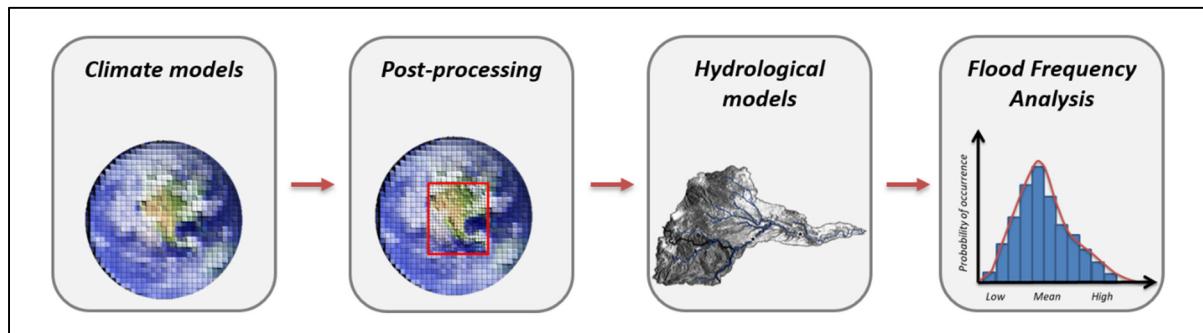
Figure 1.2 Common hydroclimatic modelling chain used in climate change impact studies in hydrology

1.3.1 Climate modelling

The first and essential stage of the hydroclimatic modelling chain at the basin scale is the climate modelling. The unprecedented rising temperatures observed during the last decades

have mobilized the research community to better understand the Earth's climate system by developing and constantly improving climate and earth system models that allow simulating potential future scenarios (AghaKouchak, Easterling, Hsu, Schubert, & Sorooshian, 2012; IPCC, 2021). A great variety of Global Climate Models (GCMs) that simulate the climate system with different levels of complexity and mathematical representations have been developed around the world (Randall et al., 2007; Trenberth, 1992). Currently, the most complex GCMs available are the atmosphere-ocean general circulation models (AOGCMs) and the earth system models (ESMs) that not only integrate physical climatic processes but also include biosphere and chemical processes interactions that provide more accurate representations of the climate system (Heavens, Ward, & Natalie, 2013; IPCC, 2021).

Climate models use and integrate large amounts of information describing the state of the Earth's surface and atmosphere, such as topography, soil types, and the atmosphere's chemical concentrations (Trenberth, 1992). Changes in these chemical concentrations, particularly in greenhouse gases and aerosols, are used to describe plausible future scenarios known as *emission scenarios*. Different emission scenarios families have been developed throughout the last decade, such as the Representative Concentration Pathways (RCP) used in the Fifth Assessment Report (AR5; IPCC, 2013a), and the most recent Shared Socio-Economic Pathways (SSPs) of the Sixth Assessment Report (AR6; IPCC, 2021). Four different RCPs including one mitigation scenario (RCP2.6), two stabilization scenarios (RCP4.5 and RCP6) and one scenario with very high greenhouse gas emissions (RCP8.5) were developed to attempt representing the largest range of plausible emissions based on the amount of radiative forcing, in W/m^2, reached at the end of the century (IPCC, 2013b; Moss et al., 2010). The new five SSPs, namely the SSP1-1.9, SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5, include new emission scenarios (1.9 and 7.0 W m^{-2}) and cover five socio-economic pathways (IPCC, 2021). To group the different GCM simulations and emission scenarios, collaboration projects as the Coordinated Modelling Intercomparison Project (CMIP) phase 3 (Meehl et al., 2007), the CMIP5 (Taylor, Stouffer, & Meehl, 2011) and the most recent CMIP6 (Haarsma et al., 2016) have been created and updated to assemble the different climate simulations available.

1.3.2 Post-processing

The spatial resolutions of most GCMs available range from 100 to 300 km in the atmosphere with few recent high-resolution GCMs part of the High-Resolution Model Intercomparison Project (HighResMIP) experiments reaching 25 to 50 km resolutions (Haarsma et al., 2016; Solomon et al., 2007; Teutschbein & Seibert, 2010). These generally coarse resolutions are one of the main drawbacks of global simulations as they are often deemed incompatible for local and regional studies (IPCC, 2013a; Solomon et al., 2007; Teutschbein & Seibert, 2010). One of their main issues is that the intensity, frequency and distribution of precipitation is often inaccurately represented, making it insufficient for studying regional hydrology (Hostetler, 1994; Randall et al., 2007; Teutschbein & Seibert, 2010). The need for higher resolutions has made post-processing methods an essential process that allows overcoming the spatial and temporal gaps while also correcting the inherent biases in climate simulations (Chen, Brissette, & Leconte, 2011; Maraun, 2016; Prudhomme, Reynard, & Crooks, 2002; Riboust & Brissette, 2015; Teutschbein & Seibert, 2010). The post-processing methods focused on overcoming the spatiotemporal scale differences are known as downscaling methods and are categorized in two main types, (1) statistical downscaling and (2) dynamical downscaling.

The statistical downscaling consists in producing future scenarios using statistical (or empirical-statistical) relationships between climate variables at larger scales and regional properties identified from recent climate observations. These post-processing methods involve various techniques such as multiple regressions, bias correction methods, and neural networks that identify statistical relationships between observed regional conditions and climate simulations (Diaz-Nieto & Wilby, 2005; Themeßl, Gobiet, & Leuprecht, 2011; Wilby et al., 1998). Dynamical downscaling, commonly referred to as Limited Area Models (LAMs) or Regional Climate Models (RCMs), consists in complex climate models that describe the atmosphere at finer resolutions and provide a more physically coherent representation of regional climates (Jones, Murphy, & Noguer, 1995; Laprise, 2008). These models are nested within coarser global models, such as a GCM or other gridded dataset such as reanalyses, that feed them with data at the boundaries of the region to simulate (Bengtsson & Shukla, 1988;

Carter, Hulme, & Lal, 2007; Charron, 2014). Similarly to the CMIP ensembles, the Coordinated Regional Downscaling Experiment (CORDEX) project (Giorgi & Gutowski, 2015) assembles RCM simulations with a variety of domains, drivers and resolutions.

Each post-processing technique has shown different advantages and disadvantages. For instance, the statistical downscaling is known for being a relatively simpler, faster and a globally available technique (Chen, Brissette, & Leconte, 2011), while RCMs have shown better representation of fundamental processes and physical coherence. However, their use has been usually hampered by the costly computing resources (Mearns, Bukovsky, Pryor, & Magaña, 2018), as well as significant biases transmitted by the driving models (Hall, 2014; Rummukainen, 2016). This issue has made bias correction techniques a more common practice, particularly for climate change impact studies that need adjusting the biases in GCM and RCM outputs (Maraun, 2016). Thus, the post-processing techniques needed to downscale and bias-correct GCM and RCM outputs, increase the series of processes involved in the hydroclimatic modelling chain.

1.3.3   Hydrological modelling

Hydrological modelling is one of the main stages of the hydroclimatic modelling chain as it allows producing the streamflow simulations for impact assessments at the basin scale. Overall, in this stage the (usually post-processed) climate model outputs are used as inputs for the hydrological models to produce the climate model-driven streamflow simulations for a given region. As observed for the climate models, there is a great variety of hydrological models available with different strategies to simulate the hydrological system. The community has identified two main categories: (1) the stochastic, and (2) the deterministic hydrological models (Te Chow, 1988).

The stochastic models consider randomness (at least partially), meaning that predictions are produced by estimating different potential outcomes. Different stochastic methods have been used in hydrology, such as perturbation and Monte Carlo methods (Sanchez-Vila & Fernàndez-

Garcia, 2016). The deterministic models, unlike the stochastic ones, are characterized by producing a single solution to a given set of inputs, meaning that randomness is not involved (Pechlivanidis, Jackson, Mcintyre, & Wheater, 2011). Their simulations are determined by initial conditions and a set of optimized parameters that serve to simplify and represent different hydrological processes depending on each model structure. Within the deterministic models, the representation of the processes can be physically-based, conceptual or empirical (Singh & Woolhiser, 2002). Other classification consists in the spatial representation of the data, where the lumped and spatially distributed models can be identified. Lumped models consider mean values all over the basin (through a single spatial simulation unit), while the distributed models consider the spatial variability of the variables over the basin (through elementary sub-basins or gridded representations; Beven, 2011; Pechlivanidis et al., 2011). In this research project, deterministic lumped conceptual models and one semi-distributed physically-based model were used.

1.3.3.1   Calibration and validation process

The calibration and validation process is an essential stage needed to find the adequate parameter values for each catchment depending on its particular properties and behaviour (Moore & Doherty, 2005; Pechlivanidis et al., 2011). The calibration of the parameters is usually performed by using optimization algorithms (Arsenault, Poulin, Côté, & Brissette, 2014) that fit the simulation with historical observations. The fitting performance is evaluated through an objective function that measures the simulation performance against the observations (Duan, Sorooshian, & Gupta, 1994; Moriasi et al., 2007). Following the calibration process, the hydrological model is validated using the same objective function over the rest of the period (i.e., the non-calibrated or validation period) available in the historical observations (Arsenault, Brissette, & Martel, 2018). One of the most commonly used types of objective functions are the quadratic error type functions, such as the Nash-Sutcliffe efficiency (NSE) metric (Nash & Sutcliffe, 1970). Gupta, Kling, Yilmaz, et Martinez (2009) proposed a derivative of the NSE metric, the Kling-Gupta Efficiency criterion (KGE). The KGE criterion, later modified by Kling, Fuchs, et Paulin (2012), describes the difference between unity and

the Euclidian distance (ED) from the ideal point in a three-dimensional space based on the bias, variance and correlation. The criterion is calculated as follows in equations 1.1-1.4:

$$KGE = 1 - ED \tag{1.1}$$

$$ED = \sqrt{(r-1)^2 + (\beta - 1)^2 + (\gamma - 1)^2} \tag{1.2}$$

$$\beta = \frac{\mu_s}{\mu_o} \tag{1.3}$$

$$\gamma = \frac{CV_s}{CV_o} = \frac{\sigma_s/\mu_s}{\sigma_o/\mu_o} \tag{1.4}$$

where r is the correlation coefficient, β is the bias ratio, and γ represents the variability ratio between the observation and the simulation. The μ represents the mean, CV is the variation coefficient, and σ is the standard deviation of the dataset. The "o" and "s" subscripts denote the observed and simulated data, respectively.

The use of the KGE criterion has increased in recent years (Beck et al., 2016; Huang et al., 2016; Knoben, Freer, & Woods, 2019; Oyerinde et al., 2017; G. Thirel et al., 2015) as it overcomes problems related to mean squared error functions (e.g., the runoff peaks underestimation) that have been identified when suing the well known NSE criterion (Gupta et al., 2009).

1.3.4 Flood frequency analysis

The frequency analysis is a fundamental and widely used tool in the analysis of extreme hydrological events in water resources. In simple terms, this tool uses historical records and/or simulated streamflows to evaluate the potential future magnitudes and occurrences of floods (Alfieri, Burek, Feyen, & Forzieri, 2015; Arnell & Lloyd-Hughes, 2014; Dankers et al., 2014;

14

Das & Umamahesh, 2018; Roudier et al., 2016; Wasko & Sharma, 2017). Frequency analyses use probability distributions to describe the relation between the magnitude of past events and their frequencies of occurrence (Hamed & Rao, 1999; Te Chow, 1988). In other words, it is a statistical method that estimates the recurrence rates of past events in order to predict their probabilities of occurrence in the future (Luke, Vrugt, AghaKouchak, Matthew, & Sanders, 2017; Musy, Meylan, & Favre, 2012).

For the analysis of floods, series of annual and/or seasonal daily maximum discharges are fitted to a distribution (e.g., Gumbel distribution) to associate extreme events at different magnitudes with their recurrence intervals. These recurrence intervals are defined by using the notion of return periods T, a probabilistic concept that describes the *average recurrence* of a given event X exceeding a known threshold X_T usually expressed in years (Hamed & Rao, 1999; Read & Vogel, 2015; Vogel & Castellarin, 2017). The return period T of an event X is thus represented as the inverse of the probability of exceeding the threshold X_T:

$$T(X_T) = \frac{1}{p(X_T)} \qquad (1.5)$$

where $p(X_T)$ is the probability of exceedance of the event and is estimated as follows:

$$p(X_T) = 1 - F(X_T) \qquad (1.6)$$

where $F(X_T)$ is the cumulative distribution function defined by the selected frequency model. As $F(X_T)$ is complementary to the probability of exceedance, it is referred to as the probability of non-exceedance and is estimated as follows:

$$F(X_T) = 1 - \frac{1}{T} \qquad (1.7)$$

This means that the rarer the event is (i.e., larger return period), the probability of exceedance decreases and the probability of non-exceedance increases (Musy et al., 2012).

In hydrology, this analysis is traditionally used to find the annual maximum discharge corresponding to a high return period (e.g., 100-year), which means a high probability of non-exceedance (e.g., 0.99). For designing infrastructures or assessing risk, the probability that such an event is exceeded at least once in a planning horizon of *N* years is referred to as *probability of failure* and is estimated as follows:

$$Pr(X_T) = 1 - \left(1 - \frac{1}{T}\right)^N \tag{1.8}$$

The complement of the probability of failure is called the *reliability* (Read & Vogel, 2015). Thus, high reliability values are usually targeted for a T-year event over an N-year planning horizon.

This traditional approach of frequency analysis of floods and the notions of return period and probability of failure are based on some critical assumptions that the series of past events are *independent* and identically distributed with *stationary* distributions (Serinaldi & Kilsby, 2015). Thus, it is implied that the statistics of the extreme discharges remain constant in time (Luke et al., 2017; Read & Vogel, 2015). This stationarity assumption is certainly one of the key bases of the flood frequency analysis. However, it has become a critical issue due to the non-stationary observed and potential effects of a changing climate (Serago & Vogel, 2018; Su & Chen, 2019). Different non-stationary flood frequency analyses have been developed to address this issue, and the number of methods has considerably increased (Condon, Gangopadhyay, & Pruitt, 2015; Luke et al., 2017; Rootzén & Katz, 2013; Salas, Obeysekera, & Vogel, 2018; Serago & Vogel, 2018; Serinaldi & Kilsby, 2015; Vogel, Yaindl, & Walter, 2011). In general, these different techniques allow the probability distribution parameters (e.g. location and scale) to vary as a function of time or as a function of one or multiple covariates (i.e. time-dependent variables) such as precipitation or temperature (Condon et al., 2015; Luke et al., 2017; Madsen, Lawrence, Lang, Martinkova, & Kjeldsen, 2013; Salas & Obeysekera, 2014; Vidrio-Sahagún & He, 2022).

1.4   Uncertainties on floods modelling

As reviewed in previous sections, hydrological projections involve different modelling steps that allow producing and analysing flooding projections. However, models and processing methods are not perfect and add uncertainty at each step of the modelling chain. So far, there is still *low confidence* on how regional-scale river flooding might be impacted by climate change in the future due to the limitations and considerable uncertainties associated with their estimations (Kundzewicz et al., 2014; Kundzewicz et al., 2017; Stocker et al., 2013). Thus, uncertainty assessments in hydrology have become an important issue that needs to be considered to improve our confidence in hydrological projections (Kundzewicz et al., 2018; Nearing et al., 2016; Schneider, 1983; Wilby & Dessai, 2010). It is important to underline that in this and the following chapters, *uncertainty* solely refers to uncertainties associated with the different modelling stages, also referred to as model response uncertainty by the IPCC (2021).

Numerous sources of uncertainties have been identified in the process of hydrological modelling and climate change impact assessments (Prudhomme, Jakob, & Svensson, 2003). The so-called *uncertainty paradigm* (Rummukainen, 2016), "*cascade of uncertainty*" (Schneider, 1983; Wilby & Dessai, 2010) or "*uncertainty explosion*" (Henderson & Sellers, 1993; Jones, 2000) is essentially based on the expansion or addition of uncertainties in each step of the hydroclimatic modelling chain. This cascade of uncertainties considers the uncertainties associated with greenhouse gas emission scenarios, global climate model structure, post-processing methods, impact (e.g., hydrological) models, and the natural climate variability among the most important ones in hydrology (Falloon et al., 2014; Giuntoli, Vidal, Prudhomme, & Hannah, 2015; Krysanova et al., 2018; Poulin, Brissette, Leconte, Arsenault, & Malo, 2011; Wilby, 2005).

Additional sources of uncertainty in the hydroclimatic modelling chain continue to be acknowledged, highlighting the importance to identify and quantify the most important sources of uncertainty for each particular study, region or season (Beven, 2016; Hawkins, Osborne, Ho, & Challinor, 2013; Vetter et al., 2017). For example, additional uncertainties have been

observed from the choice of climate model's domain size and member in seasonal simulations of precipitation extremes (Roy, Gachon, & Laprise, 2014). In RCM simulations, the use of spectral nudging, a technique applied to ensure that the RCM remains close to its driver has also been identified as a source of uncertainty (Lucas-Picher, Cattiaux, Bougie, & Laprise, 2015; Mearns et al., 2018; Sanchez-Gomez, Somot, & Déqué, 2009; Separovic, de Elía, & Laprise, 2012). Another example is the increasing spatial and temporal resolutions of climate simulations. Higher-resolution climate simulations have shown an "added value" or gain over some regions (Curry et al., 2016; Lucas-Picher, Laprise, & Winger, 2016). Consequently, these unsystematic gains add uncertainty to the climate model outputs (de Castro, Gallardo, Jylha, & Tuomenvirta, 2007; Feng et al., 2014; Sandvik, Sorteberg, & Rasmussen, 2018; Torma, Giorgi, & Coppola, 2015). Additionally, as previously described in section 1.3.4, the observed non-stationarity on floods adds uncertainty to the flood frequency models that are traditionally used to analyse climate change impact studies.

1.5   Dealing with modelling uncertainties

It is clear from the literature that studies evaluating climate change impacts on floods suffer from numerous uncertainties due to the gaps and lack of understanding of the regional complexities in the diverse hydroclimatic regimes. To deal with these modelling uncertainties, the use of *multi-model ensembles* (MMEs) has become the benchmark approach for climate change impact studies as it allows reducing the influence of a particular simulation by providing a spread of simulations (IPCC, 2021). This MME approach often relies on the models "*democracy*", meaning that equal weights are assigned to all models included in the ensemble (Knutti et al., 2017; Wang et al., 2019), particularly to deal with climate and hydrological models' uncertainty. Thus, as schematically represented in Figure 1.3, the data generated in climate change impact studies consists of envelopes of simulations where different metrics as ensemble mean or median and spread are often used to analyse the MME results.

Other studies have suggested different strategies to attempt at reducing uncertainties on flood and hydrological projections (Beven, 2016; Kundzewicz et al., 2018). Essentially, it has been proposed that adding more data from which their associated uncertainties can be precisely quantified can help reducing the global uncertainty. In other words, additional information combined with a robust assessment of their uncertainties are expected to reduce the overall uncertainty by quantifying the agreements between different results. As proposed by Kundzewicz et al. (2018) the opportunities to reduce uncertainties can be related to: (1) data and information, (2) climate models and (3) hydrological models.
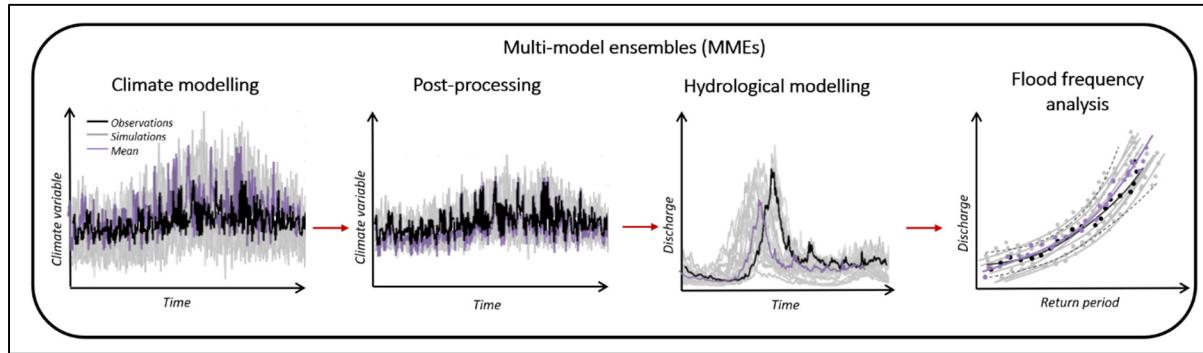
Figure 1.3 Schematic representation of the common multi-model ensembles approach over the different stages of flooding projections

Regarding the data and information, uncertainties may be reducible by finding new and more exact information from more recent and improved measuring devices. New sources of information as global datasets (e.g., Barbarossa et al., 2018), data obtained from remote sensing, or complementary data as proxies (e.g., tree rings) may enable intercomparisons to reduce uncertainties in the data (Vetter et al., 2017).

In climate modelling, improving our understanding of complex climate processes and improving the current GCMs and RCMs spatial and temporal resolutions might lead to reducing uncertainties. At the same time, improvements on the current techniques related to the scaling (i.e., statistical and dynamical downscaling) between climate and hydrological models are expected to reduce the uncertainties related to their applications. Using weighting approaches have also been proposed to reduce uncertainty (e.g., Knutti et al., 2017; Sanderson,

Wehner, & Knutti, 2017; Xu, Gao, & Giorgi, 2010). The idea behind these approaches is to favor climate model outputs with better performances instead of the common model "*democracy*" approach (Knutti, 2010). Climate-based weighting approaches have however shown different problems when used for hydrological impact studies, mainly due to the non-linear relationship between climate variables and streamflow (Chen, Brissette, Lucas-Picher, & Caya, 2017). Streamflow-based weights have consequently been suggested as an alternative (Wang et al., 2019). However, small improvements were observed when using this streamflow-based weighting approaches, suggesting to continue using the climate model democracy approach (Chen et al., 2017; Das, Treesa, & Umamahesh, 2018; Wang et al., 2019).

Hydrological modelling improvements may also lead to reducing uncertainty. Among the different efforts targeting this issue, two main approaches have been suggested: (1) establishing general guidelines for a robust selection, calibration and validation of hydrological models and (2) using multi-model frameworks (Kundzewicz et al., 2018). Regarding the latter, the multi-hydrological model ensemble "democracy" or the selection of the best-performing hydrological models are the two main approaches used (Krysanova et al., 2018). The use of different hydrological models with equal weights is a relatively common approach (Chen, Brissette, Poulin, & Leconte, 2011; Poulin et al., 2011; Roudier et al., 2016). However, some issues have been observed in this approach. For instance, removing one or two outlier hydrological models has caused significant shifts on the resulting ensemble and, in some cases, single hydrological models have shown better performance than the ensemble means (Gudmundsson, Wagener, Tallaksen, & Engeland, 2012; Zaherpour et al., 2018). Thus, users and practitioners usually favor hydrological models with good performance against observations. As suggested for climate models, the use of weighting approaches has also been suggested. For example, studies performed by Wilby et Harris (2006) and Yang et al. (2014) used a simple performance-based hydrological model weighting approach. However, few studies have used hydrological models weighting techniques for climate change impact studies. More robust weighting techniques have been tested in other applications, such as streamflow simulations over ungauged catchments (Arsenault, Gatien, Renaud, Brissette, & Martel, 2015). Presently, however, the use of different weighting techniques has not been thoroughly assessed

for climate change impact studies, especially for flood projections. The main drawback of weighting approaches is the relatively more complex and time-demanding process of the evaluation process. Additionally, it has been underlined that good performance under historical conditions does not guarantee good performance in the future, especially if different climate conditions are expected in the future under climate change (Merz, Parajka, & Blöschl, 2011). Therefore, regarding the stricter calibration and validation guidelines, it has been suggested that strategies that consider different climatic and regional conditions, such as the Differential Split-Sample Testing (DSST) might provide a more appropriate evaluation of hydrological models for climate change impact studies (Bérubé, Brissette, & Arsenault, 2022; KlemeŠ, 1986; Krysanova et al., 2020). Moreover, the use and comparison of multiple hydrological models (if possible) is always recommended (Krysanova et al., 2018).

It is observed that numerous uncertainties have been identified and highlighted in the evaluation of climate change impacts on floods. Yet, the different uncertainties associated with the hydroclimatic modelling chain have shown different impacts depending on variables such as region, catchment properties, climate type and climate/hydrological models used. For example, the increasing resolution in climate model outputs has shown improvements over GCMs in regions with complex topography, but not over flatter areas (Casanueva et al., 2016; Falco, Carril, Menéndez, Zaninelli, & Li, 2018; Soares & Cardoso, 2018). On a similar line, the different efforts that aimed at reducing the uncertainties in floods projections, are often limited by the study areas. For instance, the studies evaluating multi-model weighting approaches have only been tested over few river basins (Chen et al., 2017; Wang et al., 2019). The particular hydrological and climatic conditions of the few case-studies makes it difficult to extrapolate the obtained results and conclusions over regions with different conditions. These different issues raise the need to further evaluate the modelling uncertainties over regions with a diversity of hydrological and climate conditions to better understand the regional differences and potentially increase our confidence in regional floods projections.

1.6   Research objectives

The main objective of this research is to improve the understanding of the modelling uncertainties associated with the generation of flood projections, and to analyse different strategies to reduce these uncertainties and increase our confidence in the simulated impacts of climate change on floods. In order to address the main objective, two main research lines are followed:

1. Quantify the uncertainty contributions of different elements involved in the modelling chain used to generate flooding projections, over two regions covering a diversity of hydroclimatic conditions.

2. Evaluate the effects of different strategies to reduce uncertainties associated with hydrological and climate models on flooding projections, over the same two regions.

CHAPTER 2

METHODOLOGY AND THESIS ORGANIZATION

2.1   Experimental set-up

To investigate each of the research lines defined to address the main objective of improving the current understanding on the modelling uncertainties associated with flood projections, the first stage consisted in defining the study area and the models and methods included in the modelling chain to ensure a diversity of hydroclimatic regimes and models.

The study area covers 96 basins located in two main North American regions, (1) the Canadian province of Quebec and (2) the country of Mexico (see Figure 2.1). These two regions were selected as part of a long-term international collaboration between Quebec and Mexico where this doctoral thesis contributes through the project « *Complémentarité et diversité Québec-Mexique en matière de gestion de l'eau dans le contexte des changements climatiques: impacts sur les régimes hydriques et prise de décision en adaptation* » that aims to produce new knowledge associated with the impacts of climate change on hydrological extremes.

As displayed in Figure 2.1, each region shows contrasting mean annual temperatures and mean annual total precipitation amounts, which allows covering a wide range of both climatic and hydrological conditions. Moreover, the study basins are located over 16 different Köppen-Geiger climate types (Kottek, Grieser, Beck, Rudolf, & Rubel, 2006) that vary from subarctic and continental humid climates to arid and warm temperate regimes. Including numerous and diverse basins was one of the most important characteristics of this research project. The idea behind this is that large-scale studies allow drawing conclusions that can be relevant and used over more cases sharing similar conditions. In addition, as observed in the literature review, large-scale studies are scarce. Thus, carrying out a large-scale study can be a great contribution for the scientific community.
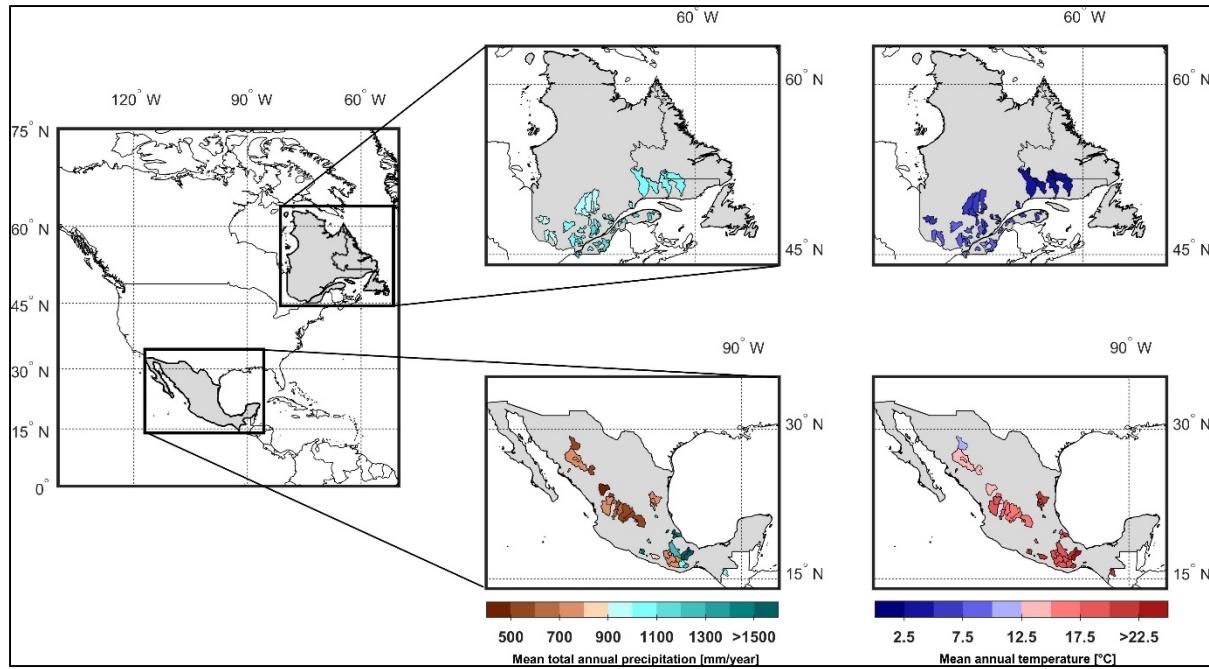
Figure 2.1 Study area of this research project. The mean annual climatology of the 50 basins in Quebec and the 46 basins in Mexico is presented in the top and bottom panels, respectively. From left to right, the mean total annual precipitation (in mm) and mean annual temperature (in °C) are presented

After defining the study area, the modelling chain used for the different lines of research was defined (see Figure 2.2). To consider the climate modelling uncertainty, a total 24 GCM simulations issued from the CMIP5 database and six RCM simulations produced by the Ouranos Consortium on Regional Climatology and Adaptation with the Canadian RCM version 5 (CRCM5) were used. These six RCM simulations consist of three CRCM5 simulations covering the Quebec domain and three covering Mexico. The three CRCM5 simulations of each region share the same spatial resolution (0.22°) and driving GCMs/ESMs (CanESM2, CNRM-CM5 and GFDL-ESM2M). All global and regional climate simulations cover the reference and future periods from 1976-2099 under the RCP 8.5 scenario. Further details on the global and regional climate model simulations are presented over the following chapters.
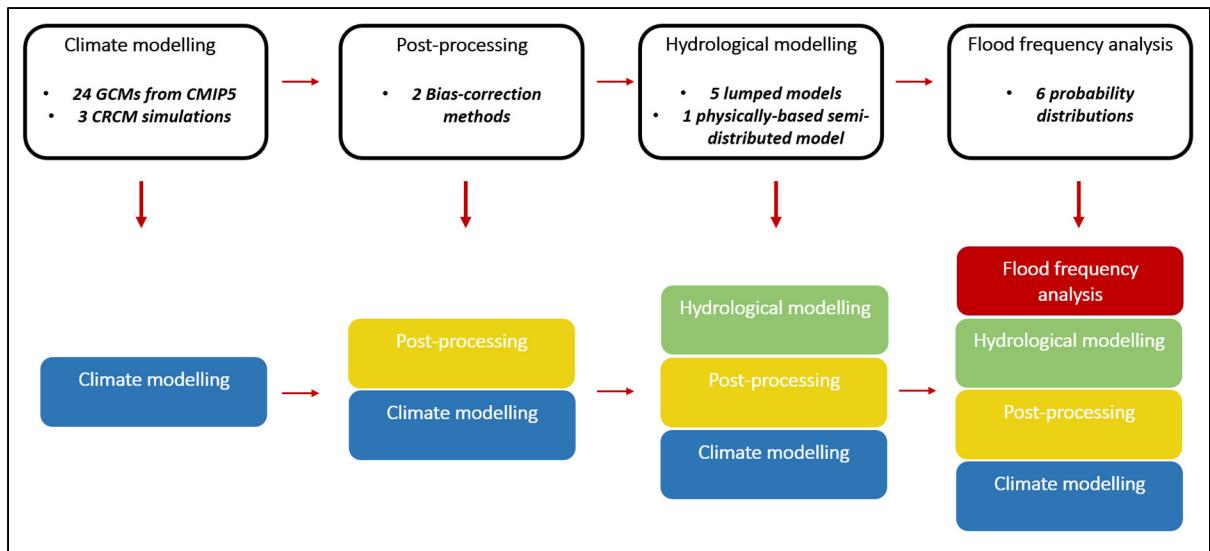
Figure 2.2 General modelling chain designed to address the different research lines. The different color blocks show the uncertainties that "accumulate" along the chain

Two post-processing methods, namely the quantile-mapping (QM; Maraun, 2016; Themeßl, Gobiet, & Heinrich, 2012; Themeßl et al., 2011) and daily bias correction (DBC; Chen, Brissette, Chaumont, & Braun, 2013b) methods were selected. These two methods have been successfully used over different regions of the world and have become a traditional process when evaluating climate change impacts in hydrology (Chen et al., 2019; Zhao, Brissette, Chen, & Martel, 2020).

Hydrological models have been identified among the most important uncertainty sources in hydrology (Poulin et al., 2011; Velázquez et al., 2013), making it important to include a diversity of models to evaluate strategies that can impact their uncertainty contributions. Thus, five lumped conceptual hydrological models and one semi-distributed and more physically-based hydrological model were used in different sections of the project. The selected lumped conceptual hydrological models consist of the GR4J (Perrin, Michel, & Andréassian, 2003), MOHYSE (Fortin & Turcotte, 2006), HMETS (Martel, Demeester, Brissette, Poulin, & Arsenault, 2017), CEQUEAU (Girard, Morin, & Charbonneau, 1972) and IHACRES (Jakeman, Littlewood, & Whitehead, 1990) models. The semi-distributed and more physically-based hydrological model HYDROTEL was selected as it has been used in different

applications showing good performance over both regions (Ibarra-Zavaleta, Landgrave, Romero-López, Poulin, & Arango-Miranda, 2017; Lucas-Picher et al., 2020).

Finally, as described in Chapter 1, the uncertainty associated with the selection of the probability distribution for flood frequency analyses is generally overlooked, as the best fitting probability distribution has been traditionally used. Thus, to consider the uncertainty arising from this modelling step, six different probability distributions were used to estimate the flooding return periods. The probability distributions included in this study are the Log-normal distribution with 2 and 3 parameters, the Gumbel distribution, the generalized Extreme Value (GEV), the Pearson type III, and the Log-Pearson type III. All these probability distributions have been commonly used and recommended in regional guidelines (Castellarin et al., 2012; Cunnane, 1989; Escalante-Sandoval & García-Espinoza, 2014).

As described in the main objective (see section 1.6), the research project was divided into two main research lines to focus on, (1) quantifying the uncertainty contributions of different modelling elements used to generate flooding projections, and (2) evaluating the effects of different strategies to reduce uncertainties associated with hydrological and climate modelling in flooding projections. To address these two research lines, the overall methodology is divided in four main stages schematically represented (and identified in red) in Figure 2.3.

The initial stage consisted in decomposing and quantifying the uncertainty contributions of modelling elements included in the modelling chain (see Figure 2.2) to address the first research line. The second research line that aims to evaluate the effects of different strategies to reduce modelling uncertainties is divided into two sections, (1) one targeting the strategies related to hydrological models, and (2) another one focusing on strategies related to climate models. From the first section, two main topics were identified. First, this project aimed at analysing the robustness of hydrological models' parameterizations under a changing climate. As described in Chapter 1, this is a topic of ongoing discussion that needs further investigations (Krysanova et al., 2018; Merz et al., 2011). Thus, in this project different tests were designed to identify the sensitivity of different hydrological models to calibration-validation strategies

under contrasted climate conditions. The second section consisted in evaluating the potential impacts of weighting hydrological models on flood projections. As described in the literature review, the effects of weighting hydrological models in climate change impact studies remain to be explored, particularly for flood projections (Wilby & Harris, 2006; Yang et al., 2014).
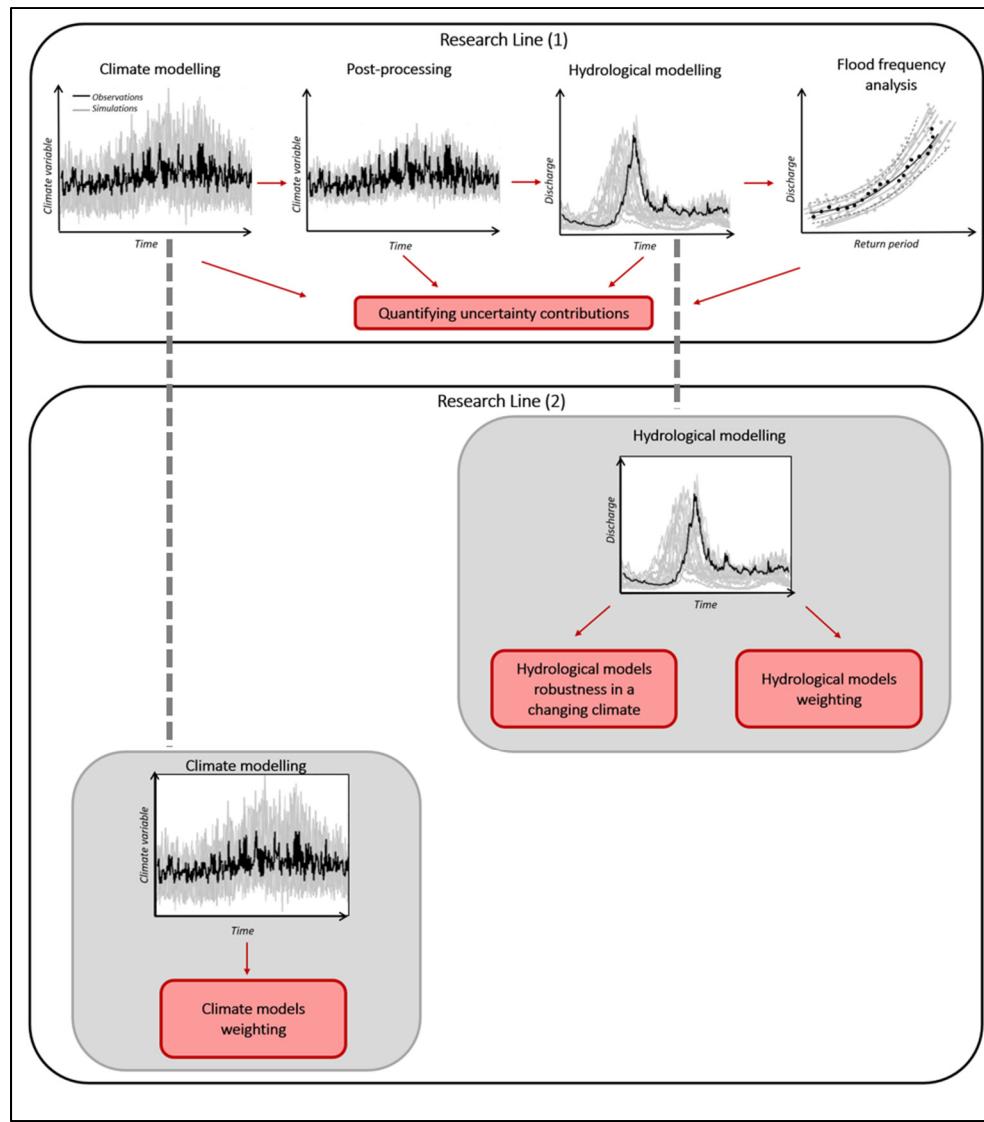
Figure 2.3 Schematic representation of the four research stages
(highlighted in red) designed for this research project

Finally, the last section focuses in assessing the effects of weighting climate models on flood projections. Different studies have shown reduced uncertainty spreads of climate simulations

when weighting strategies are applied (Knutti, 2010; Knutti et al., 2017). However, most studies have focused on uncertainty spread of climate variables, such as precipitation and temperature. Thus, this study aims at filling this research gap by investigating the potential implications of weighting climate simulations on flood projections, considering (1) the development of a new weighting approach based on seasonally simulated streamflows, and (2) comparing this approach with other existing approaches, including equal weighting or "model democracy".

2.2   Thesis organization

This thesis is composed of seven chapters. As observed in Chapter 1, a review describing the relevant literature in flood projections was presented. This Chapter 2 describes the general methodological design and organization of the research project. The following Chapters 3 to 6 present the published or submitted articles in scientific journals that served to address the main objective. Finally, a general discussion, as well as an overall conclusion and recommendations are presented.

Chapter 3 presents a first article entitled « *Uncertainty sources in flood projections over contrasting hydrometeorological regimes* » This article submitted to the *Hydrological Sciences Journal* aimed at decomposing the uncertainty contributions of different elements of the modelling chain over diverse hydrometeorological regions. The results of this study served at addressing the first axis of this research project by linking the hydrometeorological properties of the river basins with the uncertainty contributions of the different sources.

Chapter 4 presents the second article entitled « *Stress-testing hydrological models for climate change impact studies* » and submitted to the *Hydrological Sciences Journal.* This article presents the first analysis addressing the second research axis. The objective of this article was evaluating the robustness of hydrological models' parameterizations in a changing climate. Different calibration-validation strategies were defined and tested according to contrasted climate conditions: cold-warm, warm-cold, dry-humid, and humid-dry. The results contributed

to better understanding the calibration-validation elements that might bring more stress to hydrological models when used to simulate contrasting climate conditions, as well as assessing their implication for streamflow (and flood) projections.

Chapter 5 presents the third article entitled « *Hydrological models weighting for flood projections* ». This article, submitted to the *Journal of Hydrology*, aimed at evaluating the effects of weighting hydrological models based on their robustness under a changing climate (from Chapter 4's results). The results of this study allowed quantifying the potential effects of weighting hydrological models on flood projections. Thus, contributing to the second axis of this project.

Chapter 6 presents the fourth and last article entitled « *Weighting climate models for flood projections: effects on contrasting hydroclimatic regions* » submitted to *Climatic Change Journal*. This article addresses the second research axis by assessing the effects of weighting climate models on flood projections. This article analyses existing methods and also proposes a new strategy to weight climate models based on their regional adequacy. The results of this study not only showed the impacts of weighting climate models on the projected peak streamflows, but also contributed to opening the discussion on current practices to deal with climate modelling uncertainties in hydrological studies.

In addition to the articles produced during this project, different oral presentations, workshops, and collaboration articles were carried out in parallel. The ensemble of scientific contributions that derived from this project are presented in the first appendix.

CHAPTER 3

UNCERTAINTY SOURCES IN FLOOD PROJECTIONS OVER CONTRASTING HYDROMETEOROLOGICAL REGIMES

Mariana Castaneda-Gonzalez[a], Annie Poulin[b], Rabindranarth Romero-Lopez[c], Richard Turcotte[d] and Diane Chaumont[e]

[a,b] Department of Construction Engineering, École de technologie supérieure, Montréal, Canada.

[c] Faculty of Civil Engineering, Universidad Veracruzana, Xalapa, Mexico.

[d] Ministère du développement durable, Environnement et Lutte contre les changements climatiques, Québec, Canada.

[e] Ouranos Consortium on Regional Climatology and Adaptation to Climate Change, Montréal, Quebec, Canada

Abstract

This study evaluates the uncertainty of four components of the hydroclimatic modelling chain on flood projections over 96 basins covering contrasting hydrometeorological regimes located in Canada and Mexico. Two ensembles of climate simulations are considered, a large twenty-two global climate model simulations ensemble and a smaller three high-resolution regional climate model simulations ensemble. The other components consist of two post-processing techniques, three lumped hydrological models and six probability distributions. These four sources are assessed through a variance decomposition method on six flood indicators over a reference and two future periods 1976-2005, 2041-2070 and 2070-2099. Systematic differences are observed between basins with contrasting flood-generating processes. Snow-dominated basins constantly show larger variance contributions from hydrological models, while rain-dominated basins show climate simulations as their dominant source. These results

underline the need to consider the variability of each component's uncertainty contribution and its link to hydroclimatic conditions and dominant processes.

Keywords: Floods; uncertainty; hydrological modelling; climate modelling; climate change

3.1   Introduction

The observed changes in the climate system have impacted flooding around the world, and it is expected that a changing climate will continue to impact their occurrence in the future (IPCC, 2018; Vormoor et al., 2016). Hydrological projections have thus become a common approach to assess the potential impacts of climate change on floods, and to prepare adaptation and mitigation strategies for the future (Krysanova et al., 2018; Kundzewicz et al., 2017). These projections are often generated at the basin scale by coupling climate models' outputs and hydrological models under different emission scenarios (Minville, Brissette, Krau, & Leconte, 2009). These projections are often generated at the basin scale by coupling climate models' outputs and hydrological models under different emission scenarios. This process usually comprises multiple stages that create a hydroclimatic modelling chain that varies depending on the methodological choices of each study (Chen, Brissette, & Leconte, 2011; Her et al., 2019). Frequently, the hydroclimatic modelling chain for flooding projections comprises one or multiple: (1) emission scenarios, (2) Global Climate Models (GCMs) simulations, (3) statistical or dynamical downscaling (i.e., Statistical Downscaling Models [SDMs] and Regional Climate Models [RCMs]) approaches, (4) bias-adjustment methods, (5) hydrological models (HMs) and (6) statistical models for flood frequency analysis, among others. Numerous studies applying different variations of such modelling chain can be found in the literature (Bosshard et al., 2013; Chan et al., 2020; Hirabayashi, Kanae, Emori, Oki, & Kimoto, 2008; Lucas-Picher et al., 2020; Reszler, Switanek, & Truhetz, 2018; Riboust & Brissette, 2015; Taye, Ntegeka, Ogiramoi, & Willems, 2011; Teutschbein, Grabs, Karlsen, Laudon, & Bishop, 2015; Thompson, Laizé, Green, Acreman, & Kingston, 2014). However, the plethora of existing models and methods with their own structures and limitations, add uncertainty at each step of the hydroclimatic modelling chain (Beven, 2016; Kundzewicz et al., 2018; Meresa &

Romanowicz, 2017). Thus, confidence in flooding projections remains low due to the "cascade of uncertainty" arising from their simulations (AghaKouchak et al., 2012; Hall et al., 2014; Henderson & Sellers, 1993; Kundzewicz et al., 2017; Mareuil, Leconte, Brissette, & Minville, 2007; Schneider, 1983; Wilby & Dessai, 2010).

A wide range of studies has assessed the modelling uncertainty on hydrological and flood projections over basins with different hydrometeorological regimes (e.g., rain- or snow-dominated) around the world (Bosshard et al., 2013; Collet, Beevers, & Prudhomme, 2017; Falloon et al., 2014; Ho, Thompson, & Brierley, 2016; Kay, Davies, Bell, & Jones, 2009; Teng, Vaze, Chiew, Wang, & Perraud, 2012; Vetter et al., 2015). In general, these studies evaluate the uncertainty related to each source of the study's modelling chain by using ensembles that allow the analysis of their influence on the resulting hydrological projections. Table 3.1 summarizes and briefly describes the main conclusions of some relevant uncertainty analyses on hydrological simulations. This comparison reveals that the cited studies share some similarities and differences between them. For instance, it can be observed that various studies (e.g., Bosshard et al., 2013; Giuntoli, Villarini, Prudhomme, & Hannah, 2018; Kay et al., 2009; Teng et al., 2012; Vetter et al., 2017) have shown that climate simulations (i.e., GCMs and RCMs) are one of their main uncertainty sources, particularly in mean- and peak-flows simulations. This is line with similar studies showing climate simulations as the leading source of uncertainty (Chan et al., 2020; Gosling, Taylor, Arnell, & Todd, 2011; Graham, Andréasson, & Carlsson, 2007; Prudhomme & Davies, 2009; Wilby & Harris, 2006). Yet, more recent studies have shown that natural variability can play an important role in climate simulations uncertainty (Hingray, Blanchet, Evin, & Vidal, 2019; Vidal, Hingray, Magand, Sauquet, & Ducharne, 2016), an uncertainty source that it is not always considered in the modelling chains. Other studies have given particular focus to hydrological modelling uncertainties (e.g., Bosshard et al., 2013; Chegwidden et al., 2019; Giuntoli et al., 2018; Lemaitre-Basset et al., 2021) showing that hydrological models' structures can be relevant contributors of uncertainty (Hattermann et al., 2018; Poulin et al., 2011; Velázquez et al., 2013; Vetter et al., 2015), but that it depends on the hydrological variable under consideration, on the time period of the year and on the hydrological regime of the catchment considered (e.g., low-flows).

Table 3.1 Summarized description of uncertainty assessments in hydrology

Target	Uncertainty sources	Main uncertainty contributors	Limitations	Reference
Peak flows	• 4 Emission scenarios • 5 GCMs • 2 Downscaling methods • 2 HMs structures and parameterizations Natural variability	GCMs structure were found as the leading uncertainty source.	Only two rain-dominated basins considered.	Kay et al. (2009)
Mean and peak flows	• 15 GCMs 5 lumped HMs	GCMs uncertainty was larger than the conceptual HMs	Only rain-dominated basins considered.	Teng et al. (2012)
Mean and peak flows	• 3 GCMs • 4 RCMs • 2 Post-processing methods 2 distributed HMs	GCMs & RCMs on summer-fall mean and high runoff HMs and post-processing on low winter-spring runoff	Only one snow-dominated basin included	Bosshard et al. (2013)
Mean and peak streamflows	• 4 Emission scenarios • 5 RCMs • 2 Post-processing methods 2 distributed HMs	GCMs over most cases Both GCMs and HMs on some of the basins	Only large-scale basins considered	Vetter et al. (2015)
Low flows	• 4 GCMs • 3 SDMs • 6 HMs (lumped & distributed) Natural variability	· Natural variability HMs also show impacts on low-flows	Only two snow-dominated basins included	Vidal et al. (2016)
Mean, low and peak flows	• 4 Emission scenarios • 5 GCMs • 9 Global impact (hydrological) models (GIM) Natural variability	GCMs over most cases followed by GIMs	No basin-scale HMs included	Giuntoli et al. (2018)
Mean low and peak flows	• 2 Emission scenarios • 10 GCMs • 2 Downscaling methods • 4 HMs Natural variability	· Natural variability, emission scenarios and GCMs over mean runoff. HMs on low runoff.	Only one snow-dominated basin included	Chegwidden et al. (2019)
Peak and low flow	• 2 Emission scenarios • 5 RCM runs • 3 HMs (lumped &distributed) 29 HMs parameterizations	- Emission scenarios on high-flows HMs, HMs parameters and GCMs on low-flows	Only one rain-dominated basin included	Lemaitre-Basset et al. (2021)
Flood return levels	• 3 Emission scenarios • 1 distributed HMs parameterization 3 Extreme value models	· Hydrological models' parameterization, in small return periods. All sources were important on large return periods	Only one rain-dominated basin included	Zhang, Xu, et Fu (2014)
Flood return levels	6 distribution (GEV) parameterization on 2 e mission scenarios and 6 GCMs	Distributions parameterization uncertainty increased with increasing return periods	Only one rain-dominated basin included	Das et Umamahesh (2018)

It has thus often been recommended to consider hydrological model ensembles and more thorough model evaluations (Velázquez et al., 2013; Vetter et al., 2015). The uncertainty

arising from the estimation of return levels on flood frequency analyses has also been investigated in different studies (Blöschl & Montanari, 2010; Hailegeorgis & Alfredsen, 2017; Merz & Thieken, 2005; Merz, Vorogushyn, Uhlemann, Delgado, & Hundecha, 2012) showing that the uncertainty associated to historical records (e.g., presence of non-stationary trends), the selection of the "best" probability distribution to model extreme values, and the statistical models' parameterizations can be major sources of uncertainty on return levels estimation. Few studies have however included these uncertainty sources in climate change impact studies (Collet et al., 2017; Das & Umamahesh, 2018; Meresa & Romanowicz, 2017; Zhang et al., 2014). Among them, it has been highlighted, that uncertainty contributions from flood return levels statistical models and their parameterizations increased with increasing return periods, underlining the potential impact of ignoring the uncertainty associated with flood frequencies estimation (Collet et al., 2017; Meresa & Romanowicz, 2017; Merz & Thieken, 2005, 2009).

All uncertainty studies cited above have provided relevant information and insights about the leading uncertainty sources in regional streamflow projections. However, important differences and limitations are observed. For instance, each study has constructed their own hydroclimatic modelling chain over the selected basin(s) of study. The different types and number of models and methods among studies, in addition to the different climates and flood-generating processes of each region, makes it difficult to compare and extrapolate their conclusions. Regarding flood frequency analyses, studies have often relied on a single 'best' probability distribution, and fewer studies have included uncertainty related to return levels estimations and their interplay with other sources (e.g., GCMs and HMs). The regional differences, methodological choices, and limitations of each study make results hardly comparable for different contexts. Therefore, a study that not only includes various basins but also covers contrasting hydrometeorological regimes under a common methodology is still missing. Thus, this study aims to analyze the uncertainty arising from four different sources: (1) climate models, (2) post-processing techniques, (3) hydrological models and (4) probability distributions on flood simulations over 96 basins with contrasting hydroclimatic regimes.

3.2 Study regions and data

3.2.1 Study region

The study area consists of 96 basins located in two North American regions with contrasting climates and flood-generating processes. Figure 3.1 shows the location and mean total annual precipitation of each basin over both regions. The upper right panel in Figure 3.1 shows 50 snow-dominated basins located in the south of the Canadian province of Quebec. According to the Köppen-Geiger climate classification (Kottek et al., 2006), these 50 basins are in subarctic and continental humid climates (i.e., Dfc & Dfb), snowy regions characterized by cold winters and cool summers. These 50 basins have mean annual temperatures ranging from -2.6 to 11.5 °C and mean annual total precipitation varying from 900 to 1250 mm. In contrast, the lower right panel in Figure 3.1 shows the other 46 basins located in Mexico. These basins are located over 14 different Köppen-Geiger climate types (Kottek et al., 2006) including equatorial, arid, and warm temperate climates (e.g., Af, BSh & Cf). These 46 basins have mean annual temperatures ranging from 11.7 to 23.6 °C and mean annual total precipitation ranging from 475 up to 3110 mm. Regarding physical properties, the 96 catchments areas vary from 225 up to 29,584 km^2, covering a total of 419,039 km^2, with topography ranging from coastal plains and flat areas (over both regions) to mountainous regions reaching up to 2971 m.a.s.l. (in Mexico).

3.2.2 Observed data

Daily historical records of minimum temperature, maximum temperature, precipitation and streamflow were used for each basin. These meteorological and hydrometric records were obtained from the Hydrometeorological Sandbox –École de technologie supérieure (HYSETS) database (Arsenault, Brissette, Martel, et al., 2020), a large-scale hydrometeorological database of North American basins. Meteorological data consists of a grid of 1/16° (~6 km) resolution covering Mexico, the conterminous U.S. (CONUS) and southern Canada (Livneh et al., 2015; Livneh et al., 2013). The meteorological data points inside each basin's area were

averaged to obtain one time series per basin. Based on hydrometric data availability for each basin, the meteorological and hydrometric data consist of time series with a minimum length of 30 years between 1950 and 2013.
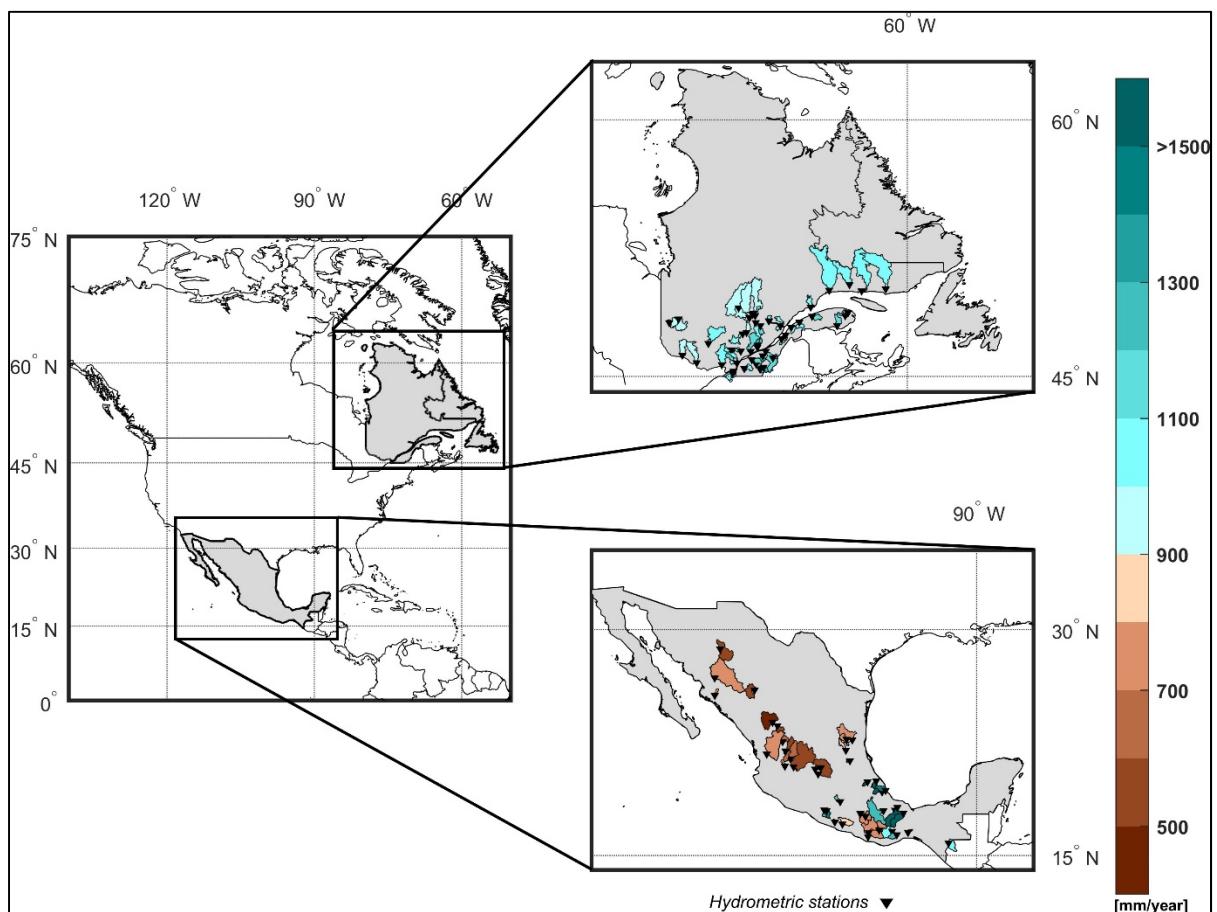
Figure 3.1 Location and mean annual total precipitation of the 50 basins in Quebec (upper right panel) and the 46 basins in Mexico (lower right panel) used in the study

3.2.3 Climate simulations

The climate model outputs used in this study consisted of daily precipitation, minimum and maximum temperatures, from two ensembles of global and regional climate model simulations. The first ensemble comprises twenty-two global climate simulations issued from the Coupled Model Intercomparison Project Phase 5 (CMIP5) database (Taylor et al., 2011) described in

Table 3.2 These simulations were selected to represent the CMIP5-ensemble diversity by including fourteen different modelling centers, five of the more complex earth system models (ESMs), and various spatial resolutions (0.75°-3.75°). The second ensemble comprises three RCM simulations issued from the Canadian Regional Climate Model version 5 (CRCM5; Separovic et al., 2013) driven by three GCMs (i.e., CanESM2, CNRM-CM5 and GFDL-ESM2M) at 0.22° spatial resolutions. The CRCM5 simulations over both regions were produced and provided by the Ouranos Consortium on Regional Climatology and Adaptation. Over Mexico, the Ouranos Consortium produced new CRCM5 simulations over a modified Central American domain (CAM) in order to extend the northern border to completely cover the country of Mexico. Simulations over both regions are part of the Coordinated Regional Downscaling Experiment (CORDEX) database (Giorgi & Gutowski, 2015). Both GCM- and CRCM5-ensembles cover the reference period of 1976-2005 and two future horizons of 2041-2070 and 2070-2099 under the Representative Concentration Pathway (RCP) 8.5. The use of these two ensembles enables an evaluation of the impacts of the choice of climate simulations ensemble (smaller-limited RCM-ensemble vs. larger GCM-ensemble) on the uncertainty contributions on flood simulations.

3.3 Methodology

3.3.1 Hydroclimatic chain overview

This study was carried out following five main stages (see Figure 2). The first stage consisted of extracting the GCM and CRCM5 outputs for each basin and period. The second stage consisted of post-processing all climate simulations using two bias-adjustment techniques. The third stage comprised the calibration and validation of the hydrological models to then couple them with all climate simulations to obtain the GCM- and CRCM5-driven streamflow simulations. The fourth stage entailed performing the flood frequency analysis on each simulated streamflow series with six probability distributions to obtain six flood indicators from each analysis. The fifth stage consisted of the analysis of mean seasonal climate and streamflow simulations, and the uncertainty analysis on six flood indicators for each basin.

Table 3.2 Description of the GCMs used in this study and the CRCM5-drivers (grey)

GCM name	Modeling group	Spatial resolution (°)
ACCESS 1.0	CSIRO & BOM	1.875 x 1.25
ACCESS 1.3	CSIRO-BOM	1.875 x 1.25
BCC-CSM1.1	BCC-CMA	1.125 x 1.125
BCC-CSM1.3	BCC-CMA	2.8 x 2.8
CanESM2	CCCMA	2.8 x 2.8
CMCC-CM	CMCC	0.7484 x 0.75
CMCC-CMS	CMCC	1.875 x 1.875
CNRM-CM5	CNRM-CERFACS	1.4 x 1.4
CSIRO-Mk3.6	CSIRO-QCCCE	1.8 x 1.8
FGOALS-g2	LASG-IAP-CAS-CESS	1.875 x 1.25
GFDL-CM3	GFDL	2.5 x 2.0
GFDL-ESM2G	GFDL	2.5 x 2.0
GFDL-ESM2M	GFDL	2.5 x 2.0
GISS-E2-H	NASA-GISS	2.5 x 2.0
GISS-E2-R	NASA-GISS	2.5 x 2.0
INM-CM4	INM	2.0 x 1.5
IPSL-CM5A-LR	IPSL	3.75 x 1.8
IPSL-CM5B-LR	IPSL	3.75 x 1.8
MIROC5	MIROC	1.4 x 1.4
MPI-ESM-LR	MPI	1.8653 x 1.875
MPI-ESM-MR	MPI	1.8653 x 1.875
MRI-CGCM3	MRI	1.1 x 1.1

3.3.2 Climate outputs

The first stage consisted of extracting and preparing the GCM and CRCM5 outputs for each basin and period. Daily precipitation, minimum and maximum temperatures of all global and

regional climate simulations were extracted in each region for the reference period of 1976-2005 and two future periods of 2041-2070 and 2070-2099. For each basin, the grid points located within or surrounding their delimitation area were averaged using the Thiessen polygons method to obtain a single data series for each of the 25 climate simulations. A minimum of four grid points was averaged using this method for the smallest basins.
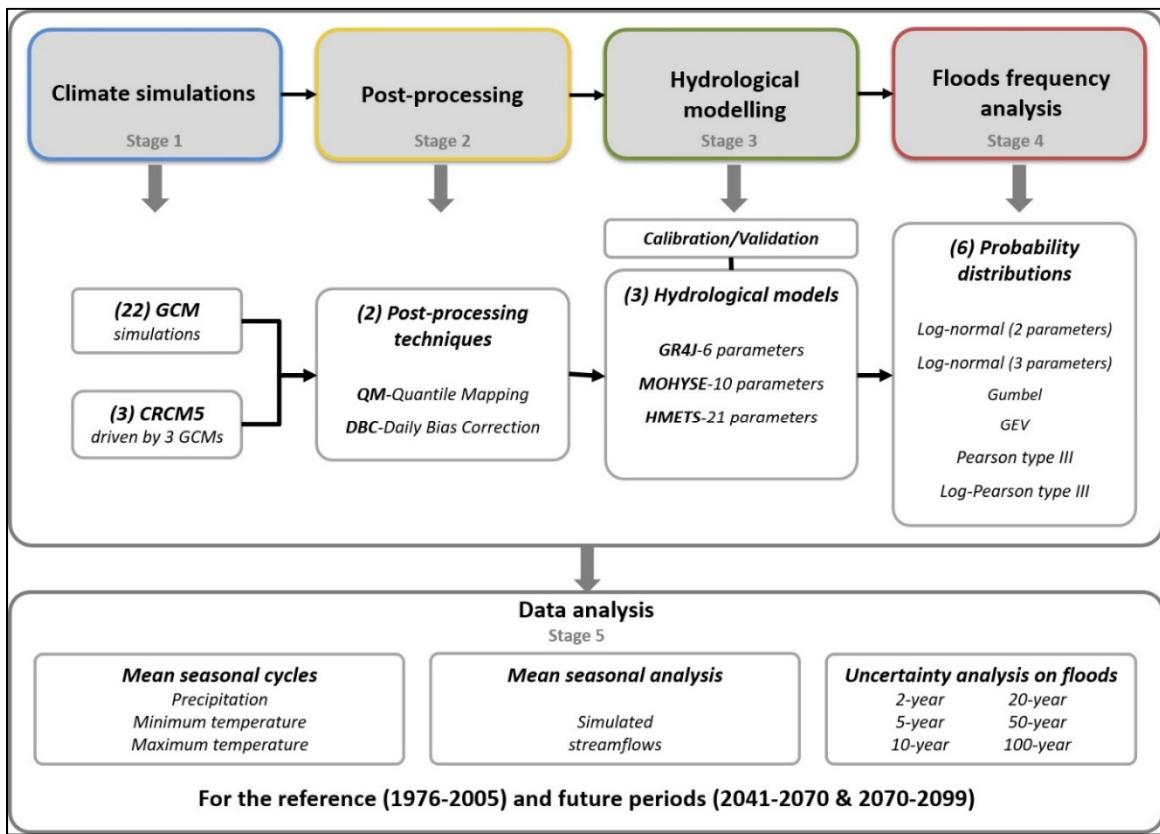
Figure 3.2 Overview of the hydroclimatic chain used for each basin

3.3.3 Post-processing methods

The second stage consisted of post-processing all climate simulations. Two post-processing techniques were included in the hydroclimatic modelling chain: (1) daily bias correction (DBC) and (2) quantile mapping (QM) methods.

The DBC is a post-processing method merging the daily translation method (DT; Mpelasoka & Chiew, 2009) and the local intensity scaling method (LOCI; Schmidli, Frei, & Vidale, 2006). The DBC starts using the LOCI method to fit simulated precipitation occurrences to the occurrences of observed precipitation data to correct the simulated wet-day frequencies. Then, the DT method is used to correct the frequency distributions of precipitation amounts and temperature simulations through a quantile-approach. The QM method (Maraun, 2016; Themeßl et al., 2012; Themeßl et al., 2011) is a simple and robust post-processing method based on the transformation of distributions by quantile. First, the correction of temperature simulations is done by fitting the quantiles of simulated data to the quantiles of observations using normal distributions. In the case of precipitation, the QM method corrects the precipitation distribution based on daily-constructed point-wise empirical cumulative distribution functions including both wet and dry days in the empirical cumulative distribution functions estimation (Chen et al., 2013b). In this way, the frequency and amounts of precipitation occurrence are simultaneously corrected. Further details on these two methods are presented in supplementary material S1. These two methods fall in the bias-adjusting approaches category, meaning that adjustment factors identified in reference periods are used to adjust the future simulations. Different studies have applied these methods for diverse applications including, comparisons of different bias-adjustment methods (e.g., Chen, Brissette, Chaumont, & Braun, 2013a; Chen et al., 2013b), and studies analyzing climate change impacts on hydrology (e.g., Chen et al., 2019; Zhao et al., 2020) endorsing their use in this study.

3.3.4   Hydrological modelling

The hydrological modelling of the third stage is carried out using three lumped conceptual hydrological models with varying structures and levels of complexity, (1) GR4J, (2) MOHYSE and, (3) HMETS, three hydrological models that have been evaluated in various regions validating their use (Arsenault, Breton-Dufour, Poulin, Dallaire, & Romero-Lopez, 2019; Arsenault, Brissette, Chen, Guo, & Dallaire, 2020).

The GR4J model is a parsimonious lumped and conceptual rainfall-runoff model developed by Perrin et al. (2003). In this hydrological model with four parameters, snow accumulation, snowmelt and evapotranspiration are not directly estimated. Thus, the snow module CemaNeige (Valéry, Andréassian, & Perrin, 2014) was added to allow for its application over the snow-dominated regions, adding two parameters to the model (6 parameters in total). To estimate the potential evapotranspiration the Oudin evaporation formulation (Oudin et al., 2005) was used. This formulation uses extraterrestrial radiation, determined from the date and latitude, and mean temperature to estimate potential evapotranspiration. The inputs required for applying the GR4J-CemaNeige-Oudin model consist of continuous series of daily precipitation, mean temperature and the previously calculated potential evapotranspiration. GR4J coupled with CemaNeige has been widely used for different applications over snow and rain dominated basins (e.g., Coron et al., 2012; Troin, Arsenault, Martel, & Brissette, 2018). The MOHYSE model is a simplified lumped conceptual model with ten parameters created by Fortin and Turcotte (2006). This model represents the main hydrological processes including potential evapotranspiration, snow melting and accumulation, vertical water budget and routing using ten different parameters to calibrate. The input data required for this hydrological model consists of mean daily temperatures, total daily rainfall and total daily snow accumulation. Some examples of recent MOHYSE applications are presented by Thiboult, Anctil, et Boucher (2016) and Arsenault, Brissette, Chen, et al. (2020). The HMETS model is a simple lumped conceptual model with 21 parameters developed by Martel et al. (2017). This hydrological model accounts for the surface flow, delayed flow, hypodermic flow and groundwater flow. The snowmelt is simulated by the melting and refreezing processes of the accumulated snowpack. For the potential evapotranspiration, the McGuinness and Bordne formulation was used (McGuinness & Bordne, 1972). The required inputs are continuous daily precipitation, maximum and minimum temperatures, and the calculated potential evapotranspiration. Recent applications cover different studies such as streamflow forecasting (Arsenault et al., 2019) and hydrological modelling (Troin et al., 2018).

The hydrological models' calibration was performed over the odd years of the period available for each of the 96 basins. Then, the validation was performed on the even (non-calibrated)

years. The idea behind this approach is to account for the climate variability of the full-time series by alternating the calibration/validation years. Based on recommendations from previous studies, the Covariance Matrix Adaptation Evolution Strategy (CMAES) was used to calibrate GR4J and MOHYSE (Arsenault et al., 2014; Hansen & Ostermeier, 1997). HMETS was calibrated using the Dynamically Dimensioned Search algorithm (Asadzadeh & Tolson, 2009; Huot, Poulin, Audet, & Alarie, 2019). The objective function used for the calibration/validation was based on the modified Kling-Gupta Efficiency criterion (KGE; Kling et al., 2012) over the odd/even years between observed and simulated streamflows, and is defined as follows:

$$KGE = 1 - \sqrt{(r-1)^2 + (\beta-1)^2 + (\gamma-1)^2} \qquad (3.1)$$

$$\beta = \frac{\mu_s}{\mu_o} \qquad (3.2)$$

$$\gamma = \frac{CV_s}{CV_o} = \frac{\sigma_s/\mu_s}{\sigma_o/\mu_o} \qquad (3.3)$$

where r, β and γ are the correlation coefficient, bias ratio and variability ratio, respectively. CV is the coefficient of variation, μ the mean, and σ is the standard deviation of the streamflow series. The "o" subscript indicates *observed* data and the "s" subscript indicates *simulated* data. KGE values range from -∞ to 1, where 1 indicates a perfect fit. According to Knoben et al. (2019), KGE values larger than ≈ -0.41 indicate that the simulation has higher skill than using the mean observations. KGE values above 0.5 were thus considered acceptable in this study. This metric was selected as it has shown to overcome issues related to mean squared error functions (e.g., Nash-Sutcliffe Efficiency), such as underestimating variability and prioritizing runoff peaks (Castaneda-Gonzalez et al., 2019; Gupta et al., 2009).

After calibrating and validating the three hydrological models, they were all coupled with the post-processed GCM- and CRCM5-ensembles. Thus, 150 streamflow simulations were

produced in total (i.e., 25 climate simulations × 2 post-processing methods × 3 hydrological models).

3.3.5 Flood frequency analysis

The fourth stage comprises the flood frequency analysis. This analysis was performed over annual and seasonal daily maximum discharges. Four seasons were considered, winter (December, January and February; DJF), spring (March, April and May; MAM), summer (June, July and August; JJA) and autumn (September, October and November; SON). Each GCM- and CRCM5-driven streamflow simulation was fitted to six different probability distributions presented in Table 3.3 Different methods were used to estimate each probability model's parameters (see Table 3.3). These six probability distributions are commonly used in hydrological applications and often recommended in regional guidelines (Castellarin et al., 2012; Cunnane, 1989; Escalante-Sandoval & García-Espinoza, 2014). For each probability distribution, six return period flows were estimated and used as flood indicators, the (1) 2-year, (2) 5-year, (3) 10-year, (4) 20-year, (5) 50-year and (6) 100-year return period flows. Performance of each statistical model was evaluated prior to their consideration in this study (results not shown).

Table 3.3 Description of probability distributions and their parameter estimation method used in this study

Probability distribution	Parameter estimation method
Log-normal with 2 parameters	Maximum likelihood
Log-normal with 3 parameters	Maximum likelihood
Gumbel	Maximum likelihood
Generalized Extreme Value (GEV)	Method of moments
Person type III	Method of moments
Log-Pearson type III	Method of moments

3.3.6   Data analysis

3.3.6.1   Climate simulations ensembles

Mean annual cycles of precipitation, minimum temperature and maximum temperature were analyzed to evaluate the uncertainty spread of all climate simulations used in this study. For each basin, the standard deviation of the 22 post-processed GCM outputs and 3 post-processed CRCM5 outputs (25 climate simulations in total) was estimated to quantify the seasonal spread of each climate variable. The seasonal analysis was performed over the winter (DJF), spring (MAM) summer (JJA) and autumn (SON) seasons.

3.3.6.2   Simulated streamflow ensembles

As described in Section 3.3.4, all post-processed GCM and CRCM5 outputs were used as inputs for the hydrological models to obtain multiple series of streamflow simulations (150 streamflow simulations in total) for each basin. To measure the seasonal streamflow simulations spread of each basin, the seasonal standard deviation of the mean annual hydrographs of all simulated streamflows was estimated for the reference and future periods. These seasonal standard deviations were then normalized by each basin's mean annual observed streamflow to obtain a customized coefficient of variation (CV). This assessment was performed for the winter, spring, summer and autumn seasons.

3.3.6.3   Flood indicators

As described in Section 3.3.5, six flood indicators were estimated and used to evaluate the uncertainty contribution of four elements of the hydroclimatic modelling chain in floods simulations. The six flood indicators (2-, 5-, 10-, 20-, 50- and 100-year return period flows) were estimated from each time series obtained from combining six probability distributions, three hydrological models, two post-processing methods and 25 climate simulations (i.e., 25 × 2 × 3 × 6 = 900 time series per basin). As performed for the climate and streamflow

simulations, the seasonal uncertainty spread of each flood indicator was quantified through a custom CV estimated from a normalized standard deviation for each basin. The analysis was also assessed for the winter, spring, summer and autumn months.

3.3.6.4   Uncertainty analysis

The uncertainty assessment on flood simulations consisted of decomposing the uncertainty contributions of (1) climate simulations, (2) post-processing methods, (3) hydrological models, and (4) probability distributions on six flood indicators using the variance decomposition method (Déqué et al., 2007; Troin et al., 2018). This method allows isolating and comparing the contribution of each element involved in the hydroclimatic chain and the interactions between them. The total variance associated to the one component, such as CRCM5 simulations (referred as V(R)), can be thus expressed as follows:

$$V(R) = R + RP + RH + RD + RPH + RHD + RPD + RPHD \qquad (3.4)$$

where R, P, H, and D represent the variance related to the CRCM5 simulations, post-processing methods, hydrological models and probability distributions, respectively. Further details on the total variance calculation and interaction terms of the variance decomposition method are given in Supplementary material S2

The assessment was divided in two main variants in terms of the climate simulations included in the hydroclimatic chain. The first configuration considers the GCM-ensemble and the second configuration only considers the CRCM5-ensemble. The rationale behind these two configurations is to evaluate the impact of a small high-resolution ensemble and a larger GCM-ensemble on the uncertainty contribution of each element. These two uncertainty evaluations were performed for the annual and seasonal (i.e., winter, spring, summer and autumn) flood indicators (i.e., 2-, 5, 10-, 20-, 50-, 100-year) for the reference and future periods (i.e., 1976-2005, 2041-2070 and 2070-2099).

3.4   Results

3.4.1   Climate simulations analysis

As previously described in Section 3.3.6 the seasonal uncertainty spread is measured in terms of standard deviation on mean seasonal precipitation, maximum and minimum temperature envelopes obtained from the GCM- and CRCM5-outputs post-processed by the QM and DBC methods. Figure 3.3 presents the seasonal standard deviation of precipitation, maximum and minimum temperature envelopes for two groups of basins based on their main flood generating process: (1) 50 Quebec snow-dominated basins (panels a) and (2) 46 Mexican rain-dominated basins (panels b). For both groups, the standard deviation of precipitation, maximum and minimum temperatures are presented in the upper, middle and lower panels, respectively. Each climate variable (each row) presents the seasonal results for the reference and future periods, from left to right. Overall, the standard deviation boxplots of GCM (in grey) and CRCM5 (in blue) climate variables show increasing values when moving toward future periods over both groups of basins. Yet the magnitude of these increasing spreads varies between seasons, climate variables, and basins. In terms of precipitation, both climate ensembles show larger standard deviations during the summer and autumn months, with the GCM-ensemble constantly showing slightly larger values. Yet, the difference between GCM- and CRCM5-ensembles is more important over the rain-dominated basins. Regarding maximum temperatures spread, the snow-dominated basins present similar standard deviations in the four seasons during the reference period, with slightly higher values during the winter. However, over future periods the GCM-ensemble presents larger standard deviations during spring and autumn months, while the CRCM5-ensemble shows larger spreads during the autumn. Some differences are also observed between ensembles over the future periods. During the 2041-2070 period, the GCM-ensemble shows slightly larger standard deviations, while during the 2070-2099 period the CRCM5-ensemble displays larger standard deviations during spring, summer and autumn months. Over the rain-dominated basins, the maximum temperature spreads show small differences between seasons and climate ensembles. Yet, during the future periods larger spreads are observed during the summer. This is especially observed during the

farthest period, where the CRCM5-ensemble shows larger standard deviations than the GCM-ensemble. The minimum temperature standard deviations show very similar behaviours to those observed for maximum temperatures, except that for the rain-dominated river basins where the GCM-ensemble spread remain important over future periods.
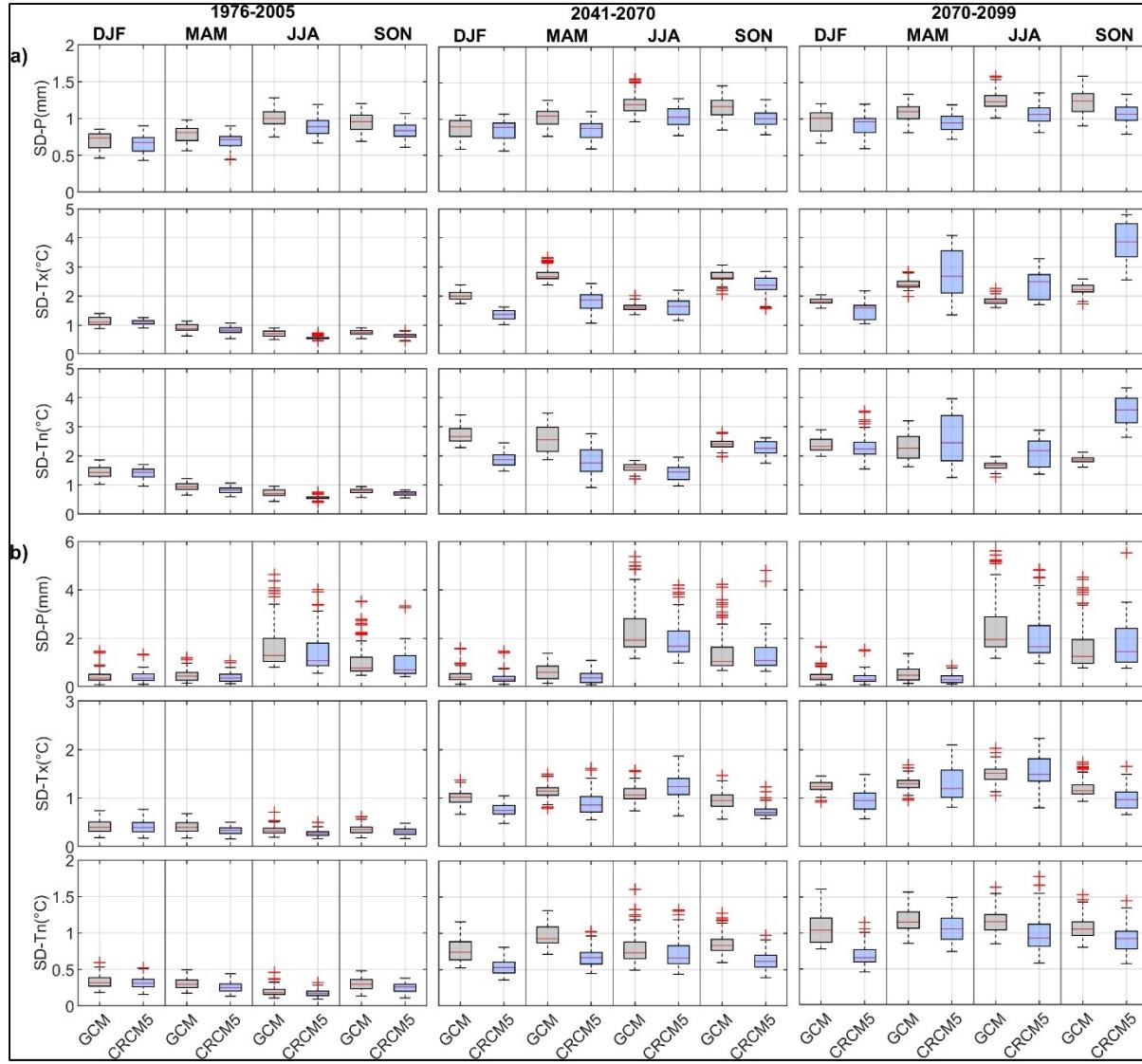
Figure 3.3 Standard deviation boxplots of mean seasonal precipitation, maximum and minimum temperature envelopes simulated with the GCM-ensemble (grey) and CRCM5-ensemble (blue) for the snow-dominated (panels a) and rain-dominated basins (panels b) over the 1976-2005, 2041-2070 and 2070-2099 periods (from left to right). The results using both post-processing methods (QM and DBC) are included in each ensemble. Each panel shows the winter (DJF), spring (MAM), summer (JJA) and fall (SON) seasons

3.4.2   Hydrological modelling performance

Table 3.4 presents the description of the hydrological modelling performance over the 96 basins. The median, 5th and 95th percentiles of the distribution of KGE-values are presented for the calibration and validation periods for the GR4J, MOHYSE and HMETS hydrological models. Overall, the 96 studied basins show a good performance with median KGE-values above 0.8 on calibration and validation periods. The distribution of KGE-values for the calibration period shows median values of 0.9, 0.87, and 0.88 for GR4J, MOHYSE and HMETS, respectively. For the validation period, the distribution of KGE-values shows median values of 0.85, 0.86 and 0.84 for GR4J, MOHYSE and HMETS, respectively. The few basins with KGE-validation values lower than 0.6 were retained in the study as at least one of the three hydrological models showed higher skill over those basins. No clear difference was observed between hydrological models' performance, yet slightly higher performances were observed from GR4J and MOHYSE over the snow- and rain-dominated basins (results not shown).

Table 3.4 Overview of the calibration and validation KGE-values distribution for GR4J, MOHYSE and HMETS hydrological models over all 96 basins

Hydrological Model	KGE-Calibration			KGE-Validation		
	5th percentile	Median	95th percentile	5th percentile	Median	95th percentile
GR4J	0.61	0.90	0.95	0.58	0.85	0.94
MOHYSE	0.60	0.87	0.93	0.61	0.86	0.92
HMETS	0.64	0.88	0.94	0.60	0.84	0.92

3.4.3   Streamflow projections

To present an overview of the hydrological diversity of the study area, Figure 3.4 shows the mean annual hydrographs simulated using observed and simulated climate data of the three

selected basins. The GCM- (grey) and CRCM5-driven (blue) mean annual hydrographs are presented for a snow-dominated basin (panels a), a dry rain-dominated basin (panels b) and a humid rain-dominated basin (panels c). The streamflow simulations obtained with all post-processing methods and hydrological models are included in the envelopes. Each basin (each row) displays the results for the reference and future periods from left to right. This figure shows that the observed mean annual streamflow is covered by the simulated envelopes among the three selected basins during the reference period. Overall, the GCM-driven streamflow ensemble shows wider spreads than the CRCM5-driven streamflow ensemble. It is also observed that the envelopes are wider during future periods over all basins. Looking at each basin, it is observed that the range of uncertainty spread clearly varies between them. The snow-dominated basin shows a smaller spread compared to the wider envelopes of the rain-dominated basins. During spring and autumn months of the reference period, it is observed that the seasonal peaks of the snow-dominated basin were not always covered by the simulated envelope. During the same seasons, the envelopes show particularly wider spreads during the future periods. This is especially observed on the GCM-driven spring peaks (in grey). In terms of spread, the dry and humid rain-dominated basins show a clearly wider uncertainty envelope during the summer and autumn months (rain season). This is especially observed over the humid basin, with values varying from ≈ 0 to 500 m^3s^{-1} during the reference period, and up to 1000 m^3s^{-1} during the farthest horizon.

To quantify the uncertainty spread of mean seasonal streamflows of each basin, a customized seasonal CV was estimated from all GCM- and CRCM-driven mean annual hydrographs. The GCM- and CRCM5-driven mean annual hydrographs include streamflow simulations using all post-processing techniques and hydrological models. As presented in Section 3.4.1 this analysis shows the results for the snow- and rain-dominated basins. Figure 3.5 shows the CVs boxplots of the mean annual streamflow envelopes simulated with the GCM- (in grey) and CRCM5-ensembles (in blue) for the reference and future periods for the snow-dominated (panels a) and rain-dominated basins (panels b).
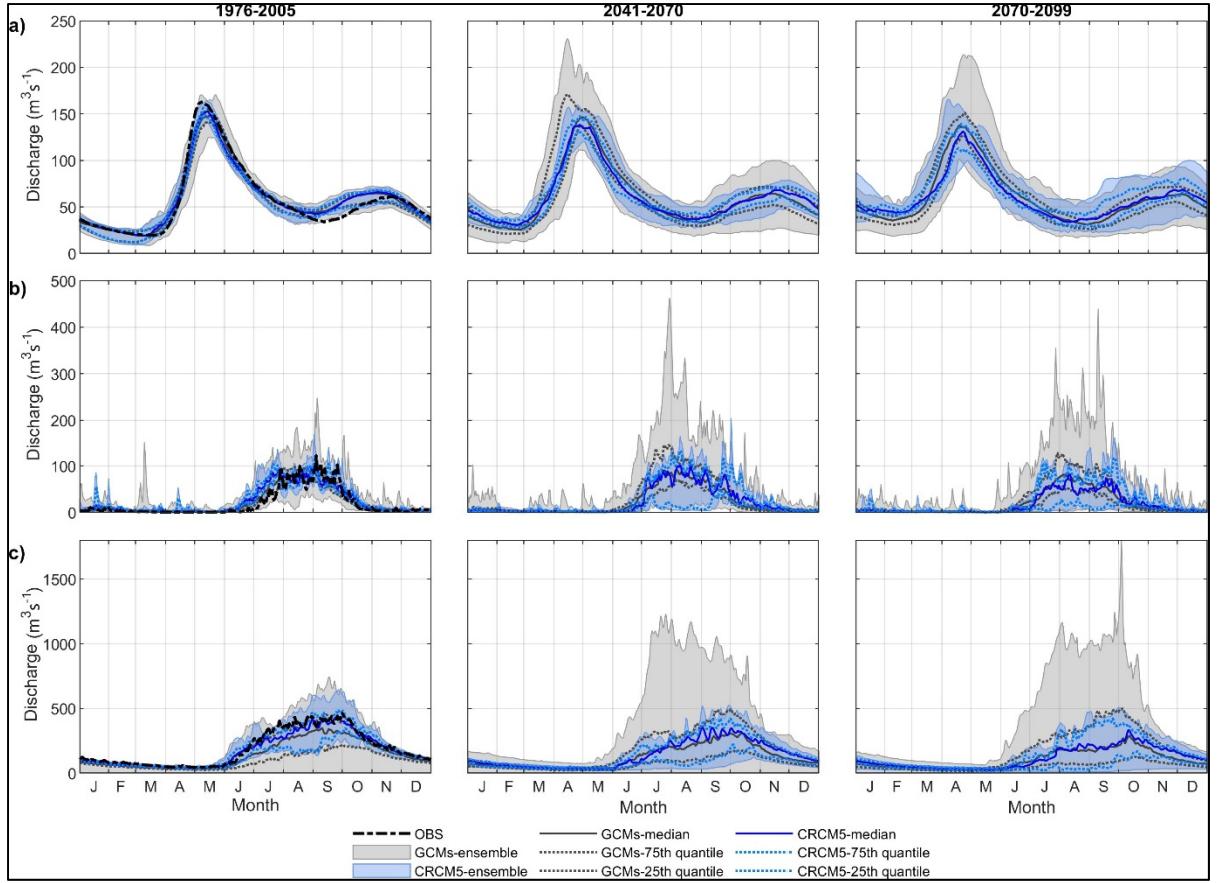
Figure 3.4 Mean annual hydrographs simulated using observed data (black), GCM (grey) and CRCM5 (blue) ensembles for three basins, a snow-dominated basin (panels a), a dry rain-dominated basin (panels b) and a humid rain-dominated basin (panels c) over the 1976-2005, 2041-2070 and 2070-2099 periods (from left to right). Both post-processing results (QM and DBC) and three hydrological models (GR4J, MOHYSE & HMETS) are included for the GCM and CRCM5 envelopes. The median (solid lines) and the 25th and 75th percentiles (dashed lines) are presented for both ensembles

Looking at the results presented between both groups of basins, three main differences are observed. First, the streamflow simulations spread using the GCM-ensemble is always wider than when using the CRCM5-ensemble. Second, the snow-dominated basins show larger spreads during the spring months. Third, rain-dominated basins show clearly larger spreads during the rainy season (summer and autumn months). Regarding temporal variations, both basin groups show generally higher CVs over the future horizons. Over the snow-dominated basins, both future periods show similar median seasonal CVs, except during spring where slightly higher values are observed in the 2041-2070 period. Contrastingly, the rain-dominated

basins show slightly larger spreads during the summer over the 2041-2070 period, while during the 2070-2099 larger CVs are presented over the autumn months. This behaviour is observed with the GCM- and CRCM5-ensembles.
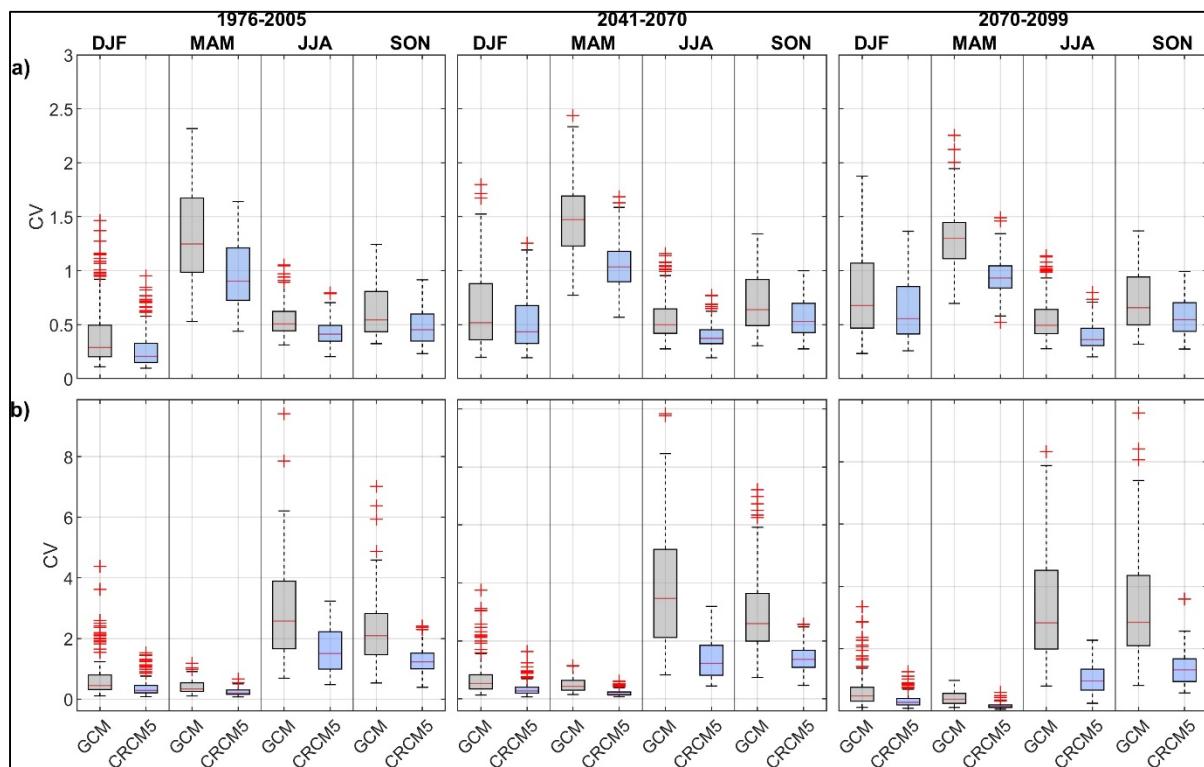
Figure 3.5 Coefficients of variation (CV) of the mean seasonal streamflow (m³/s) envelope with the GCM-ensemble (grey) and CRCM5-ensemble (blue) for the snow-dominated basins (panels a) and rain-dominated basins (panels b) over the 1976-2005, 2041-2070 and 2070-2099 periods (from left to right panels). Both post-processing results (QM and DBC) and three hydrological models (GR4J, MOHYSE & HMETS) are included for the GCM and CRCM5 envelopes. Each panel shows the winter (DJF), spring (MAM), summer (JJA) and fall (SON) seasons

3.4.4   Flood indicators

To quantify the uncertainty spread of each flood indicator, the spread of seasonal return levels estimated with the GCM- and CRCM5-ensemble, two post-processing methods, three hydrological models and six probability distributions was measured with a customized CV for each basin. This analysis is presented for the two groups of basins, as before.

Figures 3.6 and 3.7 show the seasonal CVs boxplots for the six flood indicators estimated from the GCM- (in grey) and CRCM5-driven (in blue) streamflows for the snow- and rain-dominated basins, respectively. Panels a), b), c), d), e) and f) present the results for the 2-, 5-, 10-, 20-, 50-, and 100-year return periods, respectively. The CVs boxplots over the snow-dominated basins (Figure 3.6) generally show larger uncertainty spreads of all flood indicators during the spring months, the season of annual maximum flood peaks. Over future periods, the boxplots show generally higher medians and wider distributions. This behaviour is also observed when increasing return periods. In other words, the spread of return levels increases when increasing return periods, over both the GCM- and CRCM5-ensemble. Over the rain-dominated basins (Figure 3.7), the CVs boxplots of seasonal return levels show larger values during summer and autumn, the seasons of annual maximum flood peaks. This behaviour is observed with the GCM- and CRCM5-ensemble over all flood indicators. As observed over the snow-dominated basins, the spread increases when moving to future periods and when increasing return periods. Overall, the GCM- and CRCM5-ensemble show similar temporal behaviour over the six flood indicators, with the GCM-driven flood indicators consistently showing larger CVs and wider boxplots.

3.4.5   Variance decomposition

To quantify the uncertainty contribution of each element of the hydroclimatic modelling chain on flood simulations, two main variance decomposition analyses were performed on six flood indicators estimated from each streamflow projection. As previously described in Section 3.3.6, two variance decomposition configurations were considered, one using the GCM-ensemble (22 simulations) and the other one using the CRCM5-ensemble (three simulations) as the climate simulations ensemble element. Both analyses were performed for each season (DJF, MAM, JJA and SON) and period (1976-2005, 2041-2070 and 2070-2099) over all 96 basins.
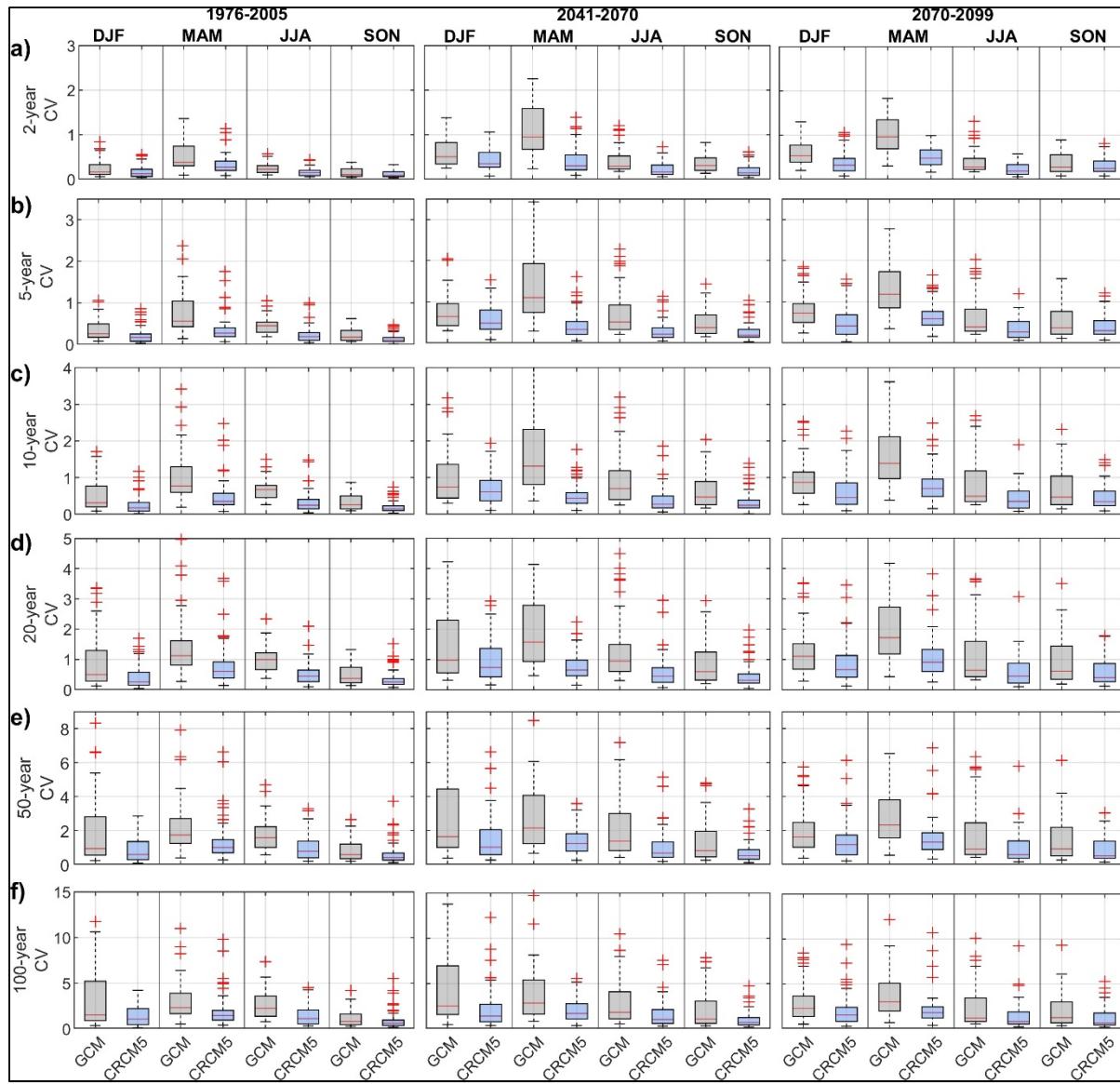
Figure 3.6 Coefficients of variation (CV) of seasonal return levels envelopes (m^3/s) simulated with the GCM-ensemble (grey) and CRCM5-ensemble (blue) for the snow-dominated basins over the 1976-2005, 2041-2070 and 2070-2099 periods (from left to right). The envelopes include return levels simulated with all post-processing methods, hydrological models and probability distributions. Each panel shows the winter (DJF), spring (MAM), summer (JJA) and fall (SON) seasons. Panels in row a), b), c), d), e) and f) present the results for the 2-, 5-, 10-, 20-, 50-, and 100-year return periods, respectively
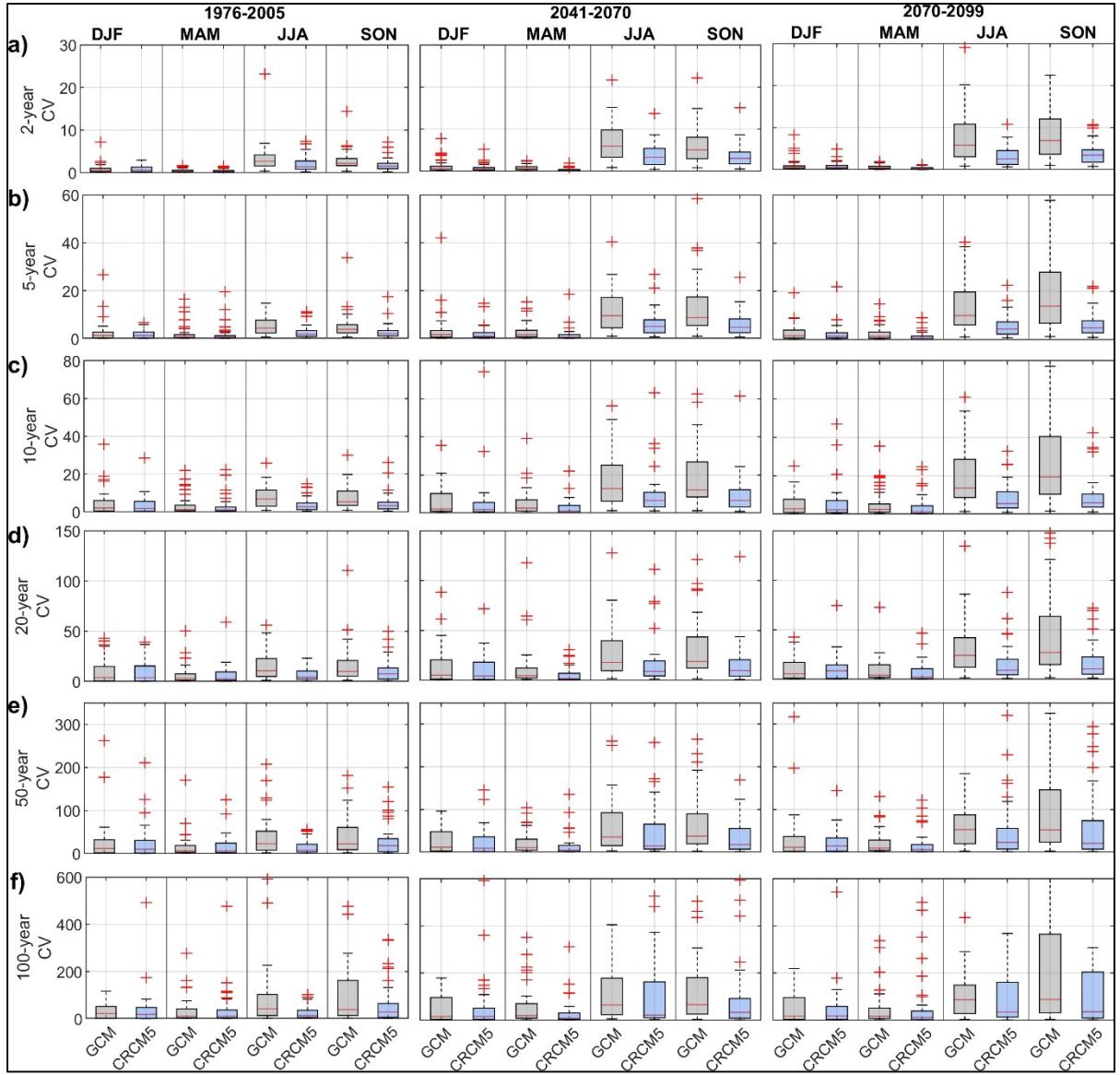
Figure 3.7 Coefficients of variation (CV) of seasonal return levels envelopes (m³/s) simulated with the GCM-ensemble (grey) and CRCM5-ensemble (blue) for the rain-dominated basins over the 1976-2005, 2041-2070 and 2070-2099 periods (from left to right). The envelopes include return levels simulated with all post-processing methods, hydrological models and probability distributions. Each panel shows the winter (DJF), spring (MAM), summer (JJA) and fall (SON) seasons. Panels in row a), b), c), d), e) and f) present the results for the 2-, 5-, 10-, 20-, 50-, and 100-year return periods, respectively

Following the same structure as in the previous sections, the results are presented per group of basins. All variance decomposition results of this section are presented in terms of the total variance associated to each element of the hydroclimatic modelling chain. As described in

Section 3.3.6.4, the total variance accounts for the variance contribution of the source alone (e.g., R for CRCM5 run) and its interactions with the other elements of the modelling chain (e.g., RP, for CRCM5 and post-processing interaction, see equation 3). Therefore, the percentages of total variance associated to each uncertainty source presented in the following figures (Figures 3.8 and 3.9) will not add up to 100%. The rationale behind this is to present the total variance contributed by each of the four elements of the hydroclimatic chain to allow ranking them by their total uncertainty contribution in this study's modelling chain. Results in this section are presented for the CRCM5-ensemble configuration only since the major conclusions for the GCM-ensemble are the same (see supplementary material). A section of the discussion is nonetheless dedicated to using CRCM5-ensemble vs. GCM-ensemble.

3.4.5.1   Uncertainty analysis with CRCM5-ensemble

Figures 3.8 and 3.9 show the boxplots of seasonal percentages of total variance contributions considering the contribution of the CRCM5-ensemble for snow- and rain-dominated basins, respectively. Figure 3.8 shows that, as observed in the uncertainty assessment using the GCM-ensemble configuration (Supplementary material - S3), the variance decomposition over the snow-dominated basins shows hydrological modelling as one of the main uncertainty sources. Hydrological modelling uncertainty contribution is constantly the first or second main contributor among most seasons, flood indicators and periods. The dominance of hydrological models on the overall variance is especially large during the spring. Overall, the CRCM5-ensemble shows the second largest uncertainty contribution over many cases. An increase of its contribution is observed when moving toward future periods, especially during the summer, where the CRCM5-ensemble shows larger contributions for return periods smaller than 50-year. The uncertainty contribution of probability distributions increases with increasing return periods over the three periods, as also observed over the variance decomposition analysis using the GCM-ensemble (Supplementary material - S3). In some cases, probability distributions become the first or second main uncertainty source, especially for the 50- and 100-year flood indicators. Post-processing methods show the smallest uncertainty contribution and small increases over the future horizons (the same is observed for the GCM-ensemble configuration).
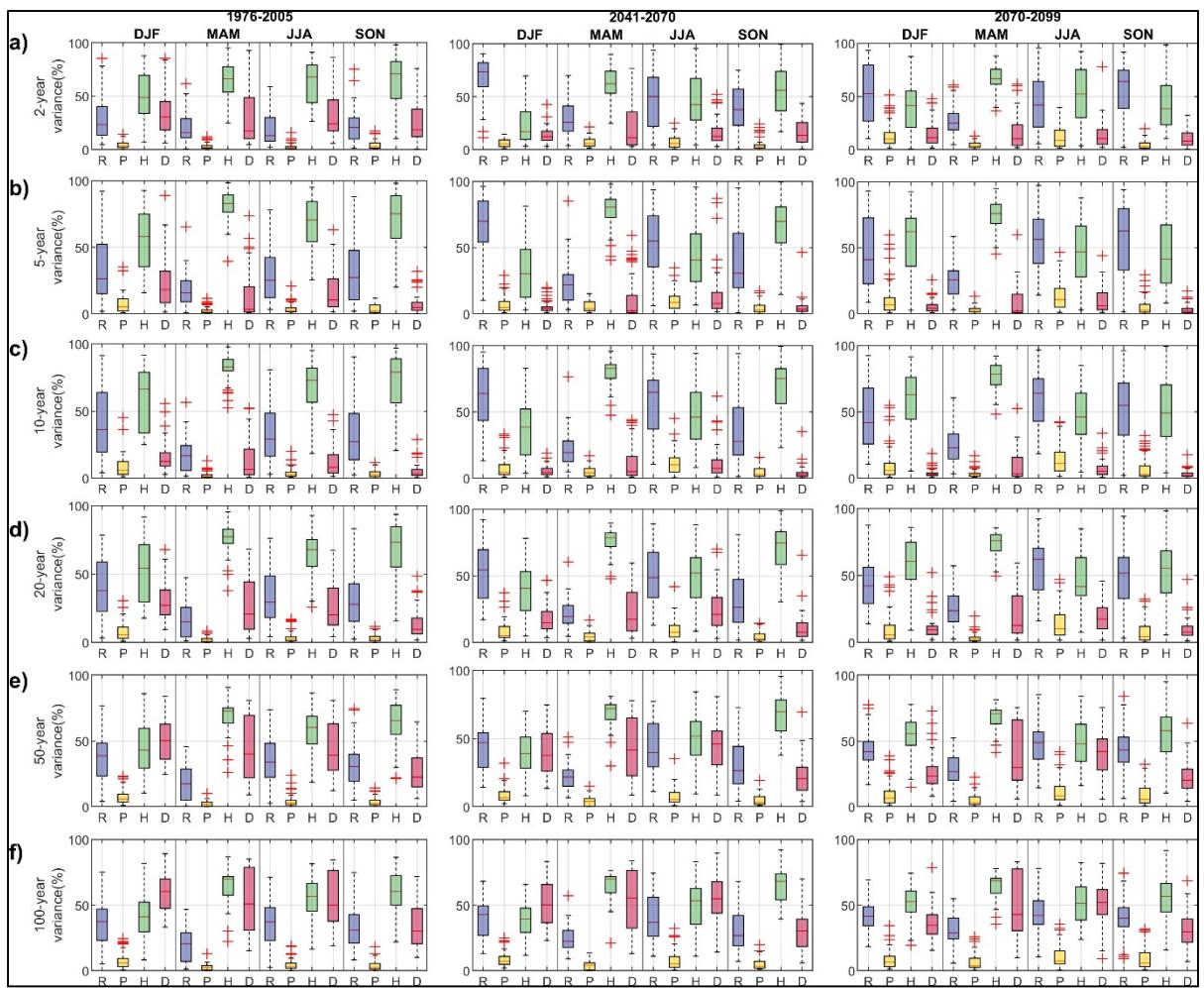
Figure 3.8 Boxplots of the seasonal percentage of total variance (%) contribution of CRCM5-runs (R, in blue), post-processing methods (P, in yellow), hydrological models (H, in green) and probability distributions in flood frequency analyses (D, in red) of the snow-dominated basins for the 1976-2005 (left panels), 2041-2070 (center panels) and 2070-2099 (right panels) periods. Panels in row a), b), c), d), e) and f) present the results for the 2-, 5-, 10-, 20-, 50-, and 100-year return periods, respectively. Winter (DJF), spring (MAM), summer (JJA) and fall (SON) results are presented for each panel
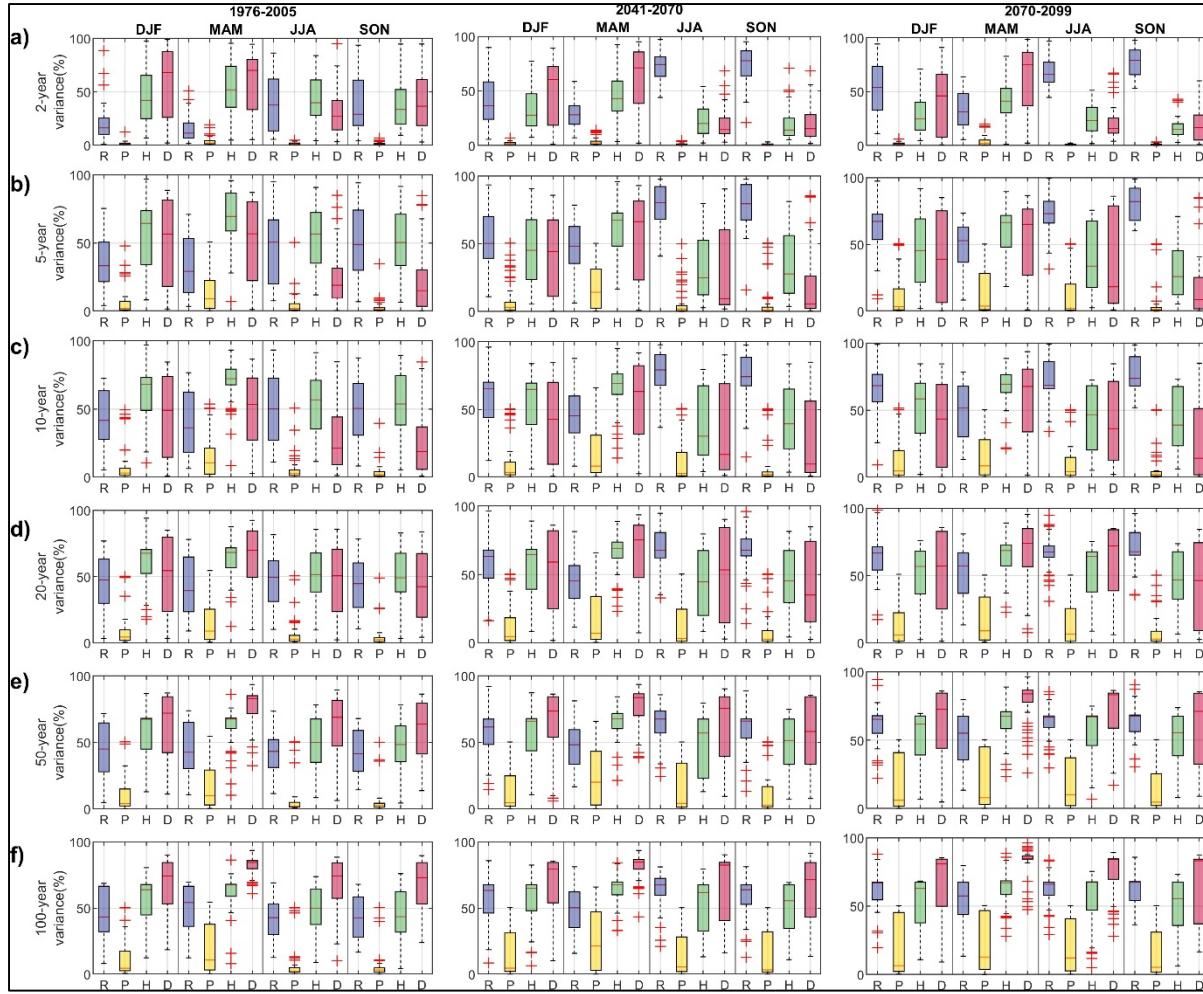
Figure 3.9 Boxplots of the seasonal percentage of total variance (%) contribution of CRCM5-runs (R, in blue), post-processing methods (P, in yellow), hydrological models (H, in green) and probability distributions in flood frequency analyses (D, in red) of the rain-dominated basins for the 1976-2005 (left panels), 2041-2070 (center panels) and 2070-2099 (right panels) periods. Panels in row a), b), c), d), e) and f) present the results for the 2-, 5-, 10-, 20-, 50-, and 100-year return periods, respectively. Winter (DJF), spring (MAM), summer (JJA) and fall (SON) results are presented for each panel

The uncertainty analysis over the rain-dominated basins, presented in Figure 3.9, reveals that the main uncertainty contributors vary between flood indicators, seasons and periods. Overall, flood indicators smaller than the 50-year return period show the CRCM5-ensemble as the main source of uncertainty during summer and autumn seasons over the future periods, while during the reference period hydrological modelling presents slightly larger median percentages of variance contributions. During winter months, no clear behaviour is observed, with the

CRCM5-ensemble, hydrological models, and probability distributions showing higher contributions over different cases. During the spring, probability distributions and hydrological models are observed as the main uncertainty contributors, with probability distributions showing higher contributions in most flood indicators over reference and future periods. In addition, uncertainty contributions of probability distributions become more important when increasing return periods, as also observed for the GCM-ensemble configuration (Supplementary material S4). In line with the GCM-ensemble analysis, post-processing methods are the smallest source of uncertainty. Yet, its contribution increases when moving toward future periods.

The results from the annual maximum flood analyses are presented in Figures 3.10 and 3.11 for the snow- and rain-dominated basins, respectively. Each figure presents the maps of the highest variance source between the CRCM5-ensemble (R, in blue), post-processing methods (P, in yellow), hydrological models (H, in green), and probability distributions (D, in red) on six return periods, and for the six flood indicators (top to bottom). From left to right, the figure shows the 1976-2005, 2041-2070 and 2070-2099 periods, respectively.

The results for snow-dominated basins (see Figure 3.10) show a larger number of basins displaying hydrological models as their major uncertainty source over most flood indicators. As observed on the seasonal analysis using the CRCM5-ensemble, hydrological modelling is constantly shown as the main uncertainty source, with few exceptions over the largest return periods. Especially over the 100-year indicator, with probability distributions identified as the largest uncertainty source over some basins. However, most of these basins show hydrological models as the main uncertainty source when moving toward future periods (as also seen with the GCM-ensemble; see supplementary material S5). Compared to the snow-dominated basins, the results over the rain-dominated basins (see Figure 3.11) are more varied, especially over the reference period. During this period, CRCM-outputs, hydrological models, and probability distributions are observed among the main uncertainty sources for different basins. Yet, the CRCM5-ensemble is constantly displayed as the largest uncertainty source on the 2-, 5- and 10-year return periods during both future periods.
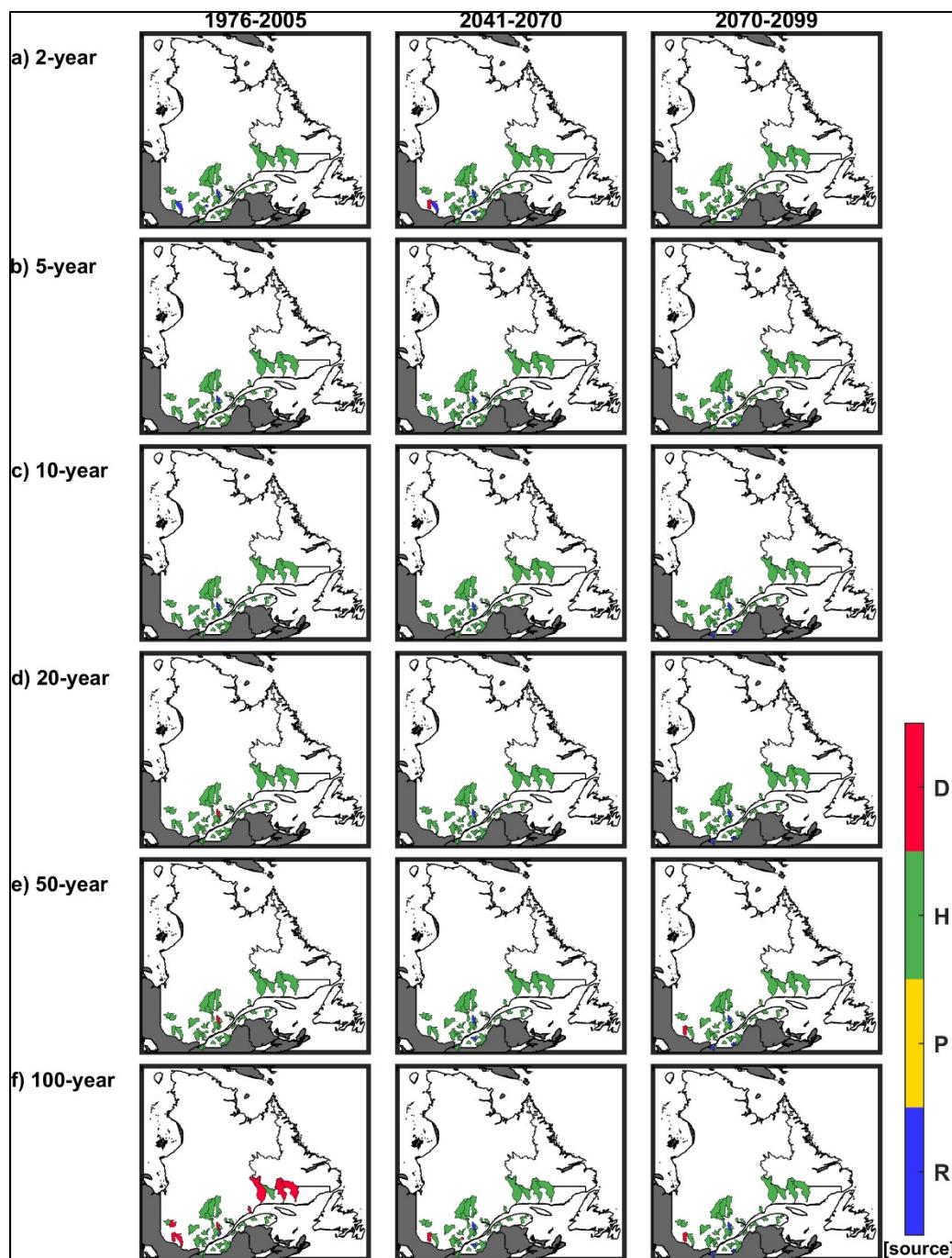
Figure 3.10 Maps of the major uncertainty contributor between CRCM5-runs (R, in blue), post-processing methods (P, in yellow), hydrological models (H, in green) and probability distributions of flood frequency analyses (D, in red) over the snow-dominated basins. Results for the 1976-2005 (left panels), 2041-2070 (middle panels) and 2070-2099 (right panels) periods are presented. In row, panels a), b), c), d), e) and f) present the results for the 2-, 5-, 10-, 20-, 50-, and 100-year return periods, respectively
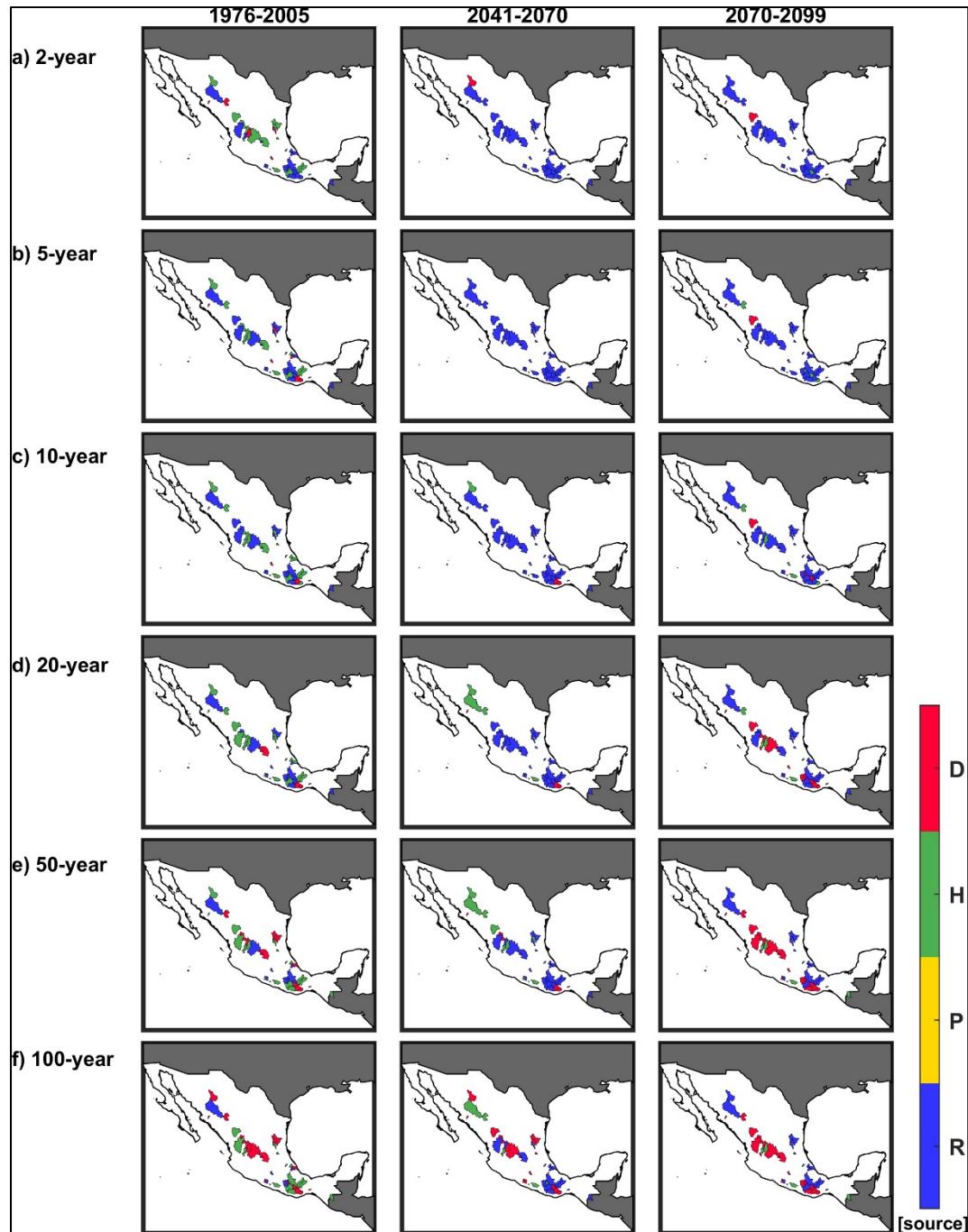
Figure 3.11 Maps of the major uncertainty contributor between CRCM5-runs (R, in blue), post-processing methods (P, in yellow), hydrological models (H, in green) and probability distributions of flood frequency analyses (D, in red) over the rain-dominated basins. Results for the 1976-2005 (left panels), 2041-2070 (middle panels) and 2070-2099 (right panels) periods are presented. In row, panels a), b), c), d), e) and f) present the results for the 2-, 5-, 10-, 20-, 50-, and 100-year return periods, respectively

As observed in previous analyses, the influence of probability distributions increases when increasing return periods. This is observed over the 20-, 50-, and 100-year return periods, where the number of basins showing probability distributions as their main uncertainty source increases over reference and future periods. The uncertainty contribution of post-processing methods, as observed in previous analyses, is the smallest compared to the other three sources. Thus, it is not observed as a dominant source over the rain-dominated basins. Similar results are also obtained for the GCM-ensemble configuration, except with a clearer domination of the GCM-ensemble as the main uncertainty source.

3.5   Discussion

3.5.1   Hydroclimatic conditions impact on uncertainty contribution

Our findings reveal that the leading uncertainty sources on flood simulations are consistently related with the basin's flood-generating process. The hydrometeorological diversity among the 96 basins included in this study allowed the identification of two main groups of basins: the snow- and rain-dominated basins. Over the snow-dominated basins, hydrological models were constantly found as the largest contributors among six flood indicators. Figure 3.10 shows a clear picture of this finding, especially over the future periods. This is in line with other studies, where hydrological models were observed as the largest uncertainty contributors on snow-dominated regimes (Giuntoli et al., 2018; Vetter et al., 2017). The uncertainty brought by hydrological models on flood simulations can be influenced by different elements. Flooding over snow-dominated regions is mainly driven by snowmelt occurring with the increasing temperatures during spring (at these latitudes). Thus, one of the reasons for the larger uncertainty contributions of hydrological models on snow-dominated basins could be the inherent structural uncertainties of hydrological modelling associated with snowmelt. This well-known uncertainty source might arise from the simplifications used to represent hydrological processes, as well as the parametric uncertainty related to hydrological models' calibration (Kuczera, Renard, Thyer, & Kavetski, 2010). It is therefore likely that different formulations used to represent the snowpack accumulation, melting and refreezing processes

could influence the different high flow simulations between hydrological models. Another possibility could be associated with the impacts of climate change on seasonality. Some studies have suggested that climate change could influence future flood seasonality of snow-dominated catchments (Vormoor, Lawrence, Heistermann, & Bronstert, 2015; Vormoor et al., 2016). This issue could affect the "parameter transferability" of hydrological models. This means that calibrating hydrological models over certain reference conditions can affect hydrological models' performance over different future conditions than the ones used to calibrate them (Seiller, Anctil, & Perrin, 2012; Zhu, Zhang, Ma, Gao, & Xu, 2016). It is therefore expected that a change in basins seasonality might influence hydrological models' response in future periods. Thus, these results not only underline the importance of using hydrological model ensembles, but also the need to evaluate models and methods adequacy for more robust projections and improved transferability under a changing climate, particularly knowing that removing low-performing models can impact the magnitude of the projected hydrological changes (Wen et al., 2020).

Contrastingly, over the rain-dominated basins, climate simulations were constantly found as the highest uncertainty contributor on flood simulations. This agrees with various studies evaluating uncertainty sources on streamflow simulations over watersheds with similar flood-generating processes (Gosling et al., 2011; Kay et al., 2009; Vormoor et al., 2015), it was thus expected that flood simulations over rain-dominated basins would be more sensitive to the climate simulations, especially over this type of basins where floods are mainly driven by extreme rainfall events. Additionally, the analysis of the climatic uncertainty spread presented in Figure 3.3 clearly shows the larger spreads of precipitation simulations over the rain-dominated basins than over snow-dominated basins. This result highlights the need of further studies on the uncertainty associated to the simulations of convective precipitation events, and on the potential added-value of using convection-permitting climate (Reszler et al., 2018).

3.5.2   Temporal and seasonal differences on uncertainty contribution

The results indicate that uncertainty contributions of the hydroclimatic chain elements was not only related with the basins' flood-generating process, but also with seasons and periods. Results over both basin groups showed that the leading uncertainty sources during the reference period sometimes changed when moving towards future periods. This temporal evolution of uncertainty has been evaluated and discussed in some studies that, in line with this study's results, have shown that the uncertainty related to a certain source is likely to vary over time (Giuntoli et al., 2018; Shen et al., 2018). Moreover, it is observed that the overall uncertainty increases with time (Shen et al., 2018). This is clearly observed in Figure 3.5, where mean seasonal streamflows show increasing spreads over future periods. In terms of seasonality, the results show that the uncertainty contribution of each source varies depending on the season. This is observed over basin groups, especially during the future periods (see Figures 3.8 and 3.9). For instance, while the leading source of uncertainty on spring flooding over the snow-dominated basins was mainly hydrological modelling, climate simulations (i.e., GCM- and CRCM5-ensemble) were constantly the leading source during other seasons (e.g., summer-autumn months). The rain-dominated basins also presented seasonal variations. During the summer-autumn flooding season, climate simulations were constantly the dominant source of uncertainty. Yet during drier winter-spring months, hydrological modelling, climate simulations, and probability distributions were all showing high uncertainty contributions. This seasonal difference could be explained by the different dominant hydrological processes during each season. For instance, during the summer all basins show climate simulations as the main uncertainty source, a season where flooding over all basins is mainly driven by rainfall events. This can be thus explained by the larger uncertainty associated with the representation of convective events on GCMs and RCMs. At the same time, the post-processing methods can also be playing a role in precipitation simulations uncertainty. Studies have shown that if a given precipitation simulation is too biased, the bias-adjustment method can be unable to overcome it, and consequently produce biased hydrology, particularly in extremes (Chen et al., 2013a). This can explain the generally larger hydrograph spreads in rain-dominated basins over the reference periods (see Figure 3.4), as well as their slightly larger post-processing

uncertainty contributions compared to snow-dominated basins (see Figure 3.9). Furthermore, as presented in the post-processing methods description (see Supplementary material S1), one of the main differences between the selected methods is the precipitation adjustment, particularly in their extremes, which can be thus linked to the observed impacts in rain-dominated basins.

Another example of seasonal variations is observed during the dry winter-spring season of the rain-dominated basins, where hydrological modelling is constantly the main source of uncertainty. During dry seasons, potential evapotranspiration estimations play an important role on streamflow estimations. Thus, the dominance of hydrological models during dry months could be explained by the uncertainty related to the estimations of potential evapotranspiration, which has been found as one of the main uncertainty contributors in hydrological modelling uncertainty (Dallaire, Poulin, Arsenault, & Brissette, 2021; Oudin et al., 2005).

3.5.3 Flood indicators impacts on uncertainty sources

This study evaluated the uncertainty contributions of four elements of the hydroclimatic chain on six flood indicators: the 2-, 5-, 10-, 20-, 50-, and 100-year return periods. Our results showed that the variance contribution of each source varied between flood indicators. This behaviour was especially observed for the uncertainty related to the choice of probability distribution for the return levels estimations. It was generally observed that the uncertainty contributions of probability distributions increased when increasing return periods. This is clearly displayed in the seasonal analyses of snow- and rain-dominated basins presented in Figures 3.8 and 3.9. This behaviour could be explained by the uncertainty associated with the length of the data series used for the flood frequency analysis. Different studies have shown that using short lengths of data (e.g., 30 years or less) to estimate larger return periods (e.g., 100 year) might result in larger uncertainties (Hu, Nikolopoulos, Marra, & Anagnostou, 2019). In this study, three 30-year streamflow series were used, the 1976-2005, 2041-2070 and 2070-2099 periods. It was thus expected to observe larger uncertainties related to their estimation on the largest

return periods (i.e., 50-year and 100-year). It can be therefore recommended to further study the uncertainties associated to flood frequency analyses, such as using peaks-over-threshold approaches to increase the length of streamflow series.

3.5.4   GCM-ensemble vs CRCM5-ensemble

The present study evaluated the uncertainty contribution of climate simulations using two different ensembles, (1) a 22 GCM simulations and (2) a three high-resolution CRCM5 simulations. The aim of these comparisons was not to evaluate GCM-ensembles over high-resolution RCM-simulations, but to compare their impact on the uncertainty analysis. It is important to recall that the size of the RCM-ensemble was limited by the availability of simulations covering Mexico (which is not the case of the regular CORDEX CAM-domain simulations). Nonetheless, this only reflects the reality of working with generally more limited RCM-ensembles with respect to larger available GCM-ensembles. The idea behind this was to evaluate the impact of the choice of climate simulations ensemble on the uncertainty contributions on flood simulations. Therefore, generalizations based on these results should be cautiously made, as a small ensemble as the CRCM5-ensemble used in this study, is not robust enough compared to the more robust GCM-ensemble results. Similarities were however observed between the two main variance decomposition configurations, especially over the snow-dominated basins. Both analyses showed hydrological models and climate simulations as the leading uncertainty sources on snow- and rain-dominated basins, respectively. However, the impact of using the GCM-ensemble against the CRCM5-ensemble on the uncertainty contributions was different between the two basin groups. The snow-dominated basins showed consistent dominance of hydrological models on both analyses (GCM or CRCM5), especially during spring floods. Over the rain-dominated basins however, the dominance of climate simulations was not as clear when using the CRCM5-ensemble as when using the GCM-ensemble (see supplementary material S6). The results using the larger GCM-ensemble showed a clear dominance of climate simulations uncertainty contribution, while the CRCM5-ensemble results showed more varied results, especially for the return periods larger than the 20-year. These results were expected since the smaller CRCM5-ensemble presented narrower

streamflow spreads. This is observed in Figure 3.6, where the seasonal CVs of the mean annual hydrographs of both basin groups showed smaller spreads with the CRCM5-ensemble than with the GCM-ensemble. This could possibly mean that the streamflow uncertainty was underestimated when using the small CRCM5-ensemble. This is especially important to highlight knowing that it has become a common approach in the literature to use a couple of high-resolution RCM simulations to evaluate climate change impacts on future flooding. It is thus important to further evaluate if a small RCM-ensemble can be representative of the streamflow uncertainty or if it is only a subsample of a more representative uncertainty spread, particularly over regions showing a larger climatic uncertainty as observed over the rain-dominated basins in this study. Contrastingly, the comparisons between the GCM- and CRCM5-outputs uncertainty spreads (i.e., precipitation, maximum and minimum temperatures) showed that the small CRCM5-ensemble sometimes displayed larger spreads than the GCM-ensemble. For instance, the seasonal uncertainty spreads of temperature envelopes over the snow-dominated basins (Figure 3.3 panel a) show that in some cases the three-member CRCM-ensemble covered a wider uncertainty envelope than the GCM-ensemble. This is observed on maximum and minimum temperatures during the spring, summer and autumn months of the 2070-2099 period. Similarly, the CRCM5-ensemble of maximal temperature displayed slightly larger spreads on the summer months over the rain-dominated basins (Figure 3.3 panel b). This behaviour could be explained by the improved representations on high-resolution RCM simulations, such as topography, underlining the need to further study the processes representations in climate simulations. On the other hand, this study did not account for the natural climate variability in climate simulations. And, as presented in recent studies (Chegwidden et al., 2019; Deser et al., 2020; Hingray et al., 2019), the GCM/CRCM5 spreads are likely to be a combination of their uncertainty and natural climate variability, particularly in extreme climate projections.

3.5.5 Limitations

This study is unavoidably limited by its different methodological choices. First, this study only covers some of the hydrometeorological regions of the world. Other hydrological regimes,

such as those located in the equatorial hydrobelt (Meybeck, Kummu, & Dürr, 2013) were not considered, thus different results can be expected. The hydroclimatic processes that drive streamflow generation and flooding are location-dependent. It can be therefore expected that the selected climate simulations, hydrological models, snow and potential evapotranspiration formulations will yield different results in different regions. Yet, applying a common methodology over numerous regions allows more robust conclusions and the identification local sensitivities.

The use of a subset of the GCM and RCM simulations available can also influence and limit the results and conclusions. For instance, high-resolution simulations issued from the latest CMIP6 were not included in this study. Different results can be expected as high-resolutions climate simulations have shown better process representation in smaller catchments (Castaneda-Gonzalez et al., 2019) and lower temperature spreads have been observed compared to the CMIP5-ensemble (Song, Chung, & Shahid, 2021).

A plethora of climate simulations post-processing methods have been and continue to be developed, yet only two methods were considered in this study. These methods were selected based on satisfactory performances obtained in previous studies. However, this choice can influence the presented results as some studies have shown that state-of-the-art methods still have difficulties in correcting variability and can misrepresent regional feedbacks (Maraun et al., 2017). This can be particularly relevant in correcting precipitation series, which it has been shown can have a greater impact on rain-driven regions and seasons (Chen et al., 2013a; Chen et al., 2013b).

This study was also limited by only using lumped hydrological models. This limitation mainly resulted from the high demand of computational power and data to apply distributed/physically-based hydrological models over the numerous basins used. Yet, it is important to highlight that the conclusions obtained might be different as distributed/physically-based hydrological models are expected to more reliably represent future climate conditions (with their own uncertainties), particularly on complex

spatiotemporal events as floods (Ajami, Gupta, Wagener, & Sorooshian, 2004; Pechlivanidis et al., 2011). At the same time, using lumped hydrological models implicated the need to average GCM and CRCM5 outputs. A process that can both impact the magnitude of the simulated climate extremes, but also have the advantage of minimizing the impact of anomalous/biased simulated points.

3.6 Conclusions

This study investigated the uncertainty contribution of four elements of the hydroclimatic modelling chain on flood simulations: (1) climate simulations, (2) post-processing techniques, (3) hydrological models and (4) probability distributions. Two main variants of the hydroclimatic modelling chain were evaluated in terms of climate simulations. One configuration used a large GCM-ensemble (22 simulations) and another one used a smaller high-resolution CRCM5-ensemble (three simulations). Two post-processing methods, three lumped hydrological models and six probability distributions were included on both configurations of the hydroclimatic modelling chain. Moreover, to cover a wide range of hydroclimatic conditions and periods, this study was performed over 96 basins with contrasting hydroclimatic and flooding regimes for a reference period (1976-2006) and two future periods (2041-2070 and 2070-2099). Using a variance decomposition approach, the annual and seasonal uncertainty contribution of each source was quantified for both configurations on six flood indicators (2-, 5-, 10-, 20-, 50- and 100-year return periods) and each period. The results showed a relation between each region's flood-generating process and the overall uncertainty contributions of each source. It was shown that hydrological models and climate simulations consistently contributed the most uncertainty to flood simulations uncertainty on snow- and rain-dominated basins, respectively. This difference is most likely due to the different flood-generating processes of the basins. Therefore, to keep assuming climate simulations as the main uncertainty source on hydrological studies should be reconsidered and further analyzed. The temporal and seasonal analyses also revealed that the uncertainty contribution of each source varied depending on seasons and periods of analysis. Yet, the main uncertainty contributors during the reference period were constantly different over the future periods. And

the uncertainty related to the choice of probability distribution consistently increased when increasing return periods. Being the first and second uncertainty contributor for the largest return periods (i.e., 50-year and 100-year). The clear variations of uncertainty contributions at the basin level discussed in this study highlight the need to further study the uncertainty contributions of various sources and possibly identify reducible uncertainties such as better processes representations in climate and hydrological models that might lead to increasing our confidence and understanding of future flood projections.

3.7 Acknowledgments

3.8 Supplementary material

3.8.1 Supplementary material – S1

Daily bias correction (DBC)

The DBC is a post-processing method that merges the daily translation method (DT; Mpelasoka & Chiew, 2009) and the local intensity scaling method (LOCI; Schmidli et al., 2006). The DBC starts using the LOCI method to fit simulated precipitation occurrences to the occurrences of observed precipitation data to correct the simulated wet-day frequencies. This process involves the following three main stages:

1. In the first stage, a monthly threshold of wet-day is estimated from the given daily precipitation simulation issued from a given climate model to match the threshold exceedance to the observed precipitation series.

2. The second stage then consists on estimating a monthly scaling factor between the mean monthly observed precipitation and the mean monthly simulated precipitation over the reference period.

3. Finally, in the third stage, the previously calculated monthly thresholds and scaling factors estimated over the reference periods, are used to adjust the monthly simulated precipitation over the future period.

After this correction, the DT method is used to correct the frequency distributions of precipitation amounts and temperature simulations through a quantile-approach as follows:

$$T_{DBC,fut} = T_{SIM,fut} + (T_{SIM,ref,Q} - T_{OBS,Q}) \qquad (3.5)$$

$$P_{DBC,fut} = P_{SIM,fut} \cdot (P_{SIM,ref,Q}/P_{OBS,Q}) \qquad (3.6)$$

where $T_{DBC,fut}$ and $P_{DBC,fut}$ are the DBC-corrected daily temperature and precipitation simulations for the future periods. $T_{SIM,fut}$ and $P_{SIM,fut}$ are the daily temperature and precipitation simulations for the future periods. $T_{SIM,fut,Q}$ and $P_{SIM,fut,Q}$ are the daily temperature and precipitation future simulations according to a given month quantile. Following the same structure, $T_{OBS,Q}$ and $P_{OBS,Q}$ are the daily temperature and precipitation according to a given month quantile. In this study 100 integral percentiles are calculated per month.

Quantile mapping (QM)

The QM method (Maraun, 2016; Themeßl et al., 2012; Themeßl et al., 2011) is a simple and robust post-processing method based on the transformation of distributions by quantile. As described in Chen et al. (2013b), the QM method corrects the precipitation distribution based on daily-constructed point-wise empirical cumulative distribution functions including both wet and dry days in the empirical cumulative distribution functions estimation. In this way, the frequency and amounts of precipitation occurrence are simultaneously corrected.

The correction of the daily precipitation is described in equations 3.3-3.6. First, as displayed in following equation 3.3 the empirical cumulative distributions (ecdfs) and their probabilities ($Pr_{SIM,ref,d}$) are determined from the simulated daily precipitation of the reference period ($P_d^{SIM,ref}$). In this stage, both dry and wet days are used to correct the frequency of precipitation occurrence but also precipitation amounts.

$$Pr_{SIM,ref,d} = ecdf_d^{SIM,ref}(P_d^{SIM,ref}) \qquad (3.7)$$

Then, as presented in the following equation 3.4, the correction factors (CF) are estimated by calculating the differences between the observed and simulated inverse edcfs (represented as ecfds $^{-1}$) at the previously probability (Pr) for the reference period.

$$CF_d^{SIM,ref} = ecdf_d^{OBS^{-1}} \cdot \left(Pr_{SIM,ref,d} \right) - ecdf_d^{SIM,ref^{-1}} \cdot \left(Pr_{SIM,ref,d} \right) \qquad (3.8)$$

The third stage presented in equation 3.5 consists in interpolating and extrapolating the CFs obtained at the reference periods to obtain the CFs of the future period. This process involves an extrapolation in order to treat the "new" extreme precipitation values of the future period. In this study, to avoid unstable CF values, the 10 CFs from the highest quantiles are averaged.

$$CF_d^{SIM,ref} \rightarrow\rightarrow\rightarrow\rightarrow CF_d^{SIM,fut} \qquad (3.9)$$

Finally, as presented in equation 3.6, the obtained CF values of the future period are added to the simulation of future precipitation to obtain the corrected daily precipitation ($P_d^{QM,fut}$).

$$P_d^{QM,fut} = P_d^{SIM,fut} + CF_d^{SIM,ref} \qquad (3.10)$$

In addition, when the frequency of dry days in the simulated precipitation is greater than the frequency in observations, the frequency is adapted as proposed by Themeßl et al. (2011). Unlike precipitation, the correction of temperature simulations is done by fitting the quantiles

of simulated data to the quantiles of observations using normal distributions. This is mainly due to larger differences between the simulated and observed temperatures, where many temperature values simulated in the future period fall beyond the ranges over the reference period.

3.8.2   Supplementary material – S2

The uncertainty analyses were carried out using the variance decomposition method. This method allows isolating and comparing the contribution of each uncertainty source involved in the hydroclimatic chain used in this study. According to Déqué et al. (2007) the variance decomposition can be represented as follows:

$$V(X_{GPHD}) = G + P + H + D + GP + GH + GD + PH + PD + HD \qquad (3.11)$$
$$+ GPH + GPD + GHD + PHD + GPHD$$

where the variance of the variable X (e.g., seasonal 20-year return period streamflow) is influenced by the uncertainty sources, G (i.e., GCMs), P (i.e., post-processing method), H (i.e., hydrological models) and D (i.e., probability distribution). Thus, assuming G varying from 1 to 22, P varying from 1 to 2, H varying from 1 to 3 and D varying from 1 to 6, the source G and its interaction with the source P, for instance, are expressed as:

$$G = \frac{1}{22} \sum_{G=1}^{22} (X_{G\circ\circ\circ} - X_{\circ\circ\circ\circ})^2 \qquad (3.12)$$

$$GP = \frac{1}{44} \sum_{G=1}^{22} \sum_{P=1}^{2} (X_{GP\circ\circ} - X_{G\circ\circ\circ} - X_{\circ P\circ\circ} - X_{\circ\circ\circ\circ})^2 \qquad (3.13)$$

where the dots (○) represent the average of the replaced source. Thus, the total variance associated to the source G is expressed as follows:

$$V(G) = G + GP + GH + GD + GPH + GHD + GPD + GPHD \qquad (3.14)$$

By following this equation, the total variance $V(X_{GPHD})$ is thus not the sum of the four uncertainty sources variance contributions V(G), V(P), V(H), and V(D), as they include interaction terms (e.g., GP or GPHD). Yet, the magnitude of each of the total variance terms (i.e., V(G), V(P), V(H), and V(D)) is indicative of the impact of each source of uncertainty to the given variable X. This approach has been used in different applications to quantify uncertainties in climatic and hydrological studies supporting its use in this study (Aryal, Shrestha, & Babel, 2018; Dallaire et al., 2021; Hosseinzadehtalaei, Tabari, & Willems, 2018; Troin et al., 2018).

3.8.3   Supplementary material – S3
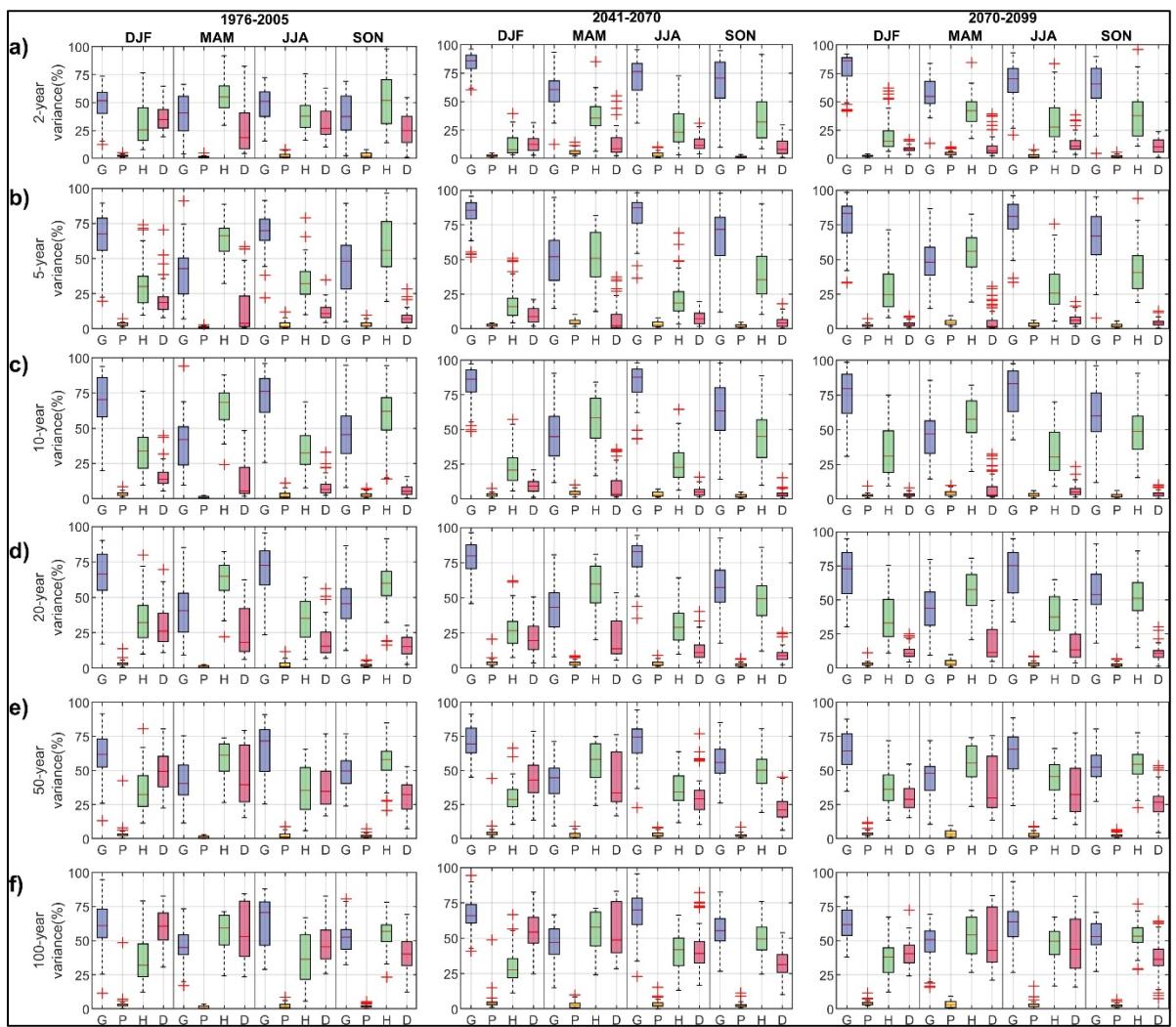
Figure S3.12 Boxplots of the seasonal percentage of total variance (%) contribution of GCM-ensemble (G, in blue), post-processing methods (P, in yellow), hydrological models (H, in green) and probability distributions in flood frequency analyses (D, in red) of the snow-dominated basins for the 1976-2005 (left panels), 2041-2070 (center panels) and 2070-2099 (right panels) periods. Panels in row a), b), c), d), e) and f) present the results for the 2-, 5-, 10-, 20-, 50-, and 100-year return periods, respectively. Winter (DJF), spring (MAM), summer (JJA) and fall (SON) results are presented for each panel
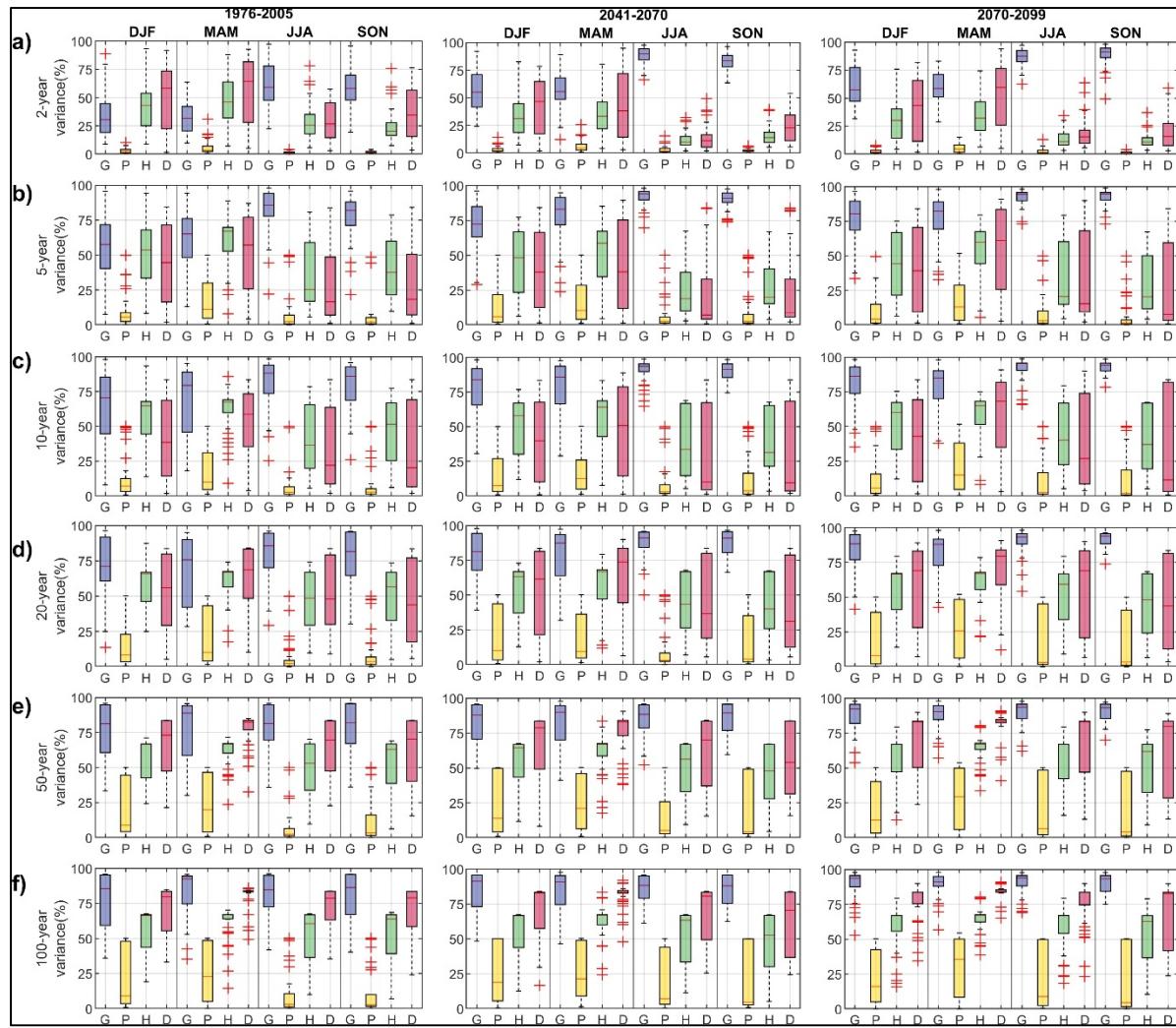
Figure S3.13 Boxplots of the seasonal percentage of total variance (%) contribution of GCM-ensemble (G, in blue), post-processing methods (P, in yellow), hydrological models (H, in green) and probability distributions in flood frequency analyses (D, in red) of the rain-dominated basins for the 1976-2005 (left panels), 2041-2070 (center panels) and 2070-2099 (right panels) periods. Panels in row a), b), c), d), e) and f) present the results for the 2-, 5-, 10-, 20-, 50-, and 100-year return periods, respectively. Winter (DJF), spring (MAM), summer (JJA) and fall (SON) results are presented for each panel
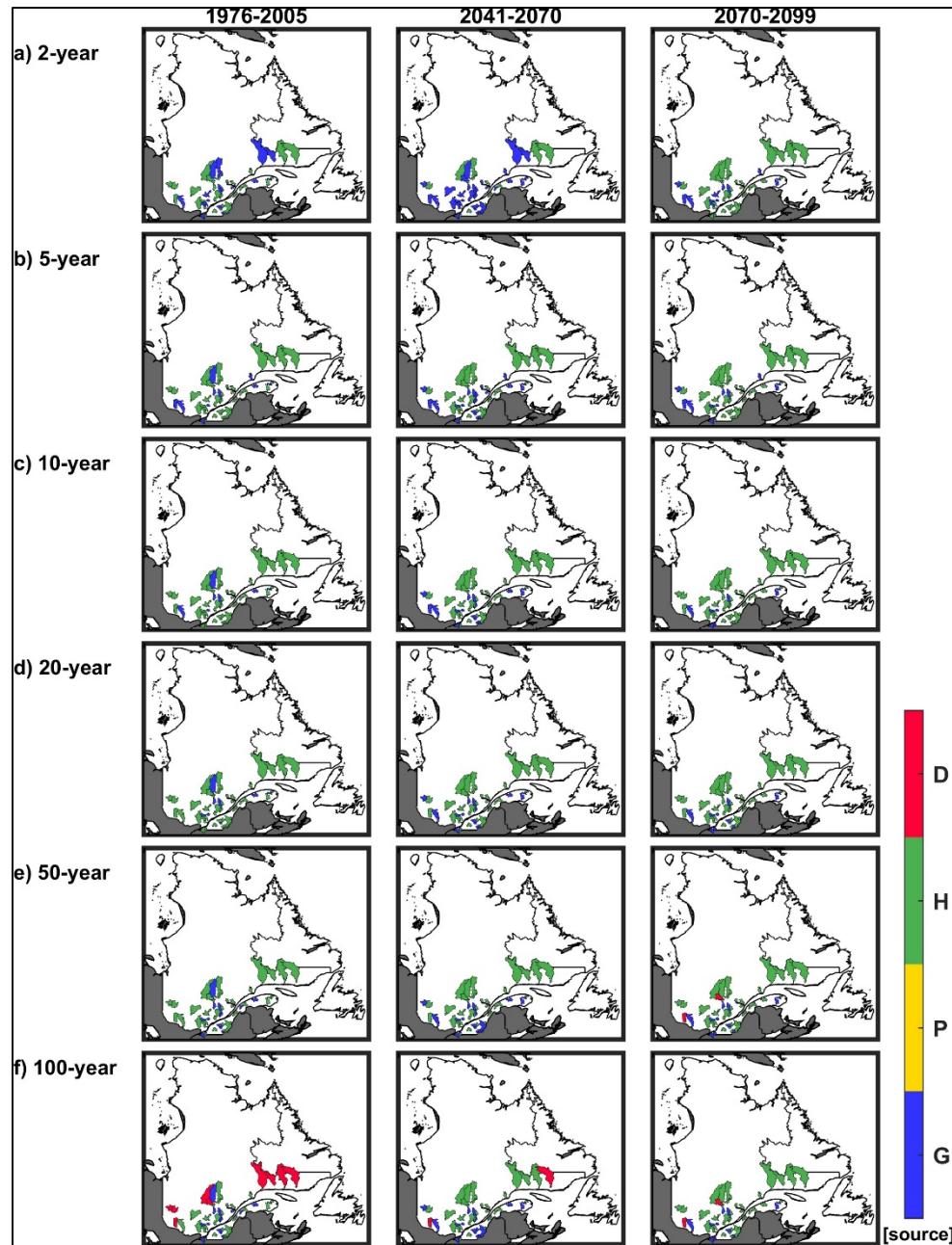
Figure S3.14 Maps of the major uncertainty contributor between GCM-
ensemble (G, in blue), post-processing methods (P, in yellow),
hydrological models (H, in green) and probability distributions of flood
frequency analyses (D, in red) over the snow-dominated basins. Results
for the 1976-2005 (left panels), 2041-2070 (middle panels) and 2070-
2099 (right panels) periods are presented. In row, panels a), b), c), d), e)
and f) present the results for the 2-, 5-, 10-, 20-, 50-, and 100-year return
periods, respectively
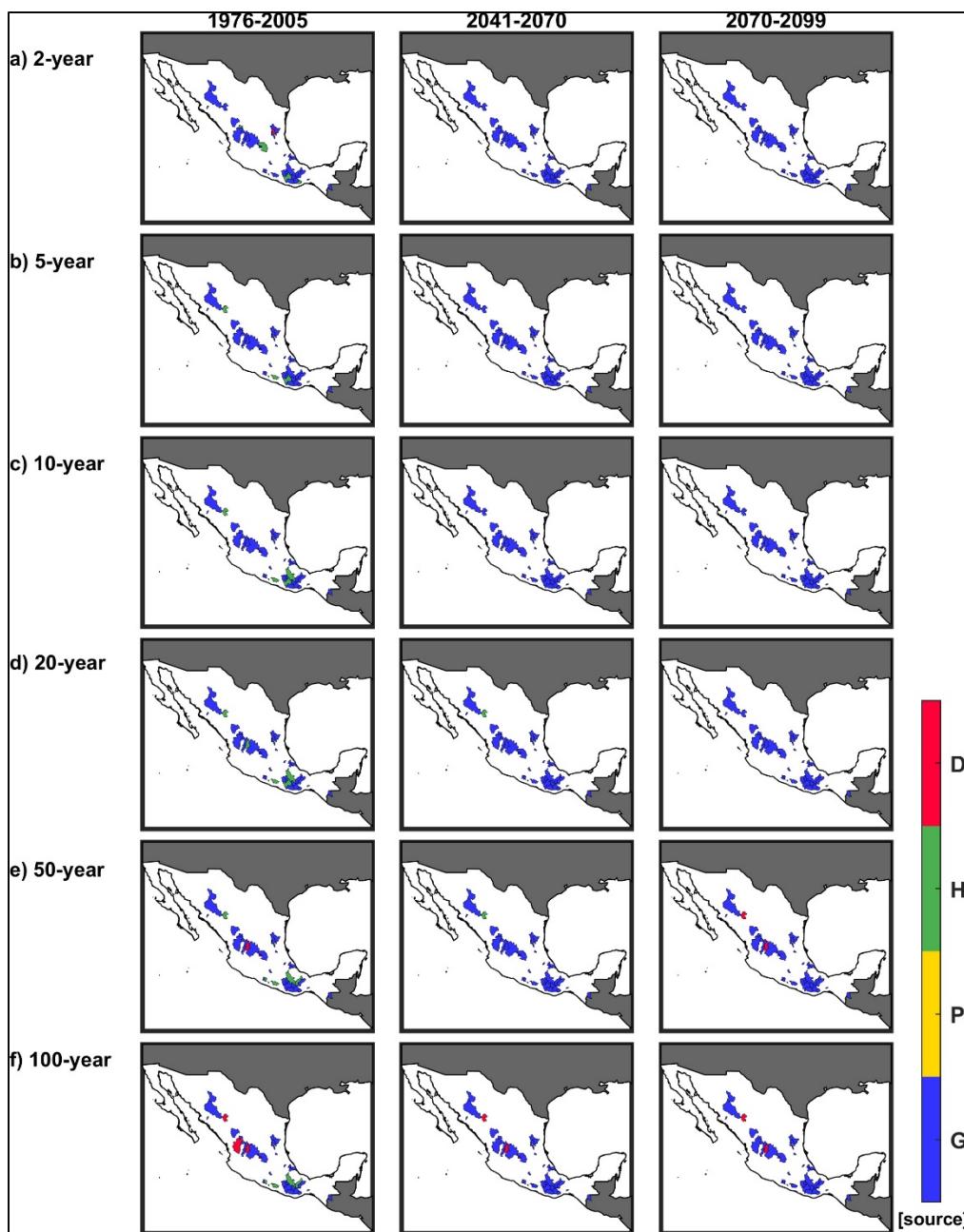
Figure S3.15 Maps of the major uncertainty contributor between GCM-ensemble (G, in blue), post-processing methods (P, in yellow), hydrological models (H, in green) and probability distributions of flood frequency analyses (D, in red) over the rain-dominated basins. Results for the 1976-2005 (left panels), 2041-2070 (middle panels) and 2070-2099 (right panels) periods are presented. In row, panels a), b), c), d), e) and f) present the results for the 2-, 5-, 10-, 20-, 50-, and 100-year return periods, respectively

CHAPTER 4

STRESS-TESTING HYDROLOGICAL MODELS FOR CLIMATE CHANGE IMPACT STUDIES

Mariana Castaneda-Gonzalez[a], Annie Poulin[b], Rabindranarth Romero-Lopez[c] and Richard Turcotte[d]

[a,b] Department of Construction Engineering, École de technologie supérieure, Montréal, Canada.
[c] Faculty of Civil Engineering, Universidad Veracruzana, Xalapa, Mexico.
[d] Ministère du développement durable, Environnement et Lutte contre les changements climatiques, Québec, Canada.

Abstract

The diminished hydrological model's performance over conditions that are different to the ones used during their calibration, raises questions on their reliability for climate change impact studies. Thus, this large-scale study evaluates the sensitivity of five lumped hydrological models to calibration data's climatic conditions and lengths over 77 basins with contrasting hydroclimatic conditions. Using five hydrological models and twenty calibration strategies, this large-scale stress-test assessed the effects of the hydrological models' calibration strategy on their validation performance and their simulated streamflow projections using two regional climate simulations. A variance decomposition analysis revealed that hydrological model structures and the calibration data's climatic conditions presented higher impacts on the hydrological models' validation performance than the calibration data's length. Moreover, the results showed that hydrological model structures had the largest impacts on

mean annual projected streamflows and mean annual peak projected streamflows, highlighting the importance of considering multi-model approaches.

Keywords: Hydrological model stress-test, hydrological model calibration, hydrological model performance, climate change impacts

4.1 Introduction

Climate change impacts on hydrology are usually investigated by coupling climate models' outputs and hydrological models to produce different hydrological scenarios. In current practices, hydrological models are first calibrated and validated to find the adequate parameters depending on each region's particular properties (Moore & Doherty, 2005; Pechlivanidis et al., 2011). This process usually consists in a split-sample test (SST), where historic data is divided into two segments, one for calibration and the other one for validation. The calibrated/validated hydrological models are then combined with climate models' outputs to simulate different climate change scenarios. Numerous studies applying this methodology can be identified in the literature (Arnell & Lloyd-Hughes, 2014; Hagemann et al., 2013; Madsen, Lawrence, Lang, Martinkova, & Kjeldsen, 2014; Roudier et al., 2016; Teutschbein et al., 2015; Wilby & Harris, 2006; Zhang, You, Chen, & Ge, 2016). It is thus observed that this approach relies on the critical assumption that hydrological model parameters identified for historic hydroclimatic conditions will remain similar and valid for the different possible future conditions (KlemeŠ, 1986). In other words, this practice assumes that hydrological model's parameterization remains stationary in time. However, the observed non-stationary behaviour in climate and hydrological systems has underlined the need to reconsider the reliability of hydrological models under a changing climate (Bérubé et al., 2022; Coron, 2013; Coron et al., 2012; KlemeŠ, 1986; Krysanova et al., 2018; Moore & Doherty, 2005; Thirel, Andréassian, & Perrin, 2015).

The ability of hydrological models to satisfactorily perform in different conditions than the ones used for calibration, often referred as "predictive capability", "parameter transferability"

or "hydrological models robustness", has gained more attention (Dakhlaoui, Ruelland, & Tramblay, 2019; Dakhlaoui, Ruelland, Tramblay, & Bargaoui, 2017; Krysanova et al., 2020; Motavita, Chow, Guthke, & Nowak, 2019; Guillaume Thirel et al., 2015; G. Thirel et al., 2015; Vormoor, Heistermann, Bronstert, & Lawrence, 2018). Different approaches have been used to evaluate the robustness of hydrological models before their use for climate change impact studies. Among them, the differential split-sample test (DSST) proposed by KlemeŠ (1986) has been identified as an option to assess hydrological models reliability under contrasting hydroclimatic conditions (Dakhlaoui et al., 2019; Dakhlaoui et al., 2017; Motavita et al., 2019; Seiller et al., 2012; Vormoor et al., 2018). The DSST approach consists of selecting calibration and validation periods based on a selected climatic and/or hydrological classification (e.g., warm or dry years) to ensure contrasting conditions between both periods (KlemeŠ, 1986; Seibert, 2003). The rationale behind this approach is thus to evaluate hydrological models' dependency to the hydroclimatic conditions of the calibration period (Seiller et al., 2012). For instance, Seiller et al. (2012) evaluated the robustness of twenty lumped conceptual hydrological models under contrasting conditions over two basins using a DSST approach. Four climatic conditions of were selected to identify contrasting 5-year periods, dry/warm, dry/cold, humid/warm and humid/cold based on annual mean precipitation and temperature. Results showed that hydrological models were generally unreliable when performing simulations under conditions that strongly differed from the ones used in calibration. Yet, hydrological models showed better performances in validation periods when calibrated on dry conditions than on humid conditions. Similarly, extended work presented in Seiller, Hajji, et Anctil (2015) showed that calibrating on dry/cold periods with the twenty hydrological model ensemble produced more robust results when validating in contrasting conditions (i.e. humid/warm) over twenty U.S. watersheds. Coron et al. (2012) carried out a study to evaluate the extrapolation capacity of three conceptual hydrological models (GR4J, MORDOR6 & SIMHYD) over 216 rainfall-dominated Australian catchments. They proposed a new methodology called Generalized Split Sample Test (GSST) that consists in multiple calibration/validation combinations of equal length (5-years) considering both similar and contrasting conditions using a sliding time window. Results showed that hydrological models' simulation error increased with increasing mean rainfall difference between calibration and

validation periods. However, hydrological models' transferability was generally better when calibrating on dry conditions and validating on wet conditions than vice versa. Additionally, it was observed that mean runoff was overestimated when calibrating on wetter years and underestimated when calibrated on dryer years. Dakhlaoui et al. (2017) evaluated the robustness of three conceptual hydrological models (GR4J, HBV & IHACRES) over five semi-arid catchments in Tunisia. In line with (Coron et al., 2012) the study showed that hydrological models lowered their performance when calibrated over wetter/colder years to simulate dryer/warmer conditions than the opposite (i.e. dry/warm to wet/cold). Based on contrasting flooding conditions, Vormoor et al. (2018) evaluated the parameter transferability of the HBV hydrological model over five Norwegian catchments. The results suggested that hydrological modelling performance decreased from 5 to 17% over the different catchments. More recently, Dakhlaoui et al. (2019) proposed a modified DSST named General Differential Split-Sample Test (GDSST). The proposed GDSST was compared with existing approaches using three hydrological models (GR4J, HBV & IHACRES) over five Tunisian basins. Results indicated that changes in mean annual temperature from -2 to 2 °C and changes in mean annual precipitation temperature from -30% to 80% were found as the limits of transferability of these three hydrological models. These limits of transferability were compared with Regional Climate Model (RCM) simulations showing that RCMs projections under the Radiative Concentration Pathway (RCP) scenario 8.5 surpassed the limits of transferability of all hydrological models. Moreover, runoff volume differences from 5% to 20% were observed in climate-change projections when calibrating on the full-time series with respect to when calibrating on smaller periods with similar climatic conditions to the ones observed in RCMs outputs.

In addition to analysing hydrological model`s sensitivities to calibration conditions, other studies have evaluated the effects of calibration data`s length on hydrological model`s robustness under climate variability. For instance, Perrin et al. (2007) assessed the sensitivity of two conceptual hydrological models (GR4J & TOPMO) to different calibration period`s lengths. Results showed that the most parsimonious hydrological model needed less calibration data to obtain robust parameter sets. Moreover, results showed that using 350 days including

dry and wet conditions were enough to estimate robust hydrological model parameters. Vaze et al. (2010) studied the validity of calibrated hydrological models in climate change studies. Four conceptual hydrological models (SIMHYD, Sacramento, SMARG & IHACRES) were calibrated using dry and wet periods of different lengths (10, 20, 30 and 40 years) over 61 Australian basins. Results indicated that hydrological model calibration using more than 20 years were robust when changes in mean annual precipitation were less than 15% drier or 20% wetter than the mean annual precipitation used in calibration. Additionally, hydrological models calibrated on wet periods showed more difficulties to predict dry periods than calibrating on dry periods and predicting wet periods. Thus, they recommended calibrating on a sub-period with conditions that are similar to the conditions of future projections. Li et al. (2010) investigated the impact of calibration periods length on hydrological model performance on ungauged or data-limited basins. The SIMHYD hydrological model was tested using random datasets with lengths varying from 1 to 10 years over 55 Australian basins. Results indicated that with 8-years calibration periods, SIMHYD showed steady results and that longer periods did not systematically improve the hydrological model performance. More recently, Arsenault et al. (2018) investigated the impact of calibration periods' length on hydrological models' performance using two hydrological models over three basins. The hydrological models were calibrated using multiple periods with varied length (from 1 to 16 years) and validated on eight independent years to assess their performance. In general, results suggested that increasing calibration period's length increased hydrological models' performance. Thus, calibrating hydrological models over the full-time series was suggested as the best strategy. Motavita et al. (2019) presented a new methodology called Comprehensive Differential Split Sample Test (CDSST) to investigate the effects of climatic conditions and dataset length of hydrological models' calibration/validation on their predictive performance. Wet, dry and mixed period and varied data lengths of 2, 4, 8, 15, 20 and 25 years were used. The HBV hydrological model was calibrated and validated on datasets with different climatic conditions and lengths over one snowmelt-dominated basin in Germany. Overall results suggested that climatic conditions had stronger impact on the HBV model predictive performance than datasets length. In line with previous studies, calibrating over dry periods was suggested as the most robust option to simulate wet, dry or mixed conditions of any length

(Coron et al., 2012; Dakhlaoui et al., 2017; Seiller et al., 2015). Moreover, the results suggested that using calibration datasets longer than 8-years might lead to overconfident simulations failing to cover the variance in future data.

Most studies presented previously, agreed that hydrological models calibrated on certain climatic conditions diminished their performance when used under strongly different climatic and/or hydrological conditions, especially in terms of precipitation (Coron et al., 2012; Dakhlaoui et al., 2019; Motavita et al., 2019; Seiller et al., 2012; Seiller et al., 2015). However, contrasting results were observed regarding calibration data length. Some studies suggest using long calibration periods (i.e., Arsenault et al., 2018; Guo, Johnson, & Marshall, 2018; Shen, Tolson, & Mai, 2022; Vaze et al., 2010) while others suggest that shorter representative periods (i.e., 7 or 8 years) might be more adequate. Additionally, few studies have extended their analyses to evaluate the effects of hydrological models' calibration on hydrological projections. These observed gaps lead us to highlight some research questions. Are hydrological models more sensitive to calibration data condition or data length? Is this sensitivity transferred to hydrological projections? What are the effects over basins with contrasting hydroclimatic regimes? Thus, to investigate these questions and provide further insights on hydrological models and their calibrations as well as their implications on hydrological projections, this paper aims to expand on the existing research (Dakhlaoui et al., 2019; Motavita et al., 2019; Seiller et al., 2012; Guillaume Thirel et al., 2015; Vormoor et al., 2018) by evaluating the sensitivity of different hydrological models performance to calibration data conditions and lengths over a large group of basins with contrasting hydroclimatic conditions. Moreover, this large-scale stress-test aims at assessing the effects of hydrological models' calibration on hydroclimatic projections.

4.2   Study area and data

To cover a variety of hydroclimatic conditions, a total of 77 basins were selected and grouped into two clusters of basins with contrasting hydrological regimes, namely a snow-dominated group of basins located in Quebec, Canada (panel a) and a rain-dominated group (panel b) of

basins located in Mexico. The location of all basins as well as their mean total annual precipitation and mean annual temperature are shown in Figure 4.1.
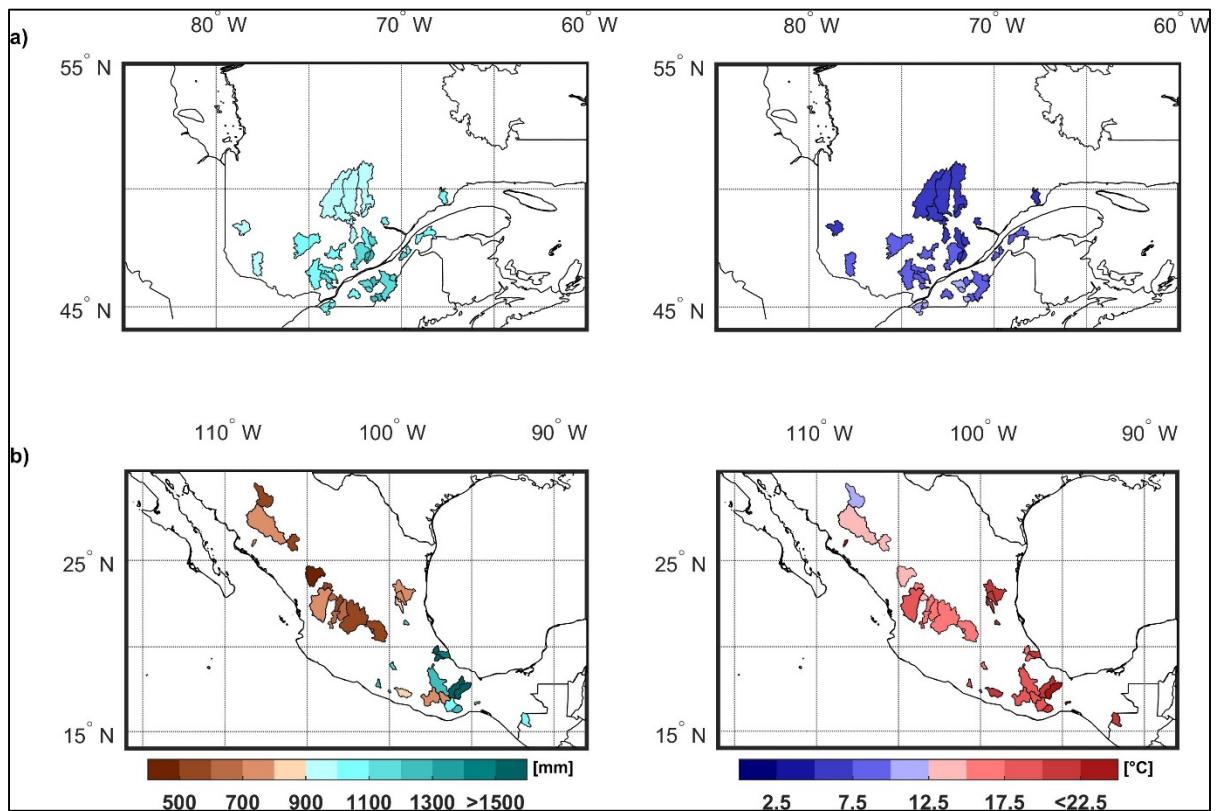
Figure 4.1 Location of the 38 basins in Quebec (upper panels) and the 39 basins in Mexico (lower panels) used in the study. The upper and lower panels show the mean total annual precipitation (mm) and mean annual temperature (°C) for the snow-dominated and rainfall-dominated basins, respectively

Overall, the figure shows that both average annual precipitation and temperature vary strongly between the two basin groups. The 38 snow-dominated basins located in the Canadian province of Quebec show that all snow-dominated basins receive on average more than 900 mm of precipitation, while among the 39 rain-dominated basins located in the country of Mexico some catchments receive, on average, less than 500 mm and others more than 1500 mm. In terms of mean annual temperature, great contrasts are observed between both regions. Over the snow-dominated basins, it is observed that all basins have mean annual temperatures below 12.5 °C with a gradient that shows increasing temperatures towards the southern basins. In contrast,

the rain-dominated basins show higher mean annual temperatures reaching values of more than 22.5 °C with a northwest-southeast gradient showing higher temperatures towards the southeastern basins.

The daily climatic and hydrometric historical records of each basin were issued from the Hydrometeorological Sandbox –École de technologie supérieure (HYSETS) database (Arsenault, Brissette, Martel, et al., 2020), a large-scale database of North American basins. The 77 basins used in this study were selected from this database based on hydrometric and climatic data availability. Minimum lengths of 30 years between 1950 and 2013 for the daily minimum and maximum temperatures, precipitation and streamflow were covered for each basin.

4.3   Methods

The methods applied for each basin consisted in three main steps as presented in Figure 4.2 First, five lumped hydrological models with different structures and levels of complexity were selected. Then, different calibration strategies with varying climatic conditions and lengths were performed and validated on climatically contrasting 8-year periods. Finally, all hydrological models' parametrizations were used jointly with regional climate simulations issued from the Canadian Regional Climate Model version 5 (CRCM5) to obtain streamflow projections for a reference and two future periods.

4.3.1   Hydrological models

To evaluate the effects of hydrological model`s calibration data climatic condition and length on their performance, five hydrological models with varying structures and complexity were used. Table 4.1 shows an overview of the five models.
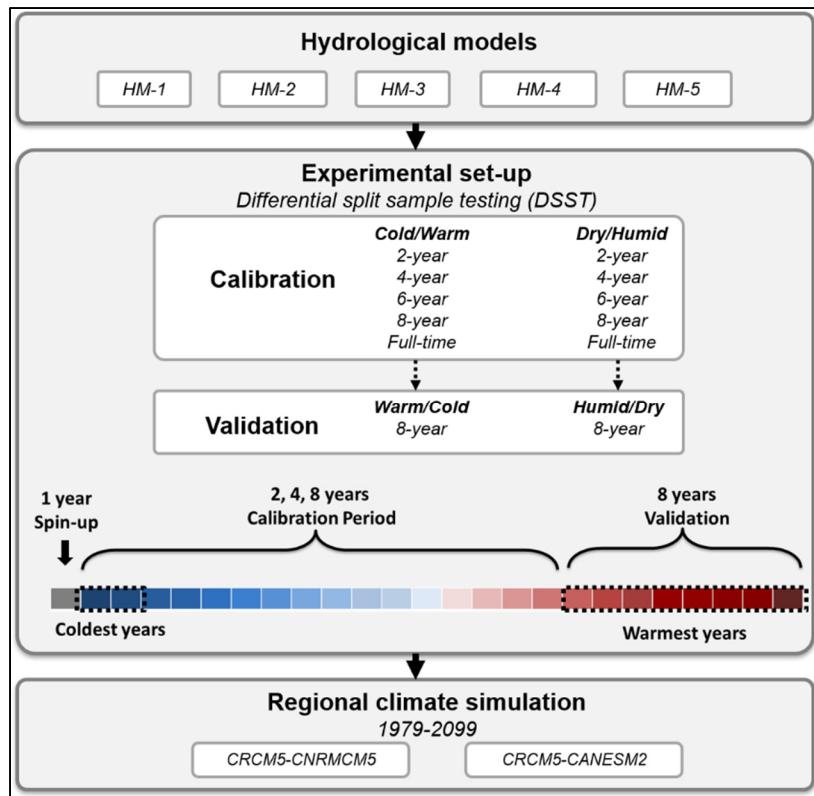
Figure 4.2 Overview of the methodology used for each basin

Table 4.1 Overview of the five hydrological models used in this study

ID	Hydrological model	Number of parameters	Reference
HM-1	GR4J	6	(Perrin et al., 2003)
HM-2	MOHYSE	10	(Fortin & Turcotte, 2006)
HM-3	HMETS	21	(Martel et al., 2017)
HM-4	CEQUEAU	11	(Girard et al., 1972)
HM-5	IHACRES	9	(Jakeman et al., 1990)

4.3.1.1 GR4J

The GR4J model (modèle du Génie Rural à 4 paramètres Journaliers meaning, daily rural engineering model with four parameters) is a simple lumped rainfall-runoff model developed by Perrin et al. (2003). This parsimonious model was initially created for water management

applications with 4 parameters to calibrate. Its structure is composed of three reservoirs that account for water production and routing. However, this model does not include a snow simulation module to represent snow accumulation and melting process. Thus, the snow module CemaNeige (Valéry et al., 2014) was added to allow applications in the 38 snow-dominated basins, adding 2 parameters that make a total of 6 parameters to calibrate. Additionally, evapotranspiration is not directly estimated within the GR4J-CEMANEIGE hydrological model, so this should be previously calculated and provided as input for the hydrological model. For this study, the Oudin potential evapotranspiration formulation (Oudin et al., 2005) was used based on recent successful applications over North American catchments (Arsenault et al., 2018; Chen et al., 2018; Troin et al., 2018; Troin, Poulin, Baraer, & Brissette, 2016). In addition to the previously calculated potential evapotranspiration, the GR4J-CEMANEIGE hydrological model requires continuous series of daily precipitation and mean temperature for its application.

4.3.1.2   MOHYSE

The MOHYSE model (Modèle Hydrologique Simplifié à l'Extrême meaning, extremely simplified hydrological model) is a simple lumped conceptual model created by Fortin et Turcotte (2006) and initially used for educational purposes. MOHYSE simulates the main hydrological processes using 10 different parameters and can be currently used for different time scales such as daily and sub-daily simulations. Its structure uses 3 reservoirs to represent the vertical water budget and routing processes. And it includes its own potential evapotranspiration calculation and snow accumulation and melting representations. This hydrological model requires continuous time series of mean daily temperature, total daily rainfall and snowfall as inputs. In addition, this conceptual model has been applied over various regions with different hydrological regimes. Some examples of recent effective MOHYSE applications covering tropical and snow-dominated regions are presented by Arsenault et al. (2019), Thiboult et al. (2016) and Troin et al. (2018)

4.3.1.3   HMETS

The HMETS (Hydrological Model – École de technologie supérieure) developed by Martel et al. (2017) is a simple lumped conceptual model with 21 parameters to calibrate. This hydrological model accounts for the main hydrological processes including surface flow, delayed flow, hypodermic flow and groundwater flow. Its structure is based on two reservoirs used to simulate the vadose and phreatic zones, and the snowmelt is simulated by the melting and refreezing process of the snowpack using the degree-day model developed by Vehviläinen (1992). To ensure a larger diversity of hydrological model structures, in terms of potential evapotranspiration the McGuinness and Bordne formulation was used (McGuinness & Bordne, 1972). The HMETS-McGuiness-Bordne has been effectively used in a recent study covering a large variety of North American basins (Castaneda-Gonzalez, Poulin, Romero-Lopez, Turcotte, & Chaumont, 2021). The required inputs for this hydrological model consist of continuous time series of daily precipitation, maximum and minimum temperature, as well as the previously calculated potential evapotranspiration.

4.3.1.4   CEQUEAU

The CEQUEAU (previously named as such at the *INRS-Eau-Institut National de la Recherche Scientifique*) hydrological model used in this study is a lumped version of the conceptual semi-distributed hydrological model initially developed by Girard et al. (1972). This lumped hydrological model with 9 parameters to calibrate requires as input daily minimum and maximum temperature, as well as daily solid and liquid precipitation time series. Its structure consists in quantifying the vertical water balance through two interconnected reservoirs and the downstream routing. The different mass transfers, such as evapotranspiration, water in unsaturated and saturated zones are accounted within the hydrological water balance. To simulate snowpack formation and melting, this lumped version was coupled with the snow module CemaNeige (Valéry et al., 2014), as used in different studies such as Velázquez, Anctil, et Perrin (2010) and Seiller et al. (2012), making a total of 11 parameters to calibrate.

4.3.1.5 IHACRES

The IHACRES (Identification of unit Hydrographs And Component flows from Rainfall, Evaporation and Streamflow data) hydrological model is a lumped model initially developed by Jakeman et al. (1990). This model consists of a non-linear production module and a linear transfer module with 7 parameters to calibrate. These modules account for the generation of effective rainfall that is then converted into quick and slow flow through unit hydrographs. The snow module CemaNeige (Valéry et al., 2014) was added to account for the snow melting and accumulation process, making a total of 9 parameters to calibrate. This model requires continuous series of daily precipitation and mean temperature. Additionally, the evapotranspiration is not directly estimated within the model, so it was previously calculated using the Hargreaves and Samini evapotranspiration formulation (Hargreaves & Samani, 1985). As previously described for the other hydrological models used in this study, similar versions of this model have been used in other studies endorsing its use over the studied basins (Perrin, Michel, & Andréassian, 2001; Seiller et al., 2012; Velázquez et al., 2010).

4.3.2 Experimental set-up: stress-testing

As described in the introduction, hydrological models' performance has shown sensitivity to changes in calibration data's climatic conditions and lengths. Thus, to assess the impacts of these two variables, this study`s methodology uses a Differential Split Sample Testing (DSST) jointly with various calibration's data series lengths as proposed in previous works (Motavita et al., 2019; Vaze et al., 2010). The idea behind this process is to stress the hydrological models through different calibration-validation strategies (hereafter called "calibration strategies", to simplify) to then evaluate their impacts on hydrological model's performance and their streamflow projections. For each of the 77 basins the following steps were applied:

1. Yearly total precipitation and average temperature were calculated to identify precipitation- and temperature-based yearly climatic conditions. From the yearly total precipitation and average temperature series all years were ranked from the driest/most

humid to the most humid/driest, and from the warmest/coldest to the coldest/warmest, respectively (as an example, see colorbars in Figure 4.2).

2. From the ranked dry (D), humid (H), cold (C) and warm (W) years, series of 2-, 4-, 6- and 8-years were used to calibrate each hydrological model. When the warm or cold years were considered, the corresponding precipitation, in those years, was used. A final calibration including all the years (except the validation years, see point (3) hereafter) was also performed. This consists of 21-year calibration data timeseries. It is important to highlight that the first year was not considered to calculate the objective function, as it is considered the warm-up period to obtain realistic internal variables (Arsenault et al., 2018).

3. All the different calibration strategies were then validated on climatically contrasting 8-year periods. For instance, as schematically presented in Figure 4.2, all the coldest 2- , 4-, 6-, 8-year and full-time calibrations (i.e., C-2, C-4, C-6, C-8 and Full) were validated on the same 8 contrasting warmest years. Thus, the five different calibration data's lengths and four contrasting climatic conditions, make a total of 20 different calibration strategies per hydrological model and per basin.

The calibration of the hydrological models was performed using two optimization algorithms: (1) the Covariance Matrix Adaptation Evolution Strategy (CMAES) was used to calibrate GR4J and MOHYSE based on recommendations from previous studies (Arsenault et al., 2014; Hansen & Ostermeier, 1997), and HMETS, CEQUEAU and IHACRES were calibrated using the Dynamically Dimensioned Search algorithm (Asadzadeh & Tolson, 2009; Huot et al., 2019). The objective function used for all calibration/validation strategies was the modified Kling-Gupta Efficiency criterion (KGE; Kling et al., 2012) between observed and simulated streamflows, and is defined as follows:

$$KGE = 1 - \sqrt{(r - 1)^2 + (\beta - 1)^2 + (\gamma - 1)^2} \tag{4.1}$$

$$\beta = \frac{\mu_s}{\mu_o} \tag{4.2}$$

$$\gamma = \frac{CV_s}{CV_o} = \frac{\sigma_s/\mu_s}{\sigma_o/\mu_o} \qquad (4.3)$$

The variables r, β and γ represent the correlation coefficient, bias ratio and variability ratio, respectively. CV denotes the coefficient of variation, μ denotes the mean, and σ represents the standard deviation of the streamflow data. For each of these variables, the "o" and "s" subscripts indicate *observed* and *simulated* streamflow time series, respectively. The ranges of the KGE evaluation metric varies from -∞ to 1, where 1 represents a perfect fit. A recent study presented by Knoben et al. (2019), showed that KGE-values larger than ≈ -0.41 indicate higher skill from the simulated series than using the observations mean, thus KGE-values higher than 0.5 are considered acceptable for this study.

4.3.3   Regional climate simulations

The regional climate model simulations used in this study were issued from the Canadian Regional Climate Model version 5 (CRCM5; Martynov et al., 2013; Separovic et al., 2013). Daily precipitation, minimum and maximum temperatures from a CRCM5 simulations driven by CNRM-CM5 global climate (GCM) model and driven by CanESM2 (referred to as CRCM5-CNRM-CM5 and CRCM5-CANESM, respectively) were used to feed the five hydrological models to obtain the RCM-driven streamflow simulations. These CRCM5 simulations were produced and provided by the Ouranos Consortium on Regional Climatology and Adaptation and is part of the Coordinated Regional Downscaling Experiment (CORDEX) database (Giorgi & Gutowski, 2015). To cover our study area, the Quebec domain (QC) and a modified Central American domain (CAM) that extends the northern border to cover Mexico were used. The CRCM5 simulations consisted of 0.22° spatial resolution simulation covering the reference period of 1976-2005 and two future horizons of 2041-2070 and 2070-2099 under the Representative Concentration Pathway (RCP) 8.5. All climate outputs were bias-corrected using the quantile mapping method (Themeßl et al. 2011, Themeßl et al. 2012, Maraun 2016). This well-known quantile-based method has shown a robust performance that justifies its use in this study (Chen et al. 2013a, Chen et al. 2013b). These four bias-corrected CRCM5

simulations (two driving GCMs, two domains) were thus combined with the five lumped hydrological models and their 20 different parameterizations (i.e., 5 different lengths x 4 different climatic conditions = 20 parameterizations), making a total of 200 streamflow simulations per basin (i.e., 5 hydrological models x 20 parameterizations x 2 climate simulations = 200 streamflow simulations).

4.3.4 Data analysis

4.3.4.1 Variance decomposition

Different calibration strategies with varying data's climatic conditions and lengths were used to test the sensitivity of hydrological models to these variables. To quantify the variance brought by the calibration data's climatic condition, the calibration's data length and the hydrological model structure on validation performance and streamflow projections, a variance decomposition analysis was performed (Déqué et al., 2007; Troin et al., 2018). This method isolates and intercompares the contribution of each of these three elements, namely hydrological model structures (H), calibration data's climatic conditions (C), and calibration data's length (L) and the interactions between them. For instance, the total variance associated to the calibration data's climatic conditions (V(C)), can be expressed as follows:

$$V(C) = C + CL + CH + CLH \qquad (4.4)$$

where C, L, and H represent the variance associated with calibration data's climatic condition (C), calibration data's length (L) and hydrological model structures, respectively.

Two main variance assessments were performed for the calibration-validation experiments, one over validation performance obtained using the historical calibration-validation strategies and the second to decompose the variance in projected streamflows. The first analysis was carried out using the validation KGE-values obtained from the different calibration strategies and hydrological models. The second analysis was performed using the mean annual and the

mean annual peak RCM-driven streamflows obtained from the different calibration strategies and hydrological models for the reference and future periods (i.e., 1976-2005, 2041-2070 and 2070-2099). The results will show the total variance contributions as described in equation 4.4. Yet, the variance contribution of each element and their interactions are presented in the second appendix.

4.3.4.2 Streamflow projections analysis

To investigate the effects of hydrological models and their different calibration strategies on streamflow projections, the mean annual streamflows obtained with the different parameterizations of each hydrological model were compared. The calibration performed over the full-time calibration series for each of the four different calibration data's climatic conditions was used as reference. The idea behind this comparison is to quantify the differences between the different calibration data lengths of each climatic condition against the commonly used method of relying on the parameters obtained during calibration. The first comparison consists of measuring the impact on mean annual streamflow magnitudes over the reference and future periods of 1976-2005, 2041-2070 and 2070-2099. This comparison is estimated in terms of relative difference (%) between the mean annual streamflow and mean annual peak streamflow estimated from the 2-year, 4-year, 6-year and 8-year calibration against the mean annual and mean annual peak streamflow estimated using the full-time calibration over all used as reference for each climatic condition and each period. The second comparison consists of comparing the climate change signals of the different hydrological models and their calibration strategies against the climate change signal of the streamflow projection obtained from full-time calibration of each climatic condition. First each climate change signal is estimated by comparing the mean annual streamflow of a given calibration strategy during a future period against its mean annual streamflow during the reference period in terms of relative difference (%). Then, the difference between climate change signals (%) is calculated using the climate change signal of the full-time calibration as a reference.

4.4   Results

4.4.1   Effects of calibration data's climatic conditions and length

To evaluate the impact of the different calibration strategies on hydrological models' validation performance, Figures 4.3 and 4.4 show the KGE validation scores for the 38 snow-dominated basins and the 39 rain-dominated basins, respectively. From top to bottom, each figure shows the results for each hydrological model with each panel displaying the boxplots of KGE-values for the four contrasting climatic conditions (Warm to Cold-WC, Cold to Warm-CW, Dry to Humid-DH, Humid to Dry-HD) and the five data lengths (2y, 4y, 6y, 8y, and Full). The results over the snow-dominated basins (Figure 4.3) show that the validation performance varies between hydrological models, climatic conditions, and data lengths. It is observed that some hydrological models show consistently higher KGE-values. This can be observed when looking at the validation results of hydrological model HM4 where median KGE values are consistently larger than or equal to 0.8 over all different calibration strategies while others as HM5 show more KGE values below 0.8. Regarding the different calibration data lengths, it is observed that KGE-scores increase with increasing length. This behaviour is observed over the four contrasting climate conditions. It is however observed that most hydrological models show generally larger sensitivity when calibrating on humid years and validating on dry years (HD, last column) as lower KGE-values are observed compared to the other calibration strategies. Calibrating on the full length does not necessarily lead to better results in terms of KGE values' median and spread (see e.g. HM4 in Figure 4.5). Figure 4.4 shows that the rain-dominated basins generally have lower validation KGE-values than the snow-dominated ones with median values of about 0.7 or less in the different calibration strategies. Yet, as observed over the snow-dominated basins the validation KGE-values also vary between hydrological model, calibration data climatic condition and length. First, as seen in Figure 4.3, the validation KGE-values increase with increasing calibration period length. These increases are especially observed when increasing the length from 2 to 4 years, and 4 to 6 years. Regarding the climatic conditions, it is observed that most hydrological models show larger sensitivity when calibrating on humid years and validating on dry years (HD) with generally lower KGE-values

than the other strategies. This is especially observed over some hydrological models, such as HM4, that shows validations with median values below 0.5 for all the different lengths of the HD calibration strategy. In contrast, the DH calibration strategy often shows higher validation KGE-values than the other strategies in most hydrological models.
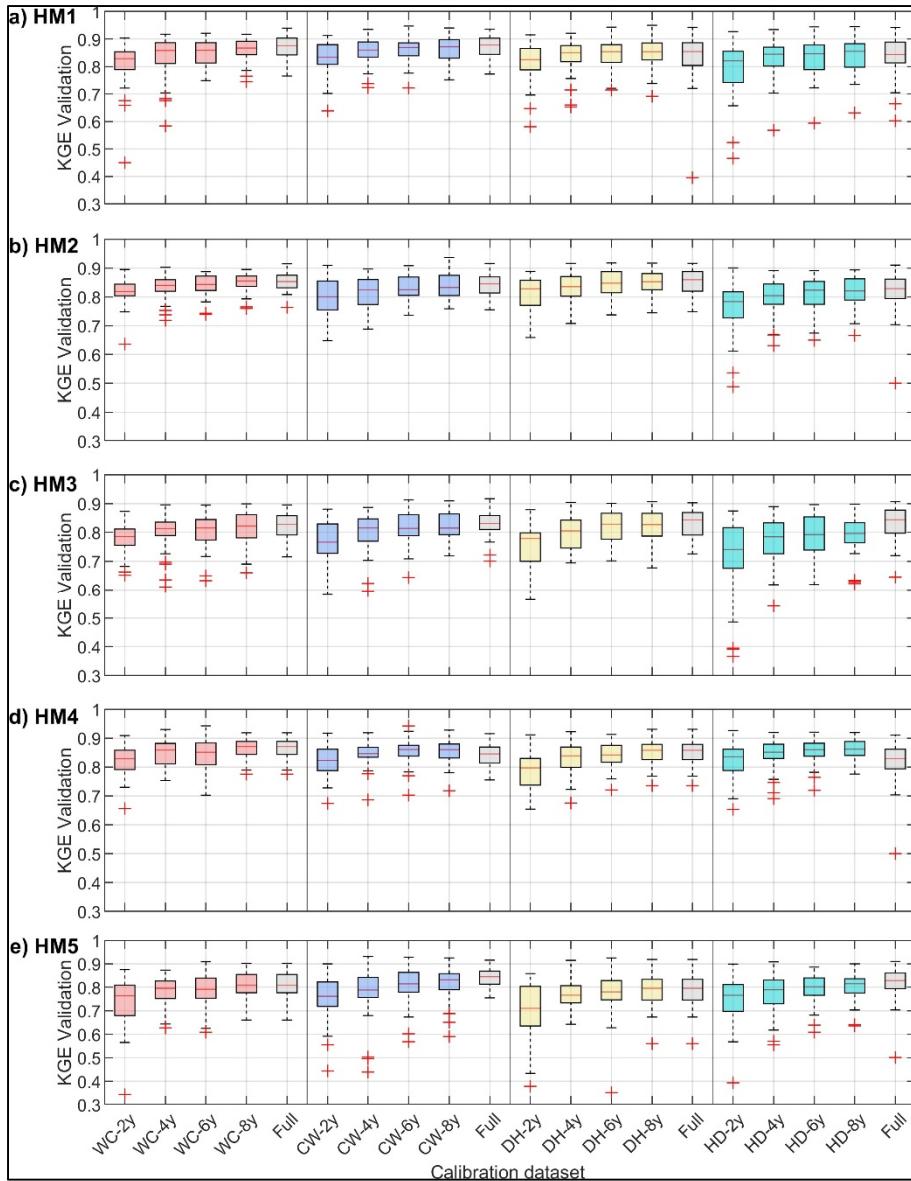
Figure 4.3 Boxplots of KGE-values obtained during validation over the snow-dominated basins. By row, the figure shows the results for each hydrological model
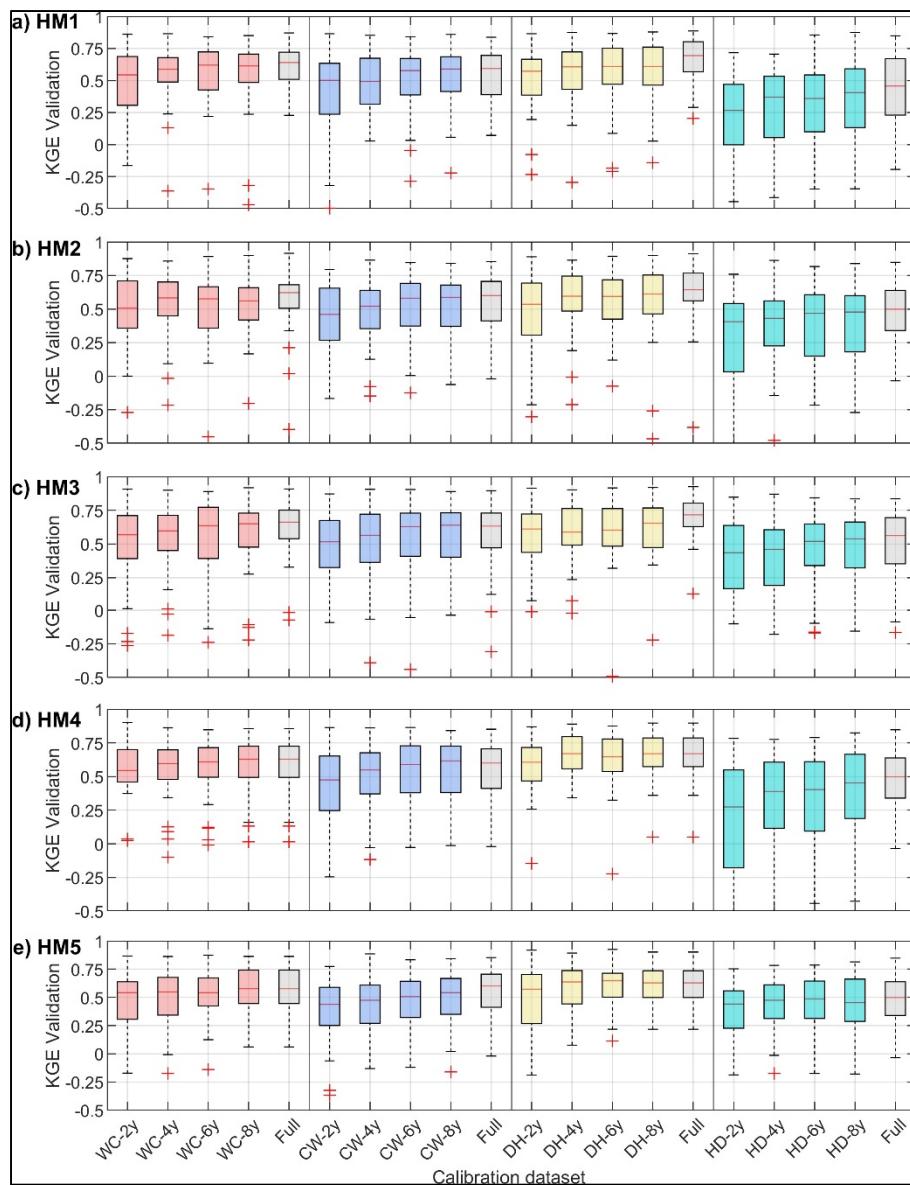
Figure 4.4 Boxplots of KGE-values obtained during validation over
the rain-dominated basins. By row, the figure shows the results for
each hydrological model

To quantify the impact of the calibration strategy on the validation performance, the difference between the calibration KGE-values obtained from each calibration strategy and the corresponding KGE-values obtained in validation are presented the snow-dominated and rain-dominated basin in Figures 5 and 6, respectively. Each figure shows the boxplots of the KGE-

difference for all calibration strategies, with negative values meaning lower performance in validation than during calibration.
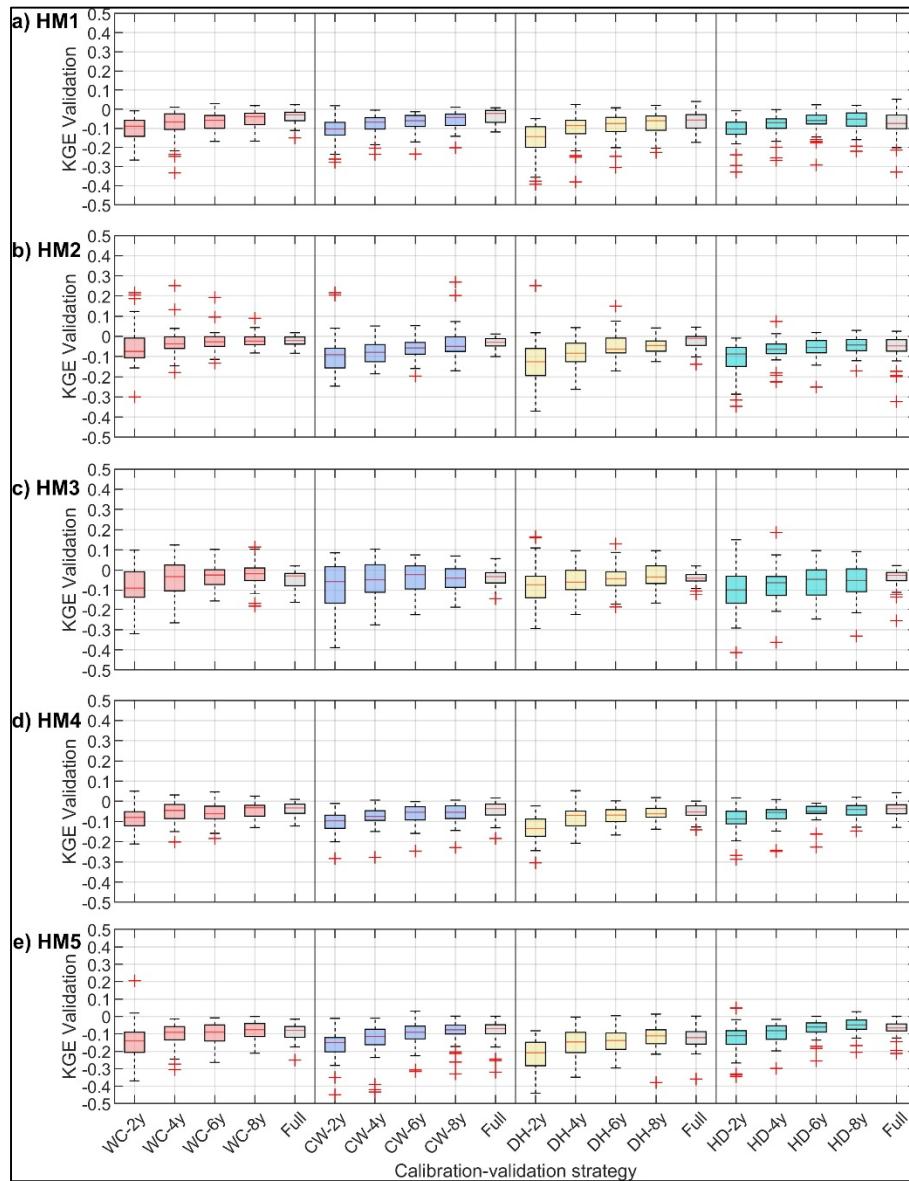
Figure 4.5 Boxplots of the difference between calibration and validation KGE-values over the snow-dominated basins. By row, the figure shows the results for each hydrological model
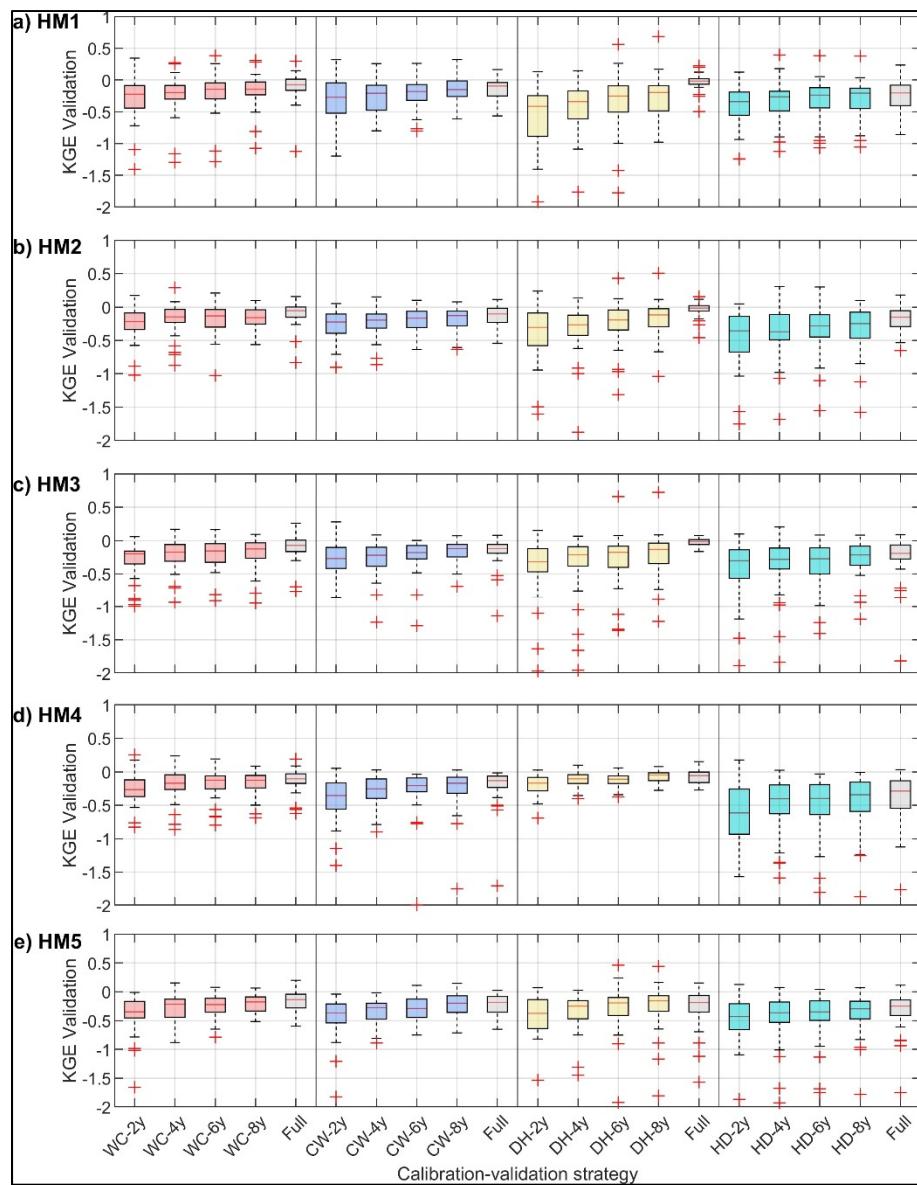
Figure 4.6 Boxplots of the difference between calibration and validation KGE-values over the rain-dominated basins. By row, the figure shows the results for each hydrological model

The results over the snow-dominated basins in Figure 4.5 show similar behaviour over the different calibration strategies, yet the magnitude varies between hydrological models. For instance, when looking at HM4 it is observed that the KGE difference between calibration and validation is generally lower than for other models for the different calibration strategies. Regarding the calibration data's climatic condition, it is observed that hydrological models are

more sensitive to some contrasted climatic conditions than others, particularly the precipitation-based contrasting conditions. This can be observed over all hydrological models with DH and HD showing generally larger spreads and sometimes larger differences between calibration and validation than the other strategies, particularly for the smallest calibration lengths. Nonetheless, some models also showed sensitivity to CW and WC conditions., as seen with HM3. In general, similar behaviour is observed regarding the calibration data's length among the four different climatic conditions used for calibration-validation. Yet, it is observed that the largest impacts on validation are mainly observed for the 2-year and 4-year lengths. For the other lengths, the medians are often similar to those for the calibration using the full-time calibration series.

In Figure 4.6, the different calibration strategies show generally larger impacts over the rain-dominated basins with medians of KGE differences down to -0.5 in some cases. When looking at the different calibration strategies in terms of climatic conditions, the results show that the DH and HD calibrations often show the largest impacts on hydrological models' performance. This can be observed over all hydrological models. Regarding the calibration data's lengths, as observed with the snow-dominated basins, the KGE differences between calibration and validation get closer to zero when increasing the length. Yet, it is observed that for some cases, such as the HD calibration, smaller impacts are generally observed when increasing length.

4.4.1.1 Variance decomposition: effects on validation performance

To decompose and quantify the impact of the different hydrological models and their calibration strategies on the validation performance (in terms of KGE values), a variance decomposition analysis was performed to evaluate the variance associated to hydrological model structures (H), calibration data's length (L) and calibration data's climatic condition (C). Figure 4.7 shows the variance contributions of these three sources on the validation performance of the snow-dominated basins (panel a) and the rain-dominated basins (panel b), with the left panel showing the percentage of total variance contributions (all interactions considered) and the left panel displaying the highest variance contributor for each basin.

Overall, the results show that the variance contributions vary between regions and basins, particularly over the snow-dominated basins where hydrological models show the largest median, meaning that hydrological models had the largest impacts on validation performance. The boxplots of total variance percentages show that calibration data's climatic conditions are the second main variance contributor over this basin group followed by the choice calibration data's length. This is also observed over the right panel showing the main variance contributor, where some basins show larger influence from climatic conditions and only two showed larger influence from calibration's length. In contrast, results over the rain-dominated basins (see panels b) consistently show that calibration data's climatic conditions bring more variance to hydrological model's validation performance. This is observed in most basins, with calibration's data length generally showing the smallest variance contributions.
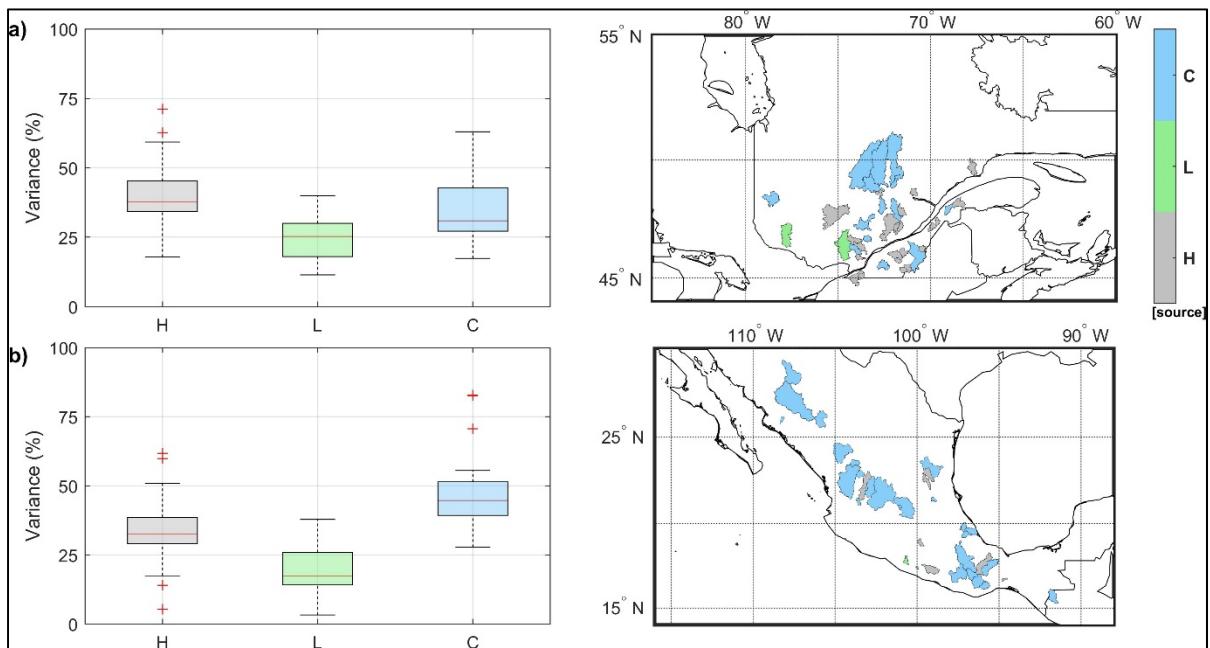
Figure 4.7 Total variance contributions (%) of hydrological models' structure (H, in grey), calibration data's length (L, in green) and calibration data's climatic condition (C, in blue) over the snow- and rain-dominated basins on panels a and b, respectively. Each row presents the boxplots of total variance (%) of the three variance contributors in the left and a map showing the main variance contributor for each basin in the right

4.4.2 Streamflow projections sensitivity to hydrological models and their calibration

The effects of the selected calibration strategies on seasonal hydrological projections were assessed by coupling temperature and precipitation simulations issued from two regional climate simulations with five hydrological models calibrated with 20 different strategies. To decompose the impact of the different hydrological models and their calibration strategies on the projected streamflows, a variance decomposition analysis was performed to evaluate the variance associated to hydrological model structures (H), calibration's data length (L) and calibration's data climatic condition (C) on the mean annual RCM-driven streamflow values. Figures 4.8 and 4.9 show the variance decomposition results on the mean annual RCM-driven streamflows of the reference and future periods of 1976-2005, 2041-2070 and 2070-2099 simulated with the CRCM5-CNRMCM5 and the CRCM5-CANESM2 simulations, respectively. Each figure shows the results for the snow- and rain-dominated basins in panels a and b, respectively.

Overall, it is observed that mean annual streamflows variance of the CRCM5-CNRMCM5 is mainly dominated by hydrological models' structure over both basin groups. The snow-dominated basins show that the influence of hydrological models' structure increases when moving toward future periods with median boxplots of total variance (%) moving from about 60% to almost 70 % in the farthest horizon. The influence of calibration data's climatic conditions and length are consequently smaller when moving toward future periods, with calibration data's climatic conditions showing slightly higher total variances (median values around 20%). Similar results are observed in the rain-dominated basins with hydrological model structures showing the largest contributions followed by calibration data's climatic conditions and length. Yet, the total variance contributions from hydrological model structures are generally lower with median values of about 55% for both reference and future periods. Total variances of calibration data's climatic conditions and lengths are slightly higher than over the snow-dominated basins with values of about 24% and 20%, respectively. And, as observed with hydrological model structures, remain very similar for all periods.

Similar results are generally observed on the variance decomposition analysis on the mean annual streamflows simulated with the CRCM5-CANESM2 simulations (Figure 4.9), particularly over the rain-dominated basins. Figure 4.9 shows that, as observed in Figure 4.8, hydrological model structures present the highest variance contributions on mean annual streamflows simulated with the CRCM5-CANESM2 simulation on snow- and rain-dominated basins. Some differences are however observed in the magnitudes compared to the results using the CRCM5-CNRMCM5 simulations. For instance, over the snow-dominated basins it is observed that the total variance contributions of hydrological model structures had a larger increase in the future horizons than in the future streamflow simulated with the CRCM5-CNRMCM5.
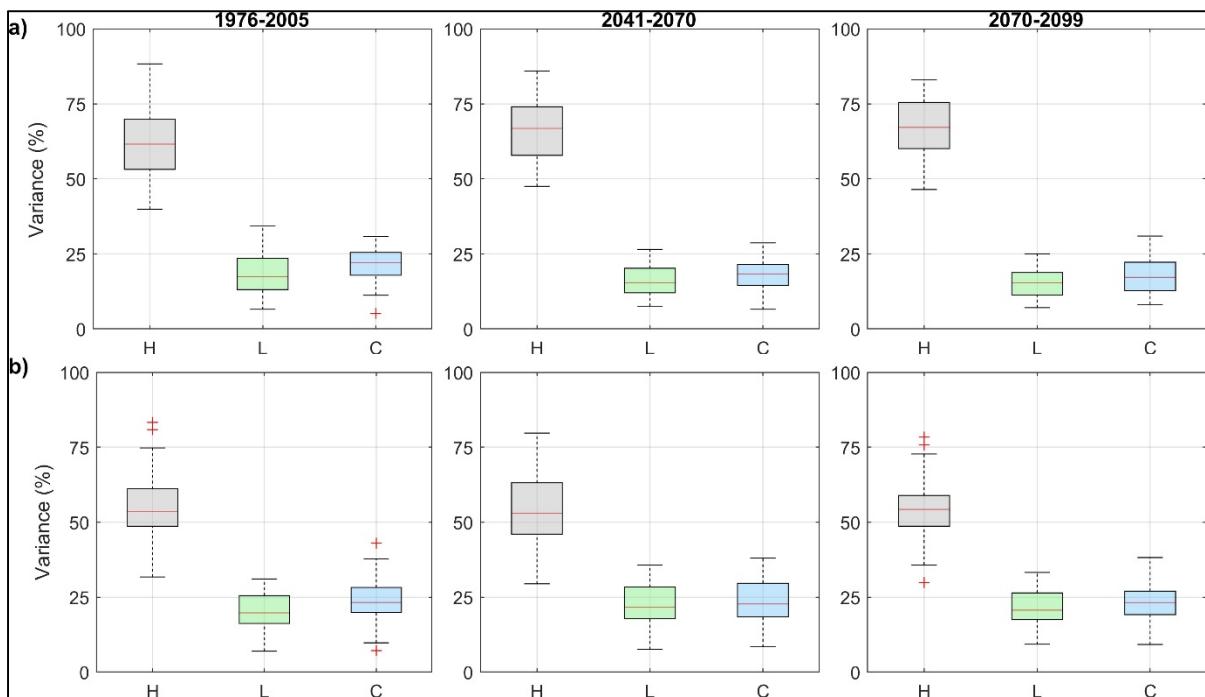
Figure 4.8 Total variance contributions (%) of hydrological models' structure (H, in grey), calibration data's length (L, in green) and calibration data's climatic condition (C, in blue) on mean annual RCM-driven streamflows simulated with the CRCM5-CNRMCM5 simulations over the snow- and rain-dominated basins on panels a and b, respectively. Each row presents the boxplots of total variance (%) of the three variance contributors for the reference and future periods of 1976-2005, 2041-2070 and 2070-2099
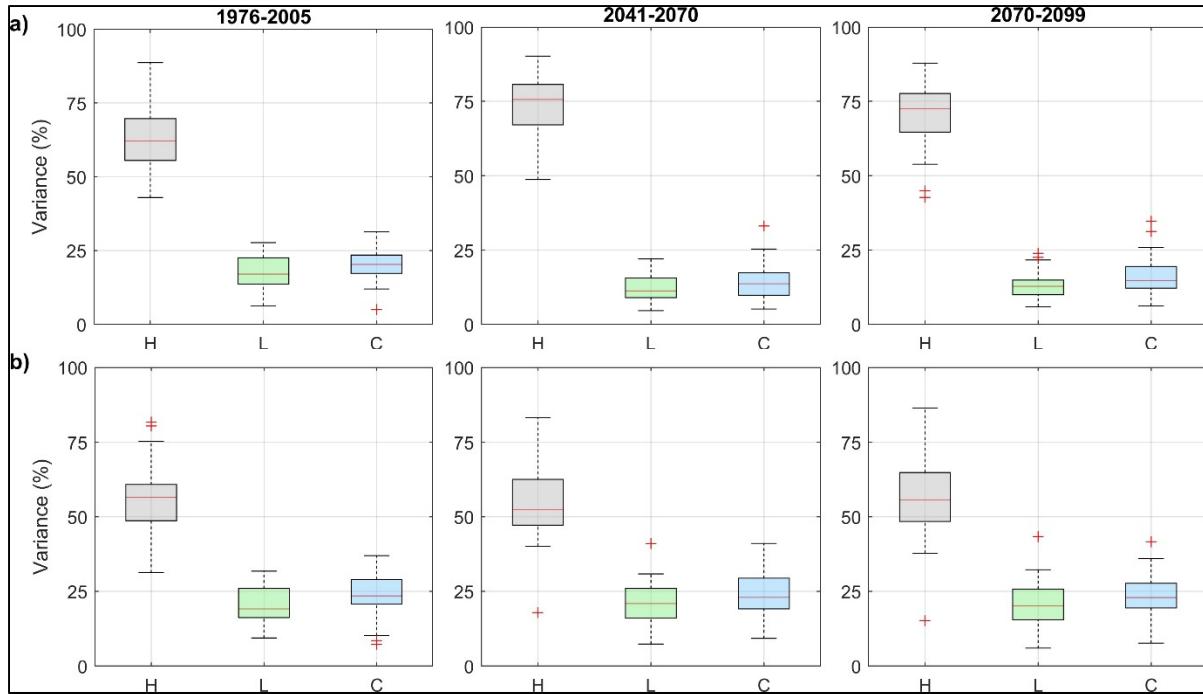
Figure 4.9 Total variance contributions (%) of hydrological models' structure (H, in grey), calibration data's length (L, in green) and calibration data's climatic condition (C, in blue) on mean annual RCM-driven streamflows simulated with the CRCM5-CANESM2 simulations over the snow- and rain-dominated basins on panels a and b, respectively. Each row presents the boxplots of total variance (%) of the three variance contributors for the reference and future periods of 1976-2005, 2041-2070 and 2070-2099

To evaluate the impact of hydrological models' structure, calibration data's length and calibration data's climatic conditions on peak streamflow projections a variance decomposition analysis was also performed on mean annual peak RCM-driven streamflows. Following the same structures as for Figures 4.8 and 4.9, Figures 4.10 and 4.11 show the variance decomposition results obtained with the CRCM5-CNRMCM5 and CRCM5-CANESM2 scenarios, respectively.

Figures 4.10 and 4.11 show that as observed on the mean annual streamflows analysis, the largest variance contribution was brought by hydrological model structures over both snow- and rain-dominated basins. However, its observed that within both CRCM5 domains, mean annual peak streamflows show more sensitivity to calibration data's climatic conditions than the sensitivity observed on mean annual streamflows analysis. Yet, as observed in Figures 4.8

and 4.9, the variance contributions of calibration data's climatic conditions decrease when moving toward future periods. This is observed on the variance results on the mean annual peak streamflows obtained with the CRMCM5 simulations on the snow- and rain-dominated basins.
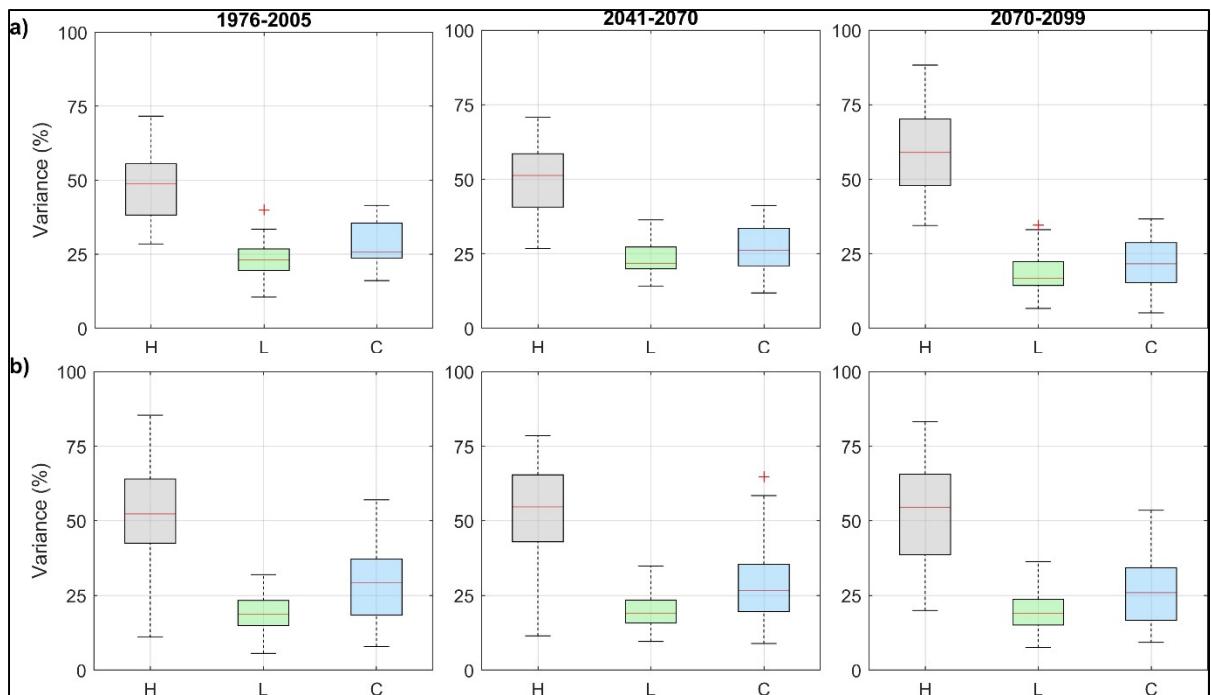
Figure 4.10 Total variance contributions (%) of hydrological models' structure (H, in grey), calibration data's length (L, in green) and calibration data's climatic condition (C, in blue) on mean annual peak RCM-driven streamflows simulated with the CRCM5-CNRMCM5 simulations over the snow- and rain-dominated basins on panels a and b, respectively. Each row presents the boxplots of total variance (%) of the three variance contributors for the reference and future periods of 1976-2005, 2041-2070 and 2070-2099
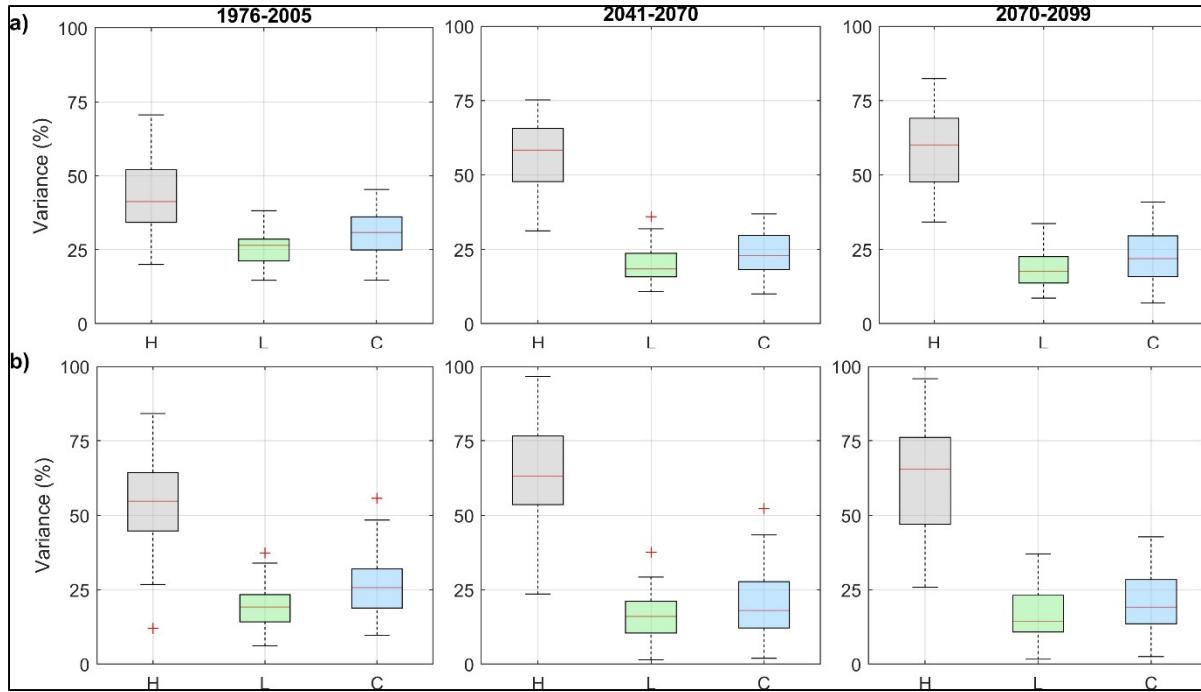
Figure 4.11 Total variance contributions (%) of hydrological models' structure (H, in grey), calibration data's length (L, in green) and calibration data's climatic condition (C, in blue) on mean annual peak RCM-driven streamflows simulated with the CRCM5-CANESM2 simulations over the snow- and rain-dominated basins on panels a and b, respectively. Each row presents the boxplots of total variance (%) of the three variance contributors for the reference and future periods of 1976-2005, 2041-2070 and 2070-2099

4.5   Discussion

Coupling hydrological models with global and/or regional climate projections has become a common process to simulate future streamflows for climate change impact studies in hydrology. An essential step prior this process is the calibration and validation of the hydrological models. However, their sometimes-diminished reliability under contrasting climatic conditions to the ones used for their calibration/validation has become a subject of discussion, particularly in the context of climate change impact studies where different climatic conditions are expected. Different studies have tested hydrological models' sensitivity to contrasting climatic conditions and others have focused on evaluating the impact of varying the calibration data's length as a way to cover more diverse climatic conditions (Brigode, Oudin, & Perrin, 2013; Guo et al., 2018). However, it remains unclear if it is the calibration

data's climatic conditions or the length that bring more instability to hydrological models' performance, particularly on streamflow projections. Thus, to evaluate the impacts of the calibration data's climatic condition as well as the calibration's data length, this study assessed the impact of different calibration strategies on the validation performance of 77 basins with a diversity of hydrometeorological regimes. Twenty different calibration strategies covering four different sets of contrasting climatic calibration-validation conditions (WC, CW, DH, HD) and five different lengths were used to stress-test five hydrological models and evaluate their impact on validation performance and on streamflow projections.

4.5.1 Impact of hydrological models and their calibration on validation performance

The different stress-tests and the variance decomposition analyses carried out on five hydrological model's validation performance revealed that the hydrological model calibration data's climatic conditions consistently showed larger impacts on their validation performance than the varying datasets lengths. This was observed over the different hydrological models and the two basin groups, the (1) 38 snow-dominated basins and the (2) 39 rain-dominated basins. The KGE-values obtained during validation in Figures 4.3 and 4.4, as well as the performance difference between calibrations and validations in Figures 4.5 and 4.6, showed that the impact of the calibration data's climatic conditions had often larger impacts. For instance, when looking at Figures 4.5 and 4.6, the resulting boxplots of snow- and rain-dominated basins often showed larger KGE-differences over DH or HD calibration-validation tests, even when increasing the calibration's datasets lengths. This was further analyzed and confirmed with the variance decomposition analysis in Figure 4.7, where calibration data's climatic conditions contribution was consistently higher than the calibration's data length over most basins. Yet, as observed in this analysis and over Figures 4.3-4.6, the five hydrological models showed different levels of sensitivity to the calibration strategies. Over the snow-dominated basins, the variance decomposition results showed that the hydrological model's structures often brought more variance to the validation performance than the calibrations choices. This can also be observed in Figures 4.3 and 4.6 where some hydrological models,

such as HM5, showed larger sensitivity to climatic conditions than others. The rain-dominated basins also showed different impacts between the five hydrological models, but as observed in Figure 4.7, most basins showed slightly larger sensitivity to the calibration data's climatic conditions than to hydrological model structures.

Regarding the hydrological models' response to the different climatic conditions included in this study, Figures 4.3-4.6 showed that some climatic conditions presented more impacts on their validation performances than others. For example, over both the snow- and rain-dominated basins, it is observed that precipitation-based contrasting conditions often displayed larger spreads and sometimes lower medians than the tests with temperature-based conditions. This is in line with previous studies showing that hydrological models consistently diminished their performance when used under strongly different precipitation conditions (Coron et al., 2012; Dakhlaoui et al., 2019; Motavita et al., 2019; Seiller et al., 2012; Seiller et al., 2015). However, the results also showed that the magnitude of the impact was systematically different between basin groups. For instance, when looking at the KGE-values during validation (see Figures 4.3 and 4.4), it is observed that rain-dominated basins show lower KGE-values than those observed on the snow-dominated basins. Moreover, the KGE-difference between calibration and validation was also larger over the rain-dominated basins, suggesting a generally larger sensitivity to contrasting climatic conditions than snow-dominated basins. Concerning calibration's data length, the results showed that changing the calibration's dataset lengths often had the smallest influence on hydrological model's validation results, which can thus suggest that calibrations on shorter periods (e.g., 6-8 years) can have similar validation performance than those calibrations performed over larger periods.

4.5.2   Potential implications on streamflow projections

The variance decomposition analyses on the streamflow projections obtained with the 20 different calibration strategies revealed that the calibration data's climatic conditions can contribute more variance to streamflow projections, particularly to peak streamflow projections (see Figures 4.10 and 4.11). This is observed over both analyses using different

CRCM5 simulations and over both basin groups. Yet, some differences are observed between basin groups. First, in contrast to the variance decomposition analyses performed on hydrological model's validation performance, the analysis on mean annual streamflow projections showed that hydrological model's structure consistently contributed more variance than the calibration strategies on both snow- and rain-dominated basins (Figures 4.8 and 4.9, respectively). This larger influence is particularly large over the snow-dominated basins, which was expected as the analyses performed on hydrological model's validation performance already showed hydrological model structures as the main variance contributor in most basins. The rain-dominated basins, in contrast to the results on validation performance, showed that hydrological model's structures also presented the highest variance contributions. Yet, compared to the snow-dominates basins, calibration data's climatic conditions showed generally larger total variance contributions.

Regarding the differences on mean annual streamflows and mean annual peak streamflows some differences are observed on the magnitude of the total variance contributions, particularly between the CRCM5 simulations with different driving GCMs. To further evaluate these differences an additional analysis was performed to compare the precipitation and temperature projected deltas from CRCM5 simulations with each driving GCM against the deltas considered in the DSST experiments. The future climate deltas of the regional climate simulations were compared by estimating the mean annual temperature delta and mean annual total precipitation relative deltas between the reference (1976-2005) and future periods (2041-2070 and 2070-2099). To compare these projected climate deltas against the contrasting climates that were considered with the different calibration strategies, Figure 4.12 shows the temperature and precipitation deltas issued from the regional climate simulations as well as the temperature and precipitation deltas from the different calibration strategies. The figure shows the analysis for the 2041-2070 and the 2070-2099 periods for the snow- and rain-dominated basins in panels a and b, respectively. For each basin group, only the calibration-validation deltas that follow the direction of change of the CRCM5 simulations are displayed. For instance, if the projections show higher temperatures over the future horizons, only the Cold to Warm deltas are displayed for an easier comparison.
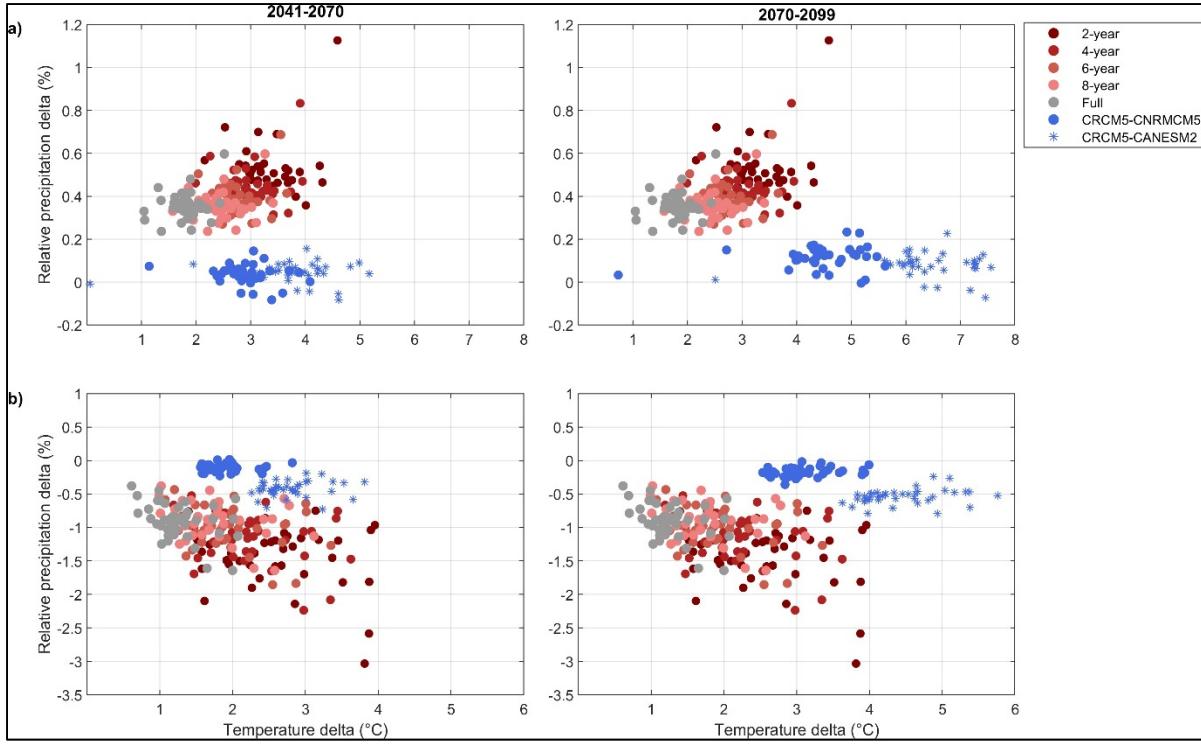
Figure 4.12 Temperature delta and relative precipitation delta of the different calibration strategies (in red and grey circles) the CRCM5-CNRMCM5 simulations (in blue circles) and the CRCM5-CANESM2 simulations (in blue asterisks) for the 2041-2070 and 2070-2099 periods

Over the snow-dominated basins, the figure shows that the projected temperature changes of both CRCM5 simulations are often larger than the deltas covered by the different DSST tests, particularly for the farthest future period of 2070-2099. This is especially observed for the CRCM5-CANESM2 simulations that show temperature changes that go up to about +7°C, while the DSST test hardly covers changes of more than +4°C with the 2-year DSST tests. This is also observed over the rain-dominated basins where most basins show that the temperature deltas are surpassed over the farthest horizon. These larger deltas can be linked to the slightly higher total variance contributions of hydrological model structures on the mean annual and mean annual peak streamflows obtained with the CRCM-CANESM2 simulations. A possible explanation can be related to the uncertainty associated with each hydrological model structure, that as shown in a recent study by Troin et al. (2018), can have a larger impact than other components as the snow melting/accumulation and potential evapotranspiration

representations. Thus, the different structures might have different ways to assimilate these strong temperature changes and consequently influence their streamflow production processes. These increasing impacts of hydrological model structures not only highlight the need to further study the uncertainties related to hydrological processes representations and their potential impacts on streamflow projections, but also the importance of including multi-model approaches in climate change impact studies (Broderick, Matthews, Wilby, Bastola, & Murphy, 2016; Seiller et al., 2012). It is however important to underline that only lumped conceptual hydrological models were included in the present study. Thus, additional works including distributed hydrological models can be of use to better understand the impacts of hydrological models process representation and their influence on streamflow projections.

The results showed clearly that the choice of hydrological model had the largest impacts on projected streamflows. Yet, regarding the calibration choices, the analyses revealed that in most cases the calibration data's climatic conditions can have a larger impact on projected streamflows than the length of the calibration dataset. These results complement the findings of previous studies, such as those of Vaze et al. (2010) and Motavita et al. (2019), suggesting that calibrating on shorter periods that include climatic conditions that are closer to the climate projections can be recommended. Moreover, the large-scale analysis showed that over basins with contrasting hydrometeorological conditions, the calibration data's climatic conditions often presented higher impacts on streamflow projections than the length of the calibration dataset.

4.6   Concluding remarks

This study investigated the impact of hydrological model structures, calibration data's climatic conditions and lengths on hydrological model's validation performance and their projected streamflows. A total of 20 different calibration strategies were tested on 77 basins covering contrasting hydroclimatic conditions (i.e. snow- and rain-dominated river basins). These different tests were evaluated with five different lumped hydrological models and four regional climate simulations issued from the CRCM5 (two diving GCMs and two domains – each

domain covering a different set of hydroclimatic conditions). Using a variance decomposition analysis, the results showed that hydrological model structures and the calibration data's climatic conditions often contributed the most to the validation performance of the snow- and rain-dominated basins, respectively. Yet, when evaluating the impacts on streamflow projections, the results show that hydrological model structures clearly become the largest variance contributor on mean annual projected streamflows and mean annual peak streamflow projections. These results not only underline the need to reconsider the implicit assumption that hydrological models calibrated on historical climatic conditions are reliable for climate change projections, but also the importance of using multi-model approaches in climate change impact studies on hydrology.

CHAPTER 5

HYDROLOGICAL MODELS WEIGHTING FOR FLOOD PROJECTIONS

Mariana Castaneda-Gonzalez[a], Annie Poulin[b], Rabindranarth Romero-Lopez[c] and Richard Turcotte[d]

[a,b] Department of Construction Engineering, École de technologie supérieure, Montréal, Canada.
[c] Faculty of Civil Engineering, Universidad Veracruzana, Xalapa, Mexico.
[d] Ministère du développement durable, Environnement et Lutte contre les changements climatiques, Québec, Canada.

Abstract

Reliable flood projections are of great importance for climate change impact studies, especially knowing that these analyses can allow identifying regional adaptation and mitigation strategies for the future. However, the literature has highlighted that hydrological models used under climate conditions that are contrasting to those used during their calibrations can lower their performance and reliability, an issue that lowers the confidence on hydrological projections and adds uncertainty to their analyses. More studies are needed to explore this issue and to evaluate potential strategies that might improve hydrological models' reliability in climate change impact studies. Thus, the present study evaluates the robustness of hydrological models under contrasting climatic conditions and investigates the use of weighting techniques to improve their combined performance and reliability. The robustness of five lumped hydrological models is analysed using a Differential Split-Sample Testing (DSST) that evaluates their performance on cold, warm, humid and dry historical contrasting conditions

over 77 basins covering different hydroclimatic conditions (two domains, one located in Quebec, Canada, and one in Mexico). Additionally, four basins were selected from the study area to evaluate the robustness of a more complex semi-distributed and more physically-based hydrological model, and to compare its simulations against the simpler lumped hydrological models. Based on the resulting performance of each hydrological model, five different weighting methods were applied to evaluate the potential improvements on the multi-model ensemble performance, and to quantify their effects on flooding projections. For each basin, these streamflow projections were produced using two regional climate simulations (one per studied domain) issued from the Canadian Regional Climate Model version 5 (CRCM5) under the Representative Concentration Pathway (RCP) 8.5 for the 1976-2006, 2041-2070 and 2070-2099 periods. The results showed that weighting hydrological models, even with the most simplistic methods, showed better performances over historical contrasting conditions than the best-performing lumped hydrological model. Between the different weighting methods, the Granger-Ramanathan type A showed the overall best performance among the different basins and climate conditions, particularly on peak streamflows. Over the CRCM5-driven flooding projections, the weighting methods Granger-Ramanathan types A and B produced the largest impacts on the projected floods magnitudes and climate change signals. On the other hand, the additional tests using the semi-distributed and more physically-based hydrological model revealed that this model showed more robust simulations than the weighted lumped hydrological models on low flows over the four selected basins. Additionally, more robust high-flow simulations were observed over a small snow-dominated basin, suggesting a potential added value of adding more complex hydrological models to simulate conditions under a changed climate.

Keywords: Hydrological models weighting; parameters robustness; flood projections; ensemble modelling

5.1 Introduction

The potential social, environmental and economic impacts of climate change in flooding has urged the scientific community to develop modelling frameworks to better understand these events. These modelling frameworks often consist in simulating hydroclimatic processes under different future scenarios by jointly using hydrological models and post-processed (bias-corrected and/or downscaled) climate simulations (Her et al., 2019; Krysanova et al., 2020). However, different studies have shown that the reliability of these hydroclimatic simulations still needs to be improved due to the modelling uncertainties associated with the different modelling steps, particularly when simulating extreme hydrological events such as floods (Castaneda-Gonzalez et al., 2021; Kundzewicz et al., 2014; Kundzewicz et al., 2017). Some studies have shown that hydrological models can be one of the main uncertainty contributors in hydrologic projections, especially when simulating extreme hydrologic events (Her et al., 2019; Krysanova et al., 2018; Vetter et al., 2017). Consequently, it has been recommended to consider ensembles of hydrological models instead of only relying in a single hydrological model (Krysanova et al., 2018; Vetter et al., 2017), particularly for impact assessments at the catchment scale.

Hydrological models are traditionally calibrated/validated on historical conditions to ensure a good performance before being used in an impact assessment. This means that, it is assumed that hydrological model's performance on observed climatic conditions will remain valid when used with the different climate scenarios. However, this practice has been questioned as hydrological models have shown lower performances under strongly different climatic condition, an issue that can consequently lower their reliability when used for climate change impact assessments (Coron et al., 2012; Guillaume Thirel et al., 2015). Therefore, different efforts have been made to better asses hydrological model's "predictive capability" also known as "transferability" by more comprehensive evaluations that can allow testing hydrological model's performance under different climatic conditions, such as the Differential Split-Sample Testing (DSST) (KlemeŠ, 1986; Krysanova et al., 2018). After this evaluation, some studies have recommended excluding the hydrological models with unsatisfactory performance

(sometimes referred to as "strict weighting") to then perform a simple ensemble mean for the impact assessment, while others have suggested using a weighted ensemble mean based on each hydrological model's performance (Huang et al., 2020; Krysanova et al., 2018; Krysanova et al., 2020).

Different weighting techniques have been tested for combining hydrological models' outputs at basin and global scales (Arsenault et al., 2015; Wan et al., 2021; Zaherpour et al., 2019), showing that weighted multi-model ensembles generally present more robust performances than single hydrological models. Most of these studies have been carried out over historical streamflow records and have particularly focused on forecasting applications. However only some climate change impact studies have applied performance-based weighting techniques on hydrological models' ensembles, and fewer have assessed the effects of different weighting approaches (Broderick et al., 2016; Huang et al., 2020; Krysanova et al., 2020; Najafi, Moradkhani, & Jung, 2011; Seiller et al., 2015). Among the former, Seiller et al. (2015) evaluated the use of two multi-model averaging techniques to improve hydrological models' transferability over twenty U.S. basins. The results showed that the weighted multi-model approach often increased the simulation performance under contrasting climate compared against individual hydrological models. Moreover, it showed that the performance-based weighing was more efficient than equally weighting all hydrological models. Similarly, Broderick et al. (2016) tested different weighting techniques on 37 Irish river basins. The results, in line with those from Seiller et al. (2015), revealed that weighted averages outperformed individual hydrological models in most cases. Yet, it also showed that among the four tested weighting techniques, some methods (e.g., Granger-Ramanathan Averaging), were more robust than others. Among the studies focused on climate change impact applications, Najafi et al. (2011) evaluated weighting three lumped and one distributed hydrological model using a Bayesian Model Averaging (BMA) method to assess the uncertainties of hydrological models in projected runoff over a rain-dominated basin in Oregon, USA. The results showed that the BMA-weighted ensemble presented generally similar projected runoff compared to the best-performing model. Yet, the BMA-weighted ensemble outperformed the best-performing model when the hydrological models with

unsatisfactory performance were excluded, suggesting that a stricter weighting method can potentially improve the weighted ensemble performance.

A recent study presented by Huang et al. (2020) evaluated the influence of a strict and a simple performance-based weighting approach on projected hydrological changes over three basins. The results showed that when the hydrological model with lowest performance was excluded (strict weighting), the total uncertainty in low flows was more significantly reduced than with the simple weighting. Following a similar approach, Wen et al. (2020) applied a simple multi-linear regression to weight three hydrological models in the Chinese Upper Yangtze river basin. The results showed that when hydrological models were more comprehensively evaluated and weighted, the magnitudes of the projected floods and droughts were less severe than with the common ensemble mean approach. The authors thus concluded that the weighted multi-model ensemble was a more reliable approach for climate change impact assessments. It was however highlighted that the degree of improvement remained unclear, suggesting that further studies exploring weighting strategies were still needed. In contrast, another study applying a simple qualitatively weighting scheme based on a comprehensive evaluation of three semi-distributed hydrological models showed generally small differences between the simple and weighted ensemble mean over the Indian Godavari river basin (Mishra, Shah, López, Lobanova, & Krysanova, 2020).

The lowered reliability of hydrological models under changing climate conditions shows the importance of reconsidering the common approach of equally weighting ensembles of hydrological models for climate change impact studies. To address this issue, some studies have suggested a stricter selection, calibration, and validation of hydrological models, while others have suggested favoring the use of multi-model weighting frameworks as each hydrological model has its own weaknesses and strengths. However, only few studies have evaluated the use of different hydrological models weighting techniques for climate change impact studies. This is of particular relevance given the fact that contrasting climatic conditions are used when coupling hydrological models with future climate scenarios. Moreover, most studies have mainly focused on mean streamflows and few have evaluated the impacts of

weighting strategies on flood projections. At the same time, while more robust weighting techniques have been used in other contexts such as regionalization and streamflow forecasting (Arsenault et al., 2015), simple performance-based weighting techniques have been used in climate change impact studies at the catchments scale. It is thus observed that the impacts of different weighting techniques that consider hydrological models' robustness to contrasting climate conditions on flood projections are yet to be explored. The aim of this study is thus to evaluate six different robustness-based weighting techniques on an ensemble of five different hydrological models to quantify their effects on flood projections over various basins to cover diverse hydroclimatic regimes.

5.2 Study area and data

5.2.1 Studied basins

Seventy-seven basins with different hydroclimatic regimes were used in this study. As observed in Figure 5.1, the study area comprises two clusters of basins, a first cluster of 38 snow-dominated basins in the Canadian province of Quebec (panel a), and a second cluster of 39 rain-dominated basins located in Mexico (panel b). For each cluster, the mean total annual precipitation and mean annual temperature are presented (from left to right) for each basin. Two basins identified with numbers 1 and 2 are presented over each region. These basins will be used for further analyses presented later in the text.

As observed in Figure 5.1, the two clusters of basins cover a great diversity of climatic and hydrologic regimes. Overall, the figure shows that a wide range of mean total annual precipitation and mean annual temperature is covered by the 77 basins. In terms of precipitation, the study area shows mean total annual precipitation values that vary from less than 500 mm to up to 2000 mm per year, with the snow-dominated basins showing more similar values among them that range from about 900 to 1200 mm. The mean annual temperature maps show that the two basin groups cover contrasting temperature conditions. The rain-dominated basins show clear warmer conditions with values ranging from about 12

to up to 23.5 °C, particularly when moving towards lower latitudes. This warm gradient is also observed over the snow-dominated basins, with overall colder temperatures ranging from about 5 to 12 °C.
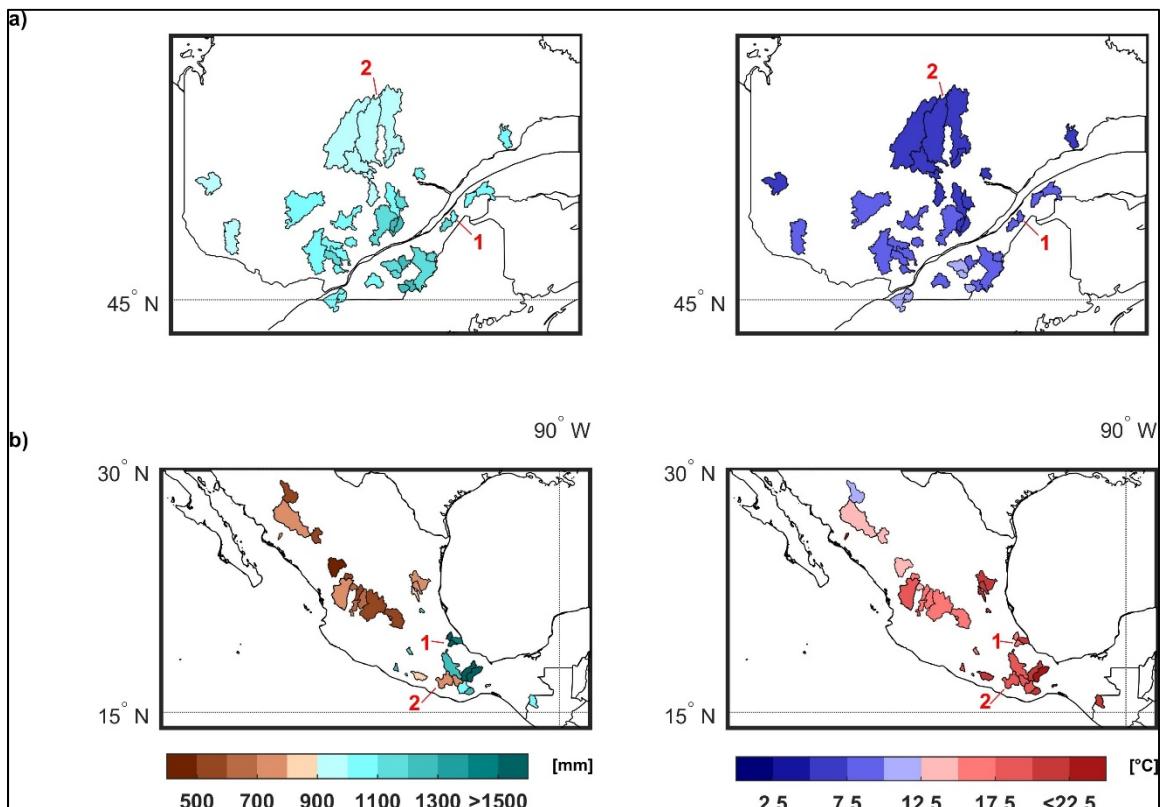
Figure 5.1 Locations and mean annual climatology of the 38 basins in Quebec and the 39 basins in Mexico in panels a and b, respectively. The left panels show the mean total annual precipitation (in mm) and the right panel shows the mean annual temperature (in °C) for both regions. Two basins selected for further analysis over each region are identified with red numbers

5.2.2   Observed and simulated data

For each of the studied basins, daily climatic and hydrometric observed data were obtained from the Hydrometeorological Sandbox –École de technologie supérieure (HYSETS) database (Arsenault, Brissette, Martel, et al., 2020). This large-scale database of North American basins includes data between 1950 and 2018 from various sources particularly selected for

hydrological and climate change impact studies. The properties, climatology, and hydrometric data of each catchment used in this study were extracted from this database by ensuring minimum lengths of 30 years between 1950 and 2013 for the daily minimum and maximum temperatures, precipitation and streamflow. From the different sources of meteorological data available in HYSETS, this study used the Livneh gridded climate data (Livneh et al., 2015; Livneh et al., 2013) a database consisting of a 1/16° (~6 km) resolution grid that covers Mexico, the conterminous U.S. (CONUS) and southern Canada.

5.2.2.1 Regional climate simulation

The regional climate model simulations used in this study were obtained from the Canadian Regional Climate Model version 5 (CRCM5; Martynov et al., 2013; Separovic et al., 2012) driven by the CanESM2, referred as CRCM5-CANESM2. These CRCM5 simulations, produced by the Ouranos Consortium on Regional Climatology and Adaptation, is part of the Coordinated Regional Downscaling Experiment (CORDEX) database (Giorgi & Gutowski, 2015). One of the simulations covers the CORDEX North American (NAM) domain and the other one covers the Central American (CAM) domain with an extension of the northern border to cover Mexico. Both simulations were run at a 0.22° horizontal resolution under the Representative Concentration Pathway (RCP) 8.5. The studied time periods include the reference period of 1976-2005 and two future horizons of 2041-2070 and 2070-2099. For each basin, time series of daily precipitation, minimum and maximum temperatures issued from the CRCM5-CanESM2 simulations were post-processed to be later combined with the five hydrological models to obtain the RCM-driven streamflow simulations. The post-processing approach used in this study is the quantile mapping method (Maraun, 2016; Themeßl et al., 2012; Themeßl et al., 2011), a robust method that adjusts the simulated distributions by quantiles. Different studies have used and recommended quantile mapping over different regions (e.g., Chen et al., 2013a; Zhao et al., 2020) endorsing its use in this study. Further details on the temperature and precipitation simulations adjustment process can be found in Chen et al. (2013b).

5.3 **Methodology**

The methodology of this study comprises different stages that allow evaluating the performance of weighting hydrological models under observed and projected contrasting climatic conditions. Five lumped hydrological models and five different weighting techniques were included in this study, and are described in the following sections.

5.3.1 **Hydrological models**

Five lumped hydrological models were selected for this study, namely GR4J, MOHYSE, HMETS, CEQUEAU and IHACRES models. All these five hydrological models have been successfully used in different studies and applications (Dallaire et al., 2021; Troin et al., 2018), supporting their use in this study. The structure of each hydrological model, identified as HM1-5 in the remaining sections of the paper, is described in the following sections.

5.3.1.1 **GR4J - HM1**

GR4J model is a parsimonious lumped rainfall-runoff model with 4 parameters initially developed for water management applications by Perrin et al. (2003). The structure of this empirical model consists of two reservoirs that simulate flow production and routing. The model itself does not include a module to represent snow accumulation and melting processes. To include these, the GR4J-4 model was coupled with the 2 parameter CemaNeige model (Valéry et al., 2014). The inputs needed for the GR4J-CEMANEIGE hydrological model are daily temperature, precipitation and evapotranspiration. Based on previous studies (Arsenault et al., 2018; Troin et al., 2016), the Oudin potential evapotranspiration (PET) formulation (Oudin et al., 2005) was used.

5.3.1.2 MOHYSE - HM2

The MOHYSE hydrological model is a simplistic lumped model that was initially developed by Fortin et Turcotte (2006) for academic applications. This model, with 10 parameters to calibrate, is composed of two main reservoirs that simulate the vertical water balance and routing. Within its own structure, it estimates potential evapotranspiration as well as snow accumulation and melting processes through simple, temperature-based, approaches. The inputs needed for MOHYSE consist of continuous time series of mean daily temperature, total daily rainfall, and snowfall (in water equivalent). This conceptual model has been effectively applied over different regions with different hydrological regimes including rain- and snow-dominated regions (Arsenault et al., 2019; Zavaleta et al., 2015).

5.3.1.3 HMETS - HM3

HMETS is a lumped conceptual hydrological model with 21 parameters to calibrate. This Matlab-based hydrological model developed by Martel et al. (2017) is composed of two reservoirs that represent the saturated and vadose zones. The hydrological processes simulated within this model's structure include surface, delayed, hypodermic and groundwater flows. The snowmelt is also simulated through a representation of the snowpack melting and refreezing processes using a temperature-based degree-day model (Vehviläinen, 1992). HMETS requires continuous times series of daily precipitation, maximum and minimum temperature, as well as potential evapotranspiration as inputs. Different studies have coupled this hydrological model with the Oudin PET formulation. However, recent studies have shown that the choice of PET formulation influences the resulting streamflow (Dallaire et al., 2021; Seiller & Anctil, 2016). Thus, to account for different potential evapotranspiration formulations and consequently have a larger diversity of hydrological model structures, the McGuinness and Bordne formulation was coupled with HMETS (McGuinness & Bordne, 1972) based on a recent successful application (Castaneda-Gonzalez et al., 2021).

5.3.1.4 CEQUEAU – HM4

CEQUEAU is a hydrological model developed by Girard et al. (1972). For this study, a lumped version of the conceptual distributed hydrological model was used. This lumped hydrological model has 9 parameters to calibrate and it represents the vertical water balance through two interconnected reservoirs and the downstream routing. Within the water balance of this hydrological model, the snowpack formation, snowmelt, evapotranspiration, flows in unsaturated and saturated zones are represented. Based on previous studies using this lumped CEQUEAU version (Seiller et al., 2012; Velázquez et al., 2010), the CemaNeige model (Valéry et al., 2014) was used. This CEQUEAU-CEMANEIGE model combination requires daily minimum and maximum temperature, as well as daily solid and liquid precipitation with a total of 11 parameters to calibrate.

5.3.1.5 IHACRES – HM5

IHACRES is a lumped hydrological model developed by Jakeman et al. (1990) with 7 parameters to calibrate. The structure of this parsimonious hydrological model is represented by a non-linear production module and a linear transfer module. Overall, these processes simulate the effective rainfall generation and its transformation to quick and slow flow using unit hydrographs. As used in previous hydrological models, the snow model CemaNeige (Valéry et al., 2014) was used jointly with IHACRES to simulate snow melting and accumulation process. Thus, the IHACRES-CEMAINEGE model resulted in 9 parameters to calibrate. Similar versions of this hydrological model have been effectively used in previous studies (Perrin et al., 2001; Seiller et al., 2012) supporting its application in this study. The inputs required for this hydrological model consist of daily precipitation, mean temperature and potential evapotranspiration time series. Potential evapotranspiration is not directly estimated within this hydrological model structure, thus the Hargreaves and Samini evapotranspiration formulation (Hargreaves & Samani, 1985) was used to ensure a variety of well-known potential evapotranspiration formulations.

5.3.2   Weighting methods

Five different weighting approaches were selected to evaluate their performance under contrasting climatic conditions, and their impacts on flooding projections. The five weighting approaches used in this study are described in the following sections.

5.3.2.1   Equal-weighting - EW

This method consists in assigning equal weights to each of the hydrological models used for each basin. The idea behind applying this simple weighting approach is to use it as reference for comparison with the other unequal weighting methods.

5.3.2.2   Simple KGE weighting – W1

This weighting approach uses the KGE criterion to assign weights to each hydrological model based on its performance. This simple weighting method estimates the performance of the simulated streamflow against the observed streamflow. For instance, the weight of HM_i is estimated as follows:

$$W1_{HMi} = \frac{KGE_{HMi}}{\sum_{i=1}^{n} KGE_{HMi}} \tag{5.1}$$

where $W1_{HMi}$ is the weight assigned to HM_i, KGE_{HMi} represent the KGE values obtained with the streamflow simulation with HM_i, and n is the number of streamflow simulations/ hydrological models.

5.3.2.3   Bates-Granger averaging – W2

The Bates and Granger averaging (BGA) method (Bates & Granger, 1969) assigns weights to the different simulations to minimize the Root Mean Squared Error (RMSE) of the combined

simulations. The weight of each simulation is estimated using the inverse of a given simulation variance. For instance, the weight of HM_i is estimated as follows:

$$W2_{HMi} = \frac{1/\sigma^2_{HM_i}}{\sum_{i=1}^{n} 1/\sigma^2_{HM_i}} \tag{5.2}$$

where $W2_{HMi}$ is the weight assigned to HM_i, $\sigma^2{}_{HMi}$ represents the variance value obtained with the streamflow simulation with HM_i, and n is the number of streamflow simulations/ hydrological models. It is important to highlight that this method relies on the hypothesis that all the simulations considered are not biased and that their errors are not correlated, conditions that are not necessarily respected in the simulations used in the present study.

5.3.2.4   Granger-Ramanathan type A – W3

The Granger Ramanathan averaging (GRA) method (Granger & Ramanathan, 1984), assigns the weights of the simulations considered in the ensemble to minimize the RMSE using the ordinary least squares (OLS) algorithm. This approach is unconstrained, meaning that the weights do not necessarily sum to unity and negative weights are possible. The weights of the ensemble of simulations n are thus estimated as follows:

$$W3_{HM} = \left(Qsim_{HM}{}^T \cdot Qsim_{HM}\right)^{-1} \cdot Qsim_{HM}{}^T \cdot Qobs \tag{5.3}$$

where $W3_{HM}$ is the vector of n weights, $Qobs$ is the observed streamflow, $Qsim_{HM}$ represents the matrix of streamflow simulations of all HMs, n is the number of simulations/hydrological models, and the superscript T represents the transpose of a given vector/matrix.

5.3.2.5   Granger-Ramanathan type B – W4

The B method of the GR approach (Granger & Ramanathan, 1984), assigns the weights of the simulations considered in the ensemble to minimize the RMSE of the combined simulations using the ordinary least squares (OLS) algorithm as described for the method A. Yet, unlike method A, this approach is constrained, meaning that the weights sum to unity. The weights of the ensemble of simulations are thus estimated as follows:

$$W4_{HM} = \left(Qsim_{HM}{}^T \cdot Qsim_{HM}\right)^{-1} \cdot Qsim_{HM}{}^T \cdot Qobs \tag{5.4}$$
$$+ 2\lambda\left(Qsim_{HM}{}^T \cdot Qsim_{HM}\right)^{-1} \cdot l$$

$$\lambda = \frac{\left(l^T \cdot \left(Qsim_{HM}{}^T \cdot Qsim_{HM}\right)^{-1} \cdot Qsim_{HM}{}^T \cdot Qobs - 1\right)}{\left(l^T \cdot \left(Qsim_{HM}{}^T \cdot Qsim_{HM}\right)^{-1} \cdot l\right)} \tag{5.5}$$

where $W4_{HM}$ is the vector of n weights assigned, $Qobs$ is the observed streamflow, $Qsim_{HM}$ represents the matrix of streamflow simulations of all *HMs*, l is a unit vector of dimension [1, n], n is the number of simulations/hydrological models, and the superscript T represents the transpose of a given vector.

5.3.3   Experimental design

As previously discussed, studies have shown that hydrological models can lower their performances when used under climatic conditions that are different than those used during their calibrations. Thus, this methodology aims to evaluate the performance of five hydrological models' weighting approaches under contrasting climatic conditions using a Differential Split Sample Testing (DSST) approach. At the same time, the effects of weighting hydrological models on flooding projections are evaluated over each of the 77 diverse basins using two RCM simulation, one per studied domain.

Figure 5.2 shows a schematic representation of the overall experimental design to test the performance of weighting methods on contrasting conditions. As observed in Figure 5.2, the experimental design consists of four main stages applied to each basin. For each basin, the first year of hydrometric/meteorological data available was excluded from the following steps as it is a warm-up period for the hydrological models, where the simulations reach more realistic internal variable states. This spin-up year is thus not included in the performance evaluation metrics. The remaining hydrometric/meteorological data of each basin were then ranked from driest to most humid, and from coldest to warmest years, based on their mean annual precipitation (for D and H years) and temperature (for C and W years). For each of these four climatic conditions the four main methodological steps are as follows:

1. The 6 driest (D), most humid (H), coldest (C) and warmest (W) years were identified to calibrate, alternately, each of the five lumped hydrological models.

2. The D, H, C and W parameterizations obtained for each hydrological model were used to simulate the streamflows over their contrasting H, D, W and C 6-year validation periods that, as observed in the schematic representation presented in Figure 5.2, exclude the 5 "most" contrasting years. The resulting simulation of each hydrological model under the four contrasting evaluation conditions were used to apply the five weighting techniques.

3. The remaining most contrasting 5-year periods were then used to evaluate the performance of the five weighting methods (hereafter called evaluation, as seen in Figure 5.2).

4. Finally, all hydrological models were coupled with the RCM simulations to produce the RCM-driven streamflows for each basin. According to the climate change signal simulated by the RCM-simulations, the most contrasting parameterizations were used. For instance, if the RCM projection shows warmer climate, the cold parametrization will be used to ensure a larger contrast between the calibration and projected data. The rationale behind this is to represent a potential situation where the climate conditions of the calibration data are contrastingly different than those projected by climate models.
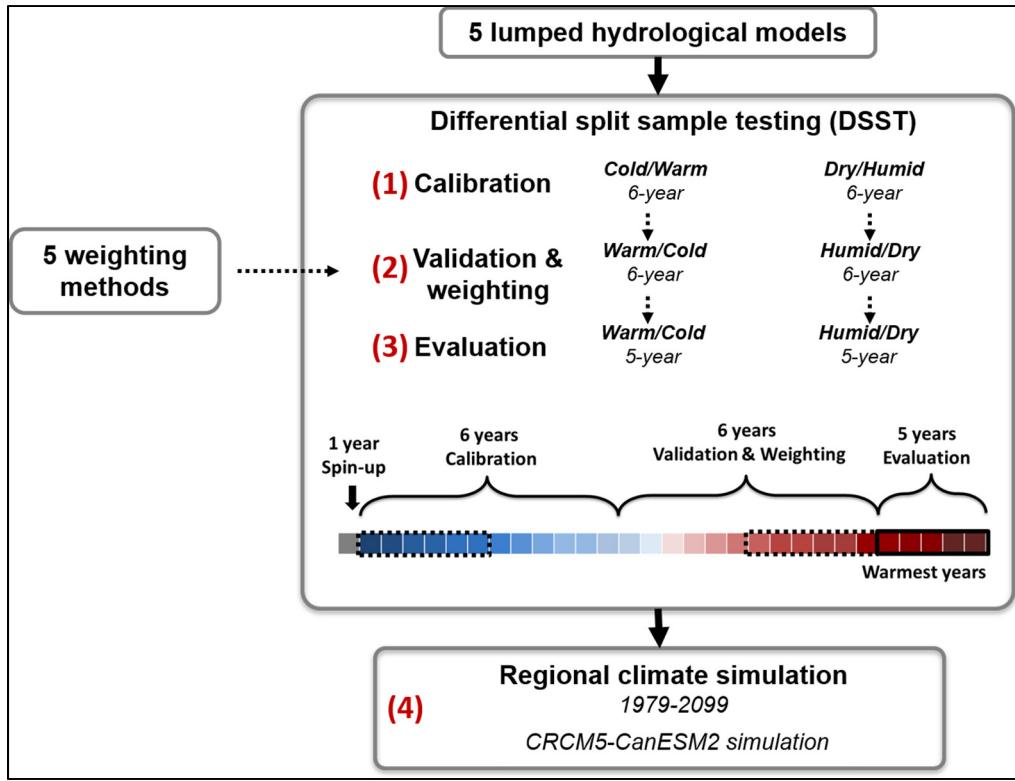
Figure 5.2 Overview of the main methodological steps used in this study

5.4   Results

5.4.1   Hydrological models and weighting methods

Figure 5.3 shows the evaluation results of the DSST experiments for all hydrological models and weighting methods over the four calibration-validation contrasting climatic conditions, namely the W-C, C-W, D-H, and H-D. The figure shows the results for the KGE and NSE criteria over panels a) and b), respectively. Each boxplot includes the results for the 77 basins of the study area. The different DSST experiments show that weighting hydrological models almost systematically outperforms individual hydrological models, in terms of the median values (most cases) and the dispersion (some cases) of the evaluated performance criteria. Between the four DSST experiments, it is observed that all hydrological models and weighting approaches present lower performances over the C-W and H-D experiments. This is observed with both the KGE and NSE criteria, meaning that the selected hydrological models had

difficulties when simulating streamflow under strongly warmer and dryer conditions than those used during calibration. Over both performance metric analyses, it is observed that some hydrological models outperform others among the different experiments. For instance, it is observed that H1 and H4 are often among the best individual models between the five lumped hydrological models. Regarding the different weighting approaches, it is noted that the W3 method generates slightly better performance among the different experiments compared to the other methods.
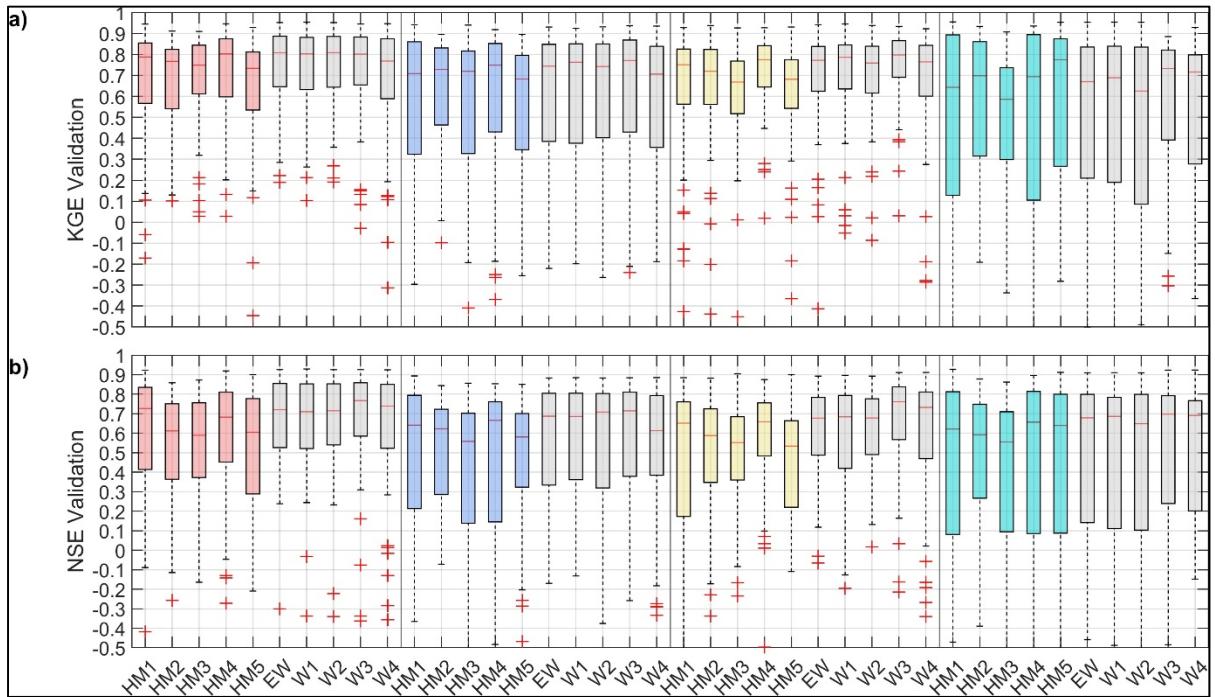
Figure 5.3 Boxplots of the KGE and NSE values in evaluation of the five hydrological models and five weighting techniques estimated from all 77 basins are presented in panels a) and b), respectively. From left to right the results for the W-C, C-W, D-H and H-D experiments are presented

For further comparison between hydrological models and weighting methods, a Kruskal-Wallis test with Bonferroni correction was performed to rank and intercompare their performances. Figure 5.4 reveals that, as observed in Figure 5.3, the different weighting approaches show similar performances with overlapping confidence intervals, meaning that they are not significantly different at the 5% level. Nonetheless, it is observed that W3 (i.e.,

GRA method) often shows slightly better performances than the other methods and, in some cases, a significant difference with respect to single hydrological models. This is particularly clear on the NSE assessment, a metric that gives stronger weight to the streamflow peaks.
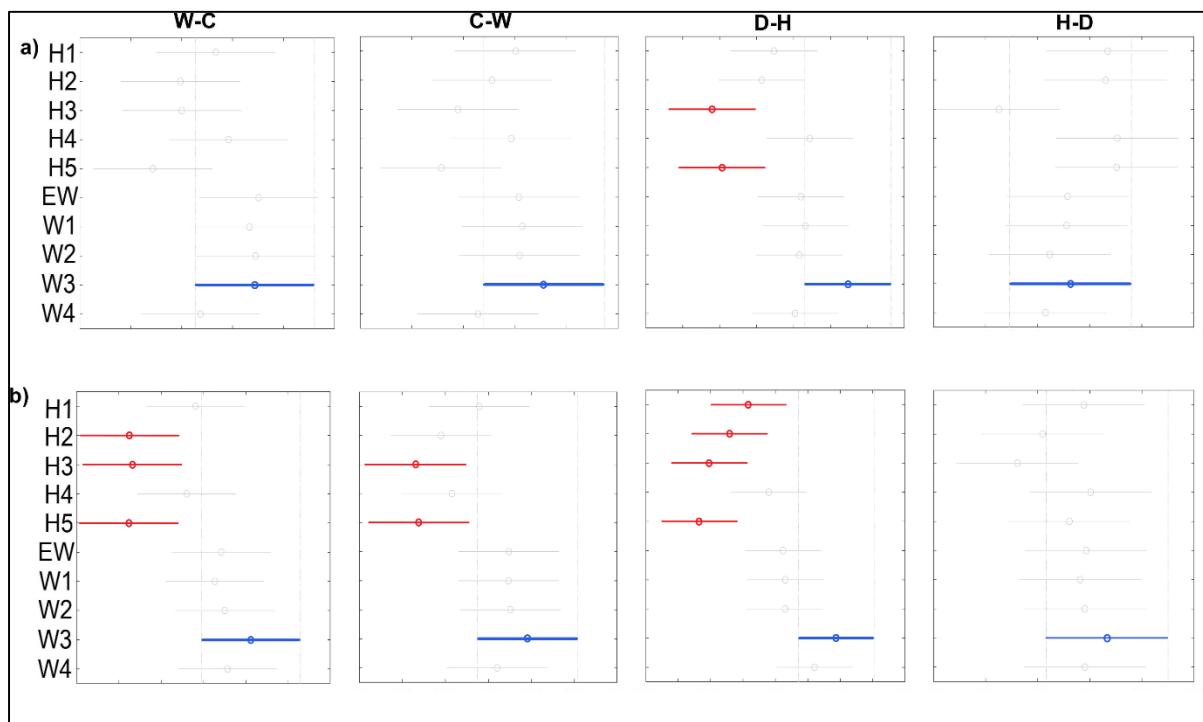
Figure 5.4 Multiple comparison Kruskal-Wallis rank test with Bonferroni correction on the KGE and NSE values are presented on panels a) and b), respectively. Over each panel the rank sum and confidence intervals, represented by horizontal lines, are shown for each hydrological model and weighting approach. The vertical lines on each panel are used as reference to indicate the W3 confidence intervals. The rank values of the x-axis are not shown as they are not used in this assessment

An additional geographical analysis with the commonly used EW approach and the W3 method is performed to compare their performance on peak streamflow simulations over the diverse study area. As previously described, the study area is divided in two groups based on their flood-generating processes, the snow- and rain-dominated basins. Thus, Figures 5.5 and 5.6 show the KGE scores estimated during the flooding months of the snow- (March, April and May, MAM) and rain-dominated basins (June, July and August, JJA), respectively.
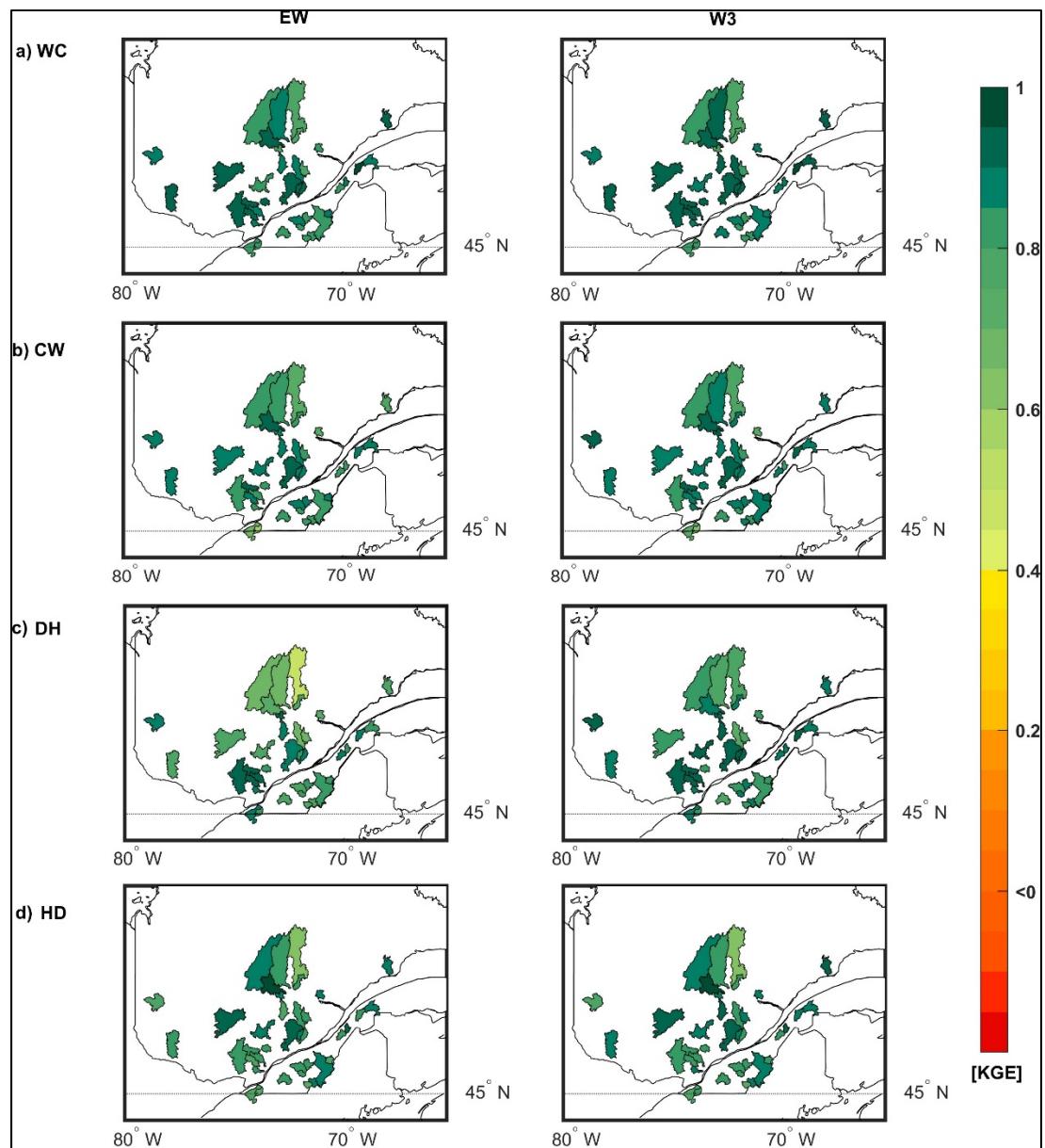
Figure 5.5 KGE values in evaluation estimated over the flooding months (March, April and May) of the snow-dominated basins. The results for the W-C, C-W, D-H and H-D experiments are presented in panels a), b), c) and d), respectively. For each experiment, the results for the EW (left column) and W3 (right column) approaches are presented
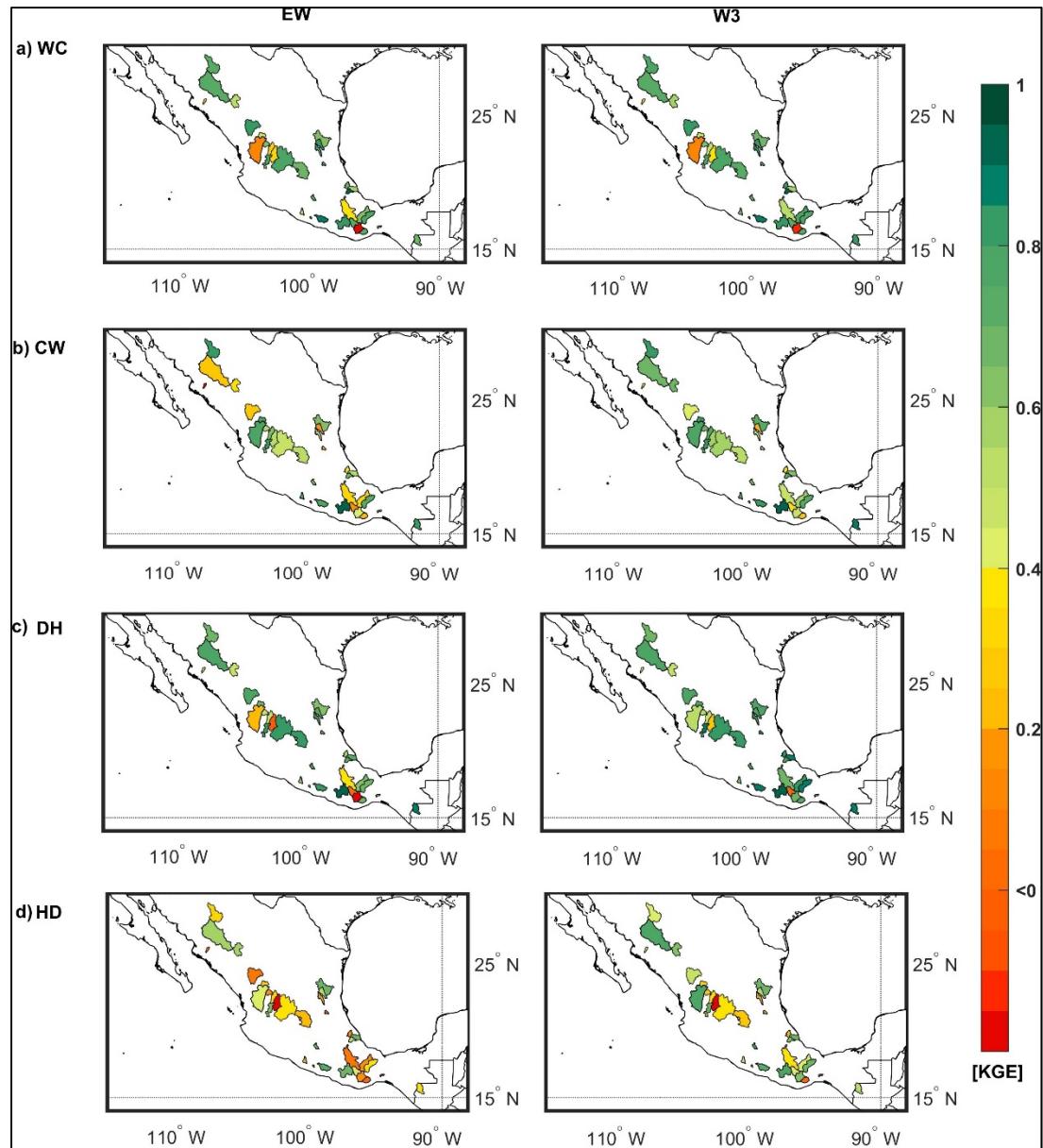
Figure 5.6 KGE values in evaluation estimated over the flooding months (June, July and August) of the rain-dominated basins. The results for the W-C, C-W, D-H and H-D experiments are presented in panels a), b), c) and d), respectively. For each experiment, the results for the EW (left column) and W3 (right column) approaches are presented

By comparing Figures 5.5 and 5.6, it is observed that some basins generally show higher performance than others. The snow-dominated basins clearly show higher KGE values than the rain-dominated ones with both weighting approaches over the four DSST experiments.

Nonetheless, it is observed that most basins keep similar or improved performance when using the W3 method. When looking at the different DSST experiments, it is noted that in some configurations the W3 method is not able to improve the performance of EW, particularly over the snow-dominated basins where smaller differences are observed between both approaches. It is only over the D-H configuration where slightly larger improvements are observed when using the W3 method. The rain-dominated basins however show larger improvements when using unequal weighting. This is observed over all DSST experiments. Yet, it is also observed that in some cases, no improvements were obtained with the W3 method, as similar low performances are observed over both methods.

5.4.2   Regional climate projection: temperature and precipitation

To evaluate the impacts of the different weighting methods on flood projections, the RCM simulations issued from the CRCM5 model driven by the CanESM2 model were combined with the different hydrological models and parameterizations estimated from the DSST experiments. To give an overview of the projected climate from the CRCM-CanESM2 bias-corrected outputs, the precipitation and temperature climate change signals are displayed in Figure 5.7 for the 2041-2070 and 2070-2099 periods.

The precipitation projections presented in Figure 5.7 (panels a) show that the basin groups display opposite climate change signals. The snow-dominated basins show generally more humid projections of about -5 to 15 %, while the rain-dominated basins show clearly dryer future conditions that go up to -63 % over both future periods. From the temperature projections (panels b), it is clear that both basin groups show warmer future climate conditions over all basins and future horizons. The rain-dominated basins show a warm climate change signal over both future horizons of about +4 °C and +5.5 °C for the 2041-2070 and 2070-2099, respectively. It is also observed that the basins located in northern Mexico show slightly warmer projections. It is also clear from these panels that snow-dominated basins show warmer future projections than the rain-dominated basins, with climate change signals that go up to +7.8 °C for the 2070-2099 horizon, particularly over basins located in higher latitudes.
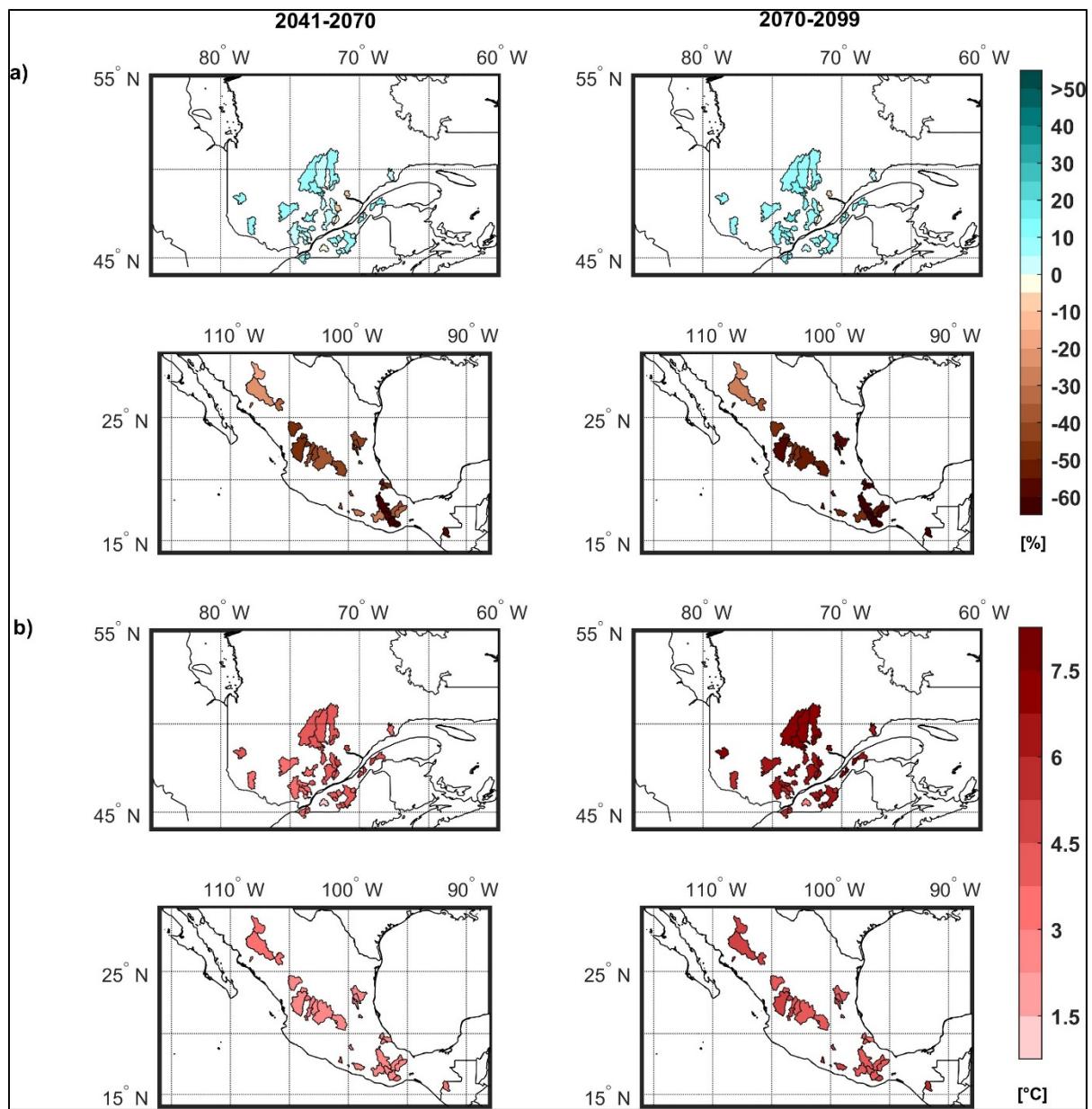
Figure 5.7 Precipitation and mean temperature projections issued from the CRCM5-CaenESM2 model on panels a) and b), respectively. Each panel shows the snow- and rain-dominated basins on the top and lower maps, respectively. From left to right, each panel shows the projections for the 2041-2070 and 2070-2099 periods

5.4.3 Weighted flood projections

In this stage, the impacts of weighting hydrological simulations based on the hydrological models' robustness to changing climatic conditions are compared to the common EW approach. The CRCM5-CanESM2 outputs analysed in section 5.4.2 showed warmer and more humid climate projections over the snow-dominated basins, while warmer and dryer conditions where observed over the rain-dominated basins. Thus, the weights and parameterizations obtained from the C-W/D-H and C-W/H-D tests were used in the following analyses for the snow- and rain-dominated basins, respectively. As described in the methodological step 4, the idea of this process is to ensure a larger contrast between the climate conditions used for calibration and the RCM projections to evaluate the effects of weighting on a changed climate, a situation that can be expected in climate change impact studies.

5.4.3.1 Impact on mean seasonal peak streamflows

Figures 5.8 and 5.9 show the relative difference between the results from the EW and each one of the weighting approaches (W1-4) for mean seasonal peak streamflows considering the C-W/D-H and C-W/H-D DSST experiments over the snow- and rain-dominated basins, respectively. Each figure shows the results for each season, namely winter (December, January and February, DJF), spring (March, April and May, MAM), summer (June, July and August, JJA) and fall (September, October and November, SON), in panels a), b), c) and d). From left to right, the results for the 1976-2005, 2041-2070 and 2070-2099 periods are presented.

By comparing the relative differences over the snow- and rain-dominated basins in Figures 5.8 and 5.9, it is clearly observed that the choice of weighting method impacts the magnitudes of the mean seasonal peak streamflow projections. More particularly, it is observed that the weighting methods W3 and W4 show the largest absolute differences against the reference approach EW. This is observed over all DSST experiments, seasons, and periods. Between the two basin groups it is also observed that the impact of weighting is generally more diverse for the rain-dominated than for the snow-dominated basins, with some outliers reaching negative

and positive relative differences of more than 100 %. Moreover, it is revealed that unequal weighting methods (W1-W4) can impact the projections toward different directions. For instance, Figure 5.8 shows that the mean seasonal peak streamflows estimated with the W4 approach are often smaller than those estimated with the EW approach over most snow-dominated basins, with median relative differences that go down to -28 % in the 2070-2099 period. Yet, the W1 and W2 methods show slightly larger mean seasonal peak streamflows than the EW approach. Among the different DSST experiments, larger impacts are observed over the D-H cases, especially over the summer and fall months of both reference and future periods. Yet, over the future periods it is observed that the relative differences are slightly larger than in the reference period. This is particularly clear over the summer months for the C-W experiments.

Over the rain-dominated basins, contrasting results are observed between the C-W and H-D experiments. For instance, generally smaller mean seasonal peak streamflows with median relative differences of about -5 to -10 % are produced with the W3 and W4 methods over the C-W experiment, while in the H-D experiment overall larger mean seasonal peak streamflows of about 10% are observed with the same methods. Among the different seasons and periods, no clear differences are observed over the rain-dominated basins, yet it is observed that the spread of relative differences increases in some cases. This is observed over all seasons, particularly with the C-W experiments.
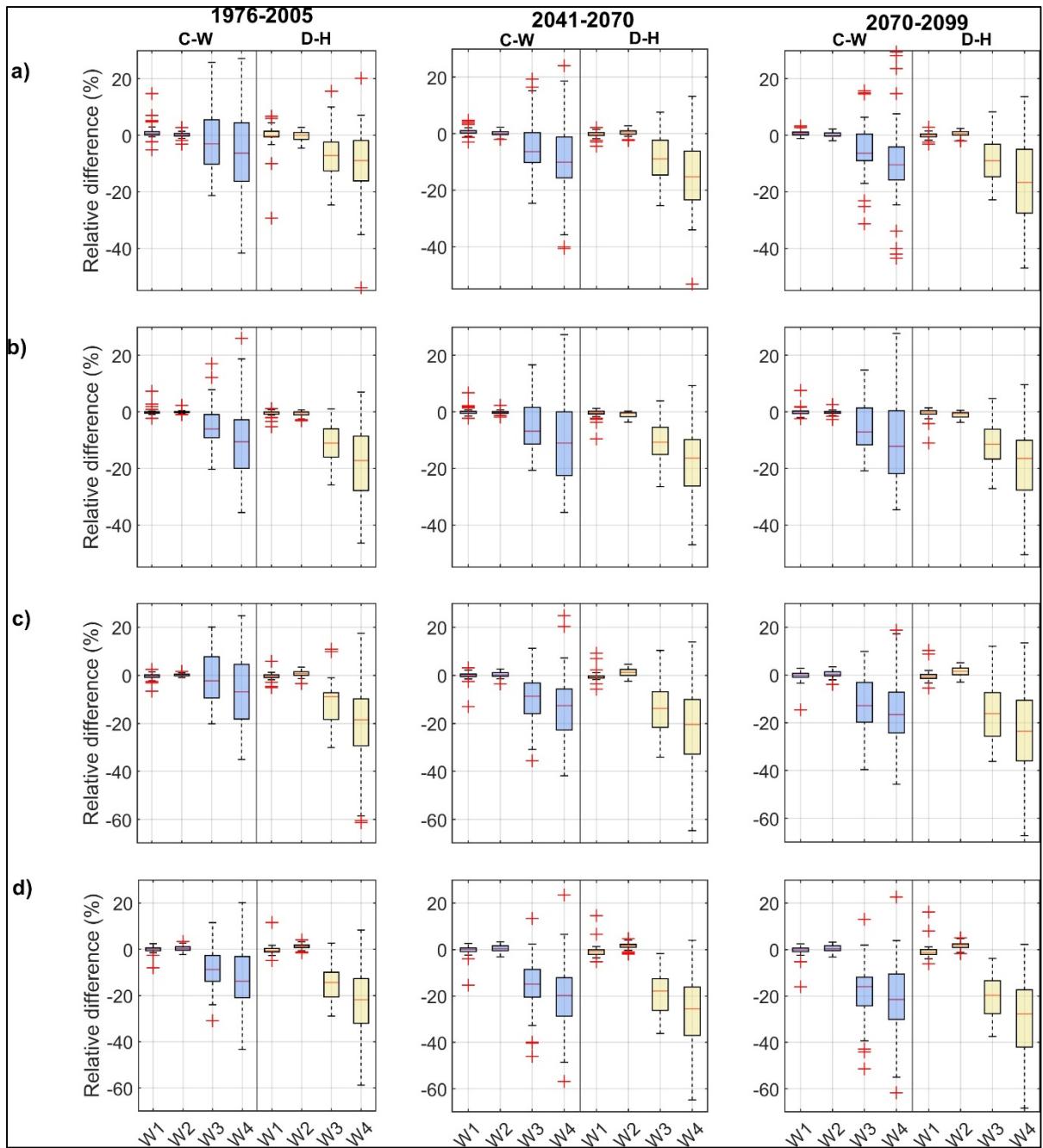
Figure 5.8 Relative difference (%) between the four unequally-weighted (W1-4) mean seasonal peak streamflows and the EW values for the snow-dominated basins. From left to right, the figure shows the results for the 1976-2005, 2041-2070 and 2070-2099 periods. By row, results for the winter (panel a), spring (panel b), summer (panel c) and fall (panel d) seasons are presented
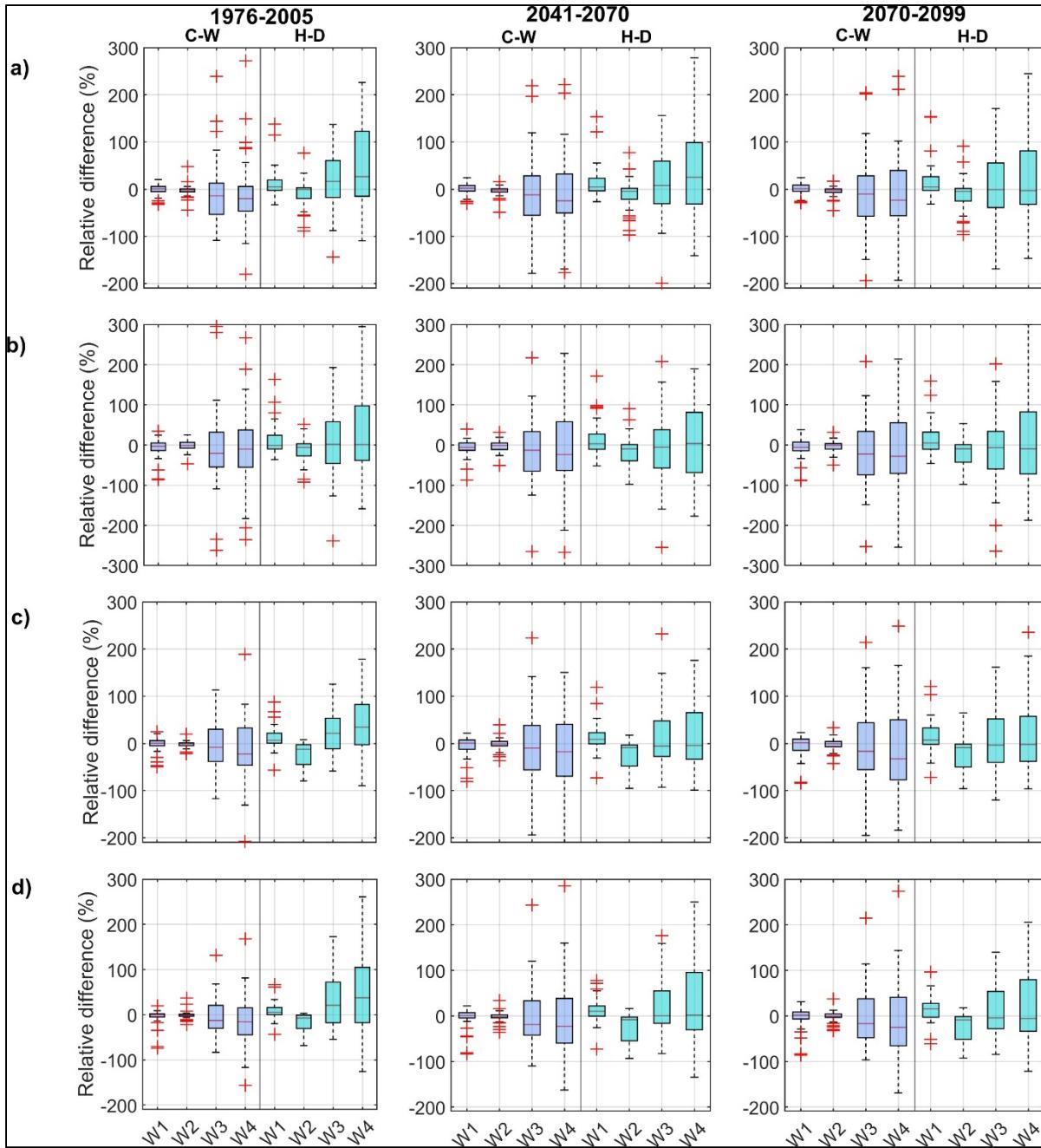
Figure 5.9 Relative difference (%) between the four unequally-weighted (W1-4) mean seasonal peak streamflows and the EW values for the rain-dominated basins. From left to right, the figure shows the results for the 1976-2005, 2041-2070 and 2070-2099 periods. By row, results for the winter (panel a), spring (panel b), summer (panel c) and fall (panel d) seasons are presented

5.4.3.2 Impacts on the climate change signal

To evaluate the impacts of weighting the projected mean projected seasonal peak streamflow values on the climate change projections, Figures 5.10 and 5.11 show the climate change signals estimated with the relative difference (%) between the reference and future mean seasonal peak streamflows of all weighting approaches over the snow- and rain-dominated basins, respectively. From top to bottom, the results for the winter, spring, summer and fall seasons are presented. From left to right, the figures show the results for the 2041-2070 and 2070-2099 periods. On each panel, the weighting approaches for the two different experiments performed for each basin group are presented on the x-axis. On the y-axis the basins of each group, sorted by ascending mean annual total precipitation (MATP, in mm/year) are presented.

The climate change signals of the snow-dominated basins in Figure 5.10 show that the different weighting approaches agree on the direction of change in many cases. Over the winter season, it is clear that higher mean peak streamflows are projected over both DSST experiments, all weighting approaches, and future horizons. In terms of magnitude, rather small differences are observed between weighting approaches. For instance, in some cases the W3 and W4 approaches show differences of about 15 to 20% smaller than the other approaches. Especially when comparing them to the EW approach. During the summer season, a clear decreasing behaviour is observed over all basins and weighting approaches. Yet, it can also be observed that slightly stronger decreases are observed with the W3 and W4 methods. This difference is especially clear over the C-W experiments where the decreases are sometimes 20 to 25% stronger than those of the EW approach. Over the spring and fall seasons, more diverse signals are observed. For instance, opposite climate change signals are observed in the spring between weighting methods, especially during the 2041-2070 period over the C-W experiment. Yet, when moving to the farthest horizon fewer cases disagree on the climate change signal. Most cases show decreasing mean annual peak streamflows, except for some of the basins with lowest MATP that show larger mean annual peak streamflows. Similarly, over the fall months the W3 and W4 methods show opposite signals with respect to the other approaches over some

basins. Yet, for the 2070-2099 horizon most cases agree on the same climate change signal, with the W3 and W4 methods showing slightly stronger decreases than the other methods.
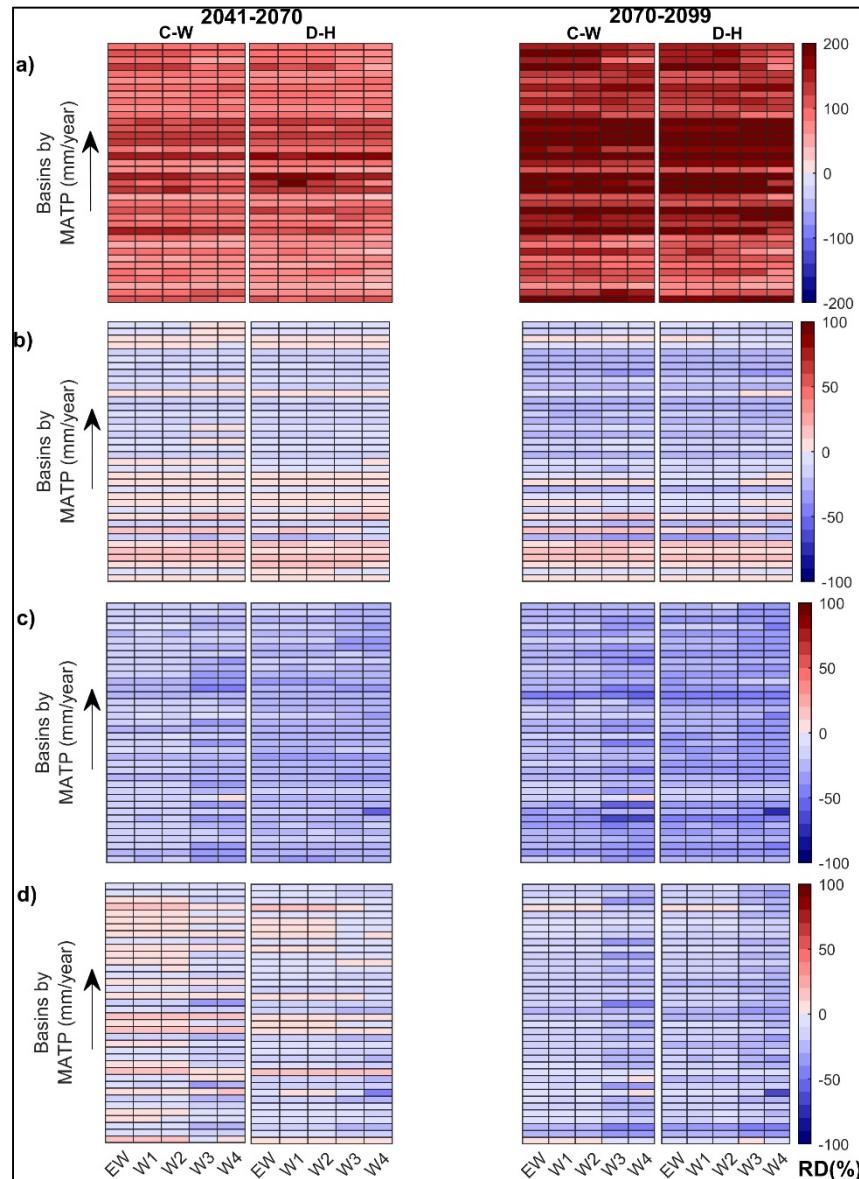
Figure 5.10 Climate change signal (%) in mean seasonal peak streamflows from all weighted hydrological model ensembles for the snow-dominated basins. From left to right, the figure shows the results for the 2041-2070 and 2070-2099 periods. By row, results for the winter (panel a), spring (panel b), summer (panel c) and fall (panel d) seasons are presented. Shades of red indicate positive relative difference and shades of blue indicate negative relative difference
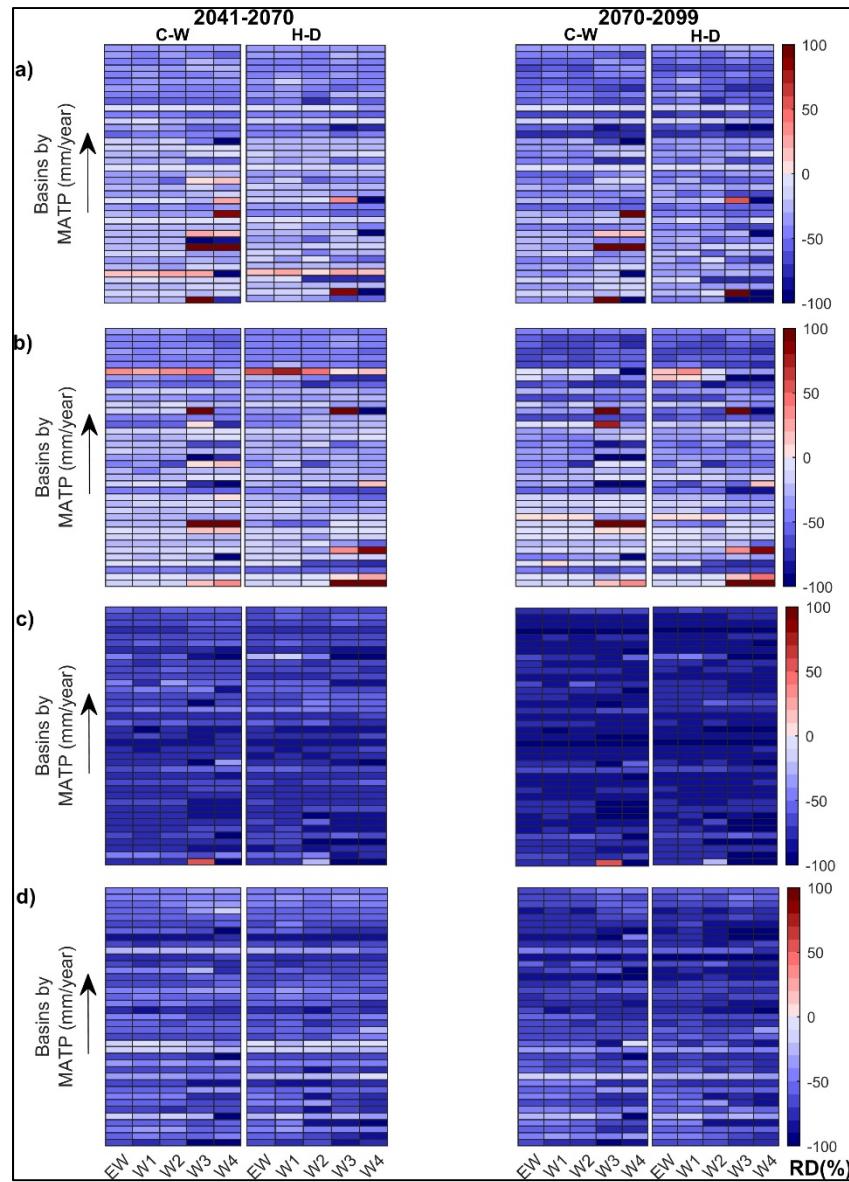
Figure 5.11 Climate change signal (%) in mean seasonal peak
streamflows from all weighted hydrological model ensembles
for the rain-dominated basins. From left to right, the figure
shows the results for the 2041-2070 and 2070-2099 periods.
By row, results for the winter (panel a), spring (panel b),
summer (panel c) and fall (panel d) seasons are presented.
Shades of red indicate positive relative difference and shades
of blue indicate negative relative difference

Figure 5.11 shows that smaller mean seasonal peak streamflows are projected in most of the
rain-dominated basins over both DSST experiments, and all seasons, and periods. This

behaviour is especially clear over the summer and fall seasons, which correspond to the flooding months. Regarding the different weighting methods, no clear systemic effects are observed during these months. Yet, it is observed that the magnitude of the decreases varies between methods. During the winter and spring seasons, differences are more clearly observed between weighting methods and, as observed in Figure 5.10, W3 and W4 are the weighting methods showing the largest differences against the other methods. This larger contrast during the dry months can be explained by the generally small streamflows during these seasons where small changes can significantly impact the signal, especially over the driest basins. Over the basins with higher values of MATP, the direction of the climate change signal is often maintained. Yet, it is observed that the magnitude of the climate change signal varies between methods and DSST experiments, especially over the farthest future time period.

5.5   Discussion

5.5.1   Hydrological model's weighting performance on observed contrasting climate conditions

Using ensembles of hydrological models, and assigning equal weights to each of them, has become a standard approach in climate change impact studies. However, the lowered performance and robustness of hydrological models under contrasting climatic conditions has underlined the need to improve the evaluating strategies used in climate change impact assessments, particularly for complex hydrological projections as floods. Some studies have proposed more comprehensive hydrological models' evaluations, such as the DSST approach, as a way to better evaluate hydrological models' reliability and robustness under different climatic conditions. After these thorough evaluations some studies have proposed discarding hydrological models with unsatisfactory performances, while others have suggested weighting the hydrological models based on their performance. Currently, few studies have explored the potential impacts of such methodologies on flooding projections. Thus, to provide further insights on this ongoing discussion, this study explored the use of five weighting approaches

to combine five lumped hydrological models based on their robustness under contrasting conditions to evaluate their performance and potential impacts on flooding projections.

As observed in Figure 5.3, the individual performance of the five lumped hydrological models, as well as their combined performances with five different weighting techniques on warm-cold, cold-warm, dry-humid, and humid-dry contrasting conditions (DSST experiments) were evaluated using the KGE and NSE criteria (panels a and b). Overall, the results revealed that the different weighting methods were able to improve the streamflow simulations performance on the four contrasting sets of climate conditions. Even the relatively simple equal-weighting (EW) approach showed slightly better results than best single hydrological model, with slightly higher medians on both the KGE and NSE analyses. Thus, in line with other studies' recommendations, the use of multiple hydrological models, if possible, is recommended (Krysanova et al., 2018). Some differences were however observed between the four sets of climatic conditions evaluated with the DSST approach. The results revealed that the basins were generally more sensitive to some conditions than others. This can be observed in Figure 5.3, where the DSST experiments C-W and H-D showed generally lower evaluation performances with both the KGE and NSE criteria. In other words, the lumped hydrological models and their weighted combinations had more difficulties to simulate warmer and dryer conditions than those used to calibrate them, particularly dryer years. This is in line with previous studies showing that hydrological models' performance was often more affected by dryer climatic conditions than the opposite, thus recommending to favor calibrating on dryer conditions (Coron et al., 2012; Dakhlaoui et al., 2019; Motavita et al., 2019; Seiller et al., 2012; Seiller et al., 2015; Yang et al., 2020).

When looking at the individual performance between the five lumped hydrological models, it was observed that some models showed generally more robust performances among the different DSST experiments. This can be observed when comparing HM1 and HM4 against the other hydrological models. These two hydrological models showed the highest medians with both the KGE and NSE analyses, suggesting that these two hydrological models were more robust under contrasting climatic conditions. Nonetheless, when looking at the weighted

streamflow combinations, the results revealed that most weighted combinations showed better performances than the individual hydrological models. In other words, the results show that even when some hydrological models showed overall worse performance than others, useful information can be extracted to slightly improve their combined performance (Figure 5.3). This is confirmed in the statistical Kruskal-Wallis tests presented in Figure 5.4, where it is observed that the often best-performing weighted combination W3 shows statistically better performance than many single hydrological models, particularly in the NSE values that focus on flooding simulations performance (panel b). Figure 5.4 also reveals that the two best-performing hydrological models H1 and H4 perform in a statistically similar manner to the best-performing weighting method W3. Nonetheless, this approach shows slightly better and constant performance among the four sets of climatic conditions. Unlike the single hydrological models that vary their performance depending the contrasting set of climatic condition.

An important part of this study was including various basins covering different hydroclimatic regimes. As observed in Figure 5.1, the diversity of the study area allowed identifying two groups based on the basins' main flood-generating processes: (1) 38 snow-dominated basins and (2) 39 rain-dominated basins. As displayed in Figures 5.5 and 5.6, some differences were observed between the hydrological modelling and weighting performance on peak streamflow simulations over the two groups. Overall, it was observed that the hydrological models had more difficulties to simulate floods on contrasting conditions over the rain-dominated basins. Meaning that the hydrological models included in this study were more sensitive to climate conditions when simulating rain-driven floods. Yet, it was also observed that the best-performing weighting approach W3 was able to improve this diminished performance over most basins. This was observed over the four DSST experiments (see Figure 5.6).

5.5.2   Effects of hydrological model's weighting on flood projections

This section of the study consisted on evaluating the effects of weighting hydrological models on flooding projections by measuring the impacts of equally (EW) and unequally weighting

(W1-4) the different hydrological simulations based on the hydrological models' robustness to contrasting climatic conditions. As observed in Figure 5.7, the CRCM5-CanESM2 simulation showed warmer/more humid projections over the snow-dominated basins, and warmer/dryer projections over the rain-dominated basins. Thus, the weights and parameterizations obtained with the cold to warm (C-W) and dry to humid (D-H) experiments were used over the snow-dominated projections, and the weights and parameterizations obtained with the cold to warm (C-W) and humid to dry (H-D) experiments were used for the rain-dominated basins. The idea behind using the most contrasting conditions was to represent a potential situation in climate change impact studies where observed conditions can be strongly different than climate projections used as inputs for hydrological modelling. Therefore, it is important to underline that this analysis was not meant to find the best-performing method on the control period, but simply to quantify the impacts of the equally and unequally weighting hydrological models on the projected floods.

Figures 5.8 and 5.9 showed the differences between the unequal weighting methods (W1-4) and the reference EW approach on the RCM-driven mean seasonal peak streamflow. These analyses showed that unequally weighting the streamflow simulations can impact the magnitude of the simulated mean seasonal peak streamflows. For instance, it was systematically observed that W3 and W4 approaches were showing the largest impacts on seasonal flood magnitudes over most basins. This can be linked to the fact that the structure of these two methods allows them to have more freedom to assign weights over wider boundaries. Yet, the magnitude of the impacts on mean seasonal peak simulations varied between basins, seasons and periods. For instance, it was observed that the W3 and W4 methods simulated generally smaller peaks in the snow-dominated basins with peaks going down to -28%, particularly over the D-H experiment. Over the rain-dominated basins however, no clear effects were observed on the projected peak streamflows. Some basins showed larger and others smaller peak streamflows over the different seasons and periods. However, it was observed that some cases showed the largest impacts reaching up to 100% relative differences against the EW approach, especially over the H-D experiments. These generally larger relative differences suggest that the hydrological models included in this study were more sensitive to

precipitation changes than to temperature changes over both snow- and rain-dominated basins. This is line with previous studies showing that hydrological simulations were generally more sensitive to changes in precipitation than temperature (Her et al., 2019; Motavita et al., 2019), suggesting that reducing precipitation projections uncertainty can help increasing our confidence in hydrological simulations.

As observed in section 5.4.3.2, the weighting methods not only impacted the magnitude of the flood projections but also their climate change signals. Figures 5.10 and 5.11 showed that unequally weighting the lumped hydrological models' projections the magnitudes and, to a lesser extent, the directions of the projected seasonal flood changes. As observed in the peak streamflow magnitudes analysis (Figures 5.8 and 5.9), the methods W3 and W4 showed the largest impacts on climate change signals. Overall, the results revealed that weighting hydrological models impacted the magnitude of the climate change signal for most cases. Over the snow-dominated basins (see Figure 5.10), slightly smaller winter peak increases and stronger summer-fall peak decreases were observed over some basins, particularly with the W3 and W4 methods. These differences were more clearly observed in the C-W experiments where the winter decreases showed values 15 to 20 % and summer-fall decreases of about 20 to 25%. As observed in Figure 5.8, the results using the C-W experiment during the summer-fall months showed that the impact on mean seasonal peak streamflows was larger in the future periods than over the reference period, while over the D-H experiment the impact was similar among the three periods. Thus, this larger impact over the future periods can explain why the climate change signal was slightly more affected when the hydrological models were calibrated on colder conditions. Over the rain-dominated basins, the impact on the climate change signal varied between weighting methods, particularly during the dry (winter-spring) months. This larger variation can be explained by the generally low streamflows during these months. Small changes on low streamflows can appear as large relative differences due to their low values. During the flooding months (summer-fall), some basins showed stronger decreases when using the W3 and W4 methods, yet no systemic effect was observed.

5.5.3   Potential impacts of hydrological models' complexity

The different DSST experiments and analyses on observed and projected streamflows revealed that the robustness of hydrological models under contrasting climatic conditions not only varied between regions but also between hydrological models. It was observed that some hydrological models, such as HM1 and HM4, showed generally better performances and higher weights (results not shown) than the other hydrological models over the different sets of climatic conditions (see Figure 5.3). However, its is important to highlight that the wide variety of hydrological models available was not covered in this study. More particularly, distributed and/or more physically-based hydrological models were not tested. This limitation can have different implications in the study's results as it is expected that hydrological models relying on more physics-based process representation can provide more reliable streamflow simulations under a changing climate (Feng, Trnka, Hayes, & Zhang, 2017; Ricard, Sylvain, & Anctil, 2020). To address this question, additional DSST experiments were performed on two snow-dominated basins and two rain-dominated basins using Hydrotel, a semi-distributed and more physically-based hydrological model. Following the same methodological steps as for the lumped hydrological models, Hydrotel was calibrated on two 6-year periods of climate conditions that are contrasting to those projected by the CRCM5-CanESM2 simulations. Therefore, the snow- and rain-dominated basins went through the C-W/D-H and C-W/H-D sets of experiments, respectively. For each experiment, Hydrotel was then coupled with the CRCM5-CanESM bias-corrected precipitation and temperature outputs to obtain the RCM-driven streamflow simulations. The resulting RCM-driven mean annual streamflow simulations produced with the C-W and D-H/H-D experiments are presented in Figures 5.12 and 5.13, respectively. Each figure shows the mean annual hydrographs for the snow- and rain-dominated basins in panels a and b, respectively. By row, each panel shows the two selected basins. Their locations are indicated in Figure 5.3 with numbers 1 and 2. From left to right, the results for the 1976-2005, 2041-2070 and 2070-2099 periods are presented. To compare the Hydrotel simulations against the lumped hydrological models, the mean annual hydrographs obtained with the observations, the EW, and W3 approaches are presented.

Over the snow-dominated basins, Figure 5.12 shows that the mean annual hydrographs estimated with Hydrotel are more similar to observations in the control period. This is particularly clear in the first basin, a small catchment of 512 km^2 that shows spring peak streamflows closer to observations followed by W3. Over the future periods, it is observed that Hydrotel simulates higher peak streamflows, as in the reference period. In the summer-fall peaks, small differences are observed. Yet, the Hydrotel simulation is closer to the one estimated with the EW approach simulation. Smaller differences are observed on the larger basin (second row) during the reference period, with the Hydrotel and W3 approach simulations showing similar mean annual hydrographs in the control period. Yet larger differences are observed during the summer months, where the Hydrotel simulations shows generally smaller streamflows. Over the rain-dominated basins in panel b, Hydrotel shows more difficulties in simulating peak streamflows. Yet, it is during the low flows that Hydrotel shows simulations closer to the observations compared to those obtained with the EW and W3 approaches. This difference against the other approaches is maintained during the future periods.

Results for the D-H experiment presented in Figure 5.13 a) show some similarities with those from the C-W experiment. The small snow-dominated basin shows that the mean annual hydrograph estimated with Hydrotel is also closer to observations, with higher spring peak streamflows. These higher peak streamflows are also maintained during the future periods. In contrast to the results with the C-W experiment, the larger basin (second row) shows that the mean annual streamflow estimated with Hydrotel is not as close to the observations, suggesting that Hydrotel was more sensitive to precipitation changes as observed with the lumped hydrological models. In the case of the H-D experiment, over the first rain-dominated basin (Figure 5.13b), it is observed that Hydrotel shows more difficulties to simulate streamflow peaks when calibrated on the most humid years. Nonetheless, as observed in the C-W experiment, the low winter streamflows are closer to the observations, suggesting that this hydrological model is less impacted during the driest months. Over the second basin, it is observed that the EW approach shows slightly larger peak streamflows than observations, Hydrotel and the W3 approach over the reference period. During the future periods, small

differences are observed between simulations, particularly between the W3 approach and Hydrotel.
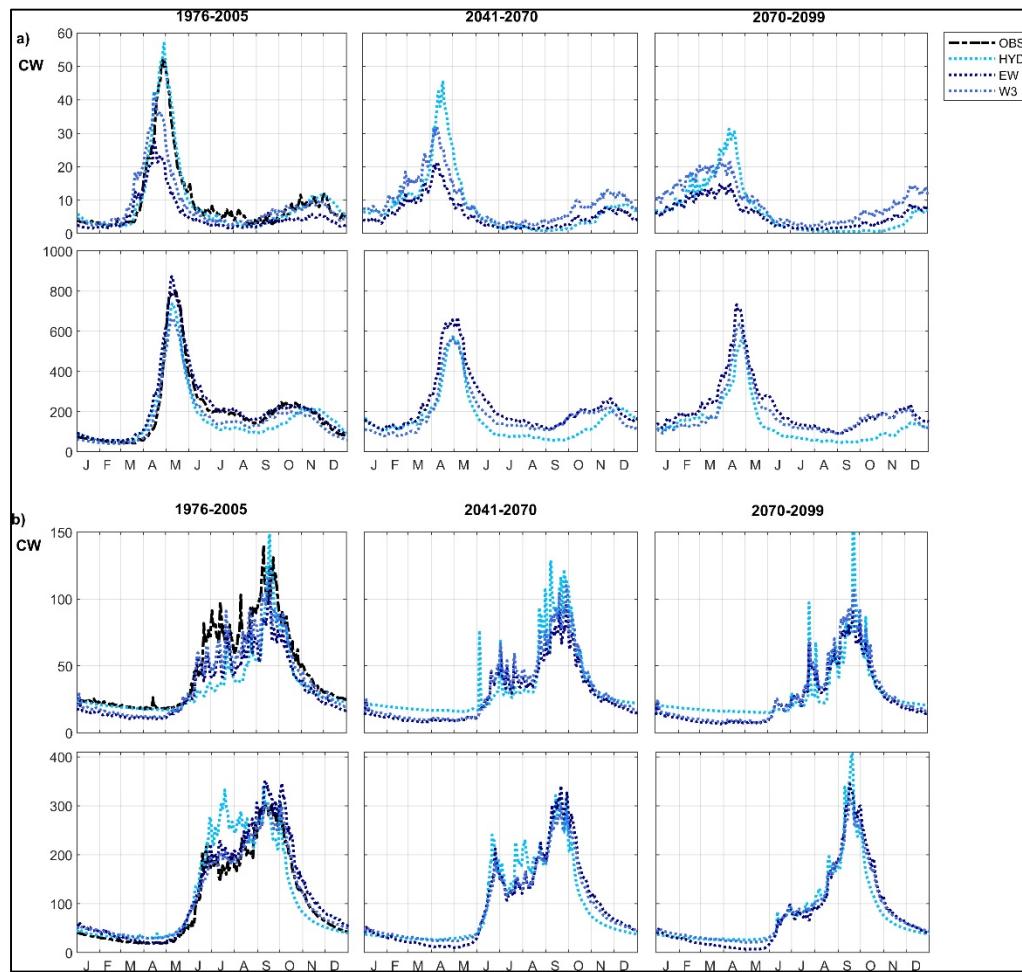
Figure 5.12 Mean annual hydrograph of two snow-dominated basins (panel a) and two rain-dominated basins (panel b) estimated with the cold calibrations. The mean annual hydrographs estimated with the historical observations (OBS), Hydrotel (HYD), the EW, and W3 approaches are presented for each basin. From left to right, the mean annual hydrographs for the 1976-2005, 2041-2070 and 2070-2099 periods are presented
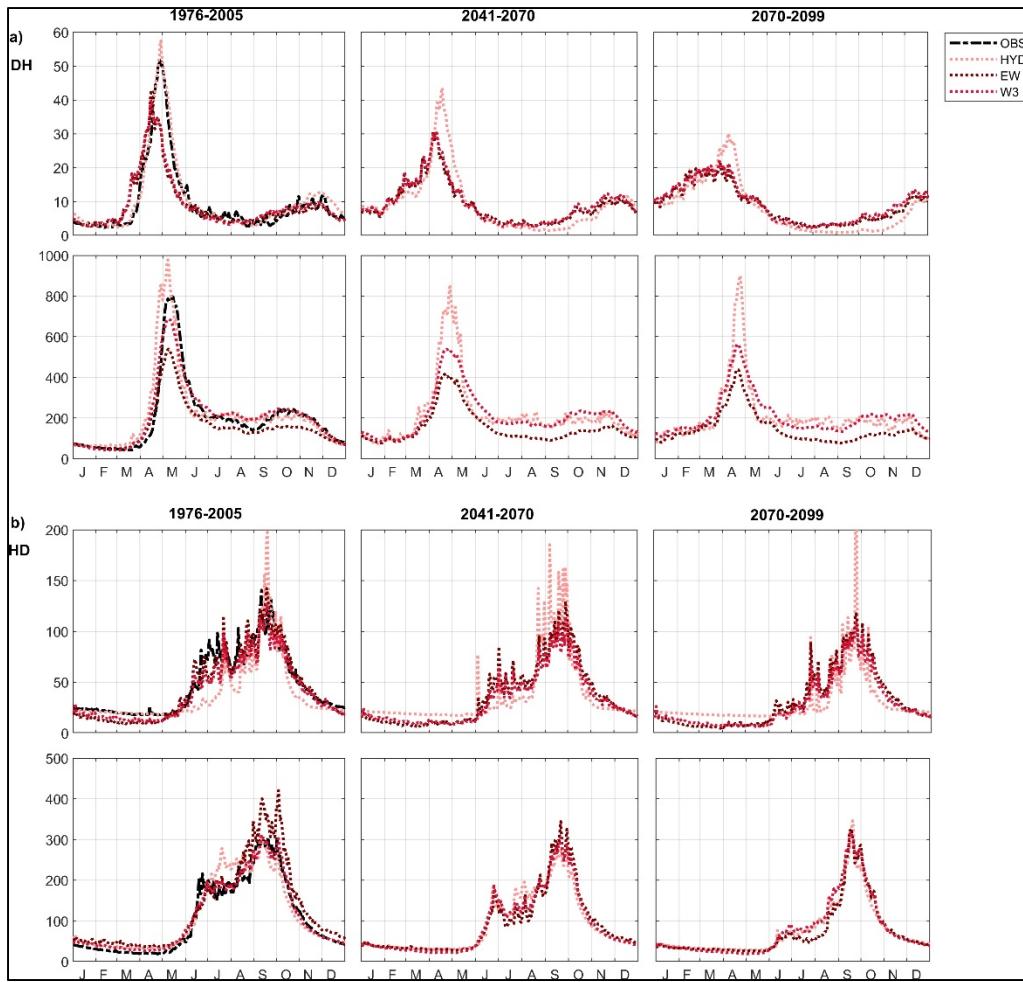
Figure 5.13 Mean annual hydrograph of two snow-dominated basins (panel a) and two rain-dominated basins (panel b) estimated with the dry and humid calibrations, respectively. The mean annual hydrographs estimated with the historical observations (OBS), Hydrotel (HYD), the EW, and W3 approaches are presented for each basin. From left to right, the mean annual hydrographs for the 1976-2005, 2041-2070 and 2070-2099 periods are presented

These additional results reveal that the semi-distributed and more physically-based hydrological model can bring some advantages compared to the ensemble of lumped hydrological models. Over the snow-dominated basins, it was observed that Hydrotel consistently showed simulations closer to the observations in both sets of contrasting climate conditions, particularly during the spring floods. This could be explained by a less sensitive snow accumulation and melting representation, an essential process in this type of basins where

floods are mainly driven by snowmelt. The lumped hydrological models however, use more simplistic degree-day modules for snow accumulation and melting, which can explain the overall difficulties to simulate in contrastingly warmer conditions. It was also observed that Hydrotel showed a more robust mean annual simulation in the smallest snow-dominated basin over both C-W and D-H experiments. As distributed models are able to simulate smaller areas in better detail than lumped models, it is expected that smaller catchments can benefit from a distributed simulation (Reed et al., 2004; Yang et al., 2020). Over the rain-dominated basins, the Hydrotel simulations showed generally more robust performance when simulating low flows in both DSST experiments. This can possibly be explained by a more adequate actual evapotranspiration representation that, as discussed in different studies (Dallaire et al., 2021; Seiller & Anctil, 2016), can be one of the main processes influencing streamflow simulations.

5.6   Conclusions

Relying on the best-performing hydrological models for a given region, or equally-weighting well-known hydrological models, has become a relatively simple and common approach to evaluate the impacts of climate change on local and regional hydrology (Chen, Brissette, Poulin, et al., 2011; Krysanova et al., 2018; Poulin et al., 2011; Roudier et al., 2016). This practice has been questioned by different studies arguing that relying on a single best hydrological model or discarding low-performing hydrological models can significantly impact the resulting ensemble (Gudmundsson et al., 2012; Zaherpour et al., 2018), and more importantly, that a good performance under historical conditions does not guarantee good performance in the future when different climate conditions are expected (Merz et al., 2011). Thus, this study aimed at providing further insights on this ongoing discussion by exploring the use of weighting techniques to combine hydrological models based on their performance over different climate conditions using a DSST approach. The effects of five different weighting methods were assessed on observed and projected flood simulations over a large sample of basins including (1) snow-dominated and (2) rain-dominated basins, according to their main flood-generating process. Five well-known lumped hydrological models were weighted using four different unequal-weighting methods to compare their performance

against the traditional equal-weighting multi-model approach. In addition, one more physically-based and semi-distributed hydrological model was tested and compared against the ensemble of five lumped hydrological model over four diverse selected basins. Finally, the effects of the different weighting approaches on flood projections were assessed using bias-corrected RCM simulations issued from the CRCM5 (one simulation for each of the studied domains covering both sets of snow- and rain-dominated basins).

The results showed that the hydrological models were generally more sensitive to changes in precipitation, particularly from humid to dry years. Yet, the weighting methods revealed that combining hydrological models, even with the relatively simplistic equal-weighting method, produced more robust performances than the best-performing hydrological models over the four DSST experiments evaluating warmer, colder, dryer and more humid conditions. Among the different weighting techniques, the unequally-weighted hydrological model ensembles showed slightly better performances than the equal-weighting approach and than the best lumped hydrological model over the different contrasting climate conditions, suggesting that weighting hydrological models based on their robustness can be useful to improve their combined performance and increase the confidence on hydrological models' transferability in time. Between the unequal-weighting techniques, the Granger-Ramanathan type A (referred to as W3) showed generally more robust performance than the other methods.

This study also showed that the choice of weighting method can have an impact on flood projections magnitudes and climate change signal. Among the different weighting methods, the Granger-Ramanathan methods (W3 and W4) had the largest impacts on the mean seasonal projected streamflows climate change signals, particularly in their magnitudes. Although these analyses were not meant to inform on the best weighting method for climate change impact studies, but to inform on the possible effects of weighting on flood projections, the additional analyses using a more physically-based semi-distributed hydrological model showed that this type of model can bring more robustness to small basins and low-flows simulations as generally less sensitivity to the different DSST experiments was observed. More work is clearly needed to improve the understanding on hydrological models' structures and reliability.

The results and methodological limitations of the present study, such as the study area, selected hydrological models and experimental design bring out additional research possibilities. For instance, including more basins with different flood-generating processes, as well as evaluating the robustness of different hydrological process formulations (e.g., evapotranspiration and snow) to contrasting climates can shed some light on the robustness of different hydrological representations under a changing climate.

CHAPTER 6

WEIGHTING CLIMATE MODELS FOR FLOOD PROJECTIONS: EFFECTS ON CONTRASTING HYDROCLIMATIC REGIONS

Mariana Castaneda-Gonzalez[a], Annie Poulin[b], Rabindranarth Romero-Lopez[c] and Richard Turcotte[d]

[a,b] Department of Construction Engineering, École de technologie supérieure, Montréal, Canada.

[c] Faculty of Civil Engineering, Universidad Veracruzana, Xalapa, Mexico.

[d] Ministère du développement durable, Environnement et Lutte contre les changements climatiques, Québec, Canada.

Abstract

Weighting climate models has recently become a more accepted approach but remains a topic of ongoing discussion, especially for analyses needed at regional scales such as flooding assessments. Few case studies have assessed the impacts of weighting climate models on streamflow projections yet the impact on flood projections still needs to be evaluated. Additionally, the methodological and location limitations of previous studies make it difficult to extrapolate their conclusions over regions with contrasting hydroclimatic regimes. Thus, this study evaluates the effects of different climate model's weighting approaches on flooding projections over multiple river basins with contrasting flood-generating processes. An ensemble of 24 global climate model (GCM) simulations coupled with a lumped hydrological model are used over 96 North America basins to generate 24 GCM-driven streamflow projections. Six weighting approaches comprising temperature- precipitation- and streamflow-based criteria were compared against the equal-weighting approach over a reference and two

156

future periods of 1976-2005, 2041-2070, and 2070-2099. Results showed that unequally-weighted ensembles can improve the mean hydrograph representation under historical conditions compared to the common equal-weighting approach. This was particularly observed with a proposed weighting approach that, unlike the common annual-based criteria, it evaluates the seasonal simulations performance. Moreover, results revealed that unequally-weighting climate models not only impacted the magnitude and climate change signal, but also reduced the uncertainty spread of flood projections, particularly over rain-dominated basins. These results underline the need to further evaluate the adequacy of equally weighting climate models, especially for variables with generally larger uncertainty at regional scale.

Keywords: Climate models weighting, flood projections, climate change impacts, contrasting hydroclimatic regions, hydrological ensemble weighting

6.1 Introduction

Flooding risk assessments are often based on projected climate change impacts on hydrology (Salman & Li, 2018). Such projections are commonly produced by using ensembles of post-processed (i.e., downscaled and/or bias-corrected) global or regional climate models' outputs to feed one or multiple hydrological models. Large ensembles of global climate models (GCMs) and regional climate models (RCMs) are often recommended to produce ensembles of GCM- or RCM-driven streamflow projections that account for the uncertainty associated to climate modelling (Giuntoli et al., 2018; Kundzewicz et al., 2018). How to manage and assess these large ensembles of hydro-climatic projections is a topic of ongoing discussion, especially for decision-making purposes (Kiesel et al., 2020; Knutti et al., 2017; Pechlivanidis, Gupta, & Bosshard, 2018). The dominant approach in hydrological impact studies, including flooding projections, is the so-called "climate models' democracy". This approach consists in giving identical weights to all the climate models of the ensemble, assuming that all members are equally plausible (Chen et al., 2017; Knutti, Furrer, Tebaldi, Cermak, & Meehl, 2010). However, different studies have questioned the climate models' democracy approach and proposed weighting or sub-selecting ensemble members depending on their adequacy to

predict a given variable for a specific purpose, especially for climate change impact studies at regional and local scales. The arguments supporting this approach include that, (1) many climate models often share or duplicate processes representations (Abramowitz et al., 2019; Eyring et al., 2019; Knutti et al., 2017) and, that (2) some climate models have shown more difficulties to represent mean regional climate than other climate models (Braconnot et al., 2012; Gleckler, Taylor, & Doutriaux, 2008), especially over extremes (Do et al., 2020; Giuntoli et al., 2018). Therefore, weighting or sub-selecting climate models has gained acceptance in recent years, particularly for climate change impact studies (Eyring et al., 2019; Kiesel et al., 2020; Knutti et al., 2017).

Different approaches to weight or sub-select climate models have been proposed and assessed on ensembles of climate projections (e.g., Knutti et al., 2017; Räisänen, Ruokolainen, & Ylhäisi, 2010; Sanderson et al., 2017; Xu et al., 2010). The main differences among them often include the criteria used to favor a given simulation such as, mean annual performance (e.g., Xu et al., 2010), independence (e.g.,Knutti et al., 2017; Sanderson et al., 2017) and convergence (e.g., Giorgi & Mearns, 2002) of one or various climate variables. In hydrological impact studies, some approaches have assessed weighting or sub-selecting methods on mean climate performance (e.g., Chen et al., 2017; Massoud, Espinoza, Guan, & Waliser, 2019; Massoud, Lee, Gibson, Loikith, & Waliser, 2020; Padrón, Gudmundsson, & Seneviratne, 2019; Ruane & McDermid, 2017), and few others have assessed weighting approaches based on mean streamflow (e.g., Wang et al., 2019; Yang et al., 2017). Among them, Padrón et al. (2019) evaluated the effects of weighting an ensemble of 36 GCM outputs on global precipitation projections. Using a modified Bayesian model averaging (BMA) method, higher weights were assigned to simulations with better performance against precipitation observations. Their results showed that projected precipitation extremes of the weighted-ensemble were less pronounced in Europe, Southern Africa, and Western North America, but more pronounced in the Amazon compared to projections using equal-weighting. Thus, performance-based weighting was recommended arguing that simulations that agree better with observations are likely to be more reliable. Focusing on GCM-driven streamflow projections at the basin scale, Chen et al. (2017) evaluated the effects of precipitation- and

temperature-based weighting on hydrological projections of a snow-dominated basin. Five weighting methods were tested on an ensemble of 28 GCM-outputs with and without post-processing. Their results showed a limited impact of climate models weighting on the GCM-driven streamflow projections. Thus, the climate model's democracy approach was suggested. More recently, Kolusu et al. (2021) evaluated the sensitivity of water resources projections to climate models weighting over two basins in eastern Africa. Four weighting methods, based on climate model's performance, independence and plausibility, were tested over an ensemble of 32 bias-corrected GCM outputs and the 32 GCM-driven streamflow projections. Similar to Chen et al. (2017), small effects of climate model's weighting were observed on their risk assessments compared to the overall ensemble spread. However, the use of climate-based weighting for streamflow studies has been questioned due to the non-linear relationship between climate variables and streamflow (Knutti et al., 2017; Wang et al., 2019). Thus, the use of streamflow-based weighting methods has been suggested as an alternative (Kiesel et al., 2020; Wang et al., 2019; Yang et al., 2017). For instance, Yang et al. (2017) compared equal-weighting against the BMA method to weight an ensemble of global monthly runoff projections issued from 31 Earth System Models (ESMs) based on decadal mean runoff performance. Significant regional differences including smaller runoff increase from the weighted projections against equal-weighting were observed over northern latitudes, yet more pronounced runoff decreases in Amazonia and sub-Saharan Africa. Using a different approach, Wang et al. (2019) evaluated the use of climate models weighting on GCM-driven hydrological projections simulated with a lumped hydrological model. Eight weighting methods, based on raw and bias-corrected mean annual climate and GCM-driven streamflow performance were tested on two different basins. The results showed that unequal weighting improved the simulation and reduced biases of the ensemble mean when using raw GCM outputs during the reference period. However, when the GCM-outputs were bias-corrected, the impact of weighting was limited. Thus, equal weighting was still recommended. More recently, Kiesel et al. (2020) compared eight different methods to weight or sub-select 16 bias-corrected climate model outputs from EURO-CORDEX over the Upper Danube basin. To evaluate the performance of weighting and sub-selection methods, the historical streamflow observations were divided into a reference and an evaluation period. The results revealed that the choice of

method influenced the streamflow projections as much as the actual climate change signal. Moreover, their results suggested that methods maintaining more information showed better performance than methods sub-selecting a single best performing model.

Contrasting conclusions are observed regarding the use of climate models weighting for hydrological projections highlighting the need for further analyses. Among them, the literature shows that there is still no agreement on the method or approach on how to assign weights to climate models for hydrological studies. Moreover, to the authors' knowledge, no study has focused on assessing the effects of different climate models weighting on flooding projections. As previously observed, studies covering various regions have shown different impacts depending on the region (Padrón et al., 2019; Yang et al., 2017). This is of particular concern for flooding projections due to the complexity and variety of dominant flooding processes at the basin scale. In light of these research gaps, the aim of this study is to evaluate the effects of weighting climate models on flooding projections. And, at the same time, to inform the ongoing discussion by evaluating the effects of different climate model's weighting criteria on flood simulations over basins with contrasting hydrometeorological regimes.

6.2   Study area and data

In this study, 96 North American basins with contrasting flood-generating process were selected (see Figure 6.1). This includes 50 snow-dominated basins located in the Canadian province of Quebec, and 46 rain-dominated basins located in the country of Mexico. The upper and lower panels show the mean total annual precipitation (mm) and mean annual temperature (°C) for the snow- and rain-dominated basins, respectively. As observed in Figure 6.1, these two groups of basins differ in climatic regimes. In terms of precipitation, the 96 basins show mean total annual precipitation that varies from about 400 up to 3110 mm per year and mean annual temperatures ranging from about -2.6 up to 23.6 °C.

The basin data used in this study included daily historical records of minimum temperature, maximum temperature, precipitation, and streamflow. All meteorological and hydrometric

data consisted of time series with a minimum length of 30 years between 1950 and 2013 obtained from the Hydrometeorological Sandbox – École de Technologie Supérieure (HYSETS) database (Arsenault, Brissette, Martel, et al., 2020).
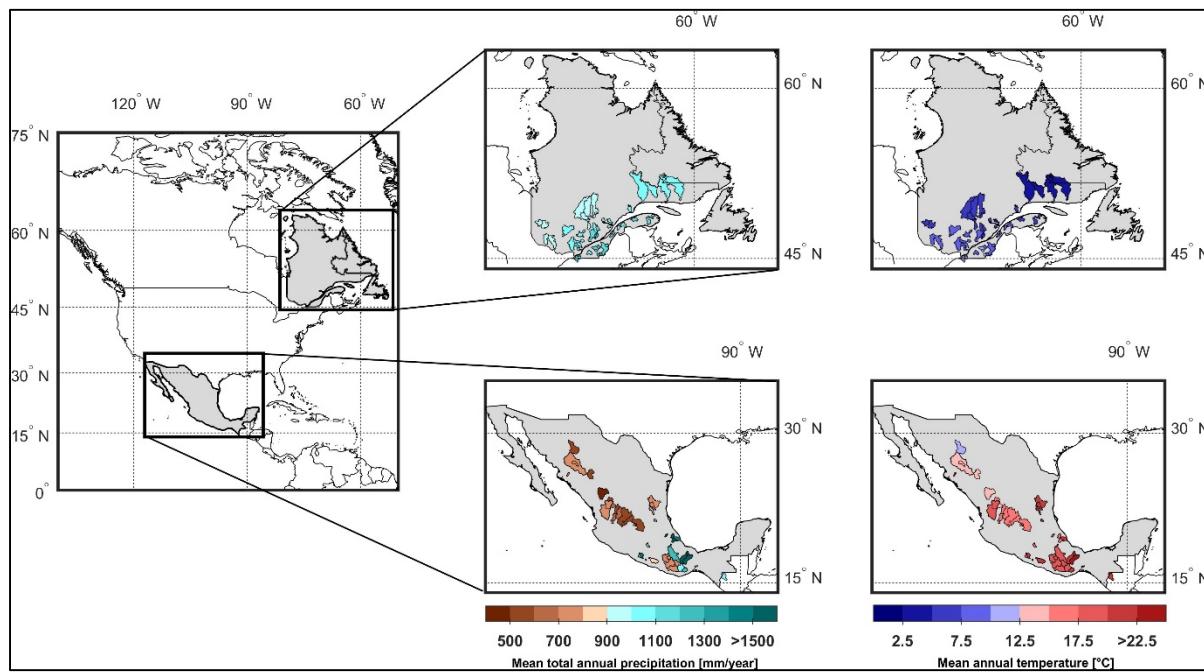
Figure 6.1 Location of the 50 basins in Quebec (upper panels) and the 46 basins in Mexico (lower panels) used in the study. The upper and lower panels show the mean total annual precipitation (mm) and mean annual temperature (°C) for the snow-dominated and rainfall-dominated basins, respectively

6.3 Methodology

The methodology used in this study consisted of four main steps: (1) building an ensemble of GCM simulations, (2) calibrating and validating the hydrological model to couple it with the GCM-ensemble outputs, (3) applying six different climate model's weighting approaches, and (4) analyzing the resulting weighted streamflows.

6.3.1 Climate simulations

The first step consisted in selecting different raw GCM simulations. The reasoning behind using raw GCM-outputs is that when fitting climate simulations to observations, the effects of weighting are directly affected as the same observations are then used to identify the climate simulations' weights. This could be thus linked with the limited impacts of weighting on bias-corrected climate simulations and streamflow projections observed in previous studies. Additionally, it has been suggested that bias-correction performance over the reference period is not preserved during future periods (Chen, Brissette, & Caya, 2020) making it problematic to interpret our results over different time frames. Thus, to avoid these issues, an ensemble of twenty-four raw GCM simulations issued from the Coupled Model Intercomparison Project Phase 5 (CMIP5) database (Taylor et al., 2011) were selected. This climate ensemble includes outputs from fourteen different modeling centers and varied spatial resolutions that allow having a diversity of GCMs from the CMIP5 database. Please refer to Table S6.2 in supplementary material – S1 for further details on the GCM-ensemble members. The outputs issued from each member of the GCM-ensemble comprised three daily variables, (1) precipitation, (2) minimum temperature and, (3) maximum temperature. Each dataset covered the reference period of 1976-2005 and two future horizons of 2041-2070 and 2070-2099 under the high-emission scenario, the Representative Concentration Pathway (RCP) 8.5.

6.3.2 Hydrological modelling

The second methodological step comprised the calibration and validation of the selected hydrological model to then couple it with the GCM-outputs and produce the ensemble of GCM-driven streamflow simulations. In this study, the lumped empirical *GR4J* hydrological model (Perrin et al., 2003) combined with the snow module CemaNeige (Valéry et al., 2014) was used. GR4J is a simple rainfall-runoff model with four parameters. However, the snow accumulation, snowmelt and evapotranspiration processes are not directly estimated by GR4J. Thus, the snow module CemaNeige and the Oudin evapotranspiration formulation (Oudin et al., 2005) were added to allow its application over the diverse study area. This means that two

162

parameters from the snow module were added making a total of 6 parameters to calibrate (i.e., 4 from GR4J and 2 from CemaNeige). The inputs required by GR4J-CemaNeige consist of continuous series of daily precipitation, mean temperature and the potential evapotranspiration calculated with the Oudin formulation. GR4J-CemaNeige combined with the Oudin formulation has shown satisfactory performances over a diversity of basins and applications (e.g., Coron et al., 2012; Dallaire et al., 2021) supporting its use in this study. The six parameters of the GR4J-CemaNige hydrological model were calibrated with the Shuffled Complex Evolution (SCE) algorithm (Duan et al., 1994) using the Kling-Gupta Efficiency (KGE) criterion (Kling et al., 2012) as objective function. According to Knoben et al. (2019), KGE values larger than ≈ -0.41 indicate that the simulation has higher skill than the mean observations. For this study, KGE values above 0.5 were considered acceptable. In this stage, an initial calibration/validation was performed to evaluate and validate the hydrological model's performance. This first calibration/validation was performed over the odd/even years during the available period of each basin (i.e., minimum of 30 years between 1950-2013). The rationale behind this approach is that by dividing the period on odd/even years instead of splitting it into two consecutive periods, the climate's change and natural variability can be considered (Castaneda-Gonzalez et al., 2019; Essou, Brissette, & Lucas-Picher, 2017). After this calibration/validation process all basins were recalibrated over the full period. This is based on a recent study showing more robust hydrological model's performance when calibrating over the full-time period (Arsenault et al., 2018). Thus, to ensure robust parametrization for the GR4J hydrological model, an additional calibration was performed over the full-time period available on each basin after a common calibration/validation process.

6.3.3   Weighting methods

The third step consisted in applying different climate- and streamflow-based weighting approaches. As previously stated, this study aims at evaluating the effects of different weighting criteria on flooding projections. Thus, six unequal-weighting approaches including three climate-based approaches comprising temperature and/or precipitation criteria, and three streamflow-based approaches were tested. All unequal-weighting approaches use two main

weighting methods, the Reliability Ensemble Averaging (REA) and the Upgraded REA (UREA). *REA* is a weighting method developed by Giorgi et Mearns (2002) that allows assigning weights to each member of a GCM simulations ensemble to minimize the contribution of members with poor performance. The performance criterion is based on reliability factors resulting from two elements, (1) each simulation's fit to historical records, and (2) a measure of the future projection convergence to the REA-weighted average. Both elements are often calculated with annual mean data. This method was initially applied over ensembles of GCM-temperature and -precipitation datasets, yet recent applications have also used it over GCM-driven streamflow ensembles (e.g., Kiesel et al., 2020; Mani & Tsai, 2017). The UREA method developed by Xu et al. (2010) proposed two major changes to the REA method. Instead of the future projection convergence criterion, the UREA method proposed adding multiple variables and statistics to the performance criteria. This allowed including two or more variables, such as precipitation and temperature, and other statistics such as interannual standard deviation and interannual coefficient of variation to define the weights. The UREA method has been applied in different regions and studies for climate projections (e.g., Chen et al., 2017; Colorado-Ruiz, Cavazos, Salinas, De Grau, & Ayala, 2018; Singh & AchutaRao, 2020) and GCM-driven streamflow projections (Wang et al., 2019). As previously presented, these two methods not only have been applied in different studies but also have the great advantage of including the calculation of uncertainty ranges. This feature will thus allow to measure and compare the uncertainty spreads for all weighting approaches. Further details on the weights and uncertainty ranges calculations with the REA and UREA methods are provided in Supplementary material - S2.

- *EW.* The equal weighting   method assigns identical weights to all climate models.
- *W1.* This approach uses the REA method to assign weights according to each climate model's historical mean annual temperature performance.
- *W2.* This approach uses the REA method to assign weights according to each climate model's historical mean annual precipitation performance.
- *W3.* This approach uses the UREA method to assign weights according to each climate model's historical mean annual precipitation and temperature performance.

- **W4.** This approach uses the REA method to assign weights according to each climate model's historical mean annual streamflow performance.

- **W5.** This approach uses the UREA method to assign weights according to each climate model's historical mean annual streamflow performance.

- **W6.** This customized approach proposes using the UREA method to assign weights according to each climate model's historical mean seasonal streamflow performance. By using a seasonal-based criterion instead of the commonly used annual-based criterion, it is expected that the varying dominance of hydrological processes between seasons will be better considered and more adequate for hydrological studies.

6.3.4   Data analysis

To investigate the effects of climate models' weighting on ensembles of flooding projections, four metrics were selected and evaluated over the winter (December, January, February, DJF), spring (March, April, May, MAM), summer (June, July, August, JJA) and fall (September, October and November, SON) seasons for a reference and two future periods of 1976-2005, 2041-2070 and 2070-2099. Further details on the calculation of each metric are given in Supplementary material – S3.

- **Mean annual and seasonal hydrograph representation.** This first metric uses the Kling-Gupta Efficiency criterion to compare each weighted-ensemble mean hydrograph against the observed mean hydrograph to evaluate the mean annual and seasonal streamflow representation during the reference period.

- **Mean seasonal peak relative bias.** This second comparison between the weighted-ensembles is measured in terms of relative bias $RB(\%)$ between the mean seasonal peak streamflow of a given unequal-weighting approach (i.e., W1-W6) and the equal-weighting approach (i.e., EW) used as reference. The aim is to evaluate the impact of unequal-weighting on mean seasonal peak streamflows compared to the equal-weighting approach over reference and future periods.

- ***Climate change signal.*** The third metric is used to compare the climate change signals of all weighting approaches. The climate change signal of each approach is estimated by comparing the mean seasonal peak streamflow of a given approach during a future period against its mean seasonal peak streamflow during the reference period in terms of relative bias (%).

- ***Seasonal streamflow-ensemble spread.*** The fourth comparison consists in measuring the seasonal spread of the mean hydrographs estimated with the different weighting approaches. The standard deviation is used to measure the seasonal spreads of the mean weighted hydrographs over the reference and future periods. The idea is to compare the impacts of unequal- and equal-weighting approaches on the seasonal uncertainty spreads of the simulated streamflows.

6.4 Results

6.4.1 Hydrological modelling performance

The calibration and validation results of the GR4J hydrological model are presented in this section. Table 6.1 presents the summary of the hydrological model performance over all 96 basins. This table presents the median, 5th and 95th percentiles of the KGE-values distribution obtained during the calibration, validation, and final calibration over the full-time period (see section 6.3.2 for details).

Table 6.1 Overview of the KGE-values distribution of the calibration /validation and full-period calibration for GR4J hydrological model over all 96 basins

KGE-Calibration			KGE-Validation			KGE-Full calibration		
5th percentile	Median	95th percentile	5th percentile	Median	95th percentile	5th percentile	Median	95th percentile
0.61	0.90	0.95	0.58	0.85	0.94	0.60	0.87	0.94

Overall, the distribution of KGE-values obtained during the initial calibration and validation periods show a satisfactory performance with median values of 0.9 and 0.85, respectively. It is observed that the KGE-values distribution obtained during calibration and validation periods are similar, indicating a consistent satisfactory performance during validation over all 96 basins. The final calibration over the full-time period also shows a satisfactory performance of 0.87 median KGE-values with all basins showing values above 0.5, validating its use for the following methodological stages.

6.4.2   Weights

As previously described in section 6.2, the studied basins were divided into two groups based on their main flood-generating processes. Thus, this and the following sections will present the results for (1) the 50 snow-dominated basins group and (2) the 46 rain-dominated basins group. Figures 6.2 and 6.3 present the weights calculated with the climate- and streamflow-based weighting approaches, respectively. Each figure shows the weights calculated over the snow- and rain-dominated basins (column a and b, respectively).

The weights calculated with climate-based approaches (Figure 6.2) show that each approach yielded different results, with W3 showing the most heterogeneous weights (i.e., most contrasted shades) over both basin groups. When comparing W1 and W2, it is observed that snow-dominated basins show stronger weight differences when using W1 (temperature-based), while rain-dominated basins show more contrasted weights with W2 (precipitation-based). It is also observed that different approaches agree on assigning higher/lower weights to some climate models. This is clearly observed when comparing W1 and W3 over snow-dominated basins, where some climate models (e.g., 5-7, 20 & 21) received lower weights on both methods over the majority of snow-dominated basins (column a). This is also observed over the rain-dominated basins (column b), where both W2 and W3 have assigned higher weights to climate models 14 and 15 over some basins.
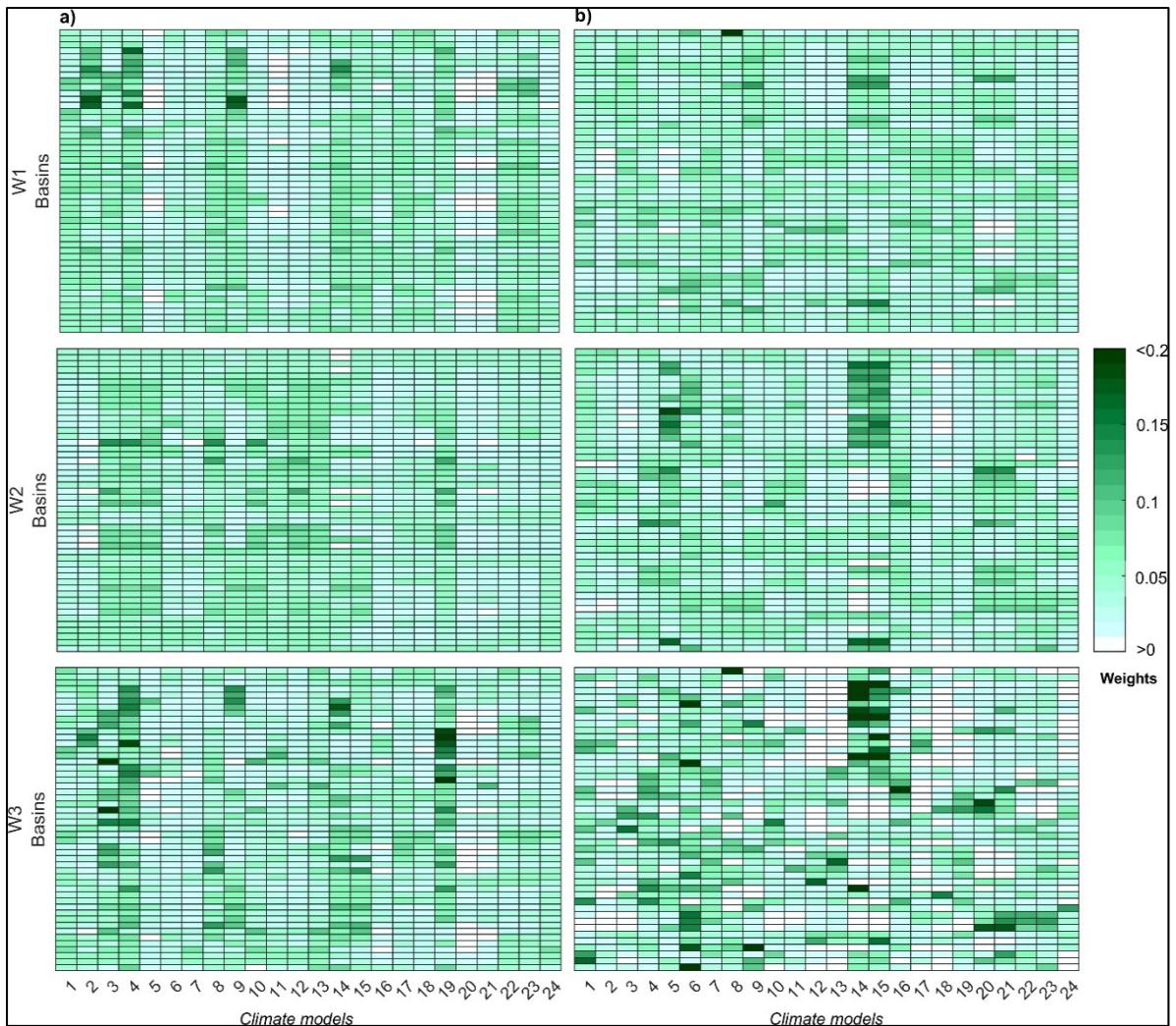
Figure 6.2 Weights obtained from the three climate-based weighting approaches (i.e., W1-3) for the snow-dominated basins (panels a) and the rainfall-dominated basins (panels b). Each panel shows climate models on x-axis and basins on y-axis. Shades of green indicate the weight assigned to the climate model, with darker shades indicating higher weights

The weights calculated with streamflow-based approaches presented in Figure 6.3 show more contrasted values between approaches. Over both basin groups, W6 shows the most contrasting weights with the highest and lowest weight values. While W4 and W5 show generally similar weights, especially over the snow-dominated basins. As observed over the climate-based weights (Figure 6.3), some climate models have been clearly disfavored over most snow-dominated basins (e.g., 1, 2, 6-8) while over the rain-dominated basins more diversity is

observed between basins. Nonetheless, it is observed over the rain-dominated basins that some climate models have been consistently favored/disfavored over the three streamflow-based weighting approaches (e.g., climate models 14 & 15). Looking at both climate- and streamflow-based weighting results, it is observed that W6, the weighting approach using the mean seasonal streamflow as evaluation criteria, shows the most contrasted weights among all methods.
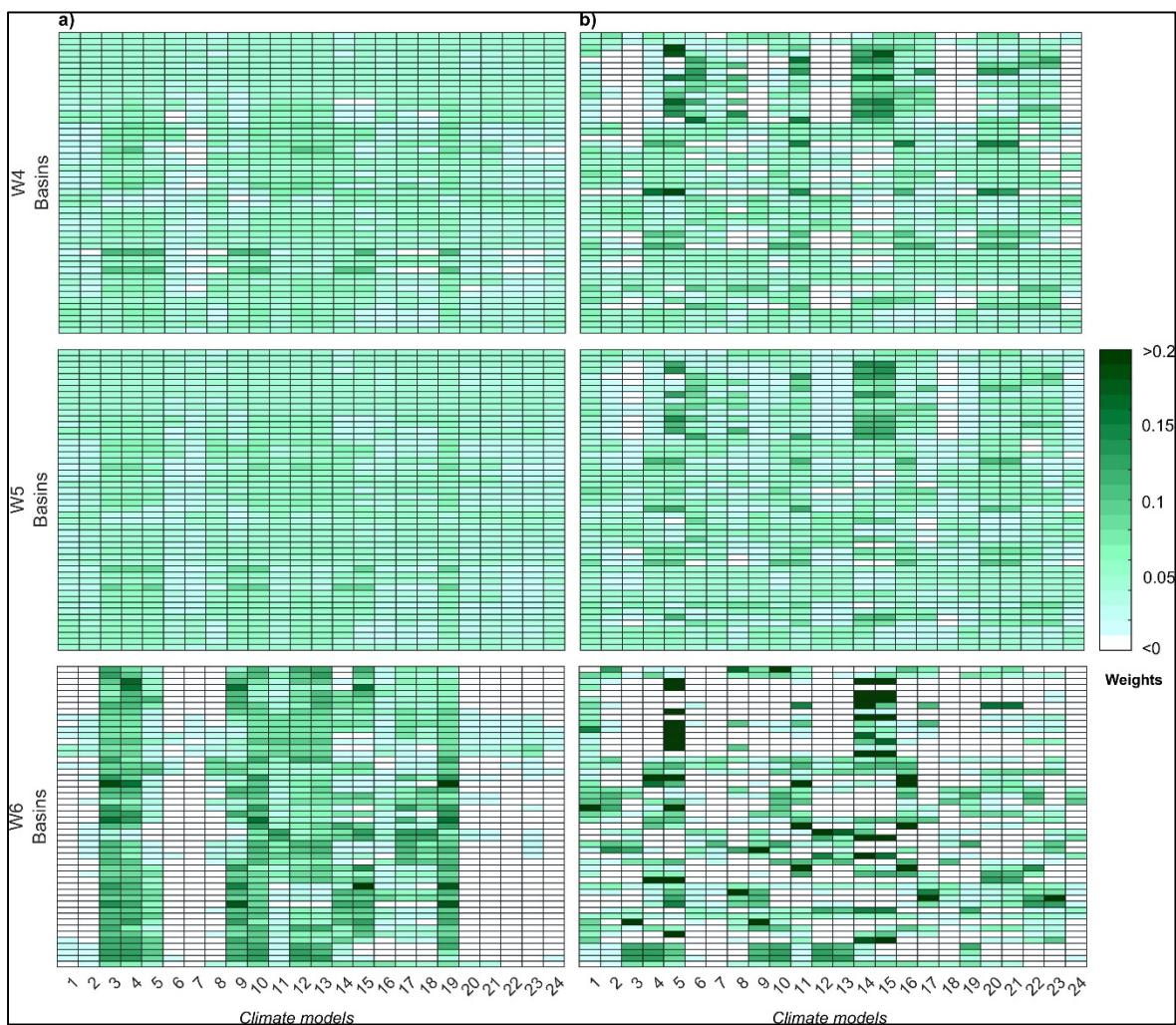

Figure 6.3 Weights obtained from the three streamflow-based weighting approaches (i.e., W4-6) for the snow-dominated basins (panels a) and the rainfall-dominated basins (panels b). Each panel shows climate models on x-axis and basins on y-axis. Shades of green indicate the weight assigned to the climate model, with darker shades indicating higher weights

6.4.3 Mean annual and seasonal hydrograph representation

The mean annual weighted and observed hydrographs were compared using the KGE criterion during the reference period. Figure 6.4 shows the KGE values obtained between the weighted and observed mean annual hydrographs for the snow- and rain-dominated on rows a and b, respectively. The KGE values are presented for the mean annual and seasonal hydrographs (from left to right). Each panel shows the weighting approaches on the x-axis and all basins sorted by ascending mean annual total precipitation (MATP, in mm/year) on the y-axis.
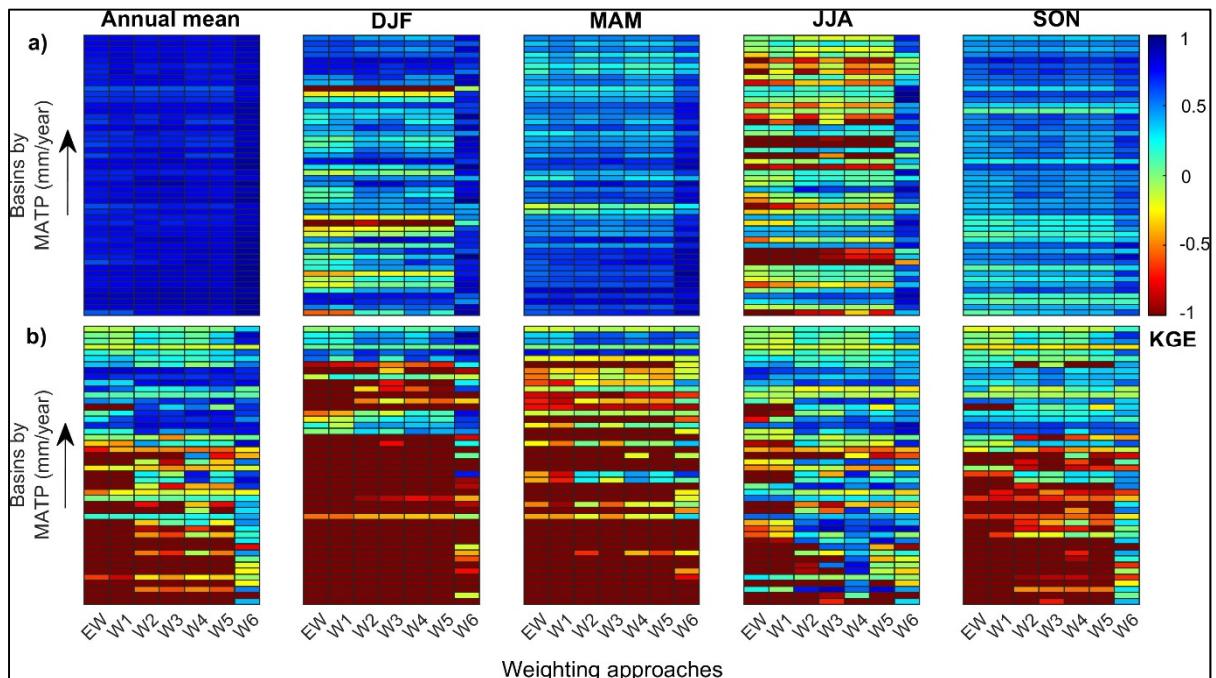
Figure 6.4 KGE values obtained from comparing weighted-ensembles mean annual hydrograph against the observed mean annual hydrograph during the reference period for the snow-dominated basins (panels a) and the rainfall-dominated basins (panels b). Each panel shows climate weighting approaches on x-axis and basins on y-axis

The results show that mean annual and seasonal hydrographs are generally better represented on snow-dominated basins with all weighting approaches. This is clearly observed over the first column where all snow-dominated basins show higher KGE values compared to the rain-dominated basins. Concerning weighting approaches, it is observed that W6 shows generally

higher KGE values than the other approaches over most snow- and rain-dominated basins. This is observed over the annual and seasonal analyses. The rain-dominated basins show more diverse KGE values, with the driest basins showing continuously negative KGE values, particularly during the dry months (DJF & MAM). These basins are also showing the overall worse mean annual hydrograph representation. Nonetheless, W6 shows improvements over the mean annual and seasonal hydrograph representation compared to the other approaches.

6.4.4   Impact on mean seasonal streamflow

Figures 6.5 and 6.6 show the relative bias between the mean seasonal peak streamflow of the unequally-weighted ensembles (W1-W6) against the equally-weighted (EW) ensemble over the snow- and rain-dominated basins, respectively. Results are presented for each season (top to bottom) and period (left to right).

Overall, most snow-dominated basins show smaller mean seasonal peak values estimated with the unequally-weighted ensembles than with the equally-weighted ensemble during winter, summer and fall seasons (panels a, c & d) over reference and future periods. The negative relative biases vary between weighting approaches with the streamflow-based approach W6 showing the largest difference against EW. During the spring flood however, it is observed that more basins show a positive relative bias. This is especially observed during the reference period where most basins show median relative biases of about +10% when using the W6. It is also observed, over all seasons, that relative differences decrease when moving toward future periods. And in the case of W6, many basins changed from positive to negative relative biases.

Over the rain-dominated basins (see Figure 6.6), most basins show smaller mean seasonal peak streamflows estimated with the six unequal-weighting approaches than with the equal-weighting approach. As observed with the snow-dominated basins, W6 shows the largest differences against EW. Yet, the rain-dominated basins show generally larger relative differences with median values of up to -50%. During winter and spring months (panels a and b), W6 shows clearly larger differences than the other approaches, while during summer and

fall months (panels c and d) both W4 and W6 show similar medians and distribution spreads. This is especially observed during the summer months.

Figure 6.5 Relative bias (%) between the unequally-weighted mean seasonal peak streamflows of the three climate-based (in green) and the three streamflow-based (in blue) ensembles and the equally-weighted ensemble for the snow-dominated basins over the 1976-2005, 2041-2070 and 2070-2099 periods (from left to right panels). By row, results for the winter (panel a), spring (panel b), summer (panel c) and fall (panel d) months are presented
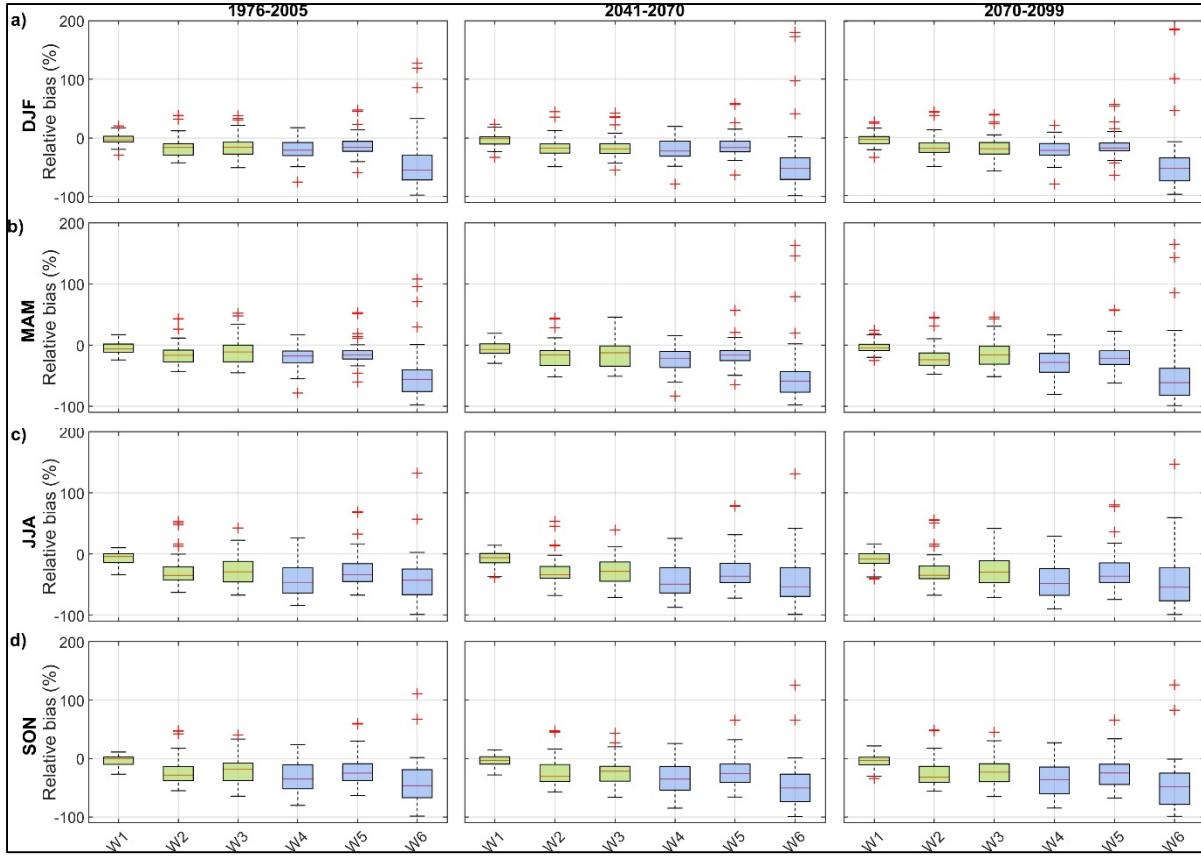
Figure 6.6 Relative bias (%) between the unequally-weighted mean seasonal peak streamflows of the three climate-based (in green) and the three streamflow-based (in blue) ensembles and the equally-weighted ensemble for the rain-dominated basins over the 1976-2005, 2041-2070 and 2070-2099 periods (from left to right panels). By row, results for the winter (panel a), spring (panel b), summer (panel c) and fall (panel d) months are presented

6.4.5   Impacts on the climate change signal

Figures 6.7 and 6.8 show the climate change signals measured in terms of relative bias (%) between the reference and future mean seasonal peak streamflows of all weighting approaches over the snow- and rain-dominated basins, respectively. The results are presented for each season (top to bottom) and future horizon (left to right). Each panel shows the weighting approaches on the x-axis and the basins sorted by ascending mean annual total precipitation (MATP, in mm/year) on the y-axis. Climate change signals over the snow-dominated basins show similar results among most weighting approaches. Yet, W6 shows slightly larger signal

differences against the other approaches over both future horizons. This is especially observed during the spring flood (panel b), where W6 shows opposite signals compared to the other approaches over some basins. During winter and summer months (panels a and c, respectively), most approaches agree on the signal direction showing overall flood increases and decreases, respectively. Yet, different magnitudes are observed as some approaches show slightly larger flood increases in the winter (panel a) and smaller flood decreases in the summer (panel c), particularly W6. During the fall (panel d), smaller effects are generally observed between weighting approaches, yet some approaches (e.g., W3) show slightly larger flood increases over both future horizons.

Over the rain-dominated basins larger diversity of climate change signals are observed between basins, periods, and weighting approaches. No clear differences are observed between unequal-weighting approaches and equal-weighting as climate change signals between approaches often disagree among many basins and future periods. This can be particularly observed during summer and fall months (panels c and d). It is however observed that unequally-weighted ensembles increased the magnitude of the climate change signal over some basins. Meaning that flood increases or decreases predicted from the equal-weighting approach are larger in magnitude for some unequal-weighting approaches such as W3 and W6. This is observed during all seasons and future horizons.
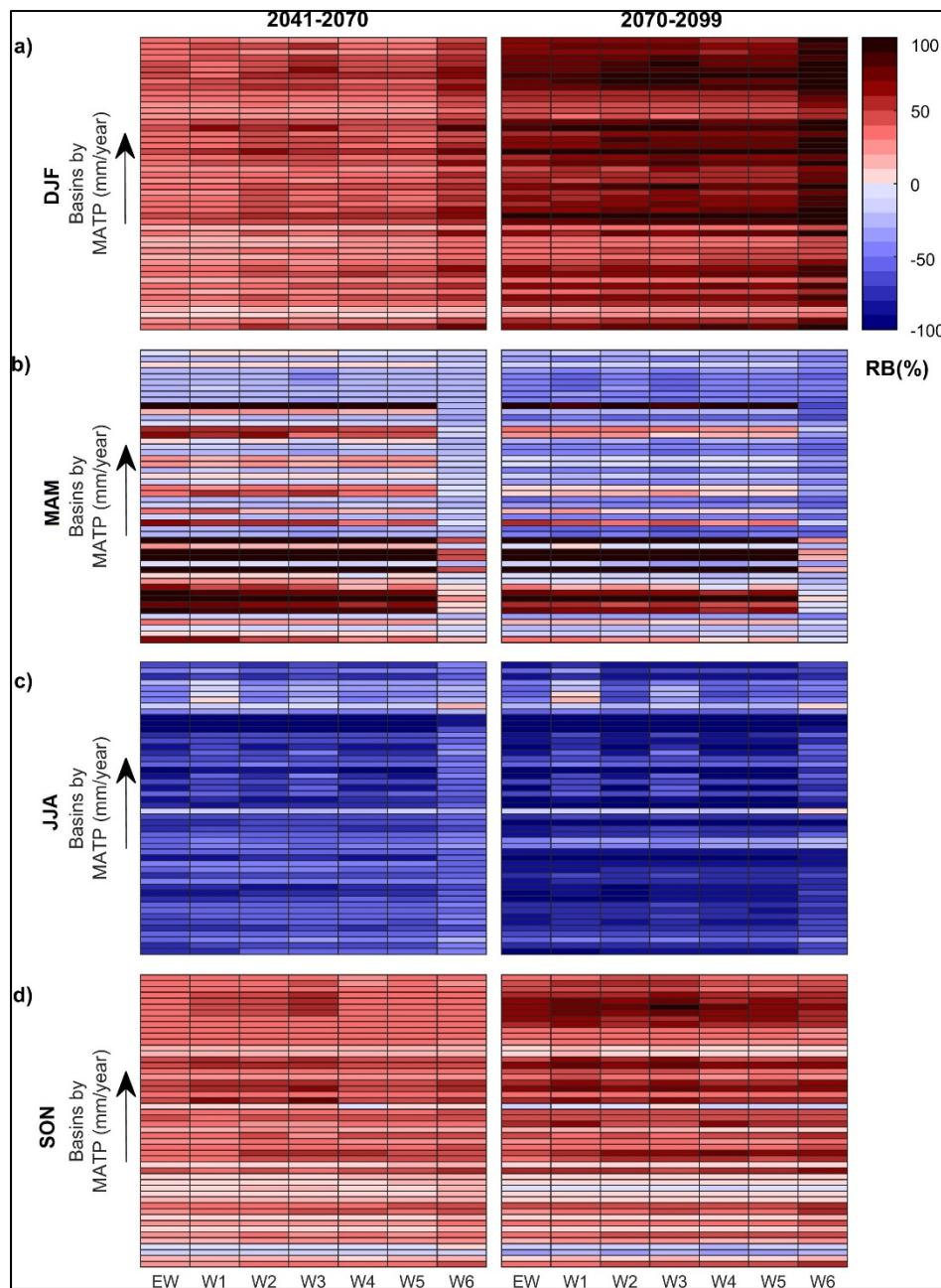
Figure 6.7 Climate change signal (%) of all equally and unequally weighted ensembles for the snow-dominated basins over the 2041-2070 and 2070-2099 periods (from left to right panels). By row, results for the winter (panel a), spring (panel b), summer (panel c) and fall (panel d) months are presented. Shades of red indicate positive relative bias and shades of blue indicate negative relative biases
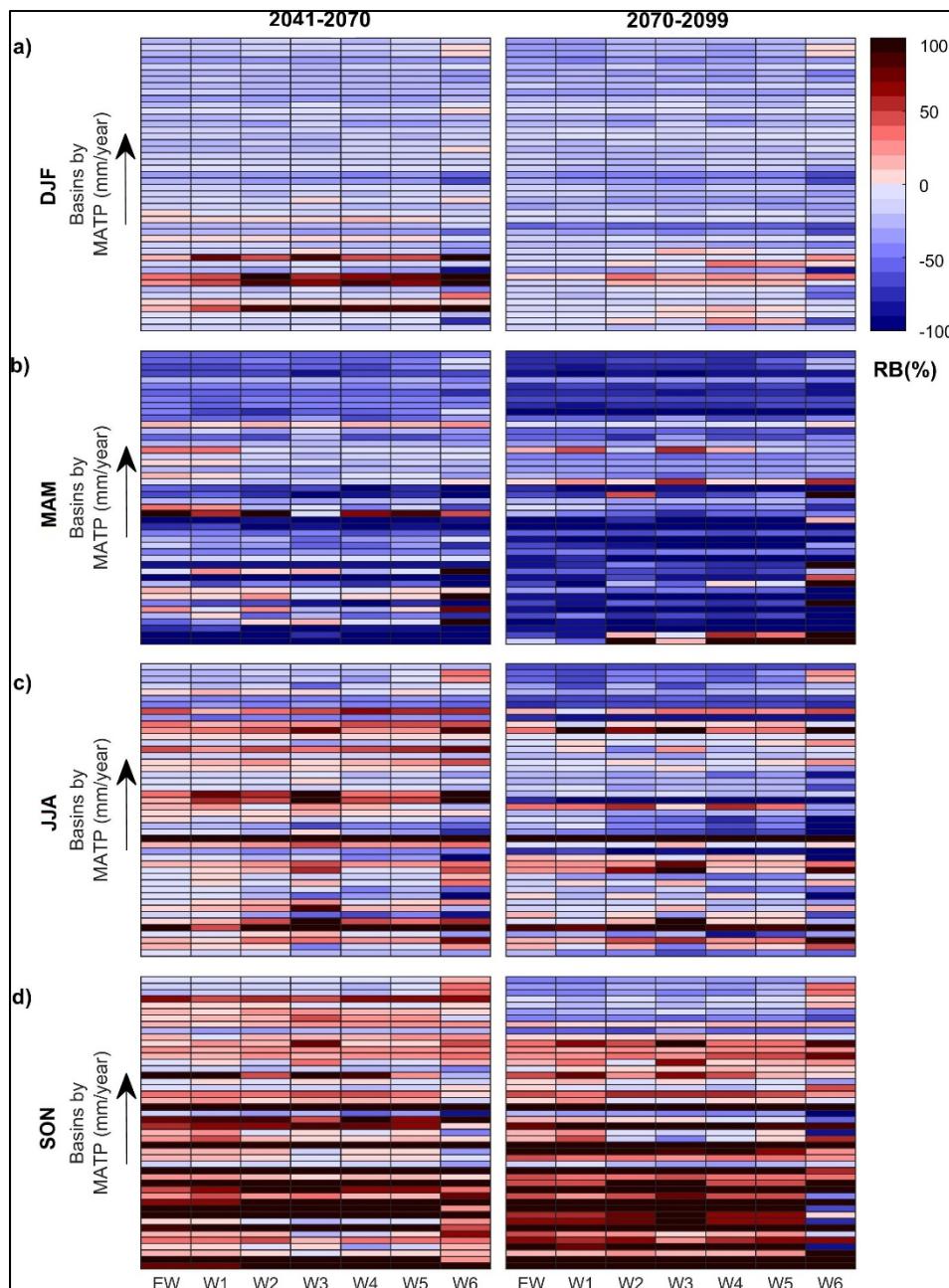
Figure 6.8 Climate change signal (%) of all equally and unequally weighted ensembles for the snow-dominated basins over the 2041-2070 and 2070-2099 periods (from left to right panels). By row, results for the winter (panel a), spring (panel b), summer (panel c) and fall (panel d) months are presented. Shades of red indicate positive relative bias and shades of blue indicate negative relative biases

6.4.6   Impacts on the streamflow-ensemble spread

Figure 6.9 and 6.10 show the boxplots of standard deviations (m^3/s) estimated for the different weighted streamflow-ensembles spread over the snow- and rain-dominated basins, respectively. Results are presented for each season (top to bottom) and period (left to right). It is generally observed over both figures that all the unequally-weighted ensembles reduced the streamflow-ensemble spread over all seasons. These effects are also observed over reference and future periods. Results over the snow-dominated basins show that unequal-weighting approaches can reduce the ensemble spread, with median standard deviations of about 50 to 67% smaller than the EW approach. These basins also show slightly smaller median standard deviations with the W6 approach, especially during summer and fall months (panels c and d). This is also observed over the rain-dominated basins. Yet, their median standard deviations show values of up to 80% smaller than the ensemble spreads using the EW approach over all seasons and periods, especially with the W6 approach.
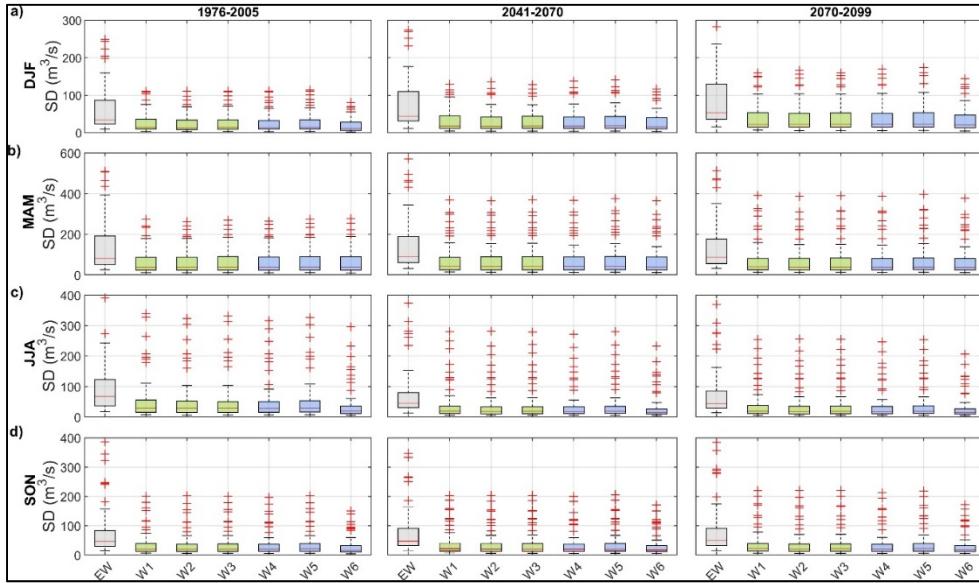
Figure 6.9 Standard deviation (m^3/s) of the equally-weighted streamflow-ensemble (in grey), the three climate-based (in green) and the three streamflow-based (in blue) unequally-weighted streamflow-ensembles for the snow-dominated basins over the 2041-2070 and 2070-2099 periods (from left to right panels). By row, results for the winter (panel a), spring (panel b), summer (panel c) and fall (panel d) months are presented
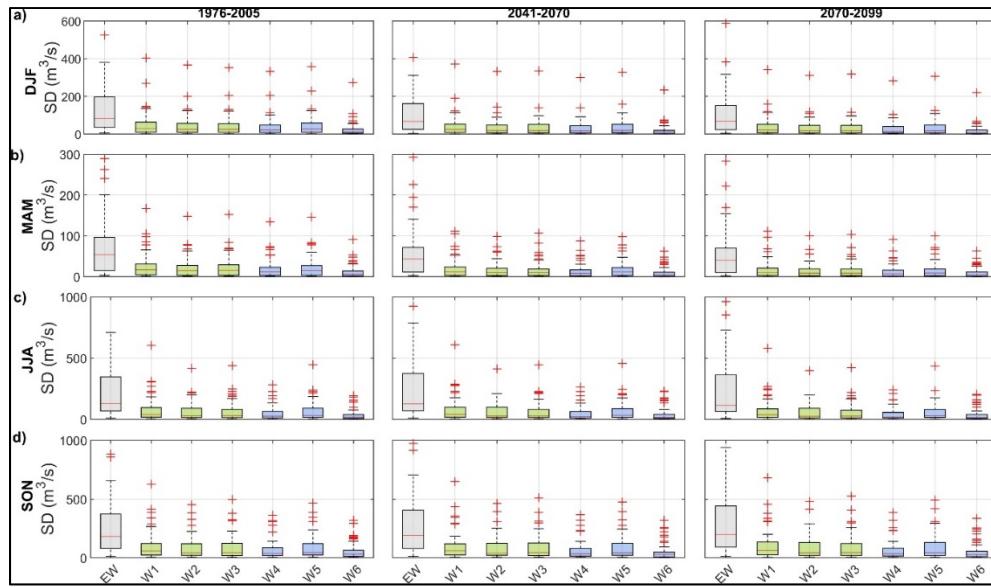
Figure 6.10 Standard deviation (m³/s) of the equally-weighted streamflow-ensemble (in grey), the three climate-based (in green) and the three streamflow-based (in blue) unequally-weighted streamflow-ensembles for the rain-dominated basins over the 2041-2070 and 2070-2099 periods (from left to right panels). By row, results for the winter (panel a), spring (panel b), summer (panel c) and fall (panel d) months are presented

6.5 Discussion

6.5.1 Impacts of climate models weighting

Using large ensembles of climate simulations has become a standard approach for assessing climate change impacts on hydrology. To deal with these large ensembles, different studies have proposed weighting the climate simulations based on their adequacy to reproduce the variable of interest. Thus, to explore the effects of climate model's weighting on flood projections, different weighting approaches were tested on a diverse ensemble of GCM-driven streamflow simulations over 96 basins with contrasting hydroclimatic regimes. The results revealed in line with other studies (e.g., Kiesel et al., 2020; Wang et al., 2019) that weighting climate models can improve the mean annual and seasonal hydrograph representation during the reference period. Figure 6.4 shows a clear picture of the hydrograph representation improvements, where some approaches showed higher annual and seasonal KGE values

compared to the equal-weighting approach. Among them, W6 (the seasonal streamflow-based weighting approach) constantly showed more robust performances over all basins and seasons, and as observed in Figure 6.3, it also showed the most rigorous weighting between climate simulations, meaning that larger weight differences are observed between "good" and "bad" performing simulations. This improvement is not only observed over snow-dominated basins where EW already shows a generally good hydrograph representation, but also over the warmer and dryer rain-dominated basins where larger differences are observed due to the long drought periods.

It was observed that weighting climate models can impact the magnitude, climate change signal and uncertainty spread of flooding projections. Unequally-weighted flood projections showed peak streamflows of about 30% and 50% different than the equal-weighting approach over snow- and rain-dominated basins, respectively. However, the direction and magnitude of these changes varied between seasons, periods, basins, and more notably between weighting approaches. This was observed over both basin groups in Figures 6.5 and 6.6, where W6 showed the largest mean peak differences against the equal-weighting approach. Consequently, the climate change signals also showed different impacts among basins and weighting approaches. The climate change signal over some snow-dominated basins changed from predicted flood increases to flood decreases with the W6 approach, especially during the spring flood. While in the winter and summer months generally larger flood increases and smaller flood decreases were observed, respectively. Over the rain-dominated basins, no systemic effect of weighting was observed. Yet, climate change signals often showed opposite predicted changes, especially during the flooding summer-fall months. And some basins showed larger flood decreases and flood increases than the ones predicted with the EW approach. These results underline the fact that the choice of weighting approach not only can impact the magnitude of the signal, but it can also change the predicted signal to an opposite direction.

Regarding the uncertainty-spread analysis, both basin groups showed clear seasonal spread reductions from all weighting approaches over all seasons and periods. This was expected as

some studies have shown global and regional uncertainty reductions when weighting climate ensembles (Exbrayat et al., 2018; Multsch et al., 2015). These reductions were especially pronounced over rain-dominated basins with seasonal uncertainty spread reductions of up to 80%. These generally larger impacts observed over rain-dominated basins could be explained by the larger uncertainty related to climate modelling observed over regions or seasons where relatively heavy rainfall occurs (Woldemeskel, Sharma, Sivakumar, & Mehrotra, 2016). Thus, larger impacts can be expected over regions or parts of the year were streamflow is mainly driven by rainfall. It is however important to highlight that this study was limited by the weighting methods that include uncertainty ranges estimation (i.e., REA and UREA). Thus, different results can be expected with different methods for estimating uncertainty ranges.

6.5.2   Effects of climate model's weighting criteria

Precipitation-, temperature- and streamflow-based criteria were used and compared among all basins, with some approaches relying on a single climate/streamflow variable and others on multiple climate/streamflow variables. The results showed that the impact level of the weighting approaches varied depending on the selected criteria. This is clearly observed when comparing W6 against the other approaches. This weighting approach not only showed the best mean annual and seasonal hydrograph representation compared to all approaches during the reference period, but also showed the largest differences against the equal-weighting approach. This behavior can be explained by the variable and multi-criteria used for evaluation. This approach is based on streamflow performance which, in line with Wang et al. (2019), generally show better annual streamflow representation than climate-based weighting due to the non-linear relationship of climate variables and streamflow. Additionally, this approach is not solely relying on annual indicators but on multiple seasonal indicators. Annual indictors have been previously used for weighting climate and streamflow projections, but this can be misleading for streamflow due to the different processes that drive streamflow throughout the year. This effect is also observed in Figure 6.4, as some basins show better seasonal representation with temperature- or precipitation-based weighting approaches. For instance, over snow-dominated basins, temperature-based approaches sometimes outperform other

approaches as flood peaks are strongly driven by temperature during spring melt. These results highlight the importance of selecting weighting criteria based on the needs of the studied variable. It is however important to underline that even though objective metrics are selected to evaluate climate models' performance, the choice of weighting criteria is unavoidably subjective as there is no generally accepted approach for such evaluation in the hydrological community. Thus, exploring different weighting criteria based on the variable of interest, as attempted with this study, is needed to shed some light regarding the adequacy of the weighting criteria for a given purpose and their associated impacts.

6.6 Conclusions

In this study, six unequal climate model's weighting approaches were tested and compared against the most common equal-weighting approach to analyze their impacts on flooding projections. An ensemble of 24 raw GCM simulations was used to feed a simple lumped hydrological model to produce 24 GCM-driven streamflows for a reference and two future periods. Different weighting methods and criteria were used to explore their effects on flooding projections over two groups of basins with contrasting hydroclimatic regimes, (1) a 46 rain-dominated basins group and (2) a 50 snow-dominated basins group.

Overall, our results suggest that weighting climate models can impact the magnitude, climate change signal and uncertainty spread of the ensemble of flooding projections. More particularly, the results revealed that:

1. Mean annual and seasonal hydrograph representation over the reference period can improve when climate models are weighted. From the different weighting approaches tested in this study, the approach considering mean seasonal floods as weighting criteria often outperformed the other approaches that rely on mean annual streamflow/precipitation/temperature data.

2. The choice of climate model's weighting approach can strongly impact the magnitude and climate change signal of flood projections. The impacts vary between seasons and more notably between regions with different hydroclimatic regimes. Yet, both snow-

ad rain-dominated basins showed that during their main flood seasons (spring and summer-fall, respectively) the climate change signal often changed to an opposite direction.

3. Rain-dominated basins showed generally larger impacts than snow-dominated basins when climate models were unequally weighted.

4. The uncertainty spread of flooding and streamflow projections can be reduced when climate simulations are weighted.

The obtained results allowed answering some questions, but the limitations of this study highlight the need of further assessments to identify adequate approaches for weighting climate simulations for hydrological applications. For instance, including more recent global/regional climate simulations at higher resolutions and/or physically-based hydrological models can lead to a better understanding of hydro-climatic simulations adequacy for hydrological applications. And also, the need to develop weighting methods that allow estimating weighted uncertainty spreads to assess the possible uncertainty spread reductions.

6.7 Supplementary materials

6.7.1 Supplementary material – S1

Table S6.2 Description of the GCMs used in this study

ID	GCM name	Modeling group	Spatial resolution (°)
1	ACCESS 1.0	CSIRO & BOM	1.875 x 1.25
2	ACCESS 1.3	CSIRO-BOM	1.875 x 1.25
3	BCC-CSM1.1	BCC-CMA	1.125 x 1.125
4	BCC-CSM1.3	BCC-CMA	2.8 x 2.8
5	CanESM2	CCCMA	2.8 x 2.8
6	CMCC-CM	CMCC	0.7484 x 0.75
7	CMCC-CMS	CMCC	1.875 x 1.875

ID	GCM name	Modeling group	Spatial resolution (°)
8	CNRM-CM5	CNRM-CERFACS	1.4 x 1.4
9	CSIRO-Mk3.6	CSIRO-QCCCE	1.8 x 1.8
10	FGOALS-g2	LASG-IAP-CAS-CESS	1.875 x 1.25
11	GFDL-CM3	GFDL	2.5 x 2.0
12	GFDL-ESM2G	GFDL	2.5 x 2.0
13	GFDL-ESM2M	GFDL	2.5 x 2.0
14	GISS-E2-H	NASA-GISS	2.5 x 2.0
15	GISS-E2-R	NASA-GISS	2.5 x 2.0
16	INM-CM4	INM	2.0 x 1.5
17	IPSL-CM5A-LR	IPSL	3.75 x 1.8
18	IPSL-CM5B-LR	IPSL	3.75 x 1.8
19	MIROC5	MIROC	1.4 x 1.4
20	MIROC ESM	MIROC	2.8 x 2.8
21	MIROC ESM-CHEM	MIROC	2.8 x 2.8
22	MPI-ESM-LR	MPI	1.8653 x 1.875
23	MPI-ESM-MR	MPI	1.8653 x 1.875
24	MRI-CGCM3	MRI	1.1 x 1.1

6.7.2   Supplementary material – S2

Reliability ensemble averaging

The reliability ensemble averaging (REA) method developed by Giorgi et Mearns (2002) allows calculating the average and uncertainty range of an ensemble of simulations to minimize the contribution of members with poor performance over a given region. The performance of each climate simulation i is estimated using a reliability measure R_i defined as follows:

$$R_i = \left[\left(R_{B,i} \right)^m \cdot \left(R_{D,i} \right)^n \right]^{1/m \cdot n} \tag{6.1}$$

where $R_{B,i}$ and $R_{D,i}$ are the reliability factors that measure the bias and model convergence, respectively. The bias and convergence factors are presented in the following equations 2 and 3:

$$R_{B,i} = \frac{\varepsilon}{|B_i|} \tag{6.2}$$

$$R_{D,i} = \frac{\varepsilon}{|D_i|} \tag{6.3}$$

where ε is the natural climate or streamflow variability estimated as the difference between the maximum and minimum values of 20-year moving averages of the observed climate or streamflow data series. $|B_i|$ represents the absolute bias of a given simulation i against the observed data series during the reference period. $|D^{(i)}|$ represents the absolute distance between the simulated climate or streamflow change of a given simulation i against the change calculated from the ensemble weighted using the REA method for the future period. The reliability factors are set to 1 when the absolute bias or absolute distance is smaller than the natural climate or streamflow variability. The parameters m and n are used to weight the bias and convergence criterion, respectively. For this study, both parameters are equal to 1.

The reliability factors calculated for each climate or streamflow simulation are then used to calculate the REA-weighted ensemble mean $\tilde{X}$ as follows:

$$\tilde{X} = \frac{\sum_{i=1}^{N} R_i \cdot X_i}{\sum_{i=1}^{N} R_i} \tag{6.4}$$

where X_i denotes the climate or streamflow simulation i and N is the number of members in the ensemble.

The uncertainty range of the REA-weighted mean is calculated by using the root mean squared error of the simulations against the REA-weighted ensemble mean. The upper and lower uncertainty ranges are thus defined by adding or subtracting the calculated root mean squared error to the REA-weighted ensemble mean. The uncertainty range calculation is presented in the following equations 6.5-6.7.

$$\tilde{\delta}_X = \left[\frac{\sum_{i=1}^{N} R_i \cdot (X_i - \tilde{X})^2}{\sum_{i=1}^{N} R_i} \right]^{1/2} \tag{6.5}$$

$$X_{up} = \tilde{X} + \tilde{\delta}_X \tag{6.6}$$

$$X_{low} = \tilde{X} + \tilde{\delta}_X \tag{6.7}$$

where $\tilde{\delta}_X$ is the root mean squared error around the REA-weighted ensemble mean. X_{up} and X_{low} are the upper and lower uncertainty limits, respectively.

Upgraded reliability ensemble averaging

The upgraded reliability ensemble averaging (UREA) method proposed by Xu et al. (2010) is a modified version of the REA method. This method proposed eliminating the convergence factor but including multiple statistics and variables in the reliability factor. In this study, for each climate or streamflow variable a bias and variability factor were included. For example, the reliability measure of the UREA method for temperature (T) and precipitation (P) variables is defined by:

$$UR_i = [f1(T)_B]^{m1} \cdot [f2(T)_V]^{m2} \cdot [f3(P)_B]^{m3} \cdot [f4(P)_V]^{m4} \tag{6.8}$$

where the bias and variability factors for a given temperature and precipitation simulation i are presented in the following equations 6.9-6.10 and 6.11-6.12, respectively.

$$f1(T)_B = \frac{\varepsilon(Ta)}{|B(T)_i|} \tag{6.9}$$

$$f2(T)_V = \frac{\varepsilon(Tv)}{|STD(T)_i - STD(T)_{obs}|} \tag{6.10}$$

$$f3(P)_B = \frac{\varepsilon(Pa)}{|B(P)_i|} \tag{6.11}$$

$$f4(P)_V = \frac{\varepsilon(Pv)}{|CV(P)_i - CV(P)_{obs}|} \tag{6.12}$$

where $\varepsilon(Ta)$ / $\varepsilon(Pa)$ and $\varepsilon(Tv)$ / $\varepsilon(Pv)$ denote the natural variability in terms of annual averages (i.e., Ta/Pa) and interannual variability (i.e., Tv/Pv), respectively. $|B(T)_i|$ and $|B(P)_i|$ represent the absolute bias of a given climate simulation i against observations during a reference period for temperature and precipitation simulations, respectively. As presented in equations 6.10 and 6.12, the absolute variability for temperature simulations is estimated with standard deviation and with the coefficient of variance for precipitation simulations. For using the UREA method with streamflow series, the coefficient of variance is also used. All factors (i.e., f1-f4) are set to 1 when the absolute bias (e.g., $|B(T)_i|$) or absolute variability (e.g., $|STD(T)_i - STD(T)_{obs}|$) are smaller than the natural climate or streamflow variability in terms or annual averages (e.g., $\varepsilon(Ta)$) or interannual variability (e.g., $\varepsilon(Tv)$). For this study, the parameters m1, m2, m3, m4 are all set to 1.

The UREA uncertainty range $\bar{\delta}_X$ modified the root mean squared error $\tilde{\delta}_X$ described in equation 6.5 as follows:

$$\bar{\delta}_X = \sqrt{\frac{N_{eff}}{N_{eff} - 1}} \cdot \tilde{\delta}_X \tag{6.13}$$

$$N_{eff} = \frac{1}{\sum_{i=1}^{N} P_i^2} \tag{6.14}$$

$$P_i = \frac{R_i}{\sum_{j=1}^{N} R_j} \tag{6.15}$$

where P_i is the contribution of a given simulation i to the overall likelihood. N_{eff} is the effective number of models that allows avoiding to artificially narrow the uncertainty range. To obtain the upper and lower uncertainty ranges, equations 6.6 and 6.7 are used by replacing the REA root mean squared error $\tilde{\delta}_X$ by the UREA root mean squared error $\bar{\delta}_X$.

6.7.3 Supplementary material – S3

Mean annual and seasonal hydrogramme representation

This first metric evaluates the mean annual and seasonal hydrogramme of the weighted-ensembles against the observed mean annual hydrogramme during the reference period using the KGE criterion defined as follows:

$$KGE = 1 - \sqrt{(r-1)^2 + (\beta-1)^2 + (\gamma-1)^2} \tag{6.16}$$

$$\beta = \frac{\mu_s}{\mu_o} \tag{6.17}$$

$$\gamma = \frac{CV_s}{CV_o} = \frac{\sigma_s/\mu_s}{\sigma_o/\mu_o} \tag{6.18}$$

where r, β and γ are the correlation coefficient, bias ratio and variability ratio, respectively. CV is the coefficient of variation, μ the mean, and σ is the standard deviation of the streamflow series. The "o" and "s" subscripts indicate *observed* and *simulated* data, respectively. KGE values range from -∞ to 1, where 1 indicates a perfect fit.

The following three metrics compare the weighted-ensembles in terms of mean seasonal peak streamflows of all seven equal- and unequal-weighting approaches. This means that for each weighting approach, season and period; a weighted-ensemble mean seasonal peak streamflow. First, the seasonal peak streamflow $Qmax$ of a given season is calculated by averaging the seasonal peak-flow over the 30-year reference or future period. Then, the weighted-ensemble mean seasonal peak streamflow $\overline{Qmax}$ is calculated as follows:

$$\overline{Qmax} = \frac{\sum_{i=1}^{N}(Qmax_i \cdot W_i)}{N} \tag{6.19}$$

where $Qmax_i$ is the mean seasonal peak streamflow of a given GCM-driven streamflow i for a given period (e.g., 1976-2005), W_i is the calculated weight for the climate model i with a given weighting approach (e.g., W2) and N is the number of GCM-driven streamflows and weights (i.e., twenty-four in this study).

Mean seasonal peak relative bias

This comparison between the weighted-ensembles is measured in terms of relative bias (%) between the mean seasonal peak streamflow of a given unequal-weighting x (i.e., W1-W6) and the equal-weighting approach y (i.e., EW) used as reference. The relative bias is thus defined as follows:

$$RB(\%) = \frac{\overline{Qmax_x} - \overline{Qmax_y}}{\overline{Qmax_y}} \tag{6.20}$$

where $\overline{Qmax_X}$ is the mean seasonal peak streamflow of a given unequal weighted-ensemble x and $\overline{Qmax_y}$ is the mean seasonal peak streamflow of the equally weighted-ensemble.

Climate change signal

This metric is used to compare the climate change signal on mean seasonal peak streamflows from each of the seven approaches. This is done by comparing the mean seasonal peak

streamflow of a given weighted-ensemble *x* during a future period *fut* (e.g., 2041-2070) against its mean seasonal peak streamflow during the reference period *ref* (i.e., 1975-2006) in terms of relative bias (%). This seasonal flooding climate change signal is calculated as follows:

$$RB(\%) = \frac{\overline{Qmax_x} - \overline{Qmax_y}}{\overline{Qmax_y}} \tag{6.21}$$

where $Qmax_x^{fut}$ represents a given weighted-ensemble *x* during a future period *fut* against its mean seasonal peak during the reference period *ref*.

Seasonal ensemble spread

This comparison consists in comparing the spread of the uncertainty ranges for the equal and unequal weighting approaches. The spreads are measured in terms of standard deviation for each season (i.e., DJF, MAM, JJA & SON) over the reference and future periods as follows:

$$STD = \sqrt{\frac{\sum_{i=1}^{N_s}(Q_i - \bar{Q})^2}{N_s}} \tag{6.22}$$

where Q_i is the mean daily streamflow of a given simulation *i* of an ensemble member comprised in the uncertainty range. $\bar{Q}$ is the mean daily streamflow of all the ensemble members comprised in the uncertainty range. N_s is the number of days of the season in study.

CHAPTER 7

GENERAL DISCUSSION

The objective of this research project was analysing the main modelling uncertainties associated with the generations of flood projections, and evaluating the use of different strategies to reduce the flood projections uncertainty spread. The obtained results and assessments not only provided further insights on the different uncertainty sources and their links to regional processes, but also allowed assessing evaluating the impacts of different evaluation and weighting strategies to deal with the modelling uncertainties on flood projections. These results directly contributed to improving our current knowledge on the modelling uncertainties associated with flood projections, and thus allowed fulfilling the aim of this project. Nonetheless, the implications of the obtained results need to be discussed and put into perspective. Thus, this chapter presents a general discussion of the main conclusions obtained in the present work.

7.1   Modelling uncertainties and the regional flood-generating processes

Multi-model ensembles (MMEs) have become essential for climate change impact studies. One of the main advantages of this approach is that it provides a range of projections, while also allowing to assess the associated modelling uncertainties (IPCC, 2021). From the different modelling elements involved in the generation of flood projections, GCMs have often been identified as one of the main uncertainty contributors (Chan et al., 2020; Giuntoli et al., 2018; Vetter et al., 2017). Consequently, using multiple GCMs in the ensembles is now considered fundamental to increase our confidence in the resulting projections (IPCC, 2021). However, the results of Chapter 3 revealed that although GCMs were indeed found as one of the main uncertainty contributors in flood projections, other elements such as hydrological models can play a more important role in certain conditions. The variance decomposition analyses described in Chapter 3 showed that the uncertainty contributions systematically varied depending the main flood-generating process of a given river basin. Rain- and snow-dominated basins showed that the main uncertainty contributors in their flood projections were climate

models and hydrological models, respectively. These results can have different implications in climate change impact studies as they suggest that uncertainty contributions are linked to the main streamflow-generating process, which varies not only between regions but also throughout the seasons. This is clearly observed in the seasonal analyses presented in Chapter 3, where seasons dominated by rain agreed on their main uncertainty contributor (i.e., climate models). While, in line with the results presented in Chapters 4 and 5, it is observed that uncertainty over snow-dominated basins was mainly influenced by hydrological models' structures. Recent studies have suggested that this can be linked to the uncertainty related to the snow models (Troin, Martel, Arsenault, & Brissette, 2022), as they have shown larger influence in high-latitude basins. Yet, further studies are needed to better understand the influence of each component and propose improvements to reduce the uncertainty associated with these models. In the case of the rain-dominated basins, the climatic uncertainty can also be associated with the regional characteristics of the different basins located in Mexico, where more complex processes as hurricanes and orographic precipitation events can occur.

The observed variations in uncertainty contributions underline the need for multi-model approaches. Yet, it particularly highlights the potential hazards of relying on a single best-performing hydrological model in climate change impact studies, especially knowing that this approach is still being used (though less and less). Nonetheless, using larger MMEs will consequently demand more important computational and human resources. Thus, knowing which elements bring more uncertainty to flood projections might also serve as a guide to modellers with limited computational resources to favor MMEs to the most important uncertainty contributors. On the other hand, these results also emphasize the importance of better understanding the link between uncertainty contributions and the main processes driving the studied variables and processes.

The analyses performed in Chapter 3 are limited by the different methodological choices. Among them, it is important to mention the potential impacts of the selected number of elements that were included for each uncertainty source. For instance, in terms of climate simulations, a total of 22 GCM simulations were used, while only three hydrological models

and two bias-correction methods were included. The idea behind this difference was using a number of elements that is often found in the literature, where including more climate simulations is a common practice (Kundzewicz et al., 2017). However, the different sample sizes can influence the variance decomposition evaluation. Although the variance decomposition method considers the number of elements of each uncertainty source (see section 3.8.2), it is more likely to include contrasting or outlier simulations/models when more elements are included. Another limitation is related to the different emission scenarios. In this study, only the RCP 8.5 was used. It can be thus expected that the variance decomposition results are impacted by the selected high-emission scenario. For instance, recent large-scale studies as the NAC2H dataset (Arsenault, Brissette, Chen, et al., 2020), allow comparing different uncertainty sources, including different emission scenarios. Thus, future large-scale studies that consider different emission scenarios can evaluate the impact of the emission scenarios selection on large study domains.

7.2   Strategies to reduce modelling uncertainties in climate change impact studies

One of the most important elements of this research project was evaluating different strategies that have been proposed to reduce the modelling uncertainties associated with the generation of flood projections. As described in the literature, strategies related to hydrological and climate modelling have been identified as potential ways to reduce the overall uncertainty in climate change impact studies. Thus, Chapters 4-6 have focused on evaluating different strategies that have been suggested to improve our confidence on flood projections.

7.2.1   Hydrological modelling under a changing climate

The reliability of hydrological models used in climate change impacts studies has been and continues to be a topic of discussion. As described in Chapter 4, evidences of the decreased performance of hydrological models used over climate conditions that are different to the ones used during their calibration have underlined the importance to reconsider the critical assumption that the parameterizations of hydrological models remain valid in time (Dakhlaoui

et al., 2019; Dakhlaoui et al., 2017; Krysanova et al., 2020; Motavita et al., 2019; Guillaume Thirel et al., 2015; G. Thirel et al., 2015; Vormoor et al., 2018). Consequently, some studies have suggested the need for more "strict" calibration/validation strategies that allow testing hydrological models under different conditions before including them in a climate change impact study, while others have suggested the use of weighting techniques based on their performance and/or robustness. Thus, this project addressed these research gaps in Chapters 4 and 5 by (1) studying the impact of different calibration strategies on hydrological model's response under contrasting climatic conditions, and (2) evaluating the effects of using weighting strategies on flood projections.

To improve our confidence on hydrological model used in climate change impact studies, some studies have suggested using longer calibration datasets (Arsenault et al., 2018; Guo et al., 2018; Vaze et al., 2010), while others recommend using shorter but more representative calibration periods (Motavita et al., 2019). Different calibration data lengths and climate conditions were thus tested to evaluate their impacts on hydrological projections. The different tests performed in Chapter 4 revealed that the climate conditions of the calibration and validation data had larger impacts on the five hydrological model's performance than the different lengths, suggesting that shorter periods that are more contrasted and likely more similar to the changes expected from climate projections might be more adequate to increase the reliability of hydrological models in climate change impact studies. Moreover, the different tests revealed that precipitation changes generally showed stronger effects on hydrological models' performance than those for temperature changes, which was also observed in different case studies around the world (Coron et al., 2012; Dakhlaoui et al., 2019; Motavita et al., 2019; Seiller et al., 2012; Seiller et al., 2015). In consequence, generally larger effects were observed on rain-dominated basins, where precipitation plays a dominant role in the flood-generating process. Although the calibration strategy impacted the uncertainty spread on streamflow projections, the results in Chapter 4 showed that the choice of hydrological model had a generally larger impact, highlighting the importance of using of ensembles of hydrological models. As presented in Chapter 4, this was particularly clear for the snow-dominated basins, which directly corresponds to the findings of Chapter 3 showing that hydrological model's

structure was identified as their main uncertainty contributor. Overall, these results provided more evidences that each hydrological model can respond differently to contrasting climate conditions (Troin et al., 2022), and that adding more calibration data is not necessarily better.

The second stage of this research line was evaluating the effects of weighting hydrological models on flood projections. As revealed in Chapter 4, changing climate conditions can have a large impact on the reliability of hydrological models. Thus, different weighting methods based on hydrological models' robustness to changing climate conditions were evaluated in Chapter 5. The results obtained in this chapter revealed that even the most simplistic method (equal-weights for all hydrological models) outperformed the best-performing hydrological model on contrasting climatic conditions, showing once again the importance of using ensembles of hydrological models. Nevertheless, weighting hydrological models based on their robustness showed some improvements over this simplistic approach, particularly on contrasting warmer and dryer conditions. These differences were more clearly observed on the flood projections, with the robustness-based weighting approaches showing different future flood magnitudes and climate change signals. This reveals that the effects of weighting hydrological models are not minor as they can influence the interpretation of climate change impacts on floods. In addition, this chapter explored the effects of a more complex semi-distributed and more physically-based hydrological model, showing that it might bring some advantages when simulating low-flows, and high-flows on small basins (of about 500 km^2). These results, as discussed in Chapter 5, can be linked to the fact that distributed hydrological models can simulate smaller areas with more detail than a lumped hydrological model. Although these analyses were only performed over four basins, the results bring out the potential use of more complex hydrological models for cases were hydrological models contribute high levels of uncertainty, as observed over the snow-dominated flooding events (Chapter 3).

The results in Chapters 3 and 4 showed that the uncertainty associated with hydrological models can play an important role in flooding projections. However, in these analyses, hydrological models were treated as a single entity. In other words, the different elements that

influence hydrological models' response and evaluation, such as the objective functions used for their calibrations, their parameterizations, and their different sub-components (e.g., snow accumulation representation, potential evapotranspiration formulations) were not evaluated. Thus, large-scale studies that decompose the impacts of the different elements that influence hydrological models' simulations can provide further insights on what elements of the hydrological modelling process might bring more uncertainty in climate change impact studies (e.g. Troin et al., 2022). Similarly, the uncertainty associated with the historical observations (Hamilton & Moore, 2012) was not considered in this study. It can be thus expected that the uncertainty associated with historical records can influence the obtained results. For instance, the selected historical records can impact the performance evaluation of the hydrological simulations during the calibration/validation process, which has shown more significant impacts in simpler lumped hydrological models (Montanari & Di Baldassarre, 2013). Thus, further studies that evaluate the impact of alternative historical data sources can shed some light on the importance of this uncertainty source in climate change impact studies.

7.2.2   Analysing climate models weighting for flood projections

Climate models have shown to be one of the main sources of uncertainty in climate change impact studies (Hattermann et al., 2018; IPCC, 2021). Using MMEs has thus become not only standard but essential when using climate models. As described in Chapter 1, the most common approach to work with MMEs is the climate model's "*democracy*", where all climate models are considered independent and equiprobable. However, at present, climate models share parts of their structures and do not perform equally, raising concerns with the use of this approach (Abramowitz et al., 2019; Knutti, 2010) and its interpretation for climate change impact studies at the regional or local scales (Palmer & Stevens, 2019; Touzé-Peiffer, Barberousse, & Le Treut, 2020). In the latest IPCC AR6 (IPCC, 2021) it is now stated with *high confidence* that ensembles of regional climate projections should be selected to discard climate models that generate unrealistic representations of processes that are relevant for a given application (IPCC, 2021). This recent statement highlights the importance of studies that further evaluate

the effects of selecting or weighting climate models on climate change impact studies, as performed in this research project.

Chapter 6 explored the effects of weighting climate models on flood projections and proposed potential improvements on the existing approaches. The results revealed that all the different basin types included in the study improved their hydrographs representation when climate models were weighted based on their performance. This was particularly observed with the proposed approach that measures the simulated streamflow performance based on seasonal performance metrics. As discussed in previous sections, the processes that drive streamflow-generation are not constant throughout the year. Thus, this proposed approach offers a better assessment by considering these seasonal changes. This improvement was also expected as this approach has more degrees of freedom, meaning that it provides more information about the hydrograph of a given basin. Nonetheless, the skill of the different weighting approaches was solely evaluated over an annual hydrograph. Further performance evaluations that consider other performance metrics or processes can lead to a more comprehensive performance evaluation of the weighting methods.

Between the different basin types, the rain-dominated basins showed the largest impacts when climate models were weighted. This can be explained by the results obtained in Chapter 3, where it was observed that the uncertainty spread on rain-dominated basins are strongly influenced by climate model's uncertainty. Thus, larger impact in these regions were expected. By comparing the different results obtained over the different chapters, it is possible to realize that the modeling uncertainties and the impacts of the strategies to reduce them are linked to the hydroclimatic processes that dominate the streamflow-generation at a given region or season. For instance, it was often observed over Chapters 3 to 6 that precipitation played an important role in both climate and hydrological models' uncertainty. Consequently, higher uncertainty contributions and impacts were often observed on rain dominated basins and also on rain-dominated seasons of the snow-dominated regions, underlining the importance of process-based analyses to identify the underlying components that contribute to uncertainty.

The observed effects of the different evaluation and weighting strategies underlined some questions related to their implications in climate change impact studies. For instance, how much can we reduce the uncertainty? And more importantly, up to which point should we reduce these uncertainties? As highlighted by different authors (Beven, 2016; Krysanova et al., 2018; Krysanova et al., 2020), the goal of reducing modelling uncertainties is to increase our confidence in projected climate change impacts, but not to provide a definitive "crisp" value. In other words, finding an optimal ensemble that allows representing the uncertainty as best as possible and increasing the credibility of the ensembles is the ultimate aim. For instance, a recent study by Wang, Chen, Cannon, Xu, et Chen (2018) compared the uncertainty spreads of hydrological projections produced with climate model ensembles of different sizes, showing that a subset of 10 climate simulations was able to cover up to 85% of the complete 50 climate simulations ensemble. Future work that focuses on evaluating the different sample-sizes can provide further insights for more robust and credible hydrological projections.

CONCLUSION

The latest AR6 of the IPCC (2021) has documented the increasing evidences of climate change impacts over every inhabited region of the world. Moreover, it has shown that changes on climate and hydrological extremes have already been observed in every continent, and it is expected that they will intensify in the future (IPCC, 2018, 2021). However, the potential effects of these observed changes on future flooding events remain unclear. The high levels of uncertainty associated with their simulations along with the strong impact of local and regional characteristics, makes their analysis a complex task. Therefore, the objective of this research project was analysing the main modelling uncertainties associated with the generation of flood projections and investigating the effects of different strategies to reduce the envelope of uncertainty on flood projections. The different results and analyses presented in Chapters 3 to 6 allowed improving our understanding of the uncertainty contributions of different elements of the modelling chain, and quantifying the potential effects of different strategies to reduce uncertainties arising from hydrological and climate models.

The first stage of this research project (Chapter 3) allowed confirming that the uncertainty contributions of the different elements of the modelling chain used to generate flood projections are not constant in time or in space. The large-scale uncertainty analysis revealed that climate models dominated the uncertainty contributions in regions where the main flood-generating process is rain, while hydrological models were the main uncertainty source in snow-dominated basins. This finding directly contributes to improving our understanding on modelling uncertainties and reveals the potential hazards of assuming that the uncertainty analyses performed over certain regions of the world are valid for other regions with different hydroclimatic regimes.

Chapters 4 and 5 focused on evaluating the effects of different strategies to reduce the uncertainties arising from hydrological models. The first strategy addressing the reliability of hydrological model's calibrations for climate change impact studies (Chapter 5) revealed that hydrological models' performance was consistently more sensitive to changes in the

calibration-validation data's climate conditions than the lengths of the calibration-validation periods, suggesting that calibrating-validating on years with climate conditions that are more similar to the expected changes in climate can increase our confidence on future hydrological projections. These results contribute to the ongoing discussion on hydrological model's reliability showing that adding more years in the calibration datasets will not necessarily improve the hydrological model's reliability unless it contains more representative climate conditions. Based on these results, the use of weighting hydrological models based on their robustness under contrasting climate conditions was tested in Chapter 5. The results revealed that weighting hydrological models based on their robustness can outperform equal-weighting techniques, particularly when simulating warmer and dryer climate conditions. Additionally, it was observed that weighting hydrological models can lead to different projected flood changes than those obtained with the traditional equal-weighting method, highlighting the potential impacts on climate change impact assessments. At the same time, few tests including a semi-distributed and more physically-based hydrological model suggest its potential added value on low-flow simulations and peak simulations on smaller basins (of about 500 km^2).

Regarding the strategies to reduce the uncertainty associated with climate models, Chapter 6 revealed that weighting climate models can lead to improvements in the basins' hydrographs representations. More particularly, this study proposed a new approach that considers the seasonal performance of the climate models-driven streamflow simulations. This proposed approach showed a generally better performance and larger effects on flood projections, showing that the choice of weighting approach can directly impact the resulting projected flood magnitudes and climate change signals. Thus, the obtained results contribute to the ongoing discussion on strategies to reduce climate model's uncertainty for impact studies.

Overall, the large-scale study performed in this thesis allowed drawing more robust conclusions related to the modelling uncertainties in flood projections. The findings of this research project showed that more attention should be given to the hydroclimatic processes influencing the studied variable, particularly when studying climate change impacts. Finally, the different works carried out in this thesis demonstrate that identifying modelling approaches

based on the objective of each particular study can help increasing our confidence in hydroclimatic projections, which is of great relevance knowing that the resulting assessments are often used to inform decision-makers and guide mitigation and adaptation strategies.

RECOMMENDATIONS

The different analyses undertaken in this research project combined with the large and diverse study area can contribute to different fields and applications, such as flood risk assessments, hydrological modelling, ensemble modelling techniques, and climate change impact studies. It is thus important to highlight the limitations of the obtained results and conclusions. The methodological choices and geographical extension of this research project unavoidably limits its scope, and reveals potential ways forward. Thus, this section describes the different limitations of this work, and identifies potential ways forward.

- The uncertainty analysis performed in Chapter 4 is limited by the selected elements included in the modelling chain. For instance, the important impact of the natural climate variability in the uncertainty spread of climate change impact studies was not included in the analysis. It is thus recommended to perform a large-scale study that includes the uncertainty associated to natural climate variability.

- Although a large-scale study was included in the study, this study is limited by its geographical extension. It is thus recommended to perform more studies including basins with different climatic and physical properties, as well as regions with different flood-generating processes.

- The choice of climate models clearly limits the reach of the obtained conclusions. More recent climate simulations issued from the latest CMIP6 can lead to different findings, especially knowing these high-resolution simulations have shown improved performances at the regional and local scales. Further studies including these high-resolution simulations are thus recommended.

- In line with the previous point, including the new convection-permitting climate simulations can be tested. As observed in this project, precipitation plays a very important role in the modelling uncertainties. Thus, the potential improvements expected from the convection-permitting climate models might lead to reduced climatic uncertainty.

- This study was also limited by the hydrological models included in the study. Thus, including more complex and process-based hydrological models, as well as different conceptual models with different structures is recommended. A larger and more varied ensemble of hydrological models can help studying the robustness of their different components and formulations to contrasting climate conditions.

- The selected post-processing methods also limited the obtained results. It is thus, recommended to test including other post-processing methods. More recent and improved post-processing methods focusing on extreme events have been developed. Thus, adding such post-processing methods can help improving the simulations of extreme hydroclimatic events.

- The probability distributions included in the study limited the study. Including the more recent non-stationary statistical models can lead to different results. It is thus recommended to test the use of these statistical models on flood projections.

- This study tested only a few weighting approaches on both climate and hydrological models. It is thus recommended to compare the presented results with other weighting approaches that consider other criteria, such as model's independence and climate sensitivity.

- Finally, it is recommended to test other strategies that can lead to reducing the uncertainties in flood projections and climate change impact studies, such as including different observation datasets.

APPENDIX I

LIST OF SCIENTIFIC CONTRIBUTIONS

Published articles-Peer-reviewed

Arsenault R, Brissette F, Martel J-L, Troin M, Lévesque G, Davidson-Chaput J, **Castaneda-Gonzalez M**, Ameli A, Poulin A (2020) A comprehensive, multisource database for hydrometeorological modeling of 14,425 North American watersheds. Nature Scientific Data 7:243.

Aubry-Wake C, Somers LD, Alcock H, Anderson AM, Azarkhish A, Bansah S, Bell NM, Biagi K, **Castaneda-Gonzalez M**, Champagne O (2020) A new flow for Canadian young hydrologists: Key scientific challenges addressed by research cultural shifts. Hydrological Processes 34:2001-2006.

Oral presentations

Castaneda-Gonzalez, M. Poulin, A. Romero-López, R. and Turcotte, R (2021). « The impact of weighting climate simulations on flood projections». Oral presentation in the Canadian Water Resources Association (CWRA) 2021 National Conference. Canmore, Canada, June 4th – June 4th 2022.

Castaneda-Gonzalez, M. Poulin, A. Romero-López, R. and Turcotte, R (2021). « Quels sont les effets de la pondération des simulations climatiques sur les projections de crues? ». Oral presentation in the Colloque RHQ 2022 – Recherche hydrologique au Québec. Quebec City, Canada, May 5th – May 6th 2022.

Castaneda-Gonzalez, M. Poulin, A. Romero-López, R. and Turcotte, R (2021). «How is the overall uncertainty on flood projections influenced by local hydroclimatic conditions? ». Oral

presentation in the Canadian Water Resources Association (CWRA) 2021 National Conference. Montreal, Canada, May 31st – June 4th 2021.

Castañeda-González, M. Poulin, A. Romero-López, R. and Turcotte, R (2021). «Dealing with climatic uncertainty for hydrological projections». Oral presentation in the « Ma présentation en 3 minutes » des Journées québécoises étudiantes CentrEau 2021 (JQEC21). Montreal, Canada, March 4th 2020.

Castañeda-González, M. Poulin, A. Romero-López, R. and Turcotte, R (2021). « Incertitudes sur l'analyse des débits de crues dans le contexte des changements climatiques ». Oral presentation in the « Atelier sur la collaboration Québec-Mexique » Montreal, Canada, November 19th 2020.

Castañeda-González, M. Poulin, A. Romero-López, R. and Turcotte, R (2019). « Evaluación de la incertidumbre climática en análisis de frecuencia de avenidas de diseño ». Oral presentation in the V Congreso nacional y I Latinoamericano de manejo de cuencas hidrograficas. Mexico City, Mexico, 29-31 October 2019.

Castañeda-González, M. Poulin, A. and Romero-López, R. (2018). « Impacts of regional climate model spatial resolution on summer flood simulation ». Paper and oral presentation in the 13th International Conference on Hydroinformatics (HIC 2018) Palermo, Italy, 1-6 July 2018.

APPENDIX II

**VARIANCE DECOMPOSITION INTERACTIONS**
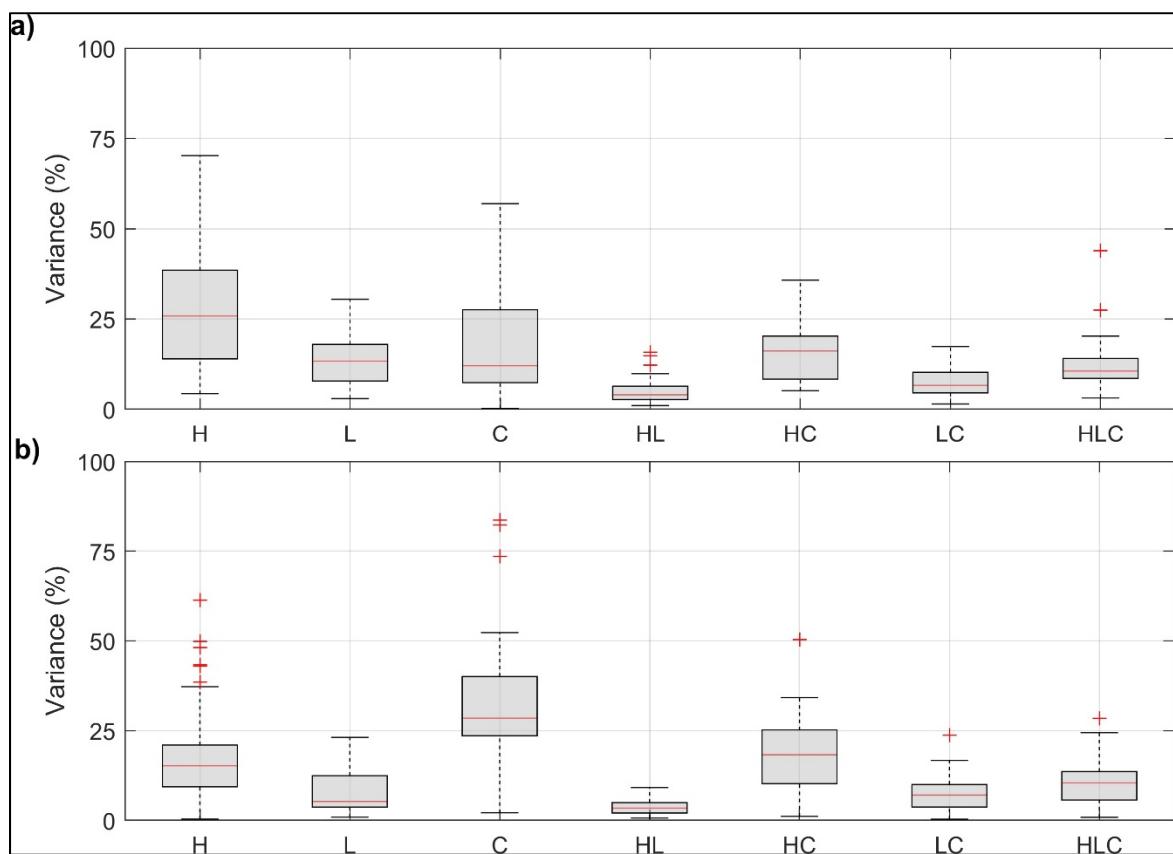
Figure A II-1 Variance contributions (%) of hydrological models' structure (H, in grey), calibration data's length (L, in green), calibration data's climatic condition (C, in blue) and their interactions over the snow- and rain-dominated basins on panels a and b, respectively.
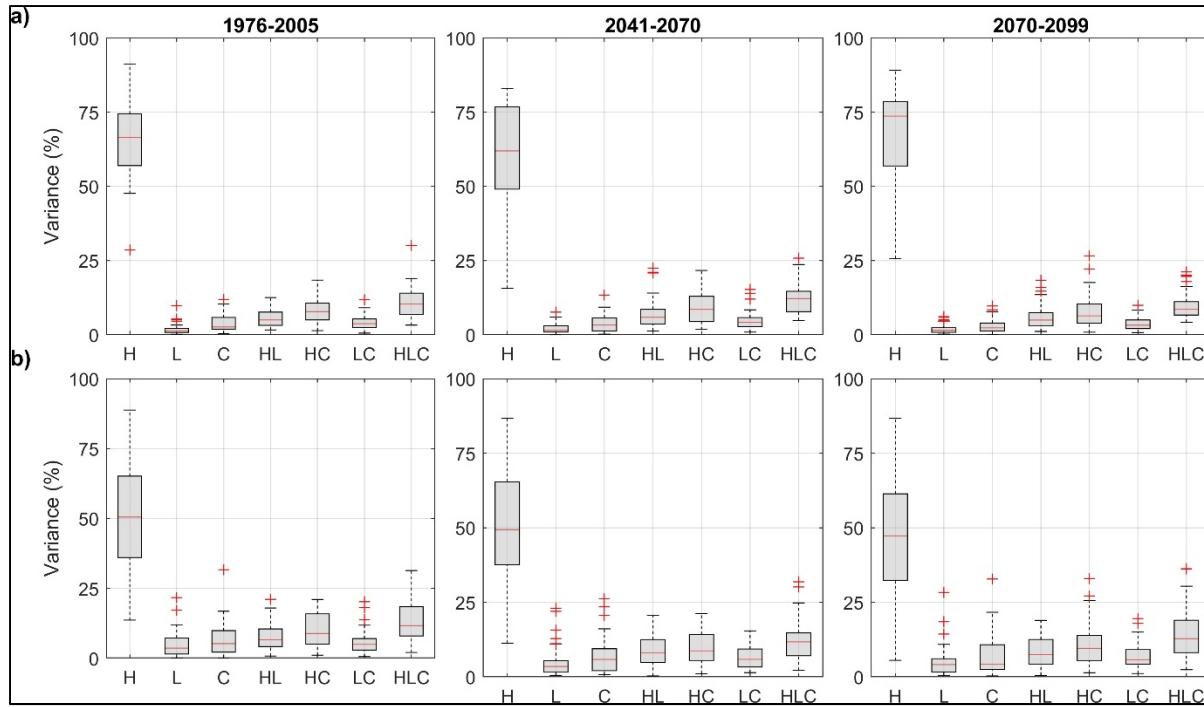
Figure A II-2 Variance contributions (%) of hydrological models' structure (H, in grey), calibration data's length (L, in green), calibration data's climatic condition (C, in blue) and their interactions on mean annual RCM-driven streamflows simulated with the CRCM5-CNRMCM5 simulations over the snow- and rain-dominated basins on panels a and b, respectively. Each row presents the boxplots of variance (%) of the three variance contributors and their interactions for the reference and future periods of 1976-2005, 2041-2070 and 2070-2099
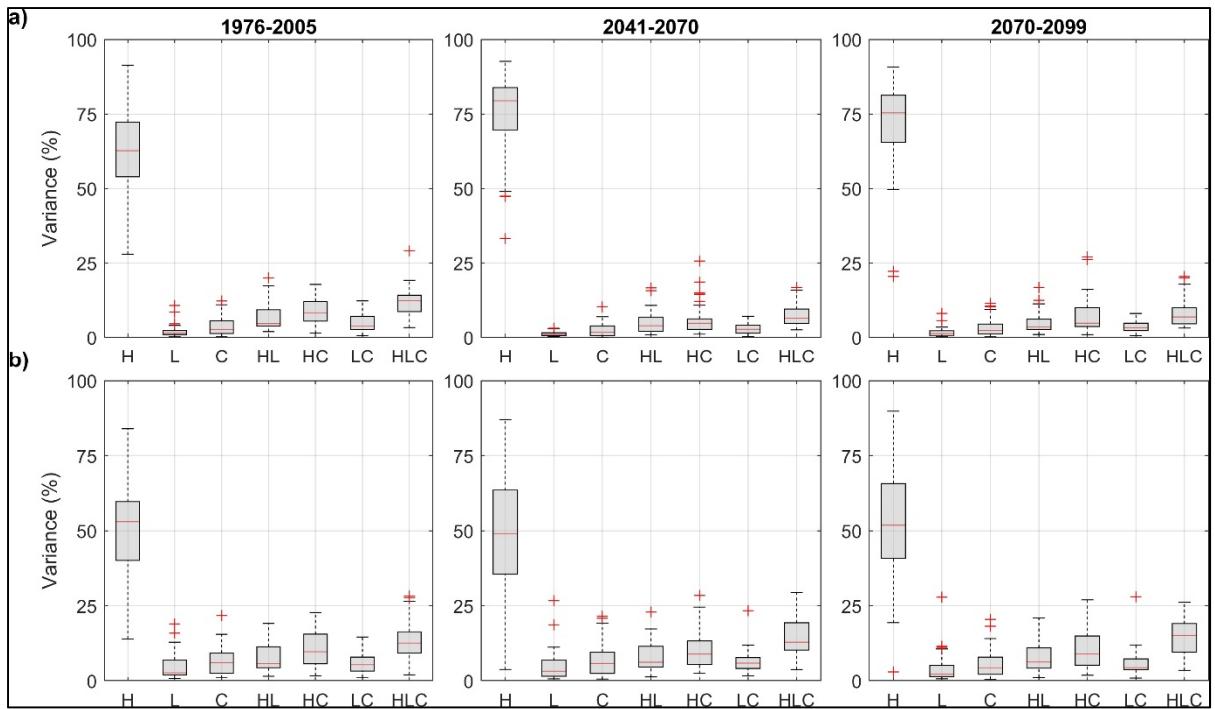
Figure A II-3 Variance contributions (%) of hydrological models' structure (H, in grey), calibration data's length (L, in green), calibration data's climatic condition (C, in blue) and their interactions on mean annual RCM-driven streamflows simulated with the CRCM5-CANESM2 simulations over the snow- and rain-dominated basins on panels a and b, respectively. Each row presents the boxplots of variance (%) of the three variance contributors and their interactions for the reference and future periods of 1976-2005, 2041-2070 and 2070-2099
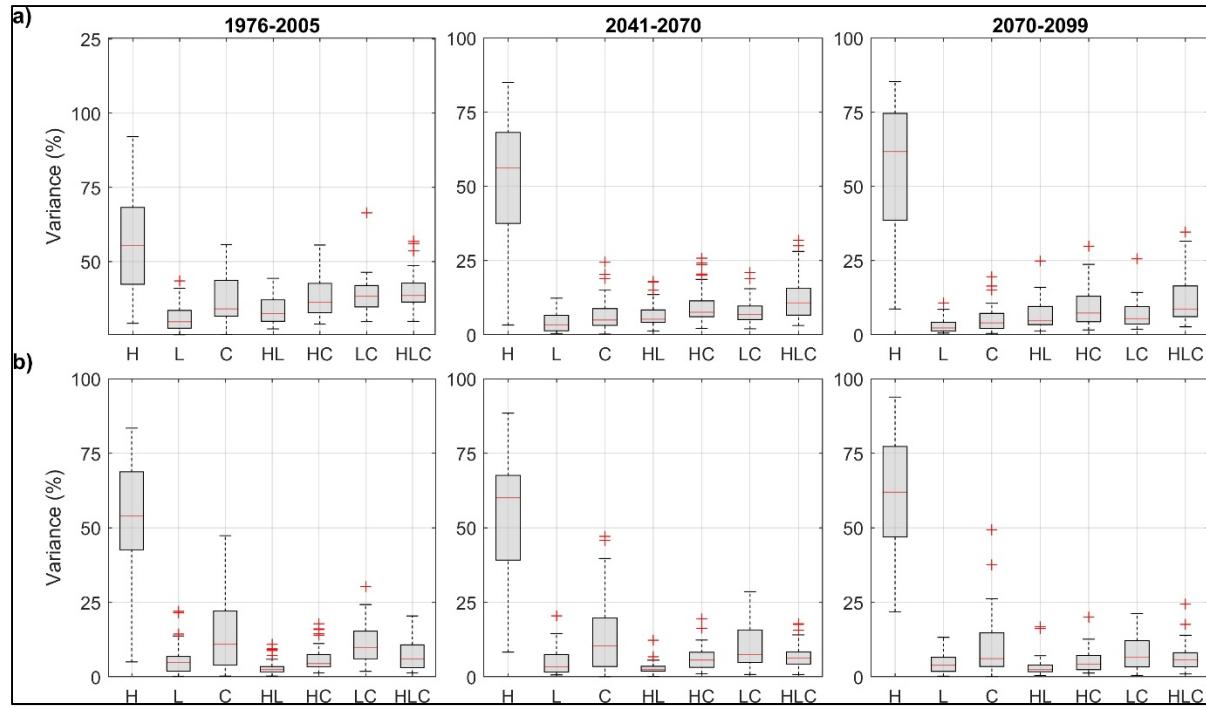
Figure A II-4 Variance contributions (%) of hydrological models' structure (H, in grey), calibration data's length (L, in green), calibration data's climatic condition (C, in blue) and their interactions on mean annual peak RCM-driven streamflows simulated with the CRCM5-CNRMCM5 simulations over the snow- and rain-dominated basins on panels a and b, respectively. Each row presents the boxplots of variance (%) of the three variance contributors and their interactions for the reference and future periods of 1976-2005, 2041-2070 and 2070-2099
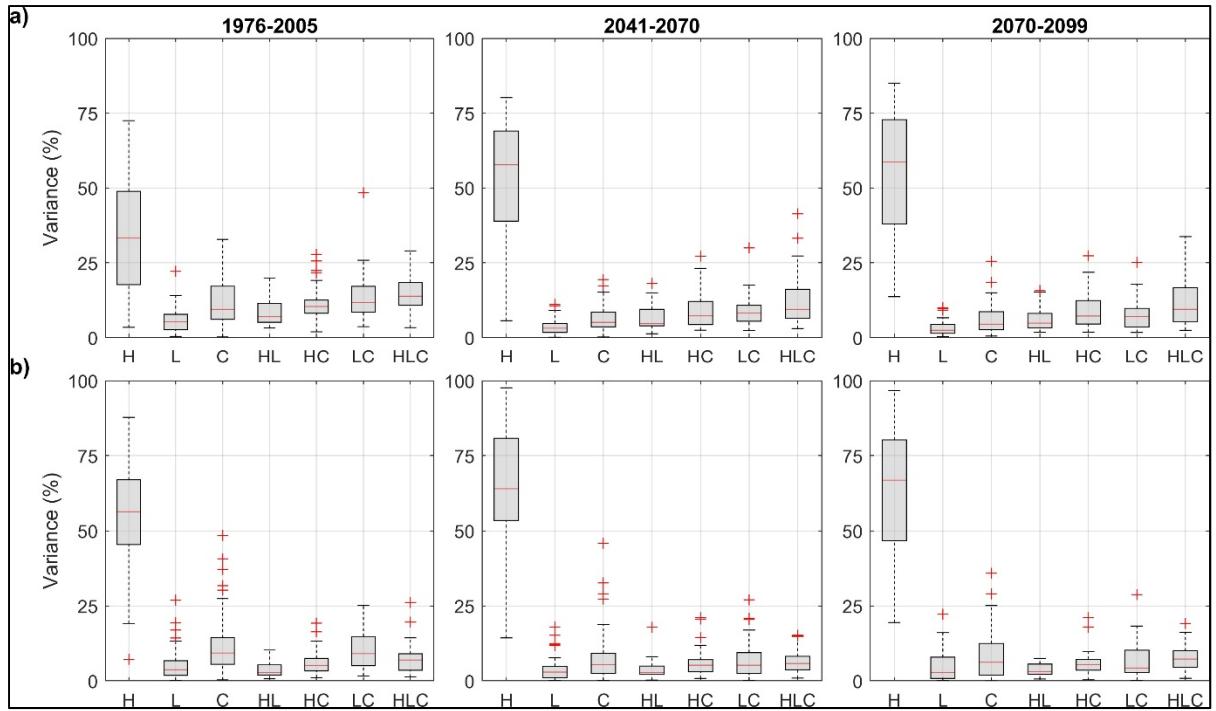
Figure A II-5 Variance contributions (%) of hydrological models' structure (H, in grey), calibration data's length (L, in green), calibration data's climatic condition (C, in blue) and their interactions on mean annual peak RCM-driven streamflows simulated with the CRCM5-CANESM2 simulations over the snow- and rain-dominated basins on panels a and b, respectively. Each row presents the boxplots of variance (%) of the three variance contributors and their interactions for the reference and future periods of 1976-2005, 2041-2070 and 2070-2099

LIST OF BIBLIOGRAPHICAL REFERENCES

Abramowitz, G., Herger, N., Gutmann, E., Hammerling, D., Knutti, R., Leduc, M., . . . Schmidt, G. A. (2019). ESD Reviews: Model dependence in multi-model climate ensembles: weighting, sub-selection and out-of-sample testing. *Earth Syst. Dynam., 10*(1), 91-105. doi: 10.5194/esd-10-91-2019. Repéré à https://esd.copernicus.org/articles/10/91/2019/

AghaKouchak, A., Easterling, D., Hsu, K., Schubert, S., & Sorooshian, S. (2012). *Extremes in a changing climate: detection, analysis and uncertainty* (Vol. 65). Springer Science & Business Media.

Ajami, N. K., Gupta, H., Wagener, T., & Sorooshian, S. (2004). Calibration of a semi-distributed hydrologic model for streamflow estimation along a river system. *Journal of hydrology, 298*(1), 112-135. doi: https://doi.org/10.1016/j.jhydrol.2004.03.033. Repéré à https://www.sciencedirect.com/science/article/pii/S0022169404002410

Alfieri, L., Burek, P., Feyen, L., & Forzieri, G. (2015). Global warming increases the frequency of river floods in Europe. *Hydrol. Earth Syst. Sci., 19*(5), 2247-2260. doi: 10.5194/hess-19-2247-2015. Repéré à https://www.hydrol-earth-syst-sci.net/19/2247/2015/

Allaire, M. (2018). Socio-economic impacts of flooding: A review of the empirical literature. *Water Security, 3*, 18-26. doi: https://doi.org/10.1016/j.wasec.2018.09.002. Repéré à https://www.sciencedirect.com/science/article/pii/S2468312418300063

Arnell, N. W., & Lloyd-Hughes, B. (2014). The global-scale impacts of climate change on water resources and flooding under new climate and socio-economic scenarios. *Climatic Change, 122*(1), 127-140. doi: 10.1007/s10584-013-0948-4. Repéré à https://doi.org/10.1007/s10584-013-0948-4

Arsenault, R., Breton-Dufour, M., Poulin, A., Dallaire, G., & Romero-Lopez, R. (2019). Streamflow prediction in ungauged basins: analysis of regionalization methods in a hydrologically heterogeneous region of Mexico. *Hydrological Sciences Journal, 64*(11), 1297-1311. doi: 10.1080/02626667.2019.1639716. Repéré à https://doi.org/10.1080/02626667.2019.1639716

Arsenault, R., Brissette, F., Chen, J., Guo, Q., & Dallaire, G. (2020). NAC2H: The North-American Climate Change and hydroclimatology dataset. *Water resources research, n/a*(n/a), e2020WR027097. doi: 10.1029/2020wr027097. Repéré à https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020WR027097

Arsenault, R., Brissette, F., & Martel, J.-L. (2018). The hazards of split-sample validation in hydrological model calibration. *Journal of hydrology, 566*, 346-362.

Arsenault, R., Brissette, F., Martel, J.-L., Troin, M., Lévesque, G., Davidson-Chaput, J., . . . Poulin, A. (2020). A comprehensive, multisource database for hydrometeorological modeling of 14,425 North American watersheds. *Scientific Data, 7*(1), 243. doi: 10.1038/s41597-020-00583-2. Repéré à https://doi.org/10.1038/s41597-020-00583-2

Arsenault, R., Gatien, P., Renaud, B., Brissette, F., & Martel, J.-L. (2015). A comparative analysis of 9 multi-model averaging approaches in hydrological continuous streamflow simulation. *Journal of hydrology, 529*, 754-767. doi: https://doi.org/10.1016/j.jhydrol.2015.09.001. Repéré à https://www.sciencedirect.com/science/article/pii/S0022169415006848

Arsenault, R., Poulin, A., Côté, P., & Brissette, F. (2014). Comparison of stochastic optimization algorithms in hydrological model calibration. *Journal of hydrologic engineering, 19*(7), 1374-1384.

Aryal, A., Shrestha, S., & Babel, M. S. (2018). Quantifying the sources of uncertainty in an ensemble of hydrological climate-impact projections. *Theoretical and Applied Climatology*. doi: 10.1007/s00704-017-2359-3. Repéré à https://doi.org/10.1007/s00704-017-2359-3

Asadzadeh, M., & Tolson, B. A. (2009). *A new multi-objective algorithm, pareto archived DDS* présentée à Proceedings of the 11th Annual Conference Companion on Genetic and Evolutionary Computation Conference: Late Breaking Papers, Montreal, Québec, Canada. doi: 10.1145/1570256.1570259

Barbarossa, V., Huijbregts, M. A., Beusen, A. H., Beck, H. E., King, H., & Schipper, A. M. (2018). FLO1K, global maps of mean, maximum and minimum annual streamflow at 1 km resolution from 1960 through 2015. *Scientific Data, 5*, 180052.

Bates, B., Kundzewicz, Z. W., Wu, S., & Palutikof, J. (2008). *Climate change and water: Technical paper vi*. Intergovernmental Panel on Climate Change (IPCC).

Bates, J. M., & Granger, C. W. (1969). The combination of forecasts. *Journal of the Operational Research Society, 20*(4), 451-468.

Beck, H. E., Dijk, A. I. J. M. v., Roo, A. d., Miralles, D. G., McVicar, T. R., Schellekens, J., & Bruijnzeel, L. A. (2016). Global-scale regionalization of hydrologic model parameters. *Water resources research, 52*(5), 3599-3622. doi: 10.1002/2015WR018247. Repéré à http:https://doi.org/10.1002/2015WR018247

Bengtsson, L., & Shukla, J. (1988). Integration of space and in situ observations to study global climate change. *Bulletin of the American Meteorological Society, 69*(10), 1130-1143.

Bérubé, S., Brissette, F., & Arsenault, R. (2022). Optimal Hydrological Model Calibration Strategy for Climate Change Impact Studies. *Journal of hydrologic engineering, 27*(3),

04021053. doi: doi:10.1061/(ASCE)HE.1943-5584.0002148. Repéré à https://ascelibrary.org/doi/abs/10.1061/%28ASCE%29HE.1943-5584.0002148

Beven, K. (2016). Facets of uncertainty: epistemic uncertainty, non-stationarity, likelihood, hypothesis testing, and communication. *Hydrological Sciences Journal, 61*(9), 1652-1665. doi: 10.1080/02626667.2015.1031761. Repéré à https://doi.org/10.1080/02626667.2015.1031761

Beven, K. J. (2011). *Rainfall-runoff modelling: the primer*. John Wiley & Sons.

Blöschl, G. (2022). Flood generation: process patterns from the raindrop to the ocean. *Hydrol. Earth Syst. Sci., 26*(9), 2469-2480. doi: 10.5194/hess-26-2469-2022. Repéré à https://hess.copernicus.org/articles/26/2469/2022/

Blöschl, G., Hall, J., Viglione, A., Perdigão, R. A. P., Parajka, J., Merz, B., . . . Živković, N. (2019). Changing climate both increases and decreases European river floods. *Nature, 573*(7772), 108-111. doi: 10.1038/s41586-019-1495-6. Repéré à https://doi.org/10.1038/s41586-019-1495-6

Blöschl, G., & Montanari, A. (2010). Climate change impacts—throwing the dice? *Hydrological processes, 24*(3), 374-381. doi: 10.1002/hyp.7574. Repéré à https://onlinelibrary.wiley.com/doi/abs/10.1002/hyp.7574

Blöschl, G., & Sivapalan, M. (1995). Scale issues in hydrological modelling: A review. *Hydrological processes, 9*(3-4), 251-290. doi: doi:10.1002/hyp.3360090305. Repéré à https://onlinelibrary.wiley.com/doi/abs/10.1002/hyp.3360090305

Bosshard, T., Carambia, M., Goergen, K., Kotlarski, S., Krahe, P., Zappa, M., & Schär, C. (2013). Quantifying uncertainty sources in an ensemble of hydrological climate-impact projections. *Water resources research, 49*(3), 1523-1536. doi: 10.1029/2011WR011533. Repéré à http:https://doi.org/10.1029/2011WR011533

Box, G. E., & Draper, N. R. (1987). *Empirical model-building and response surfaces*. John Wiley & Sons.

Braconnot, P., Harrison, S. P., Kageyama, M., Bartlein, P. J., Masson-Delmotte, V., Abe-Ouchi, A., . . . Zhao, Y. (2012). Evaluation of climate models using palaeoclimatic data. *Nature Climate Change, 2*(6), 417-424. doi: 10.1038/nclimate1456. Repéré à https://doi.org/10.1038/nclimate1456

Brigode, P., Oudin, L., & Perrin, C. (2013). Hydrological model parameter instability: A source of additional uncertainty in estimating the hydrological impacts of climate change? *Journal of hydrology, 476*, 410-425. doi: https://doi.org/10.1016/j.jhydrol.2012.11.012. Repéré à http://www.sciencedirect.com/science/article/pii/S002216941200964X

214

Broderick, C., Matthews, T., Wilby, R. L., Bastola, S., & Murphy, C. (2016). Transferability of hydrological models and ensemble averaging methods between contrasting climatic periods. *Water resources research, 52*(10), 8343-8373. doi: 10.1002/2016wr018850. Repéré à https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016WR018850

Burn, D. H., & 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. Repéré à https://doi.org/10.1080/07011784.2015.1026844

Byun, K., & Hamlet, A. F. (2020). A risk-based analytical framework for quantifying non-stationary flood risks and establishing infrastructure design standards in a changing environment. *Journal of hydrology, 584*, 124575. doi: https://doi.org/10.1016/j.jhydrol.2020.124575. Repéré à https://www.sciencedirect.com/science/article/pii/S0022169420300354

Camargo, A., & Cortesi, L. (2019). Flooding water and society. *WIREs Water, 6*(5), e1374. doi: https://doi.org/10.1002/wat2.1374. Repéré à https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/wat2.1374

Carter, T., Hulme, M., & Lal, M. (2007). General guidelines on the use of scenario data for climate impact and adaptation assessment.

Casanueva, A., Kotlarski, S., Herrera, S., Fernández, J., Gutiérrez, J. M., Boberg, F., . . . Vautard, R. (2016). Daily precipitation statistics in a EURO-CORDEX RCM ensemble: added value of raw and bias-corrected high-resolution simulations. *Climate Dynamics, 47*(3), 719-737. doi: 10.1007/s00382-015-2865-x. Repéré à https://doi.org/10.1007/s00382-015-2865-x

Castaneda-Gonzalez, M., Poulin, A., Romero-Lopez, R., Arsenault, R., Brissette, F., & Turcotte, R. (2019). Sensitivity of seasonal flood simulations to regional climate model spatial resolution. *Climate Dynamics, 53*(7), 4337-4354. doi: 10.1007/s00382-019-04789-y. Repéré à https://doi.org/10.1007/s00382-019-04789-y

Castaneda-Gonzalez, M., Poulin, A., Romero-Lopez, R., Turcotte, R., & Chaumont, D. (2021). Uncertainty sources in flood projections over contrasting hydrometeorological regimes. *Manuscript submitted for publication.*

Castellarin, A., Kohnová, S., Gaál, L., Fleig, A., Salinas, J., Toumazis, A., . . . Macdonald, N. (2012). Review of applied statistical methods for flood frequency analysis in Europe. *NERC/Centre for Ecology & Hydrology.*

Chan, W. C. H., Thompson, J. R., Taylor, R. G., Nay, A. E., Ayenew, T., MacDonald, A. M., & Todd, M. C. (2020). Uncertainty assessment in river flow projections for Ethiopia's

Upper Awash Basin using multiple GCMs and hydrological models. *Hydrological Sciences Journal, 65*(10), 1720-1737. doi: 10.1080/02626667.2020.1767782. Repéré à https://doi.org/10.1080/02626667.2020.1767782

Charron, I. (2014). A Guidebook on Climate Scenarios: Using Climate Information to Guide Adaptation Research and Decisions. Ouranous. 86.

Chegwidden, O. S., Nijssen, B., Rupp, D. E., Arnold, J. R., Clark, M. P., Hamman, J. J., . . . Xiao, M. (2019). How Do Modeling Decisions Affect the Spread Among Hydrologic Climate Change Projections? Exploring a Large Ensemble of Simulations Across a Diversity of Hydroclimates. *Earth's Future, 7*(6), 623-637. doi: 10.1029/2018ef001047. Repéré à https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018EF001047

Chen, J., Brissette, F. P., & Caya, D. (2020). Remaining error sources in bias-corrected climate model outputs. *Climatic Change, 162*(2), 563-582. doi: 10.1007/s10584-020-02744-z. Repéré à https://doi.org/10.1007/s10584-020-02744-z

Chen, J., Brissette, F. P., Chaumont, D., & Braun, M. (2013a). Finding appropriate bias correction methods in downscaling precipitation for hydrologic impact studies over North America. *Water resources research, 49*(7), 4187-4205.

Chen, J., Brissette, F. P., Chaumont, D., & Braun, M. (2013b). Performance and uncertainty evaluation of empirical downscaling methods in quantifying the climate change impacts on hydrology over two North American river basins. *Journal of hydrology, 479*, 200-214. doi: https://doi.org/10.1016/j.jhydrol.2012.11.062. Repéré à http://www.sciencedirect.com/science/article/pii/S0022169412010414

Chen, J., Brissette, F. P., & Leconte, R. (2011). Uncertainty of downscaling method in quantifying the impact of climate change on hydrology. *Journal of hydrology, 401*(3), 190-202.

Chen, J., Brissette, F. P., Lucas-Picher, P., & Caya, D. (2017). Impacts of weighting climate models for hydro-meteorological climate change studies. *Journal of hydrology, 549*, 534-546. doi: https://doi.org/10.1016/j.jhydrol.2017.04.025. Repéré à https://www.sciencedirect.com/science/article/pii/S0022169417302457

Chen, J., Brissette, F. P., Poulin, A., & Leconte, R. (2011). Overall uncertainty study of the hydrological impacts of climate change for a Canadian watershed. *Water resources research, 47*(12).

Chen, J., Brissette, F. P., Zhang, X. J., Chen, H., Guo, S., & Zhao, Y. (2019). Bias correcting climate model multi-member ensembles to assess climate change impacts on hydrology. *Climatic Change, 153*(3), 361-377. doi: 10.1007/s10584-019-02393-x. Repéré à https://doi.org/10.1007/s10584-019-02393-x

Chen, J., Li, C., Brissette, F. P., Chen, H., Wang, M., & Essou, G. R. C. (2018). Impacts of correcting the inter-variable correlation of climate model outputs on hydrological modeling. *Journal of hydrology, 560*, 326-341. doi: https://doi.org/10.1016/j.jhydrol.2018.03.040. Repéré à http://www.sciencedirect.com/science/article/pii/S0022169418302063

Collet, L., Beevers, L., & Prudhomme, C. (2017). Assessing the Impact of Climate Change and Extreme Value Uncertainty to Extreme Flows across Great Britain. *Water, 9*(2), 103. Repéré à http://www.mdpi.com/2073-4441/9/2/103

Colorado-Ruiz, G., Cavazos, T., Salinas, J. A., De Grau, P., & Ayala, R. (2018). Climate change projections from Coupled Model Intercomparison Project phase 5 multi-model weighted ensembles for Mexico, the North American monsoon, and the mid-summer drought region. *International Journal of Climatology, 38*(15), 5699-5716. doi: https://doi.org/10.1002/joc.5773. Repéré à https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.5773

Condon, L., Gangopadhyay, S., & Pruitt, T. (2015). Climate change and non-stationary flood risk for the upper Truckee River basin. *Hydrology and Earth System Sciences, 19*(1), 159-175.

Coppola, E., Raffaele, F., & Giorgi, F. (2016). Impact of climate change on snow melt driven runoff timing over the Alpine region. *Climate Dynamics*, 1-15.

Coron, L. (2013). *Les modèles hydrologiques conceptuelssont-ils robustes face à un climat en évolution? Diagnostic sur un échantillon de bassins versants français et australiens.*

Coron, L., Andréassian, V., Perrin, C., Lerat, J., Vaze, J., Bourqui, M., & Hendrickx, F. (2012). Crash testing hydrological models in contrasted climate conditions: An experiment on 216 Australian catchments. *Water resources research, 48*(5). doi: 10.1029/2011wr011721. Repéré à https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2011WR011721

CRED, U. (2020). Human Cost of Disasters. An Overview of the last 20 years: 2000–2019. Repéré à https://reliefweb.int/report/world/human-cost-disasters-overview-last-20-years-2000-2019

Cunnane, C. (1989). Statistical distributions for flood frequency analysis. *Operational hydrology report (WMO)*.

Curry, C. L., Tencer, B., Whan, K., Weaver, A. J., Giguère, M., & Wiebe, E. (2016). Searching for added value in simulating climate extremes with a high-resolution regional climate model over western Canada. *Atmosphere-Ocean, 54*(4), 364-384.

Dakhlaoui, H., Ruelland, D., & Tramblay, Y. (2019). A bootstrap-based differential split-sample test to assess the transferability of conceptual rainfall-runoff models under past and future climate variability. *Journal of hydrology, 575*, 470-486. doi: https://doi.org/10.1016/j.jhydrol.2019.05.056. Repéré à http://www.sciencedirect.com/science/article/pii/S0022169419305013

Dakhlaoui, H., Ruelland, D., Tramblay, Y., & Bargaoui, Z. (2017). Evaluating the robustness of conceptual rainfall-runoff models under climate variability in northern Tunisia. *Journal of hydrology, 550*, 201-217. doi: https://doi.org/10.1016/j.jhydrol.2017.04.032. Repéré à http://www.sciencedirect.com/science/article/pii/S0022169417302512

Dallaire, G., Poulin, A., Arsenault, R., & Brissette, F. (2021). Uncertainty of potential evapotranspiration modelling in climate change impact studies on low flows in North America. *Hydrological Sciences Journal, 66*(4), 689-702. doi: 10.1080/02626667.2021.1888955. Repéré à https://doi.org/10.1080/02626667.2021.1888955

Dankers, R., Arnell, N. W., Clark, D. B., Falloon, P. D., Fekete, B. M., Gosling, S. N., . . . Wisser, D. (2014). First look at changes in flood hazard in the Inter-Sectoral Impact Model Intercomparison Project ensemble. *Proceedings of the National Academy of Sciences, 111*(9), 3257-3261. doi: 10.1073/pnas.1302078110. Repéré à http://www.pnas.org/content/pnas/111/9/3257.full.pdf

Das, J., Treesa, A., & Umamahesh, N. V. (2018). Modelling Impacts of Climate Change on a River Basin: Analysis of Uncertainty Using REA & Possibilistic Approach. *Water Resources Management, 32*(15), 4833-4852. doi: 10.1007/s11269-018-2046-x. Repéré à https://doi.org/10.1007/s11269-018-2046-x

Das, J., & Umamahesh, N. V. (2018). Assessment of uncertainty in estimating future flood return levels under climate change. *Natural Hazards, 93*(1), 109-124. doi: 10.1007/s11069-018-3291-2. Repéré à https://doi.org/10.1007/s11069-018-3291-2

de Castro, M., Gallardo, C., Jylha, K., & Tuomenvirta, H. (2007). The use of a climate-type classification for assessing climate change effects in Europe from an ensemble of nine regional climate models. *Climatic Change, 81*(1), 329-341. doi: 10.1007/s10584-006-9224-1. Repéré à https://doi.org/10.1007/s10584-006-9224-1

Déqué, M., Rowell, D. P., Lüthi, D., Giorgi, F., Christensen, J. H., Rockel, B., . . . van den Hurk, B. (2007). An intercomparison of regional climate simulations for Europe: assessing uncertainties in model projections. *Climatic Change, 81*(1), 53-70. doi: 10.1007/s10584-006-9228-x. Repéré à https://doi.org/10.1007/s10584-006-9228-x

Deser, C., Lehner, F., Rodgers, K. B., Ault, T., Delworth, T. L., DiNezio, P. N., . . . Ting, M. (2020). Insights from Earth system model initial-condition large ensembles and future

prospects. *Nature Climate Change, 10*(4), 277-286. doi: 10.1038/s41558-020-0731-2. Repéré à https://doi.org/10.1038/s41558-020-0731-2

Diaz-Nieto, J., & Wilby, R. L. (2005). A comparison of statistical downscaling and climate change factor methods: impacts on low flows in the River Thames, United Kingdom. *Climatic Change, 69*(2-3), 245-268.

Do, H. X., Zhao, F., Westra, S., Leonard, M., Gudmundsson, L., Boulange, J. E. S., . . . Wada, Y. (2020). Historical and future changes in global flood magnitude – evidence from a model–observation investigation. *Hydrol. Earth Syst. Sci., 24*(3), 1543-1564. doi: 10.5194/hess-24-1543-2020. Repéré à https://hess.copernicus.org/articles/24/1543/2020/

Duan, Q., Sorooshian, S., & Gupta, V. K. (1994). Optimal use of the SCE-UA global optimization method for calibrating watershed models. *Journal of hydrology, 158*(3-4), 265-284.

Escalante-Sandoval, C., & García-Espinoza, E. (2014). Analysis of annual flood peak records in Mexico. *WIT Trans. Inf. Commun. Technol, 47*, 49-60.

Essou, G. R. C., Brissette, F., & Lucas-Picher, P. (2017). Impacts of combining reanalyses and weather station data on the accuracy of discharge modelling. *Journal of hydrology, 545*, 120-131. doi: https://doi.org/10.1016/j.jhydrol.2016.12.021. Repéré à http://www.sciencedirect.com/science/article/pii/S0022169416308101

Exbrayat, J. F., Bloom, A. A., Falloon, P., Ito, A., Smallman, T. L., & Williams, M. (2018). Reliability ensemble averaging of 21st century projections of terrestrial net primary productivity reduces global and regional uncertainties. *Earth Syst. Dynam., 9*(1), 153-165. doi: 10.5194/esd-9-153-2018. Repéré à https://esd.copernicus.org/articles/9/153/2018/

Eyring, V., Cox, P. M., Flato, G. M., Gleckler, P. J., Abramowitz, G., Caldwell, P., . . . Williamson, M. S. (2019). Taking climate model evaluation to the next level. *Nature Climate Change, 9*(2), 102-110. doi: 10.1038/s41558-018-0355-y. Repéré à https://doi.org/10.1038/s41558-018-0355-y

Falco, M., Carril, A. F., Menéndez, C. G., Zaninelli, P. G., & Li, L. Z. X. (2018). Assessment of CORDEX simulations over South America: added value on seasonal climatology and resolution considerations. *Climate Dynamics*. doi: 10.1007/s00382-018-4412-z. Repéré à https://doi.org/10.1007/s00382-018-4412-z

Falloon, P., Challinor, A., Dessai, S., Hoang, L., Johnson, J., & Koehler, A.-K. (2014). Ensembles and uncertainty in climate change impacts. *Frontiers in Environmental Science, 2*, 33.

Feng, S., Hu, Q., Huang, W., Ho, C.-H., Li, R., & Tang, Z. (2014). Projected climate regime shift under future global warming from multi-model, multi-scenario CMIP5 simulations. *Global and Planetary Change, 112*, 41-52. doi: https://doi.org/10.1016/j.gloplacha.2013.11.002. Repéré à http://www.sciencedirect.com/science/article/pii/S0921818113002403

Feng, S., Trnka, M., Hayes, M., & Zhang, Y. (2017). Why Do Different Drought Indices Show Distinct Future Drought Risk Outcomes in the U.S. Great Plains? *Journal of Climate, 30*(1), 265-278. doi: 10.1175/jcli-d-15-0590.1. Repéré à https://journals.ametsoc.org/view/journals/clim/30/1/jcli-d-15-0590.1.xml

Field, C. B. (2012). *Managing the risks of extreme events and disasters to advance climate change adaptation: special report of the intergovernmental panel on climate change*. Cambridge University Press.

Fortin, V., & Turcotte, R. (2006). Le modèle hydrologique MOHYSE. *Note de cours pour SCA7420, Département des sciences de la terre et de l'atmosphere, Université du Québec a Montréal*.

Gaines, J. M. (2016). Flooding: Water potential. *Nature, 531*(7594), S54-S55. doi: 10.1038/531S54a. Repéré à https://doi.org/10.1038/531S54a

Giorgi, F., & Gutowski, W. J. (2015). Regional Dynamical Downscaling and the CORDEX Initiative. *Annual Review of Environment and Resources, 40*(1), 467-490. doi: 10.1146/annurev-environ-102014-021217. Repéré à https://doi.org/10.1146/annurev-environ-102014-021217

Giorgi, F., & Mearns, L. O. (2002). Calculation of Average, Uncertainty Range, and Reliability of Regional Climate Changes from AOGCM Simulations via the "Reliability Ensemble Averaging" (REA) Method. *Journal of Climate, 15*(10), 1141-1158. doi: 10.1175/1520-0442(2002)015<1141:COAURA>2.0.CO;2. Repéré à https://doi.org/10.1175/1520-0442(2002)015<1141:COAURA>2.0.CO;2

Girard, G., Morin, G., & Charbonneau, R. (1972). Modèle précipitations-débits à discrétisation spatiale. *Cahiers ORSTOM, série hydrologie, 9*(4), 35-52.

Giuntoli, I., Prosdocimi, I., & Hannah, D. M. (2021). Going Beyond the Ensemble Mean: Assessment of Future Floods From Global Multi-Models. *Water resources research, 57*(3), e2020WR027897. doi: https://doi.org/10.1029/2020WR027897. Repéré à https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020WR027897

Giuntoli, I., Vidal, J., Prudhomme, C., & Hannah, D. (2015). Future hydrological extremes: the uncertainty from multiple global climate and global hydrological models, . *Earth Syst. Dynam. 6, 267–285*.

Giuntoli, I., Villarini, G., Prudhomme, C., & Hannah, D. M. (2018). Uncertainties in projected runoff over the conterminous United States. *Climatic Change, 150*(3), 149-162. doi: 10.1007/s10584-018-2280-5. Repéré à https://doi.org/10.1007/s10584-018-2280-5

Gleckler, P. J., Taylor, K. E., & Doutriaux, C. (2008). Performance metrics for climate models. *Journal of Geophysical Research: Atmospheres, 113*(D6). doi: https://doi.org/10.1029/2007JD008972. Repéré à https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2007JD008972

Gosling, S. N., Taylor, R. G., Arnell, N. W., & Todd, M. C. (2011). A comparative analysis of projected impacts of climate change on river runoff from global and catchment-scale hydrological models. *Hydrol. Earth Syst. Sci., 15*(1), 279-294. doi: 10.5194/hess-15-279-2011. Repéré à https://www.hydrol-earth-syst-sci.net/15/279/2011/

Graham, L. P., Andréasson, J., & Carlsson, B. (2007). Assessing climate change impacts on hydrology from an ensemble of regional climate models, model scales and linking methods – a case study on the Lule River basin. *Climatic Change, 81*(1), 293-307. doi: 10.1007/s10584-006-9215-2. Repéré à https://doi.org/10.1007/s10584-006-9215-2

Granger, C. W. J., & Ramanathan, R. (1984). Improved methods of combining forecasts. *Journal of Forecasting, 3*(2), 197-204. doi: https://doi.org/10.1002/for.3980030207. Repéré à https://onlinelibrary.wiley.com/doi/abs/10.1002/for.3980030207

Gudmundsson, L., Wagener, T., Tallaksen, L. M., & Engeland, K. (2012). Evaluation of nine large-scale hydrological models with respect to the seasonal runoff climatology in Europe. *Water resources research, 48*(11). doi: doi:10.1029/2011WR010911. Repéré à https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2011WR010911

Guo, D., Johnson, F., & Marshall, L. (2018). Assessing the Potential Robustness of Conceptual Rainfall-Runoff Models Under a Changing Climate. *Water resources research, 54*(7), 5030-5049. doi: 10.1029/2018wr022636. Repéré à https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018WR022636

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of hydrology, 377*(1), 80-91.

Haarsma, R. J., Roberts, M. J., Vidale, P. L., Senior, C. A., Bellucci, A., Bao, Q., . . . Storch, J. S. v. (2016). High Resolution Model Intercomparison Project (HighResMIP v1.0) for CMIP6. *Geoscientific Model Development, 9*(1), 4185-4208. doi: 10.5194/gmd-9-4185-2016. Repéré à https://doi.org/10.5194/gmd-9-4185-2016

Hagemann, S., Chen, C., Clark, D., Folwell, S., Gosling, S. N., Haddeland, I., . . . Voss, F. (2013). Climate change impact on available water resources obtained using multiple global climate and hydrology models. *Earth System Dynamics, 4*, 129-144.

Hailegeorgis, T. T., & Alfredsen, K. (2017). Regional flood frequency analysis and prediction in ungauged basins including estimation of major uncertainties for mid-Norway. *Journal of Hydrology: Regional Studies, 9*, 104-126. doi: https://doi.org/10.1016/j.ejrh.2016.11.004. Repéré à http://www.sciencedirect.com/science/article/pii/S2214581816301768

Hall, A. (2014). Projecting regional change. *Science, 346*(6216), 1461-1462. doi: 10.1126/science.aaa0629. Repéré à http://science.sciencemag.org/content/sci/346/6216/1461.full.pdf

Hall, J., Arheimer, B., Borga, M., Brázdil, R., Claps, P., Kiss, A., . . . Lang, M. (2014). Understanding flood regime changes in Europe: A state of the art assessment. *Hydrology and Earth System Sciences, 18*(7), 2735-2772.

Hamed, K., & Rao, A. R. (1999). *Flood frequency analysis*. CRC press.

Hamilton, A. S., & Moore, R. D. (2012). Quantifying Uncertainty in Streamflow Records. *Canadian Water Resources Journal / Revue canadienne des ressources hydriques, 37*(1), 3-21. doi: 10.4296/cwrj3701865. Repéré à https://doi.org/10.4296/cwrj3701865

Hansen, N., & Ostermeier, A. (1997). Convergence properties of evolution strategies with the derandomized covariance matrix adaptation: The CMA-ES. *Eufit, 97*, 650-654.

Hargreaves, G. H., & Samani, Z. A. (1985). Reference Crop Evapotranspiration from Temperature. *Applied Engineering in Agriculture, 1*(2), 96-99. doi: https://doi.org/10.13031/2013.26773. Repéré à https://elibrary.asabe.org/abstract.asp?aid=26773&t=3

Hartmann, D. L., Klein Tank, A. M. G., Rusticucci, M., Alexander, L. V., Brönnimann, S., Charabi, Y., . . . Zhai, P. M. (2013). Observations: Atmosphere and Surface. Dans T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex & P. M. Midgley (Éds.), *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 159–254). Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press. doi: 10.1017/CBO9781107415324.008. Repéré à www.climatechange2013.org

Hattermann, F. F., Vetter, T., Breuer, L., Su, B., Daggupati, P., Donnelly, C., . . . Krysnaova, V. (2018). Sources of uncertainty in hydrological climate impact assessment: a cross-scale study. *Environmental Research Letters, 13*(1), 015006. doi: 10.1088/1748-9326/aa9938. Repéré à http://dx.doi.org/10.1088/1748-9326/aa9938

Hawkins, E., Osborne, T. M., Ho, C. K., & Challinor, A. J. (2013). Calibration and bias correction of climate projections for crop modelling: an idealised case study over Europe. *Agricultural and forest meteorology, 170*, 19-31.

Heavens, N., Ward, D., & Natalie, M. (2013). Studying and projecting climate change with earth system models. *Nature Education Knowledge, 4*(5), 4.

Henderson & Sellers, A. (1993). An antipodean climate of uncertainty? *Climatic Change, 25*(3-4), 203-224.

Her, Y., Yoo, S.-H., Cho, J., Hwang, S., Jeong, J., & Seong, C. (2019). Uncertainty in hydrological analysis of climate change: multi-parameter vs. multi-GCM ensemble predictions. *Scientific Reports, 9*(1), 4974. doi: 10.1038/s41598-019-41334-7. Repéré à https://doi.org/10.1038/s41598-019-41334-7

Hingray, B., Blanchet, J., Evin, G., & Vidal, J.-P. (2019). Uncertainty component estimates in transient climate projections. *Climate Dynamics, 53*(5), 2501-2516. doi: 10.1007/s00382-019-04635-1. Repéré à https://doi.org/10.1007/s00382-019-04635-1

Hirabayashi, Y., Kanae, S., Emori, S., Oki, T., & Kimoto, M. (2008). Global projections of changing risks of floods and droughts in a changing climate. *Hydrological Sciences Journal, 53*(4), 754-772. doi: 10.1623/hysj.53.4.754. Repéré à https://doi.org/10.1623/hysj.53.4.754

Ho, J. T., Thompson, J. R., & Brierley, C. (2016). Projections of hydrology in the Tocantins-Araguaia Basin, Brazil: uncertainty assessment using the CMIP5 ensemble. *Hydrological Sciences Journal, 61*(3), 551-567. doi: 10.1080/02626667.2015.1057513. Repéré à https://doi.org/10.1080/02626667.2015.1057513

Hosseinzadehtalaei, P., Tabari, H., & Willems, P. (2018). Precipitation intensity–duration–frequency curves for central Belgium with an ensemble of EURO-CORDEX simulations, and associated uncertainties. *Atmospheric Research, 200*, 1-12.

Hostetler, S. (1994). Hydrologic and atmospheric models: the (continuing) problem of discordant scales. *Climatic Change, 27*(4), 345-350.

Hu, L., Nikolopoulos, E. I., Marra, F., & Anagnostou, E. N. (2019). Sensitivity of flood frequency analysis to data record, statistical model, and parameter estimation methods: An evaluation over the contiguous United States. *Journal of Flood Risk Management, 13*(1), e12580. doi: 10.1111/jfr3.12580. Repéré à https://onlinelibrary.wiley.com/doi/abs/10.1111/jfr3.12580

Huang, S., Kumar, R., Flörke, M., Yang, T., Hundecha, Y., Kraft, P., . . . Krysanova, V. (2016). Evaluation of an ensemble of regional hydrological models in 12 large-scale river

basins worldwide. *Climatic Change, 141*(3), 381-397. doi: 10.1007/s10584-016-1841-8. Repéré à https://doi.org/10.1007/s10584-016-1841-8

Huang, S., Shah, H., Naz, B. S., Shrestha, N., Mishra, V., Daggupati, P., . . . Vetter, T. (2020). Impacts of hydrological model calibration on projected hydrological changes under climate change—a multi-model assessment in three large river basins. *Climatic Change, 163*(3), 1143-1164. doi: 10.1007/s10584-020-02872-6. Repéré à https://doi.org/10.1007/s10584-020-02872-6

Huot, P.-L., Poulin, A., Audet, C., & Alarie, S. (2019). A hybrid optimization approach for efficient calibration of computationally intensive hydrological models. *Hydrological Sciences Journal, 64*(10), 1204-1222. doi: 10.1080/02626667.2019.1624922. Repéré à https://doi.org/10.1080/02626667.2019.1624922

Ibarra-Zavaleta, S. P., Landgrave, R., Romero-López, R., Poulin, A., & Arango-Miranda, R. (2017). Distributed Hydrological Modeling: Determination of Theoretical Hydraulic Potential & Streamflow Simulation of Extreme Hydrometeorological Events. *Water, 9*(8), 602. Repéré à https://www.mdpi.com/2073-4441/9/8/602

IPCC. (2013a). *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press. doi: 10.1017/CBO9781107415324. Repéré à www.climatechange2013.org

IPCC. (2013b). Summary for Policymakers. Dans T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex & P. M. Midgley (Éds.), *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 1–30). Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press. doi: 10.1017/CBO9781107415324.004. Repéré à www.climatechange2013.org

IPCC. (2018). *Global Warming of 1.5℃. An IPCC Special Report on the Impacts of Global Warming of 1.5℃ Above Pre-Industrial Levels and Related Global Greenhouse Gas Emission Pathways, in the Context of Strengthening the Global Response to the Threat of Climate Change, Sustainable Development, and Efforts to Eradicate Poverty.* Geneva, Switzerland: World Meteorological Organization.

IPCC. (2021). *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, United Kingdom and New York, NY, USA: Cambridge University Press.

Jakeman, A. J., Littlewood, I. G., & Whitehead, P. G. (1990). Computation of the instantaneous unit hydrograph and identifiable component flows with application to two small upland

catchments. *Journal of hydrology, 117*(1), 275-300. doi: https://doi.org/10.1016/0022-1694(90)90097-H. Repéré à https://www.sciencedirect.com/science/article/pii/002216949090097H

Jones, R., Murphy, J., & Noguer, M. (1995). Simulation of climate change over europe using a nested regional-climate model. I: Assessment of control climate, including sensitivity to location of lateral boundaries. *Quarterly Journal of the Royal Meteorological Society, 121*(526), 1413-1449.

Jones, R. N. (2000). Managing uncertainty in climate change projections–issues for impact assessment. *Climatic Change, 45*(3-4), 403-419.

Kay, A. L., Davies, H. N., Bell, V. A., & Jones, R. G. (2009). Comparison of uncertainty sources for climate change impacts: flood frequency in England. *Climatic Change, 92*(1), 41-63. doi: 10.1007/s10584-008-9471-4. Repéré à https://doi.org/10.1007/s10584-008-9471-4

Kiesel, J., Stanzel, P., Kling, H., Fohrer, N., Jähnig, S. C., & Pechlivanidis, I. (2020). Streamflow-based evaluation of climate model sub-selection methods. *Climatic Change, 163*(3), 1267-1285. doi: 10.1007/s10584-020-02854-8. Repéré à https://doi.org/10.1007/s10584-020-02854-8

KlemeŠ, V. (1986). Operational testing of hydrological simulation models. *Hydrological Sciences Journal, 31*(1), 13-24. doi: 10.1080/02626668609491024. Repéré à https://doi.org/10.1080/02626668609491024

Kling, H., Fuchs, M., & Paulin, M. (2012). Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *Journal of hydrology, 424–425*, 264-277. doi: https://doi.org/10.1016/j.jhydrol.2012.01.011. Repéré à http://www.sciencedirect.com/science/article/pii/S0022169412000431

Knoben, W. J., Freer, J. E., & Woods, R. A. (2019). Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. *Hydrology and Earth System Sciences, 23*(10), 4323-4331.

Knutti, R. (2010). The end of model democracy? *Climatic Change, 102*(3), 395-404. doi: 10.1007/s10584-010-9800-2. Repéré à https://doi.org/10.1007/s10584-010-9800-2

Knutti, R., Furrer, R., Tebaldi, C., Cermak, J., & Meehl, G. A. (2010). Challenges in Combining Projections from Multiple Climate Models. *Journal of Climate, 23*(10), 2739-2758. doi: 10.1175/2009jcli3361.1. Repéré à https://journals.ametsoc.org/view/journals/clim/23/10/2009jcli3361.1.xml

Knutti, R., Sedláček, J., Sanderson, B. M., Lorenz, R., Fischer, E. M., & Eyring, V. (2017). A climate model projection weighting scheme accounting for performance and

interdependence. *Geophysical research letters, 44*(4), 1909-1918. doi: doi:10.1002/2016GL072012. Repéré à https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016GL072012

Kolusu, S. R., Siderius, C., Todd, M. C., Bhave, A., Conway, D., James, R., . . . Kashaigili, J. J. (2021). Sensitivity of projected climate impacts to climate model weighting: multi-sector analysis in eastern Africa. *Climatic Change, 164*(3), 36. doi: 10.1007/s10584-021-02991-8. Repéré à https://doi.org/10.1007/s10584-021-02991-8

Kottek, M., Grieser, J., Beck, C., Rudolf, B., & Rubel, F. (2006). World Map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift, 15*(3), 259-263. doi: 10.1127/0941-2948/2006/0130. Repéré à https://www.ingentaconnect.com/content/schweiz/mz/2006/00000015/00000003/art00001

Krysanova, V., Donnelly, C., Gelfan, A., Gerten, D., Arheimer, B., Hattermann, F., & Kundzewicz, Z. W. (2018). How the performance of hydrological models relates to credibility of projections under climate change. *Hydrological Sciences Journal, 63*(5), 696-720. doi: 10.1080/02626667.2018.1446214. Repéré à https://doi.org/10.1080/02626667.2018.1446214

Krysanova, V., Zaherpour, J., Didovets, I., Gosling, S. N., Gerten, D., Hanasaki, N., . . . Wada, Y. (2020). How evaluation of global hydrological models can help to improve credibility of river discharge projections under climate change. *Climatic Change, 163*(3), 1353-1377. doi: 10.1007/s10584-020-02840-0. Repéré à https://doi.org/10.1007/s10584-020-02840-0

Kuczera, G., Renard, B., Thyer, M., & Kavetski, D. (2010). There are no hydrological monsters, just models and observations with large uncertainties! *Hydrological Sciences Journal, 55*(6), 980-991. doi: 10.1080/02626667.2010.504677. Repéré à https://doi.org/10.1080/02626667.2010.504677

Kundzewicz, Z. W., Kanae, S., Seneviratne, S. I., Handmer, J., Nicholls, N., Peduzzi, P., . . . Sherstyukov, B. (2014). Flood risk and climate change: global and regional perspectives. *Hydrological Sciences Journal, 59*(1), 1-28. doi: 10.1080/02626667.2013.857411. Repéré à http://dx.doi.org/10.1080/02626667.2013.857411

Kundzewicz, Z. W., Krysanova, V., Benestad, R. E., Hov, Ø., Piniewski, M., & Otto, I. M. (2018). Uncertainty in climate change impacts on water resources. *Environmental Science & Policy, 79*, 1-8. doi: https://doi.org/10.1016/j.envsci.2017.10.008. Repéré à http://www.sciencedirect.com/science/article/pii/S146290111730638X

Kundzewicz, Z. W., Krysanova, V., Dankers, R., Hirabayashi, Y., Kanae, S., Hattermann, F. F., . . . Schellnhuber, H. J. (2017). Differences in flood hazard projections in Europe –

226

their causes and consequences for decision making. *Hydrological Sciences Journal, 62*(1), 1-14. doi: 10.1080/02626667.2016.1241398. Repéré à https://doi.org/10.1080/02626667.2016.1241398

Laprise, R. (2008). Regional climate modelling. *Journal of Computational Physics, 227*(7), 3641-3666.

Lemaitre-Basset, T., Collet, L., Thirel, G., Parajka, J., Evin, G., & Hingray, B. (2021). Climate change impact and uncertainty analysis on hydrological extremes in a French Mediterranean catchment. *Hydrological Sciences Journal, 66*(5), 888-903. doi: 10.1080/02626667.2021.1895437. Repéré à https://doi.org/10.1080/02626667.2021.1895437

Li, C., Wang, H., Liu, J., Yan, D., Yu, F., & Zhang, L. (2010). Effect of calibration data series length on performance and optimal parameters of hydrological model. *Water Science and Engineering, 3*(4), 378-393.

Livneh, B., Bohn, T. J., Pierce, D. W., Munoz-Arriola, F., Nijssen, B., Vose, R., . . . Brekke, L. (2015). A spatially comprehensive, hydrometeorological data set for Mexico, the U.S., and Southern Canada 1950–2013. *Scientific Data, 2*, 150042. doi: 10.1038/sdata.2015.42. Repéré à http://dx.doi.org/10.1038/sdata.2015.42

Livneh, B., Rosenberg, E. A., Lin, C., Nijssen, B., Mishra, V., Andreadis, K. M., . . . Lettenmaier, D. P. (2013). A Long-Term Hydrologically Based Dataset of Land Surface Fluxes and States for the Conterminous United States: Update and Extensions. *Journal of Climate, 26*(23), 9384-9392. doi: 10.1175/JCLI-D-12-00508.1. Repéré à https://doi.org/10.1175/JCLI-D-12-00508.1

Lucas-Picher, P., Arsenault, R., Poulin, A., Ricard, S., Lachance-Cloutier, S., & Turcotte, R. (2020). Application of a High-Resolution Distributed Hydrological Model on a U.S.-Canada Transboundary Basin: Simulation of the Multiyear Mean AnnualHydrograph and 2011 Flood of theRichelieu River Basin. *Journal of Advances in Modeling Earth Systems, 12*(4), e2019MS001709. doi: 10.1029/2019ms001709. Repéré à https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019MS001709

Lucas-Picher, P., Cattiaux, J., Bougie, A., & Laprise, R. (2015). How does large-scale nudging in a regional climate model contribute to improving the simulation of weather regimes and seasonal extremes over North America? *Climate Dynamics, 46*(3-4), 929-948. doi: 10.1007/s00382-015-2623-0. Repéré à https://doi.org/10.1007/s00382-015-2623-0

Lucas-Picher, P., Laprise, R., & Winger, K. (2016). Evidence of added value in North American regional climate model hindcast simulations using ever-increasing horizontal resolutions. *Climate Dynamics*, 1-23. doi: 10.1007/s00382-016-3227-z. Repéré à http://dx.doi.org/10.1007/s00382-016-3227-z

Luke, A., Vrugt, J. A., AghaKouchak, A., Matthew, R., & Sanders, B. F. (2017). Predicting nonstationary flood frequencies: Evidence supports an updated stationarity thesis in the United States. *Water resources research, 53*(7), 5469-5494. doi: doi:10.1002/2016WR019676. Repéré à https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016WR019676

Madsen, H., Lawrence, D., Lang, M., Martinkova, M., & Kjeldsen, T. (2013). A review of applied methods in Europe for flood-frequency analysis in a changing environment.

Madsen, H., Lawrence, D., Lang, M., Martinkova, M., & Kjeldsen, T. R. (2014). Review of trend analysis and climate change projections of extreme precipitation and floods in Europe. *Journal of hydrology, 519*(PD), 3634-3650. doi: 10.1016/j.jhydrol.2014.11.003. Repéré à http://dx.doi.org/10.1016/j.jhydrol.2014.11.003

Mani, A., & Tsai, F. T.-C. (2017). Ensemble Averaging Methods for Quantifying Uncertainty Sources in Modeling Climate Change Impact on Runoff Projection. *Journal of hydrologic engineering, 22*(4), 04016067. doi: doi:10.1061/(ASCE)HE.1943-5584.0001487. Repéré à https://ascelibrary.org/doi/abs/10.1061/%28ASCE%29HE.1943-5584.0001487

Maraun, D. (2016). Bias Correcting Climate Change Simulations - a Critical Review. *Current Climate Change Reports, 2*(4), 211-220. doi: 10.1007/s40641-016-0050-x. Repéré à https://doi.org/10.1007/s40641-016-0050-x

Maraun, D., Shepherd, T. G., Widmann, M., Zappa, G., Walton, D., Gutiérrez, J. M., . . . Mearns, L. O. (2017). Towards process-informed bias correction of climate change simulations. *Nature Climate Change, 7*(11), 764-773. doi: 10.1038/nclimate3418. Repéré à https://doi.org/10.1038/nclimate3418

Mareuil, A., Leconte, R., Brissette, F., & Minville, M. (2007). Impacts of climate change on the frequency and severity of floods in the Châteauguay River basin, Canada. *Canadian journal of civil engineering, 34*(9), 1048-1060. doi: 10.1139/l07-022. Repéré à https://doi.org/10.1139/l07-022

Martel, J.-L., Brissette, F. P., Lucas-Picher, P., Troin, M., & Arsenault, R. (2021). Climate Change and Rainfall Intensity-Duration-Frequency Curves: Overview of Science and Guidelines for Adaptation. *Journal of hydrologic engineering, 26*(10), 03121001. doi: doi:10.1061/(ASCE)HE.1943-5584.0002122. Repéré à https://ascelibrary.org/doi/abs/10.1061/%28ASCE%29HE.1943-5584.0002122

Martel, J.-L., Demeester, K., Brissette, F., Poulin, A., & Arsenault, R. (2017). HMETS-A simple and efficient hydrology model for teaching hydrological modelling, flow forecasting and climate change impacts. *The International journal of engineering education, 33*, 1307-1316.

Martel, J.-L., Mailhot, A., & Brissette, F. (2020). Global and Regional Projected Changes in 100-yr Subdaily, Daily, and Multiday Precipitation Extremes Estimated from Three Large Ensembles of Climate Simulations. *Journal of Climate, 33*(3), 1089-1103. doi: 10.1175/jcli-d-18-0764.1. Repéré à https://journals.ametsoc.org/view/journals/clim/33/3/jcli-d-18-0764.1.xml

Martynov, A., Laprise, R., Sushama, L., Winger, K., Šeparović, L., & Dugas, B. (2013). Reanalysis-driven climate simulation over CORDEX North America domain using the Canadian Regional Climate Model, version 5: model performance evaluation. *Climate Dynamics, 41*(11), 2973-3005. doi: 10.1007/s00382-013-1778-9. Repéré à https://doi.org/10.1007/s00382-013-1778-9

Massoud, E. C., Espinoza, V., Guan, B., & Waliser, D. E. (2019). Global Climate Model Ensemble Approaches for Future Projections of Atmospheric Rivers. *Earth's Future, 7*(10), 1136-1151. doi: https://doi.org/10.1029/2019EF001249. Repéré à https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019EF001249

Massoud, E. C., Lee, H., Gibson, P. B., Loikith, P., & Waliser, D. E. (2020). Bayesian Model Averaging of Climate Model Projections Constrained by Precipitation Observations over the Contiguous United States. *Journal of Hydrometeorology, 21*(10), 2401-2418. doi: 10.1175/jhm-d-19-0258.1. Repéré à https://journals.ametsoc.org/view/journals/hydr/21/10/jhmD190258.xml

McGuinness, J., & Bordne, E. (1972). A comparison of lysimeter derived potential evapotranspiration with computed values, Tech. Bull., 1452. *Agric. Res. Serv., US Dep. of Agric., Washington, DC*.

Mearns, L. O., Bukovsky, M., Pryor, S. C., & Magaña, V. (2018). Downscaling of Climate Information. Dans E. A. Lloyd & E. Winsberg (Éds.), *Climate Modelling: Philosophical and Conceptual Issues* (pp. 199-269). Cham: Springer International Publishing. doi: 10.1007/978-3-319-65058-6_8. Repéré à https://doi.org/10.1007/978-3-319-65058-6_8

Meehl, G. A., Covey, C., Taylor, K. E., Delworth, T., Stouffer, R. J., Latif, M., . . . Mitchell, J. F. (2007). The WCRP CMIP3 multimodel dataset: A new era in climate change research. *Bulletin of the American Meteorological Society, 88*(9), 1383-1394.

Meresa, H. K., & Romanowicz, R. J. (2017). The critical role of uncertainty in projections of hydrological extremes. *Hydrol. Earth Syst. Sci., 21*(8), 4245-4258. doi: 10.5194/hess-21-4245-2017. Repéré à https://www.hydrol-earth-syst-sci.net/21/4245/2017/

Merz, B., Blöschl, G., Vorogushyn, S., Dottori, F., Aerts, J. C. J. H., Bates, P., . . . Macdonald, E. (2021). Causes, impacts and patterns of disastrous river floods. *Nature Reviews*

Earth & Environment, 2(9), 592-609. doi: 10.1038/s43017-021-00195-3. Repéré à https://doi.org/10.1038/s43017-021-00195-3

Merz, B., & Thieken, A. H. (2005). Separating natural and epistemic uncertainty in flood frequency analysis. *Journal of hydrology, 309*(1), 114-132. doi: https://doi.org/10.1016/j.jhydrol.2004.11.015. Repéré à http://www.sciencedirect.com/science/article/pii/S0022169404005670

Merz, B., & Thieken, A. H. (2009). Flood risk curves and uncertainty bounds. *Natural Hazards, 51*(3), 437-458. doi: 10.1007/s11069-009-9452-6. Repéré à https://doi.org/10.1007/s11069-009-9452-6

Merz, B., Vorogushyn, S., Uhlemann, S., Delgado, J., & Hundecha, Y. (2012). HESS Opinions' More efforts and scientific rigour are needed to attribute trends in flood time series'. *Hydrology and Earth System Sciences, 16*(5), 1379-1387.

Merz, R., Parajka, J., & Blöschl, G. (2011). Time stability of catchment model parameters: Implications for climate impact analyses. *Water resources research, 47*(2). doi: doi:10.1029/2010WR009505. Repéré à https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2010WR009505

Meybeck, M., Kummu, M., & Dürr, H. H. (2013). Global hydrobelts and hydroregions: improved reporting scale for water-related issues? *Hydrol. Earth Syst. Sci., 17*(3), 1093-1111. doi: 10.5194/hess-17-1093-2013. Repéré à https://hess.copernicus.org/articles/17/1093/2013/

Minville, M., Brissette, F., Krau, S., & Leconte, R. (2009). Adaptation to climate change in the management of a Canadian water-resources system exploited for hydropower. *Water Resources Management, 23*(14), 2965-2986.

Mishra, V., Shah, H., López, M. R. R., Lobanova, A., & Krysanova, V. (2020). Does comprehensive evaluation of hydrological models influence projected changes of mean and high flows in the Godavari River basin? *Climatic Change, 163*(3), 1187-1205. doi: 10.1007/s10584-020-02847-7. Repéré à https://doi.org/10.1007/s10584-020-02847-7

Montanari, A., & Di Baldassarre, G. (2013). Data errors and hydrological modelling: The role of model structure to propagate observation uncertainty. *Advances in Water Resources, 51*, 498-504. doi: https://doi.org/10.1016/j.advwatres.2012.09.007. Repéré à https://www.sciencedirect.com/science/article/pii/S0309170812002540

Moore, C., & Doherty, J. (2005). Role of the calibration process in reducing model predictive error. *Water resources research, 41*(5).

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE, 50*(3), 885-900.

Moss, R. H., Edmonds, J. A., Hibbard, K. A., Manning, M. R., Rose, S. K., Van Vuuren, D. P., . . . Kram, T. (2010). The next generation of scenarios for climate change research and assessment. *Nature, 463*(7282), 747-756.

Motavita, D. F., Chow, R., Guthke, A., & Nowak, W. (2019). The comprehensive differential split-sample test: A stress-test for hydrological model robustness under climate variability. *Journal of hydrology, 573*, 501-515. doi: https://doi.org/10.1016/j.jhydrol.2019.03.054. Repéré à http://www.sciencedirect.com/science/article/pii/S0022169419302835

Mpelasoka, F. S., & Chiew, F. H. (2009). Influence of rainfall scenario construction methods on runoff projections. *Journal of Hydrometeorology, 10*(5), 1168-1183.

Multsch, S., Exbrayat, J. F., Kirby, M., Viney, N. R., Frede, H. G., & Breuer, L. (2015). Reduction of predictive uncertainty in estimating irrigation water requirement through multi-model ensembles and ensemble averaging. *Geosci. Model Dev., 8*(4), 1233-1244. doi: 10.5194/gmd-8-1233-2015. Repéré à https://gmd.copernicus.org/articles/8/1233/2015/

Musy, A., Meylan, P., & Favre, A.-C. (2012). *Predictive hydrology: A frequency analysis approach*. CRC Press.

Najafi, M., Moradkhani, H., & Jung, I. (2011). Assessing the uncertainties of hydrologic model selection in climate change impact studies. *Hydrological processes, 25*(18), 2814-2826.

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *Journal of hydrology, 10*(3), 282-290.

Nearing, G. S., Tian, Y., Gupta, H. V., Clark, M. P., Harrison, K. W., & Weijs, S. V. (2016). A philosophical basis for hydrological uncertainty. *Hydrological Sciences Journal, 61*(9), 1666-1678. doi: 10.1080/02626667.2016.1183009. Repéré à https://doi.org/10.1080/02626667.2016.1183009

Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., & Loumagne, C. (2005). Which potential evapotranspiration input for a lumped rainfall–runoff model?: Part 2—Towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling. *Journal of hydrology, 303*(1), 290-306. doi: https://doi.org/10.1016/j.jhydrol.2004.08.026. Repéré à http://www.sciencedirect.com/science/article/pii/S0022169404004056

Oyerinde, G., Hountondji, F., Lawin, A., Odofin, A., Afouda, A., & Diekkrüger, B. (2017). Improving Hydro-Climatic Projections with Bias-Correction in Sahelian Niger Basin, West Africa. *Climate, 5*(1), 8. Repéré à http://www.mdpi.com/2225-1154/5/1/8

Padrón, R. S., Gudmundsson, L., & Seneviratne, S. I. (2019). Observational Constraints Reduce Likelihood of Extreme Changes in Multidecadal Land Water Availability. *Geophysical research letters, 46*(2), 736-744. doi: https://doi.org/10.1029/2018GL080521. Repéré à https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018GL080521

Palmer, T., & Stevens, B. (2019). The scientific challenge of understanding and estimating climate change. *Proceedings of the National Academy of Sciences, 116*(49), 24390-24395. doi: doi:10.1073/pnas.1906691116. Repéré à https://www.pnas.org/doi/abs/10.1073/pnas.1906691116

Pechlivanidis, I., Jackson, B., Mcintyre, N., & Wheater, H. (2011). Catchment scale hydrological modelling: A review of model types, calibration approaches and uncertainty analysis methods in the context of recent developments in technology and applications. *Global NEST Journal, 13*(3), 193-214.

Pechlivanidis, I. G., Gupta, H., & Bosshard, T. (2018). An Information Theory Approach to Identifying a Representative Subset of Hydro-Climatic Simulations for Impact Modeling Studies. *Water resources research, 54*(8), 5422-5435. doi: https://doi.org/10.1029/2017WR022035. Repéré à https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2017WR022035

Peng, S., Piao, S., Ciais, P., Friedlingstein, P., Zhou, L., & Wang, T. (2013). Change in snow phenology and its potential feedback to temperature in the Northern Hemisphere over the last three decades. *Environmental Research Letters, 8*(1), 014008.

Perrin, C., Michel, C., & Andréassian, V. (2001). Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments. *Journal of hydrology, 242*(3), 275-301. doi: https://doi.org/10.1016/S0022-1694(00)00393-0. Repéré à https://www.sciencedirect.com/science/article/pii/S0022169400003930

Perrin, C., Michel, C., & Andréassian, V. (2003). Improvement of a parsimonious model for streamflow simulation. *Journal of hydrology, 279*(1), 275-289. doi: https://doi.org/10.1016/S0022-1694(03)00225-7. Repéré à http://www.sciencedirect.com/science/article/pii/S0022169403002257

Perrin, C., Oudin, L., Andreassian, V., Rojas-Serna, C., Michel, C., & Mathevet, T. (2007). Impact of limited streamflow data on the efficiency and the parameters of rainfall—runoff models. *Hydrological Sciences Journal, 52*(1), 131-151. doi: 10.1623/hysj.52.1.131. Repéré à https://doi.org/10.1623/hysj.52.1.131

Poulin, A., Brissette, F., Leconte, R., Arsenault, R., & Malo, J.-S. (2011). Uncertainty of hydrological modelling in climate change impact studies in a Canadian, snow-dominated river basin. *Journal of hydrology, 409*(3), 626-636.

Prudhomme, C., & Davies, H. (2009). Assessing uncertainties in climate change impact analyses on the river flow regimes in the UK. Part 2: future climate. *Climatic Change, 93*(1), 197-222. doi: 10.1007/s10584-008-9461-6. Repéré à https://doi.org/10.1007/s10584-008-9461-6

Prudhomme, C., Jakob, D., & Svensson, C. (2003). Uncertainty and climate change impact on the flood regime of small UK catchments. *Journal of hydrology, 277*(1), 1-23.

Prudhomme, C., Reynard, N., & Crooks, S. (2002). Downscaling of global climate models for flood frequency analysis: where are we now? *Hydrological processes, 16*(6), 1137-1150.

Räisänen, J., Ruokolainen, L., & Ylhäisi, J. (2010). Weighting of model results for improving best estimates of climate change. *Climate Dynamics, 35*(2-3), 407-422.

Randall, D. A., Wood, R. A., Bony, S., Colman, R., Fichefet, T., Fyfe, J., . . . Srinivasan, J. (2007). Climate models and their evaluation. Dans *Climate Change 2007: The physical science basis. Contribution of Working Group I to the Fourth Assessment Report of the IPCC (FAR)* (pp. 589-662). Cambridge University Press.

Read, L. K., & Vogel, R. M. (2015). Reliability, return periods, and risk under nonstationarity. *Water resources research, 51*(8), 6381-6398. doi: doi:10.1002/2015WR017089. Repéré à https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015WR017089

Reed, S., Koren, V., Smith, M., Zhang, Z., Moreda, F., Seo, D.-J., & Dmip Participants, a. (2004). Overall distributed model intercomparison project results. *Journal of hydrology, 298*(1), 27-60. doi: https://doi.org/10.1016/j.jhydrol.2004.03.031. Repéré à https://www.sciencedirect.com/science/article/pii/S0022169404002380

Reszler, C., Switanek, M. B., & Truhetz, H. (2018). Convection-permitting regional climate simulations for representing floods in small and medium sized catchments in the Eastern Alps. *Nat. Hazards Earth Syst. Sci. Discuss*.

Rhoades, A. M., Jones, A. D., & Ullrich, P. A. (2018). The Changing Character of the California Sierra Nevada as a Natural Reservoir. *Geophysical research letters, 45*(23), 13,008-013,019. doi: https://doi.org/10.1029/2018GL080308. Repéré à https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018GL080308

Riboust, P., & Brissette, F. (2015). Climate change impacts and uncertainties on spring flooding of Lake Champlain and the Richelieu River. *JAWRA Journal of the American Water Resources Association, 51*(3), 776-793.

Ricard, S., Sylvain, J.-D., & Anctil, F. (2020). Asynchronous Hydroclimatic Modeling for the Construction of Physically Based Streamflow Projections in a Context of Observation Scarcity. *Frontiers in Earth Science, 8*. doi: 10.3389/feart.2020.556781. Repéré à https://www.frontiersin.org/article/10.3389/feart.2020.556781

Rootzén, H., & Katz, R. W. (2013). Design Life Level: Quantifying risk in a changing climate. *Water resources research, 49*(9), 5964-5972. doi: doi:10.1002/wrcr.20425. Repéré à https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/wrcr.20425

Roudier, P., Andersson, J. C. M., Donnelly, C., Feyen, L., Greuell, W., & Ludwig, F. (2016). Projections of future floods and hydrological droughts in Europe under a +2°C global warming. *Climatic Change, 135*(2), 341-355. doi: 10.1007/s10584-015-1570-4. Repéré à https://doi.org/10.1007/s10584-015-1570-4

Roy, P., Gachon, P., & Laprise, R. (2014). Sensitivity of seasonal precipitation extremes to model configuration of the Canadian Regional Climate Model over eastern Canada using historical simulations. *Climate Dynamics, 43*(9-10), 2431-2453.

Ruane, A. C., & McDermid, S. P. (2017). Selection of a representative subset of global climate models that captures the profile of regional changes for integrated climate impacts assessment. *Earth Perspectives, 4*(1), 1. doi: 10.1186/s40322-017-0036-4. Repéré à https://doi.org/10.1186/s40322-017-0036-4

Rummukainen, M. (2016). Added value in regional climate modeling. *Wiley Interdisciplinary Reviews: Climate Change, 7*(1), 145-159. doi: doi:10.1002/wcc.378. Repéré à https://onlinelibrary.wiley.com/doi/abs/10.1002/wcc.378

Salas, J., & Obeysekera, J. (2014). Revisiting the Concepts of Return Period and Risk for Nonstationary Hydrologic Extreme Events. *Journal of hydrologic engineering, 19*(3), 554-568. doi: doi:10.1061/(ASCE)HE.1943-5584.0000820. Repéré à https://ascelibrary.org/doi/abs/10.1061/%28ASCE%29HE.1943-5584.0000820

Salas, J., Obeysekera, J., & Vogel, R. (2018). Techniques for assessing water infrastructure for nonstationary extreme events: a review. *Hydrological Sciences Journal, 63*(3), 325-352.

Salman, A. M., & Li, Y. (2018). Flood Risk Assessment, Future Trend Modeling, and Risk Communication: A Review of Ongoing Research. *Natural Hazards Review, 19*(3), 04018011. doi: doi:10.1061/(ASCE)NH.1527-6996.0000294. Repéré à https://ascelibrary.org/doi/abs/10.1061/%28ASCE%29NH.1527-6996.0000294

Sanchez-Gomez, E., Somot, S., & Déqué, M. (2009). Ability of an ensemble of regional climate models to reproduce weather regimes over Europe-Atlantic during the period 1961–2000. *Climate Dynamics, 33*(5), 723-736. doi: 10.1007/s00382-008-0502-7. Repéré à https://doi.org/10.1007/s00382-008-0502-7

Sanchez-Vila, X., & Fernàndez-Garcia, D. (2016). Debates—Stochastic subsurface hydrology from theory to practice: Why stochastic modeling has not yet permeated into practitioners? *Water resources research, 52*(12), 9246-9258. doi: https://doi.org/10.1002/2016WR019302. Repéré à https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016WR019302

Sanderson, B. M., Wehner, M., & Knutti, R. (2017). Skill and independence weighting for multi-model assessments. *Geosci. Model Dev., 10*(6), 2379-2395. doi: 10.5194/gmd-10-2379-2017. Repéré à https://www.geosci-model-dev.net/10/2379/2017/

Sandvik, M. I., Sorteberg, A., & Rasmussen, R. (2018). Sensitivity of historical orographically enhanced extreme precipitation events to idealized temperature perturbations. *Climate Dynamics, 50*(1), 143-157. doi: 10.1007/s00382-017-3593-1. Repéré à https://doi.org/10.1007/s00382-017-3593-1

Schmidli, J., Frei, C., & Vidale, P. L. (2006). Downscaling from GCM precipitation: a benchmark for dynamical and statistical downscaling methods. *International Journal of Climatology, 26*(5), 679-689. doi: 10.1002/joc.1287. Repéré à https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.1287

Schneider, S. H. (1983). CO2, climate and society: a brief overview. Dans *Social Science Research and Climate Change* (pp. 9-15). Springer.

Schulze, R. (2000). Transcending scales of space and time in impact studies of climate and climate change on agrohydrological responses. *Agriculture, Ecosystems & Environment, 82*(1), 185-212. doi: https://doi.org/10.1016/S0167-8809(00)00226-7. Repéré à http://www.sciencedirect.com/science/article/pii/S0167880900002267

Seibert, J. (2003). Reliability of Model Predictions Outside Calibration Conditions: Paper presented at the Nordic Hydrological Conference (Røros, Norway 4-7 August 2002). *Hydrology Research, 34*(5), 477-492. doi: 10.2166/nh.2003.0019. Repéré à https://doi.org/10.2166/nh.2003.0019

Seiller, G., & Anctil, F. (2016). How do potential evapotranspiration formulas influence hydrological projections? *Hydrological Sciences Journal, 61*(12), 2249-2266.

Seiller, G., Anctil, F., & Perrin, C. (2012). Multimodel evaluation of twenty lumped hydrological models under contrasted climate conditions. *Hydrology and Earth System Sciences, 16*(4), p. 1171 - p. 1189. doi: 10.5194/hess-1116-1171-2012. Repéré à https://hal.archives-ouvertes.fr/hal-00801716

Seiller, G., Hajji, I., & Anctil, F. (2015). Improving the temporal transposability of lumped hydrological models on twenty diversified U.S. watersheds. *Journal of Hydrology: Regional Studies, 3*, 379-399. doi: https://doi.org/10.1016/j.ejrh.2015.02.012. Repéré à http://www.sciencedirect.com/science/article/pii/S2214581815000166

Separovic, L., Alexandru, A., Laprise, R., Martynov, A., Sushama, L., Winger, K., . . . Valin, M. (2013). Present climate and climate change over North America as simulated by the fifth-generation Canadian regional climate model. *Climate Dynamics, 41*(11), 3167-3201. doi: 10.1007/s00382-013-1737-5. Repéré à https://doi.org/10.1007/s00382-013-1737-5

Separovic, L., de Elía, R., & Laprise, R. (2012). Impact of spectral nudging and domain size in studies of RCM response to parameter modification. *Climate Dynamics, 38*(7), 1325-1343. doi: 10.1007/s00382-011-1072-7. Repéré à https://doi.org/10.1007/s00382-011-1072-7

Serago, J. M., & Vogel, R. M. (2018). Parsimonious nonstationary flood frequency analysis. *Advances in Water Resources, 112*, 1-16. doi: https://doi.org/10.1016/j.advwatres.2017.11.026. Repéré à http://www.sciencedirect.com/science/article/pii/S0309170817305134

Serinaldi, F., & Kilsby, C. G. (2015). Stationarity is undead: Uncertainty dominates the distribution of extremes. *Advances in Water Resources, 77*, 17-36. doi: https://doi.org/10.1016/j.advwatres.2014.12.013. Repéré à http://www.sciencedirect.com/science/article/pii/S0309170815000020

Shen, H., Tolson, B. A., & Mai, J. (2022). Time to Update the Split-Sample Approach in Hydrological Model Calibration. *Water resources research, 58*(3), e2021WR031523. doi: https://doi.org/10.1029/2021WR031523. Repéré à https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021WR031523

Shen, M., Chen, J., Zhuan, M., Chen, H., Xu, C.-Y., & Xiong, L. (2018). Estimating uncertainty and its temporal variation related to global climate models in quantifying climate change impacts on hydrology. *Journal of hydrology, 556*, 10-24. doi: https://doi.org/10.1016/j.jhydrol.2017.11.004. Repéré à http://www.sciencedirect.com/science/article/pii/S0022169417307588

Singh, R., & AchutaRao, K. (2020). Sensitivity of future climate change and uncertainty over India to performance-based model weighting. *Climatic Change, 160*(3), 385-406. doi: 10.1007/s10584-019-02643-y. Repéré à https://doi.org/10.1007/s10584-019-02643-y

Singh, V. P., & Woolhiser, D. A. (2002). Mathematical modeling of watershed hydrology. *Journal of hydrologic engineering, 7*(4), 270-292.

Sivakumar, B. (2017). Characteristics of Hydrologic Systems. Dans *Chaos in Hydrology* (pp. 29-62). Springer.

Soares, P. M. M., & Cardoso, R. M. (2018). A simple method to assess the added value using high-resolution climate distributions: application to the EURO-CORDEX daily precipitation. *International Journal of Climatology, 38*(3), 1484-1498. doi: 10.1002/joc.5261. Repéré à http:https://doi.org/10.1002/joc.5261

Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K., . . . Miller, H. (2007). Contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change, 2007: Cambridge University Press, Cambridge.

Song, Y. H., Chung, E.-S., & Shahid, S. (2021). Spatiotemporal differences and uncertainties in projections of precipitation and temperature in South Korea from CMIP6 and CMIP5 general circulation models. *International Journal of Climatology, 41*(13), 5899-5919. doi: https://doi.org/10.1002/joc.7159. Repéré à https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.7159

Stocker, T. F., Qin, D., Plattner, G.-K., Alexander, L. V., Allen, S. K., Bindoff, N. L., . . . Xie, S.-P. (2013). Technical Summary. Dans T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex & P. M. Midgley (Éds.), *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 33–115). Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press. doi: 10.1017/CBO9781107415324.005. Repéré à www.climatechange2013.org

Su, C., & Chen, X. (2019). Assessing the effects of reservoirs on extreme flows using nonstationary flood frequency models with the modified reservoir index as a covariate. *Advances in Water Resources, 124*, 29-40. doi: https://doi.org/10.1016/j.advwatres.2018.12.004. Repéré à http://www.sciencedirect.com/science/article/pii/S0309170818305475

Tanoue, M., Hirabayashi, Y., & Ikeuchi, H. (2016). Global-scale river flood vulnerability in the last 50 years. *Scientific Reports, 6*(1), 36021. doi: 10.1038/srep36021. Repéré à https://doi.org/10.1038/srep36021

Taye, M. T., Ntegeka, V., Ogiramoi, N., & Willems, P. (2011). Assessment of climate change impact on hydrological extremes in two source regions of the Nile River Basin. *Hydrology and Earth System Sciences, 15*(1), 209-222.

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2011). An Overview of CMIP5 and the Experiment Design. *Bulletin of the American Meteorological Society, 93*(4), 485-498.

doi: 10.1175/BAMS-D-11-00094.1. Repéré à https://doi.org/10.1175/BAMS-D-11-00094.1

Te Chow, V. (1988). *Applied hydrology*. Tata McGraw-Hill Education.

Tellman, B., Sullivan, J. A., Kuhn, C., Kettner, A. J., Doyle, C. S., Brakenridge, G. R., . . . Slayback, D. A. (2021). Satellite imaging reveals increased proportion of population exposed to floods. *Nature, 596*(7870), 80-86. doi: 10.1038/s41586-021-03695-w. Repéré à https://doi.org/10.1038/s41586-021-03695-w

Teng, J., Vaze, J., Chiew, F. H. S., Wang, B., & Perraud, J.-M. (2012). Estimating the Relative Uncertainties Sourced from GCMs and Hydrological Models in Modeling Climate Change Impact on Runoff. *Journal of Hydrometeorology, 13*(1), 122-139. doi: 10.1175/jhm-d-11-058.1. Repéré à https://journals.ametsoc.org/doi/abs/10.1175/JHM-D-11-058.1

Teutschbein, C., Grabs, T., Karlsen, R. H., Laudon, H., & Bishop, K. (2015). Hydrological response to changing climate conditions: Spatial streamflow variability in the boreal region. *Water resources research, 51*(12), 9425-9446. doi: 10.1002/2015WR017337. Repéré à http:https://doi.org/10.1002/2015WR017337

Teutschbein, C., & Seibert, J. (2010). Regional climate models for hydrological impact studies at the catchment scale: a review of recent modeling strategies. *Geography Compass, 4*(7), 834-860.

Themeßl, M. J., Gobiet, A., & Heinrich, G. (2012). Empirical-statistical downscaling and error correction of regional climate models and its impact on the climate change signal. *Climatic Change, 112*(2), 449-468. doi: 10.1007/s10584-011-0224-4. Repéré à https://doi.org/10.1007/s10584-011-0224-4

Themeßl, M. J., Gobiet, A., & Leuprecht, A. (2011). Empirical-statistical downscaling and error correction of daily precipitation from regional climate models. *International Journal of Climatology, 31*(10), 1530-1544.

Thiboult, A., Anctil, F., & Boucher, M. A. (2016). Accounting for three sources of uncertainty in ensemble hydrological forecasting. *Hydrol. Earth Syst. Sci., 20*(5), 1809-1825. doi: 10.5194/hess-20-1809-2016. Repéré à http://www.hydrol-earth-syst-sci.net/20/1809/2016/

Thirel, G., Andréassian, V., & Perrin, C. (2015). On the need to test hydrological models under changing conditions. *Hydrological Sciences Journal, 60*(7-8), 1165-1173. doi: 10.1080/02626667.2015.1050027. Repéré à https://doi.org/10.1080/02626667.2015.1050027

238

Thirel, G., Andréassian, V., Perrin, C., Audouy, J. N., Berthet, L., Edwards, P., . . . Vaze, J. (2015). Hydrology under change: an evaluation protocol to investigate how hydrological models deal with changing catchments. *Hydrological Sciences Journal, 60*(7-8), 1184-1199. doi: 10.1080/02626667.2014.967248. Repéré à http://dx.doi.org/10.1080/02626667.2014.967248

Thompson, J. R., Laizé, C. L. R., Green, A. J., Acreman, M. C., & Kingston, D. G. (2014). Climate change uncertainty in environmental flows for the Mekong River. *Hydrological Sciences Journal, 59*(3-4), 935-954. doi: 10.1080/02626667.2013.842074. Repéré à https://doi.org/10.1080/02626667.2013.842074

Torma, C., Giorgi, F., & Coppola, E. (2015). Added value of regional climate modeling over areas characterized by complex terrain—Precipitation over the Alps. *Journal of Geophysical Research: Atmospheres, 120*(9), 3957-3972. doi: 10.1002/2014JD022781. Repéré à http:https://doi.org/10.1002/2014JD022781

Touzé-Peiffer, L., Barberousse, A., & Le Treut, H. (2020). The Coupled Model Intercomparison Project: History, uses, and structural effects on climate research. *WIREs Climate Change, 11*(4), e648. doi: https://doi.org/10.1002/wcc.648. Repéré à https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/wcc.648

Trenberth, K. E. (1992). *Climate system modeling*. Cambridge ; New York: Cambridge University Press.

Trenberth, K. E. (2011). Changes in precipitation with climate change. *Climate Research, 47*(1-2), 123-138.

Troin, M., Arsenault, R., Martel, J.-L., & Brissette, F. (2018). Uncertainty of hydrological model components in climate change studies over two Nordic Quebec catchments. *Journal of Hydrometeorology, 19*(1), 27-46.

Troin, M., Martel, J.-L., Arsenault, R., & Brissette, F. (2022). Large-sample study of uncertainty of hydrological model components over North America. *Journal of hydrology, 609*, 127766. doi: https://doi.org/10.1016/j.jhydrol.2022.127766. Repéré à https://www.sciencedirect.com/science/article/pii/S0022169422003419

Troin, M., Poulin, A., Baraer, M., & Brissette, F. (2016). Comparing snow models under current and future climates: Uncertainties and implications for hydrological impact studies. *Journal of hydrology, 540*, 588-602. doi: http://dx.doi.org/10.1016/j.jhydrol.2016.06.055. Repéré à http://www.sciencedirect.com/science/article/pii/S0022169416304176

Valéry, A., Andréassian, V., & Perrin, C. (2014). 'As simple as possible but not simpler': What is useful in a temperature-based snow-accounting routine? Part 2–Sensitivity analysis

of the Cemaneige snow accounting routine on 380 catchments. *Journal of hydrology, 517*, 1176-1187.

Vaze, J., Post, D. A., Chiew, F. H. S., Perraud, J. M., Viney, N. R., & Teng, J. (2010). Climate non-stationarity – Validity of calibrated rainfall–runoff models for use in climate change studies. *Journal of hydrology, 394*(3), 447-457. doi: https://doi.org/10.1016/j.jhydrol.2010.09.018. Repéré à http://www.sciencedirect.com/science/article/pii/S0022169410005986

Vehviläinen, B. (1992). Snow cover models in operational watershed forecasting. Repéré à http://hdl.handle.net/10138/25706

Velázquez, J., Anctil, F., & Perrin, C. (2010). Performance and reliability of multimodel hydrological ensemble simulations based on seventeen lumped models and a thousand catchments. *Hydrology and Earth System Sciences, 14*(11), 2303-2317.

Velázquez, J. A., Schmid, J., Ricard, S., Muerth, M. J., Gauvin St-Denis, B., Minville, M., . . . Turcotte, R. (2013). An ensemble approach to assess hydrological models' contribution to uncertainties in the analysis of climate change impact on water resources. *Hydrol. Earth Syst. Sci., 17*(2), 565-578. doi: 10.5194/hess-17-565-2013. Repéré à https://www.hydrol-earth-syst-sci.net/17/565/2013/

Vetter, T., Huang, S., Aich, V., Yang, T., Wang, X., Krysanova, V., & Hattermann, F. (2015). Multi-model climate impact assessment and intercomparison for three large-scale river basins on three continents. *Earth Syst. Dynam., 6*(1), 17-43. doi: 10.5194/esd-6-17-2015. Repéré à https://www.earth-syst-dynam.net/6/17/2015/

Vetter, T., Reinhardt, J., Flörke, M., van Griensven, A., Hattermann, F., Huang, S., . . . Krysanova, V. (2017). Evaluation of sources of uncertainty in projected hydrological changes under climate change in 12 large-scale river basins. *Climatic Change, 141*(3), 419-433. doi: 10.1007/s10584-016-1794-y. Repéré à https://doi.org/10.1007/s10584-016-1794-y

Vidal, J. P., Hingray, B., Magand, C., Sauquet, E., & Ducharne, A. (2016). Hierarchy of climate and hydrological uncertainties in transient low-flow projections. *Hydrol. Earth Syst. Sci., 20*(9), 3651-3672. doi: 10.5194/hess-20-3651-2016. Repéré à https://hess.copernicus.org/articles/20/3651/2016/

Vidrio-Sahagún, C. T., & He, J. (2022). The decomposition-based nonstationary flood frequency analysis. *Journal of hydrology, 612*, 128186. doi: https://doi.org/10.1016/j.jhydrol.2022.128186. Repéré à https://www.sciencedirect.com/science/article/pii/S0022169422007594

Vogel, R. M., & Castellarin, A. (2017). Risk, reliability, return periods, and hydrologic design.

Vogel, R. M., Yaindl, C., & Walter, M. (2011). Nonstationarity: Flood magnification and recurrence reduction factors in the united states1. *JAWRA Journal of the American Water Resources Association, 47*(3), 464-474.

Vormoor, K., Heistermann, M., Bronstert, A., & Lawrence, D. (2018). Hydrological model parameter (in)stability – "crash testing" the HBV model under contrasting flood seasonality conditions. *Hydrological Sciences Journal, 63*(7), 991-1007. doi: 10.1080/02626667.2018.1466056.   Repéré   à https://doi.org/10.1080/02626667.2018.1466056

Vormoor, K., Lawrence, D., Heistermann, M., & Bronstert, A. (2015). Climate change impacts on the seasonality and generation processes of floods--projections and uncertainties for catchments with mixed snowmelt/rainfall regimes. *Hydrology & Earth System Sciences, 19*(2).

Vormoor, K., Lawrence, D., Schlichting, L., Wilson, D., & Wong, W. K. (2016). Evidence for changes in the magnitude and frequency of observed rainfall vs. snowmelt driven floods in Norway. *Journal of hydrology, 538*(Supplement C), 33-48. doi: https://doi.org/10.1016/j.jhydrol.2016.03.066.   Repéré   à http://www.sciencedirect.com/science/article/pii/S0022169416301810

Wan, Y., Chen, J., Xu, C.-Y., Xie, P., Qi, W., Li, D., & Zhang, S. (2021). Performance dependence of multi-model combination methods on hydrological model calibration strategy and ensemble size. *Journal of hydrology, 603*, 127065. doi: https://doi.org/10.1016/j.jhydrol.2021.127065.   Repéré   à https://www.sciencedirect.com/science/article/pii/S002216942101115X

Wang, H. M., Chen, J., Cannon, A. J., Xu, C. Y., & Chen, H. (2018). Transferability of climate simulation uncertainty to hydrological impacts. *Hydrol. Earth Syst. Sci., 22*(7), 3739-3759.   doi: 10.5194/hess-22-3739-2018.   Repéré   à https://hess.copernicus.org/articles/22/3739/2018/

Wang, H. M., Chen, J., Xu, C. Y., Chen, H., Guo, S., Xie, P., & Li, X. (2019). Does the weighting of climate simulations result in a more reasonable quantification of hydrological impacts? *Hydrol. Earth Syst. Sci. Discuss., 2019*, 1-29. doi: 10.5194/hess-2019-24. Repéré à https://www.hydrol-earth-syst-sci-discuss.net/hess-2019-24/

Wasko, C., & Sharma, A. (2017). Global assessment of flood and storm extremes with increased temperatures. *Scientific Reports, 7*(1), 7945. doi: 10.1038/s41598-017-08481-1. Repéré à https://doi.org/10.1038/s41598-017-08481-1

Wehner, M., Arnold, J., Knutson, T., Kunkel, K., & LeGrande, A. (2017). Droughts, floods, and hydrology.

Wen, S., Su, B., Wang, Y., Zhai, J., Sun, H., Chen, Z., . . . Jiang, T. (2020). Comprehensive evaluation of hydrological models for climate change impact assessment in the Upper Yangtze River Basin, China. *Climatic Change, 163*(3), 1207-1226. doi: 10.1007/s10584-020-02929-6. Repéré à https://doi.org/10.1007/s10584-020-02929-6

Westra, S., Alexander, L. V., & Zwiers, F. W. (2013). Global Increasing Trends in Annual Maximum Daily Precipitation. *Journal of Climate, 26*(11), 3904-3918. doi: 10.1175/jcli-d-12-00502.1. Repéré à https://journals.ametsoc.org/view/journals/clim/26/11/jcli-d-12-00502.1.xml

Wilby, R. L. (2005). Uncertainty in water resource model parameters used for climate change impact assessment. *Hydrological processes, 19*(16), 3201-3219.

Wilby, R. L., & Dessai, S. (2010). Robust adaptation to climate change. *Weather, 65*(7), 180-185. doi: 10.1002/wea.543. Repéré à http:https://doi.org/10.1002/wea.543

Wilby, R. L., & Harris, I. (2006). A framework for assessing uncertainties in climate change impacts: Low-flow scenarios for the River Thames, UK. *Water resources research, 42*(2).

Wilby, R. L., Wigley, T., Conway, D., Jones, P., Hewitson, B., Main, J., & Wilks, D. (1998). Statistical downscaling of general circulation model output: a comparison of methods. *Water resources research, 34*(11), 2995-3008.

Woldemeskel, F. M., Sharma, A., Sivakumar, B., & Mehrotra, R. (2016). Quantification of precipitation and temperature uncertainties simulated by CMIP3 and CMIP5 models. *Journal of Geophysical Research: Atmospheres, 121*(1), 3-17. doi: https://doi.org/10.1002/2015JD023719. Repéré à https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015JD023719

Xu, Y., Gao, X., & Giorgi, F. (2010). Upgrades to the reliability ensemble averaging method for producing probabilistic climate-change projections. *Climate Research, 41*(1), 61-81. Repéré à https://www.int-res.com/abstracts/cr/v41/n1/p61-81/

Yang, H., Zhou, F., Piao, S., Huang, M., Chen, A., Ciais, P., . . . Zeng, Z. (2017). Regional patterns of future runoff changes from Earth system models constrained by observation. *Geophysical research letters, 44*(11), 5540-5549. doi: https://doi.org/10.1002/2017GL073454. Repéré à https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017GL073454

Yang, T., Wang, X., Yu, Z., Krysanova, V., Chen, X., Schwartz, F. W., & Sudicky, E. A. (2014). Climate change and probabilistic scenario of streamflow extremes in an alpine region. *Journal of Geophysical Research: Atmospheres, 119*(14), 8535-8551. doi: 10.1002/2014JD021824. Repéré à https://doi.org/10.1002/2014JD021824

242

Yang, W., Chen, H., Xu, C.-Y., Huo, R., Chen, J., & Guo, S. (2020). Temporal and spatial transferabilities of hydrological models under different climates and underlying surface conditions. *Journal of hydrology, 591*, 125276. doi: https://doi.org/10.1016/j.jhydrol.2020.125276. Repéré à https://www.sciencedirect.com/science/article/pii/S0022169420307368

Zaherpour, J., Gosling, S. N., Mount, N., Schmied, H. M., Veldkamp, T. I. E., Dankers, R., . . . Wada, Y. (2018). Worldwide evaluation of mean and extreme runoff from six global-scale hydrological models that account for human impacts. *Environmental Research Letters, 13*(6), 065015. doi: 10.1088/1748-9326/aac547. Repéré à http://dx.doi.org/10.1088/1748-9326/aac547

Zaherpour, J., Mount, N., Gosling, S. N., Dankers, R., Eisner, S., Gerten, D., . . . Wada, Y. (2019). Exploring the value of machine learning for weighted multi-model combination of an ensemble of global hydrological models. *Environmental Modelling & Software, 114*, 112-128. doi: https://doi.org/10.1016/j.envsoft.2019.01.003. Repéré à https://www.sciencedirect.com/science/article/pii/S1364815217309817

Zavaleta, S. P. I., Gonzalez, M. C., López, R. R., Poulin, A., Glaus, M., Bandala, E. E. M., & Gonzalez, E. C. (2015). Global model MOHYSE, a new tool to assess the effect of hydro-meteorological phenomena in the tropics. Dans *2015 International Conference on Computing Systems and Telematics (ICCSAT)* (pp. 1-7). doi: 10.1109/ICCSAT.2015.7362957

Zhang, X., Xu, Y.-P., & Fu, G. (2014). Uncertainties in SWAT extreme flow simulation under climate change. *Journal of hydrology, 515*, 205-222. doi: https://doi.org/10.1016/j.jhydrol.2014.04.064. Repéré à http://www.sciencedirect.com/science/article/pii/S0022169414003382

Zhang, Y., You, Q., Chen, C., & Ge, J. (2016). Impacts of climate change on streamflows under RCP scenarios: A case study in Xin River Basin, China. *Atmospheric Research, 178–179*, 521-534. doi: http://dx.doi.org/10.1016/j.atmosres.2016.04.018. Repéré à http://www.sciencedirect.com/science/article/pii/S0169809516301077

Zhao, C., Brissette, F., Chen, J., & Martel, J.-L. (2020). Frequency change of future extreme summer meteorological and hydrological droughts over North America. *Journal of hydrology, 584*, 124316. doi: https://doi.org/10.1016/j.jhydrol.2019.124316. Repéré à http://www.sciencedirect.com/science/article/pii/S0022169419310510

Zhu, Q., Zhang, X., Ma, C., Gao, C., & Xu, Y.-P. (2016). Investigating the uncertainty and transferability of parameters in SWAT model under climate change. *Hydrological Sciences Journal, 61*(5), 914-930. doi: 10.1080/02626667.2014.1000915. Repéré à https://doi.org/10.1080/02626667.2014.1000915