

Stratégies optimales de calibration de modèles hydrologiques en contexte de changements climatiques

par

Samuel BÉRUBÉ

MÉMOIRE PAR ARTICLE PRÉSENTÉ À L'ÉCOLE DE TECHNOLOGIE
SUPÉRIEURE COMME EXIGENCE PARTIELLE À L'OBTENTION
DE LA MAÎTRISE AVEC MÉMOIRE EN GÉNIE DE LA CONSTRUCTION
M. Sc. A.

MONTREAL, LE 3 DÉCEMBRE 2020

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC



Samuel Bérubé, 2020



Cette licence [Creative Commons](https://creativecommons.org/licenses/by-nc-nd/4.0/) signifie qu'il est permis de diffuser, d'imprimer ou de sauvegarder sur un autre support une partie ou la totalité de cette œuvre à condition de mentionner l'auteur, que ces utilisations soient faites à des fins non commerciales et que le contenu de l'œuvre n'ait pas été modifié.

PRÉSENTATION DU JURY

CE MÉMOIRE A ÉTÉ ÉVALUÉ

PAR UN JURY COMPOSÉ DE :

M. François Brissette, directeur de mémoire
Département de génie de la construction à l'École de technologie supérieure

M. Michel Baraër, président du jury
Département de génie de la construction à l'École de technologie supérieure

Mme Annie Poulin, membre du jury
Département de génie de la construction à l'École de technologie supérieure

IL A FAIT L'OBJET D'UNE SOUTENANCE DEVANT JURY ET PUBLIC

LE 26 NOVEMBRE 2020

À L'ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

REMERCIEMENTS

Les efforts et les sacrifices nécessaires à la réussite de cette recherche n'auraient jamais été possibles sans le support de certaines personnes. Des remerciements s'imposent. Tout d'abord, je tiens à remercier mon directeur de recherche, François Brissette. Tu as su répondre à mes nombreuses questions, tu m'as permis d'alimenter ma curiosité et, bien entendu, tu m'as orienté vers d'innombrables réflexions scientifiques. Merci. D'autres professeurs méritent également une mention : Annie, Michel et Richard. Vous m'avez tous, sans peut-être même le savoir, fait découvrir les aspects les plus intéressants de votre domaine. Merci.

Je désire également souligner l'implication sans égale de deux personnes en or : mes parents, Johanne et Victor. Sans ces êtres généreux et attentionnés, mes rêves n'auraient été que des songes improbables. Merci de m'avoir transmis la passion, la persévérance et la discipline dont j'ai eu besoin tout au long de mon parcours scolaire. Merci aussi pour tous les services rendus, même à 538 km. Je vous aime et vous en suis reconnaissant.

Enfin, ma Louise. Quand j'ai voulu remettre mes objectifs en question, quand des difficultés se sont présentées, tu étais là pour me supporter. Par ton amour et ta bonne humeur, tu m'as permis de vivre des moments heureux, même lors de passages plus ardues. Je te remercie de croire en moi chaque jour, je t'aime.

Stratégies optimales de calibration de modèles hydrologiques en contexte de changements climatiques

Samuel BÉRUBÉ

RÉSUMÉ

Depuis le début de l'anthropocène, jamais des changements aussi drastiques n'ont été observés quant à la concentration de gaz à effet de serre (GES) dans l'atmosphère. Ce phénomène est l'instigateur de perturbations actuelles et futures des processus du cycle hydrologique, aux échelles globale et régionale. À ce jour, les modèles hydrologiques utilisés pour simuler les débits requièrent une calibration basée sur les données hydrométéorologiques du passé. Dans un avenir où le climat serait considérablement altéré, ces modèles risquent donc d'être inefficaces pour étudier les impacts des changements climatiques si les processus de calibration actuels demeurent en place. Ce mémoire vise à étudier diverses *stratégies de calibration* potentiellement exploitables lors d'études d'impacts des changements climatiques.

L'analyse de ces stratégies fut réalisée à partir du modèle pluie-débit GR4J-Cemaneige, sur un échantillon de 921 bassins versants d'Amérique du Nord. Un total de cinq stratégies furent testées sur la base de 46 années non-consécutives de données hydrométéorologiques. Dans le but de cibler des périodes aux anomalies distinctes, les périodes de validation étaient composées des cinq années sélectionnées selon cinq critères différents : températures moyennes les plus élevées, précipitations annuelles les plus élevées/faibles et débits moyens les plus élevés/faibles. Les 41 années restantes furent utilisées pour la calibration, sous forme d'échantillons allant de 5 à 40 ans (par multiples de 5). Par l'entremise de cette méthodologie, les recherches ont également permis de tester les performances du modèle lorsque calibré sur des périodes de diverses longueurs. Le présent mémoire comporte également une étude de la stabilité des paramètres face aux différents climats expérimentés en calibration, qui permet d'approfondir les connaissances sur la *transposabilité temporelle* des modèles.

Bien que de nombreux écarts de performances apparaissent entre les différents bassins étudiés pour chaque stratégie, certaines conclusions générales peuvent être tirées. Les résultats démontrent que d'écarter quelques-unes des années les plus froides de la période de calibration est une pratique fiable lorsque la période de validation est constituée des années les plus chaudes. Dans le cas des quatre autres stratégies de calibration, où les conditions de validation sont fortement influencées par les précipitations, retirer un bon nombre d'années aux anomalies moindres est la solution la plus efficace. D'un autre côté, une approche de calibration basée sur toutes les années disponibles (40 ans dans ce cas-ci) procure de bonnes performances dans la plupart des cas, alors que l'utilisation d'un très petit nombre d'années (5 à 10 ans) n'est pas une approche souhaitable.

Sur le plan de la stabilité paramétrique, les résultats démontrent que les évolutions que subissent les paramètres X1, X2 et X3 à mesure que des années de calibration sont ajoutées, sont généralement de directions opposées pour des climats de calibration contrastant. Par exemple, le paramètre X2 présente majoritairement une diminution pour la stratégie des années

VIII

sèches, alors qu'il subit une augmentation dans le cas des années humides. Peu importe la stratégie de calibration utilisée, une performance élevée du modèle sur un bassin versant est garante d'une stabilité accrue des quatre paramètres de GR4J. Les deux paramètres de CemaNeige semblent plus stables pour les bassins où la précipitation neigeuse annuelle moyenne excède 200 mm et que les années de calibration incluent davantage d'années plus froides. Cette stabilité est toutefois plus évidente pour le coefficient de pondération de l'état thermique du manteau (X5) que pour le facteur degré-jour (X6).

Mots-clés : modélisation hydrologique; changements climatiques; stratégies de calibration; stabilité des paramètres.

Optimal hydrological model calibration strategies in a context of climate change

Samuel BÉRUBÉ

ABSTRACT

Since the beginning of the Anthropocene, never have such drastic changes been observed in the concentration of greenhouse gases (GHGs) in the atmosphere. This phenomenon is the instigator of current and future disturbances of the hydrological cycle, at global and regional scales. To date, the hydrological models used to simulate streamflows require calibration based on historical hydrometeorological data. In a future where the climate would be considerably altered, these models may therefore be ineffective in studying the impacts of climate change if the current calibration processes remain in place. This thesis aims to study various calibration strategies useful for climate change impact studies.

The analysis of these strategies was carried out using the GR4J-Cemaneige rainfall-runoff model, on a sample of 921 North American watersheds. A total of five calibration strategies were tested based on 46 non-consecutive years of hydrometeorological data. In order to target periods with distinct anomalies, validation periods consisted of five years selected according to five different criteria: highest average temperatures, highest/lowest annual precipitation and highest/lowest average streamflows. The remaining 41 years were used for calibration, with samples ranging from 5 to 40 years (increments of 5). Using this methodology, the research also tested the performance of the model when calibrated over time periods of various lengths. This thesis also includes a study of parameter stability through different climates experienced in calibration, which allows us to deepen our knowledge on the temporal transposability of hydrological models.

Although there is a lot of across-catchment variability in the results, some general conclusions can nonetheless be drawn. Results show that removing some of the coldest years from the calibration period is a reliable practice when the validation period consists of the warmest years. For the remaining four calibration strategies, where the validation conditions are heavily influenced by precipitation, removing a large number of most dissimilar years is the most effective solution. A calibration approach based on all available years (40 years in this case) provides good performance in most cases but is not optimal. On the other hand, using a small number of years (5 to 10 years) is not an optimal approach.

In terms of parameter stability, results show that the changes undergone by the parameters X1, X2 and X3 as years of calibration are added, are generally in opposite directions for contrasting calibration climates. For example, the parameter X2 mainly shows a decrease for the dry years strategy, while it undergoes an increase in the case of wet years. Regardless of the calibration strategy used, a high calibration performance of the model on a watershed guarantees increased stability of the four parameters of GR4J. Globally, the two CemaNeige parameters seem more stable for catchments where the average annual snowfall exceeds 200 mm and the calibration

X

years include more cold years. This stability is however more obvious for the coefficient of weighting of the thermal state of the mantle (X5) than for the degree-day factor (X6).

Keywords: hydrological modeling; climate change; calibration schemes; parameter stability.

TABLE DES MATIÈRES

| | Page |
|---|------|
| INTRODUCTION | 1 |
| CHAPITRE 1 REVUE CRITIQUE DE LA LITTÉRATURE..... | 3 |
| 1.1 Transposabilité temporelle des modèles | 3 |
| 1.2 Stabilité des paramètres | 4 |
| CHAPITRE 2 PROCÉDURE DE TRAVAIL | 7 |
| 2.1 Objectif | 7 |
| 2.2 Contenu du mémoire..... | 7 |
| 2.3 Structure de l'article..... | 8 |
| CHAPITRE 3 OPTIMAL HYDROLOGICAL MODEL CALIBRATION STRATEGY FOR CLIMATE CHANGE IMPACT STUDIES | 9 |
| 3.1 Abstract..... | 9 |
| 3.2 Introduction..... | 10 |
| 3.2.1 Purpose..... | 10 |
| 3.2.2 Previous studies | 11 |
| 3.2.3 Main objective | 14 |
| 3.3 Data and methods..... | 15 |
| 3.3.1 Data gathering..... | 15 |
| 3.3.2 Hydrological model | 17 |
| 3.3.3 Calibration schemes | 18 |
| 3.4 Results..... | 20 |
| 3.4.1 All approaches | 20 |
| 3.4.2 Consistency across watersheds | 23 |
| 3.4.3 Gains in KGE achieved during the validation period | 26 |
| 3.5 Discussion..... | 26 |
| 3.6 Conclusion | 31 |
| CHAPITRE 4 ÉTUDE DE STABILITÉ DES PARAMÈTRES | 33 |
| 4.1 Étude préliminaire..... | 33 |
| 4.2 Survol des paramètres de GR4J-CemaNeige..... | 33 |
| 4.3 Méthodologie | 34 |
| 4.4 Résultats..... | 36 |
| 4.4.1 Test de Mann-Kendall..... | 36 |
| 4.4.2 Croissance et décroissance des paramètres..... | 37 |
| 4.4.3 Stabilité paramétrique selon la performance en calibration..... | 38 |
| 4.4.4 Bassins sous influence nivale..... | 41 |
| 4.5 Discussion..... | 44 |
| CONCLUSION..... | 51 |

| | | |
|------------|---|----|
| ANNEXE I | ÉVOLUTION DES VALEURS DE PARAMÈTRES (X1 À X6) | 55 |
| ANNEXE II | STRUCTURE DU MODÈLE PLUIE-DÉBIT GR4J-CemaNeige | 59 |
| ANNEXE III | COORDONNÉES DES BASSINS INFLUENCÉS PAR LA NEIGE | 61 |
| | LISTE DE RÉFÉRENCES BIBLIOGRAPHIQUES..... | 62 |

LISTE DES TABLEAUX

| | | Page |
|-------------|---|------|
| Table 3.1 | Distribution of key attributes for the 921-catchment ensemble..... | 17 |
| Table 3.2 | Calibration strategies | 19 |
| Table 3.3 | Calibration layout for the warm years experiment (W)..... | 19 |
| Table 3.4 | Statistical significance of validation results for various calibration lengths: Conover's nonparametric test (top right); Wilcoxon rank sum test (bottom left)..... | 22 |
| Table 3.5 | Statistic analysis of KGE gains/losses in validation compared to using all 40 years | 26 |
| Tableau 4.1 | Résumé du rôle des paramètres du modèle GR4J couplé au module CemaNeige..... | 34 |
| Tableau 4.2 | Nombre de bassins présentant une tendance significative | 37 |
| Tableau 4.3 | Nombre de bassins présentant une pente de Sen (significative ou non) ascendante ou descendante | 38 |
| Tableau 4.4 | Nombre de bassins présentant une pente de Sen statistiquement significative ($\alpha = 0,05$) ascendante ou descendante..... | 38 |
| Tableau 4.5 | Nombre de bassins (%) présentant une tendance significative au sein des 307 membres du sous-groupe 1 (S1)..... | 46 |
| Tableau 4.6 | Nombre de bassins présentant une tendance significative ascendante ou descendante pour chaque paramètre au sein des membres du sous-groupe 1 (S1) | 46 |

LISTE DES FIGURES

| | | Page |
|------------|---|------|
| Figure 3.1 | Location (centroid) of the 921 study catchments with their KGE calibration performance bracket: yellow (0.6 to 0.7); red (0.7 to 0.8); black (0.8 to 0.9) and blue (above 0.9). | 16 |
| Figure 3.2 | KGE validation results for each calibration strategy. [W] Warm years, [D] Dry years, [H] Wet/humid years, [LQ] Low-runoff years, [HQ] High-runoff years, with digits representing the Nth warmest (wettest, driest ...) years used for the calibration period | 20 |
| Figure 3.3 | Number of occurrences of best (blue bars) and worst (red bars) calibration strategies/years out of 921 catchments | 23 |
| Figure 3.4 | Occurrences (%) for which KGE for specific calibration years exceeds KGE using the 40-year calibration strategy. (dT, dP D, dP, dQ LQ and dQ represent temperature, precipitation and runoff anomalies between the 5-year validation period and the remaining 41 years available for calibration) | 25 |
| Figure 3.5 | Distribution of validation KGE (blue boxes) values stemming from random calibration years along with the performances of the 'W' (red circles) and 40-year (red dotted line) strategies (Temperature anomalies (dT) are given above each whisker plot and catchment locations are displayed as red markers on the map)..... | 29 |
| Figure 4.1 | Stabilité du paramètre X1 pour trois sous-groupes de bassins versants | 39 |
| Figure 4.2 | Stabilité du paramètre X2 pour trois sous-groupes de bassins versants | 39 |
| Figure 4.3 | Stabilité du paramètre X3 pour trois sous-groupes de bassins versants | 40 |
| Figure 4.4 | Stabilité du paramètre X4 pour trois sous-groupes de bassins versants | 40 |
| Figure 4.5 | Évolution du paramètre X5 par rapport à la chute de neige annuelle moyenne (stratégie des années chaudes) pour les bassins où la pente de Sen est significative (orange) et non significative (bleu)..... | 42 |
| Figure 4.6 | Évolution du paramètre X6 par rapport à la chute de neige annuelle moyenne (stratégie des années chaudes) pour les bassins où la pente de Sen est significative (orange) et non significative (bleu)..... | 43 |

LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES

| | |
|-------|--|
| CMAES | Covariance Matrix Adaptation Evolution Strategy |
| DSST | Differential Split-Sample Testing |
| EEN | Équivalent en eau de la neige |
| ESM | Earth System Model |
| GCM | General Circulation Model |
| GES | Gaz à effet de serre |
| GIEC | Groupe d'experts intergouvernemental sur l'évolution du climat |
| HU | Hydrogramme unitaire |
| KGE | Kling-Gupta Efficiency |
| RCP | Representative Concentration Pathway |

INTRODUCTION

L'hydrologie, ou la science qui étudie les aspects du cycle de l'eau en général, est un domaine qui comporte plusieurs incertitudes. Au cours de sa carrière, un hydrologue fait face à des incertitudes liées aux prévisions climatiques futures, aux modélisations simplifiées de processus naturels et aux données hydro-climatiques compilées au fil des années. Bien que tous ces aspects nécessitent une attention particulière, un enjeu important demeure au centre des réflexions depuis quelques décennies : l'impact du changement climatique, observé actuellement et dans les futurs proche et éloigné, sur la performance des modèles hydrologiques. En effet, le climat subit des changements attribués à deux types de variabilités. La variabilité interne est principalement causée par la circulation de nature chaotique observable dans l'atmosphère et les océans (Lorenz, 1995). Elle occasionne des perturbations régionales d'ampleurs annuelle, interannuelle, décennale, et multi décennale (Mayewski et al., 2004). La variabilité externe peut être engendrée, notamment, par les variations de positionnement du globe ou par une augmentation, attribuée à l'activité humaine, de la concentration des gaz à effet de serre dans l'atmosphère (IPCC, 2013). À ce jour, Hawkins et Sutton (2012) ont démontré que les changements de températures causés par l'activité humaine excèdent déjà les anomalies occasionnées par la variabilité interne. En ce qui concerne les précipitations extrêmes, le consensus scientifique pointe vers un changement de variabilité (Martel et al., 2020). Considérant que les données historiques sont actuellement utilisées pour calibrer les modèles, le contraste entre ces données et celles du climat futur (IPCC, 2013) pourrait compromettre la performance des modèles. En d'autres termes, les modèles ne seront peut-être plus aussi efficaces pour simuler des débits s'ils sont calibrés sur une période ayant ses propres caractéristiques pour ensuite être transposés sur une période aux conditions climatiques distinctes. Dans l'optique où les modèles hydrologiques sont aujourd'hui fréquemment utilisés pour prendre des décisions par rapport aux infrastructures (conception) et aux opérations (barrages, digues, irrigation, etc.), il est essentiel d'étudier la transposabilité temporelle des modèles pluie-débit.

L'article introduit dans ce mémoire a pour objectif principal l'évaluation de stratégies de calibration de modèles hydrologiques qui seraient potentiellement appropriées dans un climat

présentant un contraste non négligeable avec les conditions observées au cours du dernier siècle. Ces stratégies se base sur une méthode étudiée précédemment par d'autres chercheurs: l'utilisation de périodes de calibration ciblées, constituées d'un nombre restreint d'années qui contrastent ou non avec le climat de validation, afin de possiblement améliorer la robustesse des modèles en conditions non-stationnaires (i.e. Gharari et al. (2013)). Parallèlement, une comparaison de ces stratégies avec une approche de calibration impliquant toutes les années de données disponibles est présentée dans cet article, pour valider la robustesse de cette méthode en climat non-stationnaire, tel que démontré dans la littérature pour un climat stationnaire (Arsenault et al., 2018). Dans le but d'étendre les recherches au sujet de la transposabilité temporelle des jeux de paramètres, ce mémoire contient un chapitre axé sur l'améliorations des connaissances quant à la stabilité des jeux de paramètres optimaux découlant de la calibration.

Le chapitre 1 constitue une revue des travaux réalisés sur le sujet au cours des dernières années. Par la suite, une description de la démarche de travail et une introduction de l'article sur lequel porte ce mémoire sont présentées au chapitre 2. Suivant ce préambule, l'article est joint au chapitre 3. L'étude concernant la stabilité temporelle des paramètres se trouve au chapitre 4. Une conclusion et des annexes complètent également l'information vers la fin du document.

CHAPITRE 1

REVUE CRITIQUE DE LA LITTÉRATURE

1.1 Transposabilité temporelle des modèles

Dans un contexte où le climat terrestre subit des bouleversements accentués par l'activité humaine, la transposabilité des modèles représente un atout important afin de simuler les débits dans des conditions différentes de celles expérimentées lors du dernier siècle. En effet, en plus de générer une forte concordance entre les débits simulés et les débits observés (performance), la transposabilité temporelle est une caractéristique que les modèles hydrologiques doivent absolument posséder (Thirel et al., 2015). Pour ce faire, la structure du modèle hydrologique ainsi que le jeu de paramètres utilisé doivent être transposables. Des auteurs ont élaboré diverses méthodes d'échantillonnage pour tester des conditions climatiques variées en calibration et en analyser les répercussions sur les performances. La plupart de ces approches se basent sur la méthode du *Differential Split-Sample Testing* (DSST) (Klemeš, 1986), qui permet d'effectuer des calibrations sur un maximum de périodes contrastantes. Ayant utilisé le DSST, Brigode et al. (2013) mentionnent que des améliorations sont nécessaires à l'intérieur même de la structure des modèles, en plus d'affirmer que le manque de robustesse des deux modèles testés (GR4J et TOPMO) constitue la source principale d'incertitude des projections de débits futurs. Une étude réalisée par Merz et al. (2011) avec 6 échantillons distincts de 5 années consécutives démontre qu'une relation claire existe entre la période de calibration utilisée et les paramètres liés à la neige et à la teneur en eau du sol. Coron et al. (2012) ont également fait intervenir une version dérivée de la méthode DSST pour démontrer que « la calibration réalisée sur une période plus humide (sèche) mène à une surestimation (sous-estimation) des débits simulés moyens » [traduction libre]. Cette étude souligne un besoin d'améliorer la structure des modèles afin de minimiser les erreurs, ce qui va de pair avec les recherches de Brigode et al. (2013). Toujours au sujet des périodes de calibration, Vaze et al. (2010) expliquent que les débits en période sèche sont simulés plus efficacement lorsqu'un modèle est calibré sur des années humides, en comparaison avec la situation inverse. Vansteenkiste et al. (2014) soulignent que le choix de la stratégie de calibration génère

d'avantage d'impacts sur les débits d'étiage que sur les crues. Puis, Motavita et al. (2019) ont montré que les conditions climatiques expérimentées en calibrations sur un bassin allemand ont eu un impact sur les performances plus important que la longueur de la période de calibration sélectionnée.

Certaines études permettent de tracer le portrait global d'une période de calibration idéale, considérant un climat stationnaire : un nombre suffisant d'années (Arsenault et al., 2018) et une variété de conditions climatiques (e.g. périodes sèches et périodes humides; Anderson (2002)). Également, Li et al. (2015) ont démontré que les modèles aux performances initialement convaincantes sont plus enclins à conserver cette efficacité pour des climats aux contrastes élevés.

1.2 Stabilité des paramètres

Bien souvent, le jeu de paramètres optimal obtenu suite au processus de calibration d'un modèle hydrologique peut générer des performances inférieures lorsqu'utilisé sur une période démontrant un climat trop peu similaire. En effet, tel que démontré par Coron et al. (2011), « les paramètres issus de la calibration sont parfois dépendants du climat expérimenté en calibration » [traduction libre]. La relation qui existe entre ces deux entités n'est toutefois pas parfaitement comprise à ce jour. En ce qui concerne les périodes présentant des crues saisonnières (printemps et automne), des diminutions de performance ont été observées lorsque la validation s'effectue sur une période aux crues en contraste avec les conditions de calibration (Vormoor et al., 2018). Toutefois, rien n'indique que les extrêmes saisonniers amplifient ces diminutions d'efficacité. Le Lay et al. (2007) ont tiré des conclusions similaires en étudiant les contrastes climatiques d'un bassin béninois : l'ampleur des variations des paramètres découlant de la calibration n'est pas un indicateur précis des changements hydrologiques à l'intérieur du bassin versant.

Dans le même ordre d'idée, il a été observé que l'utilisation de plusieurs jeux de paramètres différents peut engendrer des performances similaires (Beven & Freer, 2001). Cet aspect se

nomme l'équifinalité. Elle fut approfondie par de nombreuses études dans le domaine de l'hydrologie. Her et al. (2019) stipulent que les projections des modèles concernant la teneur en eau du sol et les eaux souterraines sont plus affectées par l'équifinalité que par l'incertitude qui réside dans les projections des *General Circulation Models* (GCM) et des *Earth System Models* (ESM). Certains modèles hydrologiques risquent davantage d'être impactés par l'équifinalité puisqu'ils renferment plusieurs paramètres, ce qui augmente le nombre de combinaisons potentiellement optimales.

Sommes toutes, les recherches antérieures démontrent que le lien entre la stabilité des paramètres optimaux, le climat de calibration et la structure du modèle est complexe. L'usage de stratégies de calibration conçues spécifiquement pour les études d'impacts des changements climatiques serait une solution avantageuse, qui sera étudiée dans ce mémoire.

CHAPITRE 2

PROCÉDURE DE TRAVAIL

2.1 Objectif

Dans l'optique de poursuivre les recherches sur de nouvelles stratégies de calibrations à utiliser pour évaluer les impacts des changements climatiques, le présent mémoire a pour objectif de traiter de deux sujets distincts, qui portent sur la même problématique globale. Dans un premier temps, l'article a pour mission de procurer aux hydrologues de nouvelles connaissances quant aux pratiques de calibration qui amélioreraient les performances des modèles pluie-débit dans des conditions climatiques variées. Puis, dans un second temps, le chapitre qui suivra celui de l'article constitue une étude complémentaire sur la stabilité des paramètres, qui demeure à ce jour peu comprise lorsque le climat sur lequel un modèle est employé diffère considérablement du climat expérimenté en calibration. Ce chapitre est complémentaire au sujet de l'étude et fait usage des mêmes bassins versants, des mêmes stratégies de calibration, ainsi que des mêmes résultats en validation que ceux de l'article.

Les deux aspects de cette recherche sont, tel que mentionné précédemment, des sujets d'étude qui se juxtaposent et forment une investigation à part entière du problème énoncé dans les premières lignes de ce mémoire.

2.2 Contenu du mémoire

Le chapitre qui suivra celui-ci est constitué de l'article dans son intégralité (références incorporées au mémoire), rédigé en anglais et soumis pour publication en date du 16 septembre 2020 au *Journal of Hydrology*. Cette référence dans le domaine des sciences hydrologiques est une revue qui vise à publier des avancées importantes, réalisées par une multitude d'auteurs au fil des ans. Ses articles font l'objet de révisions avant publication. L'article présenté dans ce mémoire vient s'inscrire dans une suite logique qui regroupe un ensemble d'autres ouvrages portant sur la question de la transposabilité temporelle des modèles hydrologiques en contexte

de changements climatiques. Certains se sont attardé à la stabilité des jeux de paramètres découlant de calibrations sur différentes périodes, alors que d'autres ont permis de comprendre les lacunes de certains modèles devant des données hydrométéorologiques aux anomalies importantes, par rapport à la période préindustrielle. La majorité de ces recherches se basent sur les particularités du climat sur lequel les calibrations s'effectuent (températures, précipitations, débits, etc.) pour faire des liens entre ces variables, les paramètres obtenus et la performance des modèles.

2.3 Structure de l'article

Afin d'être conforme aux exigences de mise en page, l'article fut adapté au format du mémoire, tout comme les figures et tableaux qui en font partie. Une liste des références bibliographiques utilisées lors de la rédaction de cet article se trouve à la fin du présent document, avec les références propres au mémoire. Un résumé (*abstract*) permet de comprendre le sujet de l'étude ainsi que les principales conclusions qui en découlent dès le début de la lecture. L'introduction est divisée de façon à résumer le contexte, la problématique et l'objectif de l'article. La section suivante décrit la méthodologie employée dans le cadre de cette recherche, en plus de présenter un aperçu des caractéristiques et de l'emplacement des bassins versants à l'étude. La présentation des résultats démontre quels ont été les produits de cette recherche, qui sont ensuite analysés en détail dans la section *Discussion*. Finalement, une conclusion permet de synthétiser les résultats, les connaissances acquises et de revenir sur l'objectif initial, tout en laissant une ouverture pour d'autres angles d'étude du sujet.

CHAPITRE 3

OPTIMAL HYDROLOGICAL MODEL CALIBRATION STRATEGY FOR CLIMATE CHANGE IMPACT STUDIES

Samuel Bérubé ^a, François Brissette ^a, Richard Arsenault ^a

^a *Département de génie de la construction, École de Technologie Supérieure,
1100 rue Notre-Dame Ouest., Montréal, Québec, Canada H3C 1K3*

Cet article a été soumis pour publication au *Journal of Hydrology* en Septembre 2020

3.1 Abstract

Increases in greenhouse gas (GHG) concentration are driving changes in current and future hydrological cycles, both at the global and regional scales. To assess future impacts on water resources, hydrological models are the most commonly used tools to simulate future flows. Most hydrological models require a calibration procedure based on historical hydrometeorological data, which may not be fully representative of the future climate, and which may therefore induce biases in estimated future flows. This paper assesses various calibration strategies to minimize this issue. The impact of these calibration strategies is assessed on 921 North American catchments using a lumped hydrological model. Forty-six years of hydrometeorological data were available for each of the 921 catchments. Five calibration strategies (warm, high and low rainfall, high- and low-flow years) were investigated, each using a 5-year (non-continuous) independent validation dataset maximizing all five studied climate anomalies. The remaining 40 years were used as a pool of calibration years, using targeted subsets of years in multiples of 5 (from 5 to 40 years) to assess the impact of the number of calibration years versus the climate anomaly of each calibration subset. Results showed large cross-catchment variability, indicating that no single calibration strategy and number of calibration years was optimal for all watersheds. However, the large number of catchments used in this study nonetheless allows for some general conclusions to be drawn. For the calibration strategy for warm years, using a large number of years was the approach most likely to be successful, indicating that dropping a small subset of cold years was

preferable to keeping a small subset of warm years. For the other four calibration strategies, the approach most likely to succeed was the one in which about half of the years in the 40-year historical record were kept. These findings are probably attributable to the relative contribution of internal climate variability to both temperature and precipitation, with the former being relatively larger than the latter. For the warm year calibration strategy, keeping a larger number of years for calibration ensures better model robustness to account for precipitation variability in the validation set, whereas for the other four calibration strategies, which relied exclusively or largely on precipitation, a larger number of years had to be dropped in order to account for the much larger differences between wet and dry years. In all cases, using a small number of calibration years was not likely to be the most successful approach.

3.2 Introduction

3.2.1 Purpose

Most hydrological models require a certain degree of parameter adjustment - or calibration - to provide accurate streamflow simulations. The calibration process ideally requires long time series of historical hydrometeorological data. This data is gathered from hydrometeorological records representing years, and sometimes decades, of natural past events. Long time series are widely thought to be beneficial as they provide a wider range of past hydrometeorological conditions against which the hydrological model parameter space can be explored. In theory, this should result in a more robust model. High quality historical data has thus traditionally been considered as the backbone of reliable simulations (Perrin et al., 2007). However, the global warming observed over the past decades due to anthropogenic forcing (Pachauri et al., 2014) raises questions as to the reliability of historical data. Future greenhouse gas (GHG) emission scenarios point to increasing climate non-stationarity and potentially important impacts on the hydrological cycle. Since hydrological models are now routinely used to make decisions relating to operations (e.g., flood forecasts, reservoir management) and design (e.g., new infrastructures, climate change adaptation strategies), it is critical to assess model stability in a non-stationary climate.

3.2.2 Previous studies

Temporal transposability of model parameters is an important topic that is often only an afterthought, even when running a hydrological model over a period long enough to display patterns of climate variability (Thirel et al., 2015). Climate variability/change results from a combination of unforced and forced processes (Deser et al., 2012; Folland et al., 2001; Goosse et al., 2005). Unforced or internal variability results from the chaotic nature of the atmospheric and oceanic circulation (Lorenz, 1995). It manifests itself through regional climate anomalies, which may persist at the annual, interannual, decadal and multidecadal scales (Mayewski et al., 2004). Forced or external variability encompasses many sources of variability (e.g., solar cycles, orbital change), including anthropogenic change brought about by increases in atmospheric greenhouse gas concentration (Pachauri et al., 2014; Vogel & o'Brien, 2017). Hydrological model calibration has always had to contend with the problem of internal climate variability, which is normally addressed by using long calibration periods to encompass varying hydroclimatic conditions. However, anthropogenic forcing has already brought temperature changes larger than those related to internal variability in many regions of the world (Hawkins & Sutton, 2012; Lehner et al., 2017). The picture is more complex for precipitation, because in this case, internal variability is much larger than for temperature, and the time of emergence of the anthropogenic contribution is expected much later in this century in most regions of the world (Martel et al., 2018). Even though mean precipitation changes may be small in some regions, there is a strong consensus regarding changes in variability, and especially so for precipitation extremes (Martel et al., 2020). A relatively clear picture of the likely future climate has emerged over the past two decades (IPCC, 2013): the future climate will see higher temperatures well outside the current range of internal variability and increased variability across all scales for both precipitation and temperature. The reliability of future hydrological projections obtained from hydrological models calibrated on past hydroclimatic data is therefore potentially problematic, and especially so if the projections are used as the basis for designing adaptation strategies, or to design infrastructures with long service life.

In order to assess the impacts of climate change on hydrology, both the hydrological model structure and parameter sets need to be transposable in a contrasting climate. An investigation of the impact of parameter calibration and model structure selection for climate change impact studies has demonstrated noticeable runoff projection discrepancies between four distinct model structures (Mendoza et al., 2015). Poulin et al. (2011) found the impact of the model structure to be greater than the uncertainty from model parameters for climate change impact assessments on a snow-dominated river basin. In a similar study, Brigode et al. (2013) found that the two hydrological models used in their study were the main source of uncertainty for future streamflow projections. They stated that improvements were necessary in the hydrological model structure for climate change impact studies.

Since free parameters are implemented into hydrological models to add “more dimensions of flexibility” (Kirchner, 2006), calibration allows to rectify errors generated by these simplified renderings of hydrological processes. The problem is that optimal parameter sets provided by a calibration process can sometimes perform more poorly when applied to other time periods or contrasted climates (Gharari et al., 2013). This well-known issue can be attributed to the fact that parameters occasionally generate attributes dependent on their calibration climate, thus compromising stability over dissimilar conditions. Coron et al. (2011) noted that previous studies still do not agree on a clear explanation regarding this matter. The relationship between optimal parameter sets and the calibration period from which they stem is therefore still not perfectly understood.

In fact, many different parameter sets can oftentimes lead to relatively similar model performances (equifinality) (Beven & Freer, 2001). While this phenomenon has been studied for a long time, it has become critical for climate change impact studies. Poulin et al. (2011) demonstrated that parameter equifinality remains a source of uncertainty in a context of climate change, albeit not the prevailing one. Her et al. (2019) demonstrated that groundwater and soil moisture model projections could be more impacted by equifinality than by the uncertainty stemming from General Circulation and Earth System models. On the other hand, in some cases, uncertainty arising from equifinality appears to be less important than that of

hydrological model structure, and especially in low-flow conditions (Najafi et al., 2011). The above results underline the importance of having a robust parameter set, and therefore, point to the criticality of hydrological model calibration for climate change impact studies.

Even though GHG emission scenarios and climate model simulations have traditionally been believed to enclose greater uncertainty than that related to the hydrological modelling process itself (Kay et al., 2009; Prudhomme & Davies, 2009), recent studies (Giuntoli et al., 2018; Krysanova et al., 2018) have shown that in some cases, the hydrological model structure may dominate global uncertainty of future streamflow projection. Since calibrated parameter sets can somewhat compensate for model structure deficiencies, the issue of hydrological model calibration for climate change impact studies remains an important concern for modellers. In a stationary climate, some calibration practices are considered to be optimal. Superior validation performances usually occur with longer calibration periods (Arsenault et al., 2018), suggesting that a proper calibration period should “contain a variety of hydrological conditions” (e.g., wet and dry periods) to achieve better model robustness (Anderson, 2002). On the other hand, Motavita et al. (2019) showed, over a single German watershed, that hydrological conditions over the calibration period have more influence on performance than does the time series length. To evaluate calibration strategies that could grant better robustness in a non-stationary context, split-sample procedures have often been used to simulate as many fictive climatic scenarios as possible, employing a smaller but more targeted number of years to define the calibration period. An investigation of Differential Split-Sample Testing (DSST) (see Klemeš, 1986) was conducted by Merz et al. (2011) for 6 consecutive 5-year samples to assess the parameter stability of a semi-distributed conceptual model (similar to HBV) involving 273 catchments in Austria. They showed a clear relationship between snow and soil-related model parameters, depending on the calibration period. Seiller et al. (2015) applied a multi-model approach using 20 different lumped hydrological models over 20 U.S. catchments. They demonstrated that using a combination of weighted rainfall-runoff models improved the performance and reliability of simulations, as compared to a single model, in contrasting climates. Using an Australian ensemble of 216 catchments, Coron et al. (2012) implemented a split-sample approach called Generalized Split-Sample Testing (GSST), which

consisted in testing all possible combinations of 5-year sub-periods for calibration. They found that "calibration over a wetter (drier) climate than the validation climate leads to an overestimation (underestimation) of the mean simulated runoff". Furthermore, they also found that model structure should be assessed to minimize performance losses in climate change impact studies, corroborating previous work. Li et al. (2015) showed that models initially displaying high performance are more likely to maintain good performance across contrasted conditions. They also concluded that the performance and parameter stability of hydrological models are better on humid catchments than on dry catchments. Vaze et al. (2010) drew similar conclusions over 61 Australian catchments. They showed that runoff during dry periods is simulated more accurately when using a model calibrated over a wet period, as compared to the reverse. This conclusion is in line with those of Vansteenkiste et al. (2014), who found that the calibration strategy had a more critical impact on low flows than on peak flows.

Previous work point to a complex relationship between parameter sets, the climate over the calibration period, and the model structure. Identifying time-stable parameter sets would be one way to minimize performance losses, while adapting the model structure is another potential direction. Another possibility is to apply 'custom-built' calibration strategies, which should increase model robustness for various climates. This last approach is the focus of this work.

3.2.3 Main objective

The main objective of this work is to investigate custom calibration strategies in order to improve model robustness for climate change impact studies. Two competing specific sub-objectives will also be assessed:

- Validate if longer historical time series translate into better model robustness in non-stationary conditions, as demonstrated by Arsenault et al. (2018) under stationary conditions;
- Assess if model robustness in non-stationary conditions can be enhanced by focusing on smaller, more contrasted calibration periods (e.g., Gharari et al., 2013).

This work uses a small independent subset of anomalous years (warm, humid and dry) as a target for model performance using all remaining years in the historical record as the pool for calibration years. These calibration years will be used in different combinations of climate anomalies and durations to test how the number of selected years and climate anomalies influence the model performance (and therefore, the robustness) over the independent validation period. Work conducted here was based on the hypothesis that an efficient calibration period should include all available years of meteorological data, regardless of the actual climate stationarity. This work will aggregate results from a very large number of catchments from various climate zones to provide a better understanding of climate change impacts and insights on accurate calibration strategies.

3.3 Data and methods

3.3.1 Data gathering

The study region is composed of 921 North American catchments selected from the NAC2H database (Arsenault et al., 2020a), which is a subset of pre-processed catchments for climate change impact studies from the larger HYSETS database containing 14,425 catchments (Arsenault et al., 2020b). Watersheds were selected based on size (surface area $> 500 \text{ km}^2$), data availability, and using a minimal calibration performance threshold ($\text{KGE} > 0.6$) over the whole 45-year period. Data availability was based on the following criterion: minimum of 46 non-consecutive years of full year data between 1968 and 2018. The 921 selected watersheds cover most North American climate zones.

Figure 3.1 presents the location of the centroid of all watersheds as well as the range of the calibration KGE coefficients over all available years.



Figure 3.1 Location (centroid) of the 921 study catchments with their KGE calibration performance bracket: yellow (0.6 to 0.7); red (0.7 to 0.8); black (0.8 to 0.9) and blue (above 0.9)

Catchments selected for this study cover a wide range of hydroclimatic conditions, as shown in Table 3.1.

Table 3.1 Distribution of key attributes for the 921-catchment ensemble

| Variable | 5 th percentile | 25 th percentile | Median | 75 th percentile | 95 th percentile |
|---|-------------------------------|--------------------------------|--------|--------------------------------|--------------------------------|
| Catchment surface (km ²) | 574 | 908 | 1 748 | 4 235 | 26 627 |
| Mean annual temperature (°C) | 2.47 | 6.85 | 10.01 | 13.10 | 19.18 |
| Mean annual rainfall (mm) | 371 | 626 | 902 | 1 150 | 1 541 |
| Mean annual snowfall (mm) | 1 | 36 | 69 | 111 | 279 |
| Mean annual potential evaporation (mm) | 456 | 628 | 771 | 907 | 1 220 |
| Aridity index (Pma / PEm) (%) | 68 | 107 | 126 | 140 | 212 |
| Mean annual observed flow (m ³ /s) | 2.71 | 9.45 | 17.73 | 40.01 | 189.33 |

3.3.2 Hydrological model

The GR4J model (*modèle du Génie Rural à 4 paramètres Journalier*; see Perrin et al., 2003, for model description) coupled with the CemaNeige snow module (Valéry, 2010) was chosen for streamflow simulations. The GR4J model is a daily lumped rainfall-runoff model with four free parameters. It is a parsimonious model, and is therefore less subject to parameter equifinality. It has been shown to provide good modelling performance over several climatic regions, making it a good choice for studies involving numerous watersheds (see, e.g., Andréassian et al., 2004; Oudin et al., 2008; Zhao et al., 2020). CemaNeige is a 2-parameter parsimonious snow model specifically developed to work with lumped hydrological models (Valéry et al., 2014a, 2014b). It has been coupled with GR4J in many studies (see, e.g., Coron et al., 2017; Coron et al., 2014; Tarek et al., 2019; Zhao et al., 2020). Both models require daily precipitation and potential evapotranspiration (PET) as inputs. The Oudin PET formula (Oudin et al., 2005) was chosen for this work, and is a temperature-based formula specifically developed for lumped hydrological models. To this end, it has been shown to outperform dozens of other formulations, including more complex radiation-based approaches.

3.3.3 Calibration schemes

Five main calibration approaches were chosen, each highlighting different hydroclimatic conditions: warm years, dry and humid years, high- and low-flow years, as shown in Table 2. A ‘cold years’ approach was not considered as it does not reflect future climate conditions anywhere around the globe (see Lehner et al., 2017; IPCC, 2013). Lowest and highest streamflow years were used as proxies for ‘dry and warm’ and ‘wet and cold’ years, which tend to respectively minimize and maximize streamflow. These proxies were used since upstream selection of appropriate years was made difficult by intra-annual variability.

The experimental layout is presented below using the ‘warm years’ scheme as an example. The procedure is kept the same for the other cases. In all five approaches, the first year of the 46-year long historical record was included solely for hydrological model warm-up, and was otherwise not used.

The 5 warmest years on record (consecutive or not) are first selected and set aside to be used as an independent validation set. The remaining 40 years are then ranked from warmest to coldest, to be used in the calibration process. This 40-year sample is then divided into sub-samples of 5 to 40 years, with an increment of 5 years. By selecting non-consecutive years, it is possible to have an independent validation set of 5 years, with conditions contrasting significantly with the mean climate, and therefore, more representative of the future climate. This approach has the disadvantage of having all calibration experiments run over the entire 45-year duration of the historical record (for streamflow continuity), while only using the relevant years to calculate the value of the objective function over the calibration (or independent validation) period. The layout of the experiment is presented in Table 3. For each catchment, eight calibrations of the hydrological model are therefore required for each of the five calibration strategies. Overall, a total of 36,840 calibrations ($921 \times 5 \times 8$) were performed in this study. As mentioned above, for each calibration, the model was run for the entire 46-year duration of the historical record to preserve streamflow continuity. The objective function was computed uniquely on each N-year subset. In this paper, ‘calibration strategy’ refers to the

letter code of Table 3.2, while ‘calibration years’ refers to the number of years selected for the calibration, as per Table 3.3.

Table 3.2 Calibration strategies

| | |
|-------------|--|
| [W] | Warm years (set aside the 5 warmest years for validation) |
| [H] | Wet/humid years (set aside the 5 wettest years for validation) |
| [D] | Dry years (set aside the 5 driest years for validation) |
| [LQ] | Low runoff years (set aside the 5 lowest runoff years): proxy for high T and low P |
| [HQ] | High runoff years (set aside the 5 highest runoff years): proxy for low T and high P |

Table 3.3 Calibration layout for the warm years experiment (W)

| Validation period | | Warmth rank |
|----------------------------|--|--------------------------------------|
| 5 years | | 1 st to 5 th |
| Calibration periods | | Warmth rank |
| 5 years | | 6 th to 10 th |
| 10 years | | 11 th to 15 th |
| 15 years | | 16 th to 20 th |
| 20 years | | 21 th to 25 th |
| 25 years | | 26 th to 30 th |
| 30 years | | 31 th to 35 th |
| 35 years | | 36 th to 40 th |
| 40 years (full cal. set) | | |

For each method and over all catchments, calibrations were performed with the Covariance-Matrix Adaptation Evolution Strategy (CMA-ES) (Hansen & Ostermeier, 1996, 2001) algorithm, using the Kling-Gupta Efficiency criterion (KGE) (Gupta et al., 2009) objective function. The CMA-ES algorithm was chosen based on the work of (Arsenault et al., 2014).

3.4 Results

3.4.1 All approaches

Figure 3.2 presents boxplots of the KGE validation results for the five different calibration strategies and 8 values of calibration years. Each boxplot is composed of the 921 KGE values corresponding to each catchment under study. The blue rectangles show the 25th and 75th quantiles, with the median as a red line. The upper and lower whiskers represent the largest and lowest KGE results. The red crosses are considered statistical outliers.

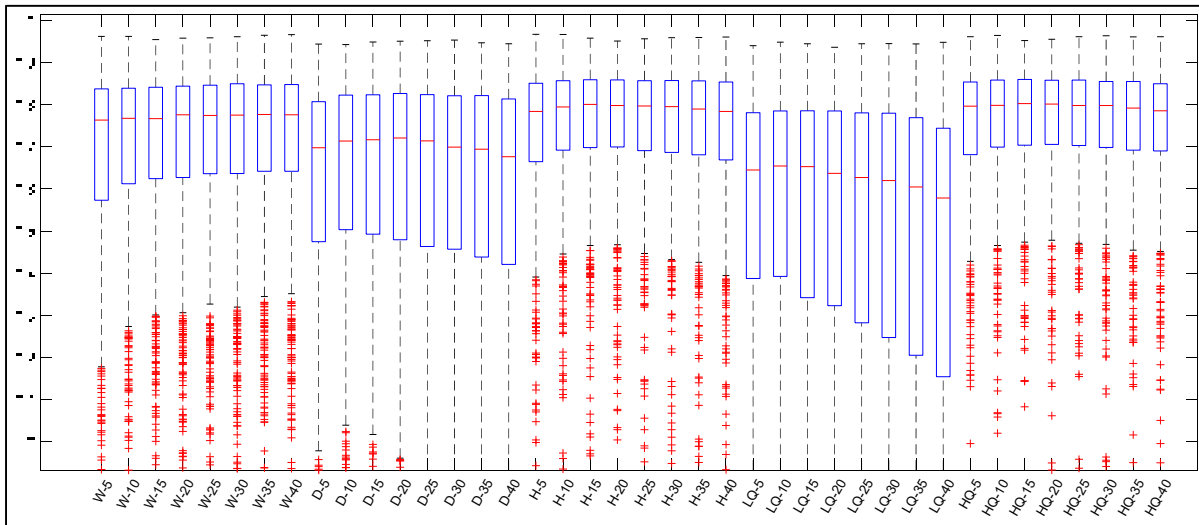


Figure 3.2 KGE validation results for each calibration strategy. [W] Warm years, [D] Dry years, [H] Wet/humid years, [LQ] Low-runoff years, [HQ] High-runoff years, with digits representing the Nth warmest (wettest, driest ...) years used for the calibration period

Figure 3.2 shows relatively small differences between all calibration year values. The largest differences are observed for the dry (D) and low-flow (LQ) calibration strategies, which represent somewhat similar climatic conditions. The overall KGE values are also lower for these two calibration strategies. Arid and semi-arid catchments often display lower performances and lower parameter stability than wet watersheds due to smaller runoff values (Li et al., 2015). The calibration year patterns are similar for both calibration strategies, with progressively larger variances (see Table 3.4) as wet years are added to the calibration pool. In

comparison, the W, H and HQ strategies are much less affected by longer calibration year periods.

In all cases, a 5-year calibration sample does not appear to be an optimal solution. Employing the whole calibration dataset could be the most efficient option when implementing the ‘W’ strategy, while using roughly half of the calibration time series (15 to 20 years) looks more optimal for all other strategies. However, for all five calibration strategies, the results do not vary significantly across all calibration years value.

To make better sense of the differences between the different subsets of N years, Table 3.4 presents results of the non-parametric tests for the equality of medians (Wilcoxon rank sum test; Hollander et al. (2013)) and equality of variance (Conover squared rank sum test; Conover (1981)). The lower diagonal presents the test significance for the equality of medians whereas the upper diagonal presents the results for the equality of variance tests. A value of 1 implies that the null hypothesis (equality) is rejected.

For the warm year (W) calibration strategy, the 5-year results are the only ones that differ from all other durations, for both the median and variance. The observed difference between the 7 other durations are not statistically significant. Results for the wet (H) and high flow (HQ) are largely similar to W, although there are a few more statistically significant differences, especially for the wet years’ variance. Conversely, there are many more statistically significant differences for the low flow (LQ) and, to a lesser extent, dry year (D) calibration strategies. This is especially true for the variance.

Table 3.4 Statistical significance of validation results for various calibration lengths:
 Conover's nonparametric test (top right); Wilcoxon rank sum test (bottom left)

| | | | | | | | | |
|--------------|-------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| | W-5 | W-10 | W-15 | W-20 | W-25 | W-30 | W-35 | W-40 |
| W-5 | | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| W-10 | 0 | | 0 | 0 | 0 | 0 | 0 | 0 |
| W-15 | 0 | 0 | | 0 | 0 | 0 | 0 | 0 |
| W-20 | 1 | 0 | 0 | | 0 | 0 | 0 | 0 |
| W-25 | 1 | 0 | 0 | 0 | | 0 | 0 | 0 |
| W-30 | 1 | 0 | 0 | 0 | 0 | | 0 | 0 |
| W-35 | 1 | 0 | 0 | 0 | 0 | 0 | | 0 |
| W-40 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | |
| | D-5 | D-10 | D-15 | D-20 | D-25 | D-30 | D-35 | D-40 |
| D-5 | | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| D-10 | 0 | | 0 | 0 | 1 | 1 | 1 | 1 |
| D-15 | 0 | 0 | | 0 | 1 | 1 | 1 | 1 |
| D-20 | 1 | 0 | 0 | | 0 | 0 | 1 | 1 |
| D-25 | 0 | 0 | 0 | 0 | | 0 | 0 | 1 |
| D-30 | 0 | 0 | 0 | 0 | 0 | | 0 | 0 |
| D-35 | 0 | 0 | 0 | 0 | 0 | 0 | | 0 |
| D-40 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | |
| | H-5 | H-10 | H-15 | H-20 | H-25 | H-30 | H-35 | H-40 |
| H-5 | | 1 | 1 | 1 | 1 | 1 | 0 | 0 |
| H-10 | 1 | | 0 | 0 | 0 | 0 | 1 | 1 |
| H-15 | 1 | 0 | | 0 | 0 | 0 | 1 | 1 |
| H-20 | 1 | 0 | 0 | | 0 | 0 | 1 | 1 |
| H-25 | 1 | 0 | 0 | 0 | | 0 | 0 | 1 |
| H-30 | 1 | 0 | 0 | 0 | 0 | | 0 | 1 |
| H-35 | 0 | 0 | 0 | 0 | 0 | 0 | | 0 |
| H-40 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | |
| | LQ-5 | LQ-10 | LQ-15 | LQ-20 | LQ-25 | LQ-30 | LQ-35 | LQ-40 |
| LQ-5 | | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| LQ-10 | 0 | | 1 | 1 | 1 | 1 | 1 | 1 |
| LQ-15 | 0 | 0 | | 1 | 1 | 1 | 1 | 1 |
| LQ-20 | 0 | 0 | 0 | | 1 | 1 | 1 | 1 |
| LQ-25 | 0 | 1 | 0 | 0 | | 1 | 1 | 1 |
| LQ-30 | 1 | 1 | 1 | 0 | 0 | | 1 | 1 |
| LQ-35 | 1 | 1 | 1 | 1 | 0 | 0 | | 1 |
| LQ-40 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | |
| | HQ-5 | HQ-10 | HQ-15 | HQ-20 | HQ-25 | HQ-30 | HQ-35 | HQ-40 |
| HQ-5 | | 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| HQ-10 | 0 | | 0 | 0 | 0 | 0 | 0 | 0 |
| HQ-15 | 1 | 0 | | 0 | 0 | 0 | 0 | 0 |
| HQ-20 | 1 | 0 | 0 | | 0 | 0 | 0 | 0 |
| HQ-25 | 0 | 0 | 0 | 0 | | 0 | 0 | 0 |
| HQ-30 | 0 | 0 | 0 | 0 | 0 | | 0 | 0 |
| HQ-35 | 0 | 0 | 0 | 0 | 0 | 0 | | 0 |
| HQ-40 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | |

3.4.2 Consistency across watersheds

Figure 3.2 gives a general portrait of all results aggregated within a single boxplot for each calibration strategy and calibration year. While this portrait shows small but consistent changes across calibration years, there could still be significant variability across catchments. To better explore this issue, Figure 3.3 shows the number of catchments (y-axis), out of 921, for which each calibration year (x-axis) provides the best validation KGE (blue bars). It also shows the number of catchments for which it provides the worst validation KGE results (red bars). This demonstrates the reliability of each number of calibration years. These results should be interpreted with care as they are not true indicators of performance.

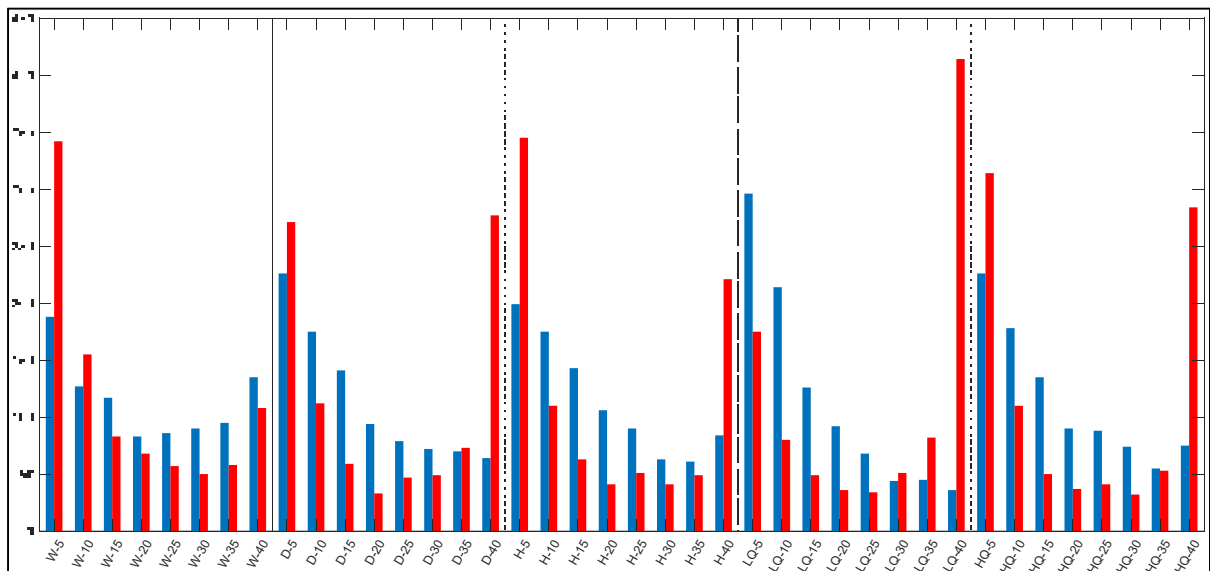


Figure 3.3 Number of occurrences of best (blue bars) and worst (red bars) calibration strategies/years out of 921 catchments

Figure 3.3 shows that for all 5 calibration strategies, each calibration year duration provides the best option for a certain number of watersheds. The 5-year duration is most likely to be the best one or the worst one, depending on the watershed considered. Using all 40 years does not appear to be a good approach, with the exception of the warm year (W) strategy, where it appears to be a reasonable choice. Focusing on regions where the blue bars are above the red ones, we see that longer calibration periods should be used for W, whereas intermediate

durations of 15 to 25 years appear to be the best compromise for the other four calibration strategies. Overall, the results clearly outline that there is considerable variability from one watershed to the next, and that finding a consistent best calibration approach applicable to all watersheds is unlikely.

Since one of the hypotheses of this study was that longer calibration periods should be preferred, a plot of the percentage of catchments for which each calibration year duration results in a better validation performance than using all 40 years as a benchmark for calibration is presented in Figure 3.4. For each calibration strategy, the 921 catchments are separated into sub-groups which identify the departure from the mean climate for the 5 years set aside for the validation period. For example, for the warm year calibration strategy (W), the watersheds were classified using the annual mean temperature anomaly of the 5 warmest years used for validation, compared to the remaining 41 years. The 50% threshold (horizontal dashed line) represents the target that has to be exceeded to conclude that a given number of validation years is better than using all of the 40 years of available data. For the dry (D) and wet (H) year calibration strategies, the catchments are classified by their mean annual precipitation anomalies, whereas the mean annual streamflow anomalies are used for the low-flow (LQ) and high-flow (HQ) calibration strategies.

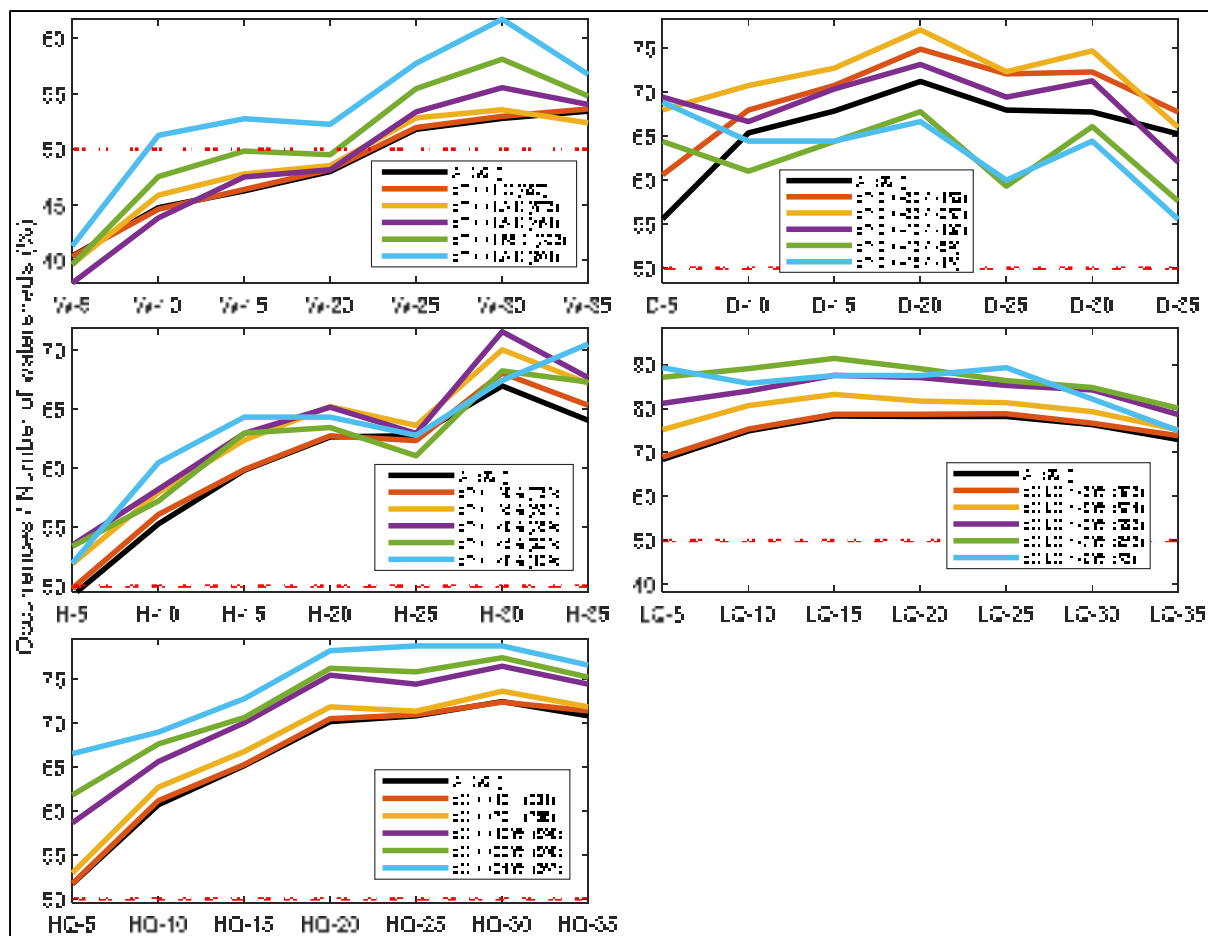


Figure 3.4 Occurrences (%) for which a validation KGE for specific calibration years exceeds KGE using the 40-year calibration strategy. (dT, dP D, dP, dQ LQ and dQ represent temperature, precipitation and runoff anomalies between the 5-year validation period and the remaining 41 years available for calibration)

Figure 3.4 clearly shows that using all 40 years for calibration is not a good approach for all calibration strategies, with the exception of warm years (W). However, even in this case, this approach becomes less and less useful as the temperature anomaly increases. There is a strong dependence on the strength of the anomaly and on the probability of a shorter calibration year duration being more successful than using all available years. The only exception to this is for the dry year (D) calibration strategy, where the trend is less clear.

3.4.3 Gains in KGE achieved during the validation period

The results presented above show that some calibration year durations are more likely to be efficient for calibration than keeping all available years. The results do not however show how much efficiency is gained over the validation years. Table 3.5 provides a few statistics on the KGE validation efficiency gains (losses) by using the optimum duration of calibration years, and compares them to using the entire 40-year period. The optimum number of calibration years is established using the highest percentage value from Figure 3.4 for each calibration strategy. Overall, Table 3.5 shows weak median/mean performance benefits even when using the optimum number of calibration years. The KGE gains can however be quite large for some catchments, again outlining the difficulty of finding a ‘one-size fits all’ calibration strategy.

Table 3.5 Statistic analysis of KGE gains/losses in validation compared to using all 40 years

| METHODS | W-30 | D-20 | H-30 | LQ-15 | HQ-30 |
|-----------------------------|-------------|-------------|-------------|--------------|--------------|
| Number of catchments | 897 | 864 | 910 | 808 | 916 |
| Mean gain | 0.0030 | 0.0396 | 0.0136 | 0.1115 | 0.0094 |
| Median gain | 0.0075 | 0.0229 | 0.0048 | 0.0492 | 0.0073 |
| Standard deviation | 0.0517 | 0.0878 | 0.0456 | 0.2096 | 0.0352 |
| Maximum gain | 0.4791 | 0.5047 | 0.7129 | 1.5548 | 0.4268 |
| Worst loss | -0.6045 | -0.4746 | -0.3753 | -0.3613 | -0.2931 |

Finally, an analysis of the spatial patterns across climatic regions was done for all of the results presented above. No conclusive relationship was identified between catchment location and calibration years for all 5 calibration strategies. Those results are therefore not shown in this paper. In addition, no relationship to catchment size was identified.

3.5 Discussion

The results presented in this work point to small median differences between the number of calibration years for all 5 tested calibration strategies, as shown in Figure 3.2, which displays

weak optimum values. These results nonetheless suggest that the use of all available data for calibration, in this case 40 years, likely does not constitute the most efficient calibration strategy, but that choosing this approach is not likely to result in a large decrease in validation performance. However, our results, and in particular those presented in Figure 3.3 and Table 3.5, show important cross-catchment variability. A subset of 18 catchments was therefore selected to explore this variability in more detail for the warm (W) year calibration strategy. These catchments all had a minimum temperature anomaly of +1.1 °C (median of +2.32°C; max. of +2.61°C) for the 5-year validation period. For each calibration year value, the use of the N warmest years was compared to a set of 30 random selections of N years. For example, for N=15, 31 calibrations were performed, with one using the warmest 15 years, and all the other 30 using 15 randomly selected years. The choice of 30 random selections was large enough to provide repeatable results while keeping the calibration computing budget manageable.

Results for three catchments are presented in Figure 3.5. No results are presented for N=40, since there is only a single combination of 40 years ($C_{40}^{40} = 1$). The geographical location of the three catchments is shown at the bottom right. Figure 3.5 shows that the success of the warmest year calibration strategy differs largely, depending on the catchment considered. For watershed 12 (upper left), using the N warmest years for calibration is a very good strategy as it is systematically near the upper whisker of the boxplot of random calibrations, indicating that it beats most other possible combinations of N years (as approximated by our 30 random selections of N years). For watershed 469 (upper right), the warm year calibration strategy is successful up to N=15, and becomes worse than the random approach for larger values of calibration years, as shown by the red circles being below the median of the boxplots for the random years. For watershed 899, which has a large temperature anomaly (lower left), using the warm year strategy is detrimental as it systematically ranks well below the median of the boxplots of the random year selection (with the exception of N=5). These results illustrate the clear limitations of using a single criterion (mean annual temperature in this case) to select calibration years. Using the N warmest years may appear to be a reasonable strategy, but the

complex interactions between temperature and precipitation, as well as the interannual variability of both variables, may ultimately result in a bad selection of calibration years.

Results presented in this study on a large number of catchments allows to draw some general conclusions on the likelihood of a calibration strategy being successful, but offer no guarantee for any individual catchment. It is therefore only possible to offer some general suggestions as to the best possible calibration approach for hydrological climate change impact studies.

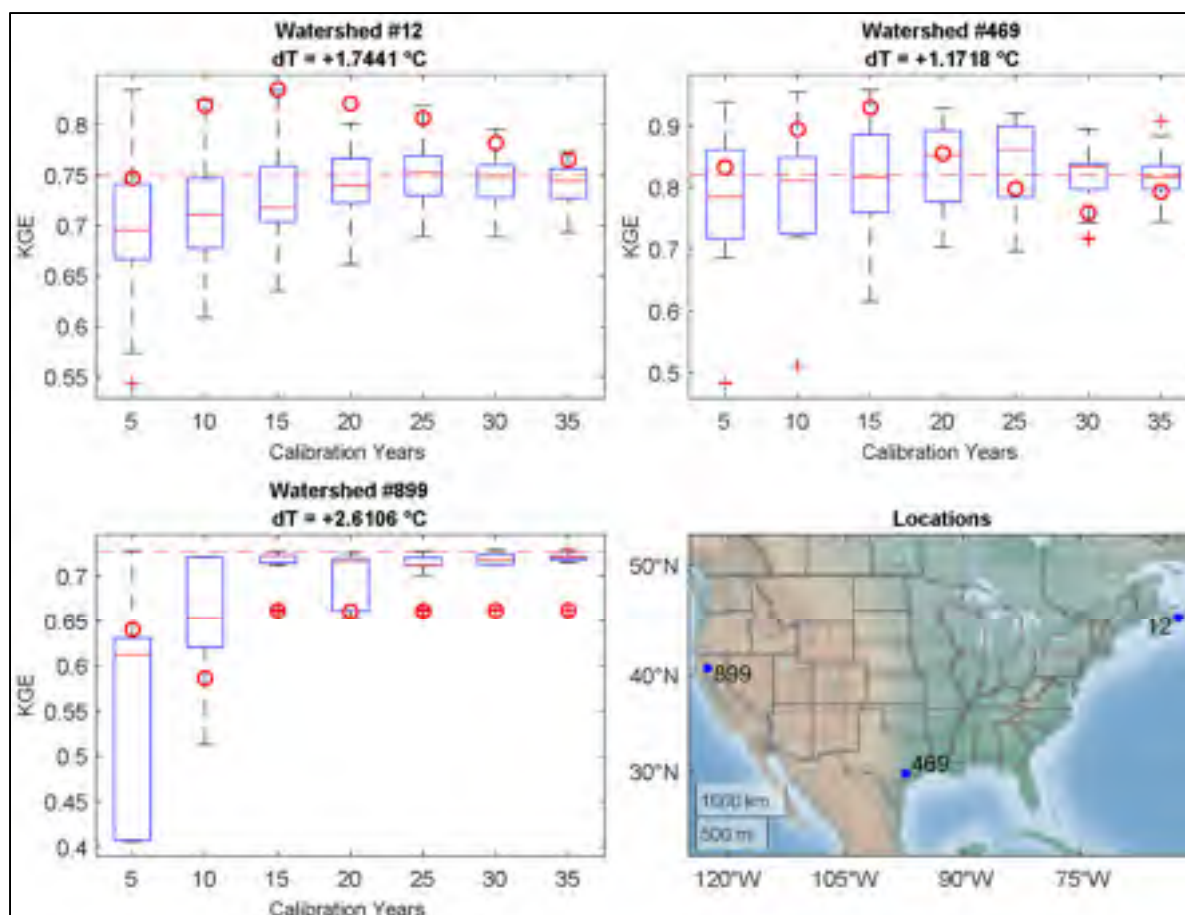


Figure 3.5 Distribution of validation KGE (blue boxes) values stemming from random calibration years along with the performances of the ‘W’ (red circles) and 40-year (red dotted line) strategies (Temperature anomalies (dT) are given above each whisker plot and catchment locations are displayed as red markers on the map)

To strictly adapt hydrological model calibration to a warmer climate, results indicate that using a large number of calibration years leads to more robust performances. Using 15 years or less for the warm year (W) calibration strategy results in statistically significant weaker median performance (Table 3.4). Ultimately, results from Figure 3.4 indicate that dropping the coldest years on record (5 years out of 40, in this case) is a better option than keeping a shorter but warmer subset of years. The warm year calibration strategy is the one for which keeping a large number of years is most efficient. In particular, the large variance associated with keeping a limited number of years (5 years, in particular) makes this a potentially precarious solution.

For the other 4 calibration strategies, results indicate that using 15 to 25 years, roughly half of the 40-year sorted record, provides the best strategy. These 4 calibration strategies are more strongly related to precipitation than is the warm year strategy, which is solely temperature-based. The relative impact of interannual (internal) variability of precipitation and temperature may explain our results. The internal variability of total annual precipitation is quite significant, with very large differences seen between dry and wet years. In comparison, the internal variability of the mean annual temperature is much lower. Therefore, when calibrating for a warmer period, a sufficiently long calibration period may build robustness to precipitation variability in the validation set. When targeting dry or wet years, the large internal variability of precipitation makes dropping more dissimilar years more reasonable for precipitation, while a shorter number of years may not be impacted as much by smaller potential temperature differences in the validation set. If correct, this could have implications for climate change impact studies. As mentioned in the introduction, climate projections show large changes in mean temperature and comparatively much smaller changes for precipitation by the end of the century, as evidenced by the much earlier time of emergence for temperature as compared to precipitation (e.g., Hawkins & Sutton, 2012; Lehner et al., 2017; Martel et al., 2018). The key problem is therefore the temperature sensitivity of the calibration process since end-of century mean annual temperatures will be significantly larger than the ones in the current historical record. The temperature sensitivity of the hydrological model, and in particular, of temperature-dependent processes such as evapotranspiration and snowmelt, must therefore be carefully evaluated. Future dry/wet conditions will be a lot more comparable to what most catchments are currently experiencing, and climate change should not be expected to significantly exacerbate the dry/wet calibration issues we already face in the current historical record.

To put temperature change into context, an average temperature anomaly of 1.54°C was obtained for the five warmest year validation set of the 921 catchments used in this study. Arsenault et al. (2020a) showed (for 10 climate models) that a 1.5 degrees warming for North American catchments would be expected as early as 2032 under RCP4.5, and 2026 under RCP8.5. By the end of the century, an average warming of roughly 6°C is expected (Pachauri

et al. (2014), Fig. SPM.7) and up to 10°C for high-latitude watersheds under RCP8.5. Clearly, the temperature anomalies generated by using the 5 warmest years in the historical record only provide a limited window into the end-of-century future climate. In comparison, precipitation anomalies analyzed in this paper (average of +39 % and -32%) are much larger than the mean projected change for 2100 (Pachauri et al. (2014), Fig. SPM.7).

In all cases, using less than 10 years (out of 40) generates less robust and more variable results. This is in line with the work of Razavi et Tolson (2013), who concluded that “model calibration solely to a short data period may lead to a range of performances from poor to very well depending on the representativeness of the short data period which is typically not known a priori”.

Ultimately, it must be kept in mind that median differences between the different calibration year values remain small for all five calibration strategies, and that the uncertainty linked to this choice may be small as compared to other uncertainties related to the traditional top-down approach for modelling streamflows in a changed climate (e.g., GCM, GHGES, downscaling, bias correction), although this remains to be investigated.

3.6 Conclusion

Hydrologists rely mostly on historical hydrometeorological data to calibrate hydrological models, with relatively limited knowledge on the transferability of such parameter sets to different climate conditions. This is a particularly important concern for climate change impact studies, with cases where the mean future climate may be well outside the scope of the internal variability present in the historical calibration period. This study proposed a methodological framework to find viable calibration strategies to assess the relevance of selecting calibration periods with climate characteristics similar to those of an independent validation period chosen to maximize relevant climate anomalies. The impact of the number of calibration years was also evaluated as an aside. This study was conducted over 921 North American catchments, a number that allowed general conclusions to be drawn, despite the large variability seen across

the catchments involved, and despite the absence of a ‘one size fits all’ solution applicable to all the catchments. Results showed that the robustness of calibration to increasing temperature was maximized by keeping a large subset of years and that dropping a small subset of cold years was preferable to keeping a small subset of warmer years. Calibration strategies targeting precipitation-dependent targets (wet/dry and high-/low-flow years) were better employed using roughly half the available calibration years whose climate conditions were closest to the validation target. In all cases, using a small number of calibration years (10 years or less) was least likely to be successful.

CHAPITRE 4

ÉTUDE DE STABILITÉ DES PARAMÈTRES

4.1 Étude préliminaire

Une analyse exploratoire des jeux de paramètres obtenus suite aux calibrations du modèle sur les 921 bassins versants a permis d'identifier une certaine tendance pour quelques paramètres. Ce constat provient des valeurs moyennes et médianes de chacun des paramètres, qui laissent à croire que le climat de calibration influence la valeur des paramètres, dans certains cas. L'ANNEXE I représente la totalité des valeurs prises par les six paramètres (traits gris) du modèle pour les 921 bassins versants, accompagnée des valeurs médianes (rouge) et moyennes (vert) de ce paramètre. Les années sélectionnées pour la calibration sont les mêmes que celles de l'article, classées selon la température moyenne annuelle, la précipitation moyenne annuelle et le débit moyen annuel. Ces périodes permettent de cibler des anomalies parmi les années de calibration et de mesurer les variations que subissent les paramètres à travers celles-ci.

La poursuite de cette analyse s'effectue dans le but d'en apprendre davantage sur la relation qui lie la valeur de certains paramètres aux conditions hydrométéorologiques expérimentées en calibration et de vérifier si des conclusions plus spécifiques peuvent être tirées. Ces travaux d'approfondissement vont de pair avec l'étude de Merz et al. (2011), qui stipule que les paramètres optimaux ont un lien de dépendance avec le climat expérimenté lors de la calibration. Le chapitre qui suit comprend : un résumé de la structure du modèle hydrologique, une description de la méthodologie employée, ainsi que des résultats et une discussion.

4.2 Survol des paramètres de GR4J-CemaNeige

Afin de mieux comprendre le rôle des paramètres à l'intérieur de la structure du modèle GR4J-CemaNeige, il est essentiel de présenter sa structure. Le modèle hydrologique du *Génie Rural à 4 paramètres Journalier* (GR4J, voir Perrin et al. (2003)) possède, tel qu'indiqué par son nom, quatre paramètres. Le module de neige CemaNeige (Valéry, 2010) comporte quant à lui

deux paramètres. L'ANNEXE II illustre la structure du modèle pluie-débit GR4J-CemaNeige. Voici la liste de chaque paramètre et de son rôle au sein du processus de modélisation du cycle hydrologique :

Tableau 4.1 Résumé du rôle des paramètres du modèle GR4J couplé au module CemaNeige

| Identifiant | Nom complet | Processus hydrologique |
|----------------|--|---|
| X1 (GR4J) | Capacité du réservoir de production (mm) | Ruissellement/Infiltration |
| X2 (GR4J) | Coefficient d'échanges souterrains (mm) | Écoulements souterrains |
| X3 (GR4J) | Capacité à un jour du réservoir de routage (mm) | Volume disponible pour le ruissellement |
| X4 (GR4J) | Temps de base de l'HU1 (jours) | Durée du ruissellement |
| X5 (CemaNeige) | Coefficient de pondération de l'état thermique du manteau (sans unité) | Température du manteau et accumulation de neige |
| X6 (CemaNeige) | Facteur degré-jour (mm/°C) | Fonte de neige |

La structure des modèles est souvent une source d'incertitude lors des simulations de débit. Dans le cas présent, la structure du modèle comprend relativement peu de paramètres, ce qui limite l'incertitude liée à l'équifinalité (Poulin et al., 2011). La combinaison du modèle GR4J et du module neige CemaNeige est fréquente dans le domaine de la modélisation hydrologique (Riboust et al., 2019; Tarek et al., 2019; Zhao et al., 2019).

4.3 Méthodologie

La section précédente permet de bien comprendre l'impact que peuvent avoir chacun des paramètres sur la simulation du débit. Pour la suite, il est essentiel d'analyser les six paramètres de façon individuelle puisque chacun aura un impact et un comportement unique à l'intérieur du modèle. Initialement, une vérification globale quant à la présence de tendance fut réalisée, en supposant que les données ne suivent pas une loi statistique en particulier. Le test non-paramétrique de tendance de Mann-Kendall (Burkey, 2006; Yue et al., 2002) fut donc appliqué aux six paramètres, pour l'ensemble des bassins étudiés. Pour chaque paramètre, huit valeurs

découlent du processus de calibration, donc le test est réalisé sur ces huit résultats. Il permet d'établir si une tendance statistiquement significative (seuils de confiance de $\alpha = 0,05$ et $\alpha = 0,01$) existe dans l'évolution des valeurs d'un paramètre. Lorsque c'est le cas, le test produit le chiffre un (1) pour signifier que l'hypothèse nulle (absence de tendance significative sous α) est rejetée.

Bien que des tendances puissent apparaître au sein d'un paramètre, il est primordial de vérifier si ce dernier subit une augmentation ou une diminution au fil des différentes périodes de calibration afin de tirer des conclusions relatives à l'hydrologie. Pour se faire, la pente de Sen découlant du test de Mann-Kendall est prise en compte afin de vérifier si une direction ascendante ou descendante se présente, pour les bassins où une tendance significative existe ainsi que tous les autres. Puisque les huit valeurs de paramètres peuvent varier énormément pour une même stratégie, cette analyse ne garantit pas que les valeurs présentent une augmentation ou une diminution évidente, mais elle permet tout de même de démontrer la direction de la tendance (significative ou non) pour une majorité de bassins versants.

Dans le but de faire l'étude de la stabilité des paramètres sur différentes périodes, l'échantillon global de 921 bassins fut divisé en trois sous-groupes de 307 bassins. De cette façon, il est possible de vérifier si la performance élevée du modèle est synonyme d'une plus grande stabilité paramétrique, tel que démontré par (Li et al., 2015). Le premier sous-groupe contient les 307 bassins ayant obtenu les valeurs de *Kling-Gupta Efficiency* (KGE) les plus élevées en calibration sur les 45 années disponibles, le second sous-groupe contient ceux aux KGE intermédiaires, puis le troisième sous-groupe contient ceux aux KGE les plus faibles. Cette séparation permet d'approfondir la stabilité des paramètres pour des résultats de calibration situés dans différentes fourchettes de KGE. Pour chaque sous-groupe (3) et chaque stratégie de calibration (5), les écarts entre la valeur minimale que prend un paramètre et sa valeur maximale sont calculés, en considérant un seul paramètre à la fois. Par l'entremise de ces écarts, il est possible d'identifier si un sous-groupe se démarque en terme de stabilité des paramètres. Puisque les paramètres X5 et X6 ont pour objectif de tenir compte des processus liés respectivement à la fonte et au volume de la neige présente sur un bassin versant, les

calibrations réalisées sur les bassins ne présentant aucune précipitation neigeuse utilisent ces deux paramètres de façon aléatoire, ce qui ne fait aucunement intervenir la base physique des processus hydrologiques. Pour cette raison, cette portion de l'analyse ne comprend que les paramètres X1 à X4.

Afin de tenir compte des bassins influencés par la présence de neige, une sélection de bassins fut effectuée en ayant comme seuil de sélection une précipitation de neige annuelle moyenne de 50 millimètres équivalents en eau de la neige (EEN). Donc, les paramètres plus fortement influencés par un couvert de neige peuvent être évalués en considérant que, spécifiquement pour ces bassins, les paramètres 5 et 6 apportent un aspect logique au processus de modélisation.

4.4 Résultats

4.4.1 Test de Mann-Kendall

Le Tableau 4.2 présente le nombre de bassins, pour chaque paramètre et stratégie de calibration, où une tendance statistiquement significative a été observée grâce au test de Mann-Kendall. Les colonnes alignées vers la gauche démontrent les nombres de bassins pour un niveau de confiance alpha de 0,05, alors que celles placées vers la droite présentent les résultats pour un alpha plus restrictif de 0,01. La portion inférieure du tableau énumère les résultats équivalents sous forme de pourcentage de bassins versants par rapport à la taille de la population étudiée (921 bassins).

Tableau 4.2 Nombre de bassins présentant une tendance significative

| $\alpha =$ | W | | D | | H | | LQ | | HQ | |
|------------|------------|------|------------|------|------------|------|------------|------|------------|------|
| | 0,05 | 0,01 | 0,05 | 0,01 | 0,05 | 0,01 | 0,05 | 0,01 | 0,05 | 0,01 |
| X1 | 195 | 103 | 227 | 103 | 262 | 161 | 243 | 129 | 282 | 190 |
| X2 | 286 | 165 | 285 | 170 | 368 | 230 | 370 | 244 | 371 | 261 |
| X3 | 188 | 92 | 276 | 158 | 326 | 225 | 244 | 132 | 383 | 260 |
| X4 | 216 | 110 | 229 | 120 | 267 | 163 | 218 | 129 | 257 | 156 |
| X5 | 135 | 52 | 118 | 41 | 109 | 46 | 118 | 38 | 96 | 34 |
| X6 | 61 | 12 | 74 | 21 | 53 | 18 | 47 | 15 | 51 | 11 |
| | W | | D | | H | | LQ | | HQ | |
| X1 | 21% | 11% | 25% | 11% | 28% | 17% | 26% | 14% | 31% | 21% |
| X2 | 31% | 18% | 31% | 18% | 40% | 25% | 40% | 26% | 40% | 28% |
| X3 | 20% | 10% | 30% | 17% | 35% | 24% | 26% | 14% | 42% | 28% |
| X4 | 23% | 12% | 25% | 13% | 29% | 18% | 24% | 14% | 28% | 17% |
| X5 | 15% | 6% | 13% | 4% | 12% | 5% | 13% | 4% | 10% | 4% |
| X6 | 7% | 1% | 8% | 2% | 6% | 2% | 5% | 2% | 6% | 1% |

On constate qu'un grand nombre de bassins présentent, pour toutes les stratégies, une tendance auprès des paramètres X2 et X3 (en rouge). Un portrait opposé se dessine pour les paramètres X5 et X6, qui visent normalement à représenter les processus liés au manteau nival.

4.4.2 Croissance et décroissance des paramètres

Toujours dans l'optique de percevoir certaines tendances dans l'évolution des paramètres, le Tableau 4.3 et le Tableau 4.4 permettent de cibler quels paramètres subissent une augmentation ou une diminution au fil des différentes durées de calibration, respectivement pour tous les bassins et uniquement pour ceux où une tendance significative est présente. Pour évaluer cette propriété, on considère le signe de la pente de Sen calculée pour chaque bassin versant. Cette pente est nulle pour les paramètres qui demeurent inchangés à travers les différentes périodes de calibration, donc ces occurrences (très peu nombreuses) ne sont pas comptabilisées dans le Tableau 4.3. Bien que les résultats ne pointent pas vers des conclusions robustes, les périodes caractérisées par des climats contrastants ont démontré une opposition (majoritairement) en ce qui concerne les signes des pentes, mise en évidence par les nombres en rouge. Ces oppositions ne concernent que les paramètres X1 à X3.

Tableau 4.3 Nombre de bassins présentant une pente de Sen (significative ou non) ascendante ou descendante

| | W | | D | | H | | LQ | | HQ | |
|-----------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| | As. | Des. | As. | Des. | As. | Des. | As. | Des. | As. | Des. |
| X1 | 370 | 459 | 367 | 470 | 498 | 336 | 231 | 571 | 570 | 245 |
| X2 | 452 | 427 | 340 | 541 | 557 | 316 | 660 | 193 | 216 | 655 |
| X3 | 489 | 379 | 645 | 233 | 242 | 634 | 614 | 249 | 219 | 661 |
| X4 | 455 | 420 | 398 | 485 | 505 | 382 | 413 | 464 | 509 | 373 |
| X5 | 403 | 493 | 418 | 481 | 381 | 521 | 368 | 509 | 404 | 488 |
| X6 | 396 | 525 | 415 | 506 | 469 | 452 | 403 | 504 | 423 | 498 |

Tableau 4.4 Nombre de bassins présentant une pente de Sen statistiquement significative ($\alpha = 0,05$) ascendante ou descendante

| | W | | D | | H | | LQ | | HQ | |
|-----------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| | As. | Des. | As. | Des. | As. | Des. | As. | Des. | As. | Des. |
| X1 | 85 | 110 | 93 | 134 | 177 | 85 | 42 | 201 | 228 | 54 |
| X2 | 163 | 123 | 84 | 201 | 256 | 112 | 341 | 29 | 44 | 327 |
| X3 | 110 | 78 | 252 | 24 | 45 | 281 | 209 | 35 | 37 | 346 |
| X4 | 123 | 93 | 84 | 145 | 181 | 86 | 106 | 112 | 177 | 79 |
| X5 | 57 | 78 | 48 | 69 | 40 | 69 | 36 | 82 | 47 | 49 |
| X6 | 16 | 45 | 25 | 48 | 26 | 27 | 20 | 27 | 18 | 33 |

L'évaluation des bassins aux tendances non significatives dispense une seconde série de résultats pertinents, notamment en raison de l'efficacité modérée du test de Mann-Kendall sur un ensemble de seulement huit valeurs. La somme des deux nombres obtenus pour chaque paramètre et chaque stratégie correspond au nombre énoncé dans la première colonne du Tableau 4.2 ($\alpha = 0,05$).

4.4.3 Stabilité paramétrique selon la performance en calibration

Les figures suivantes présentent l'état de la stabilité des paramètres X1 à X4 par l'entremise des deltas calculés entre la valeur supérieure que prend un paramètre et sa valeur inférieure. Ces figures démontrent l'ensemble des valeurs de deltas possibles pour chacun des sous-groupes. Les KGE médians des trois sous-groupes, composés de 307 membres chacun, sont

respectivement de 0,9011, 0,8521 et 0,7490. Par mesure de simplification, les sous-groupes seront notés S1, S2 et S3.

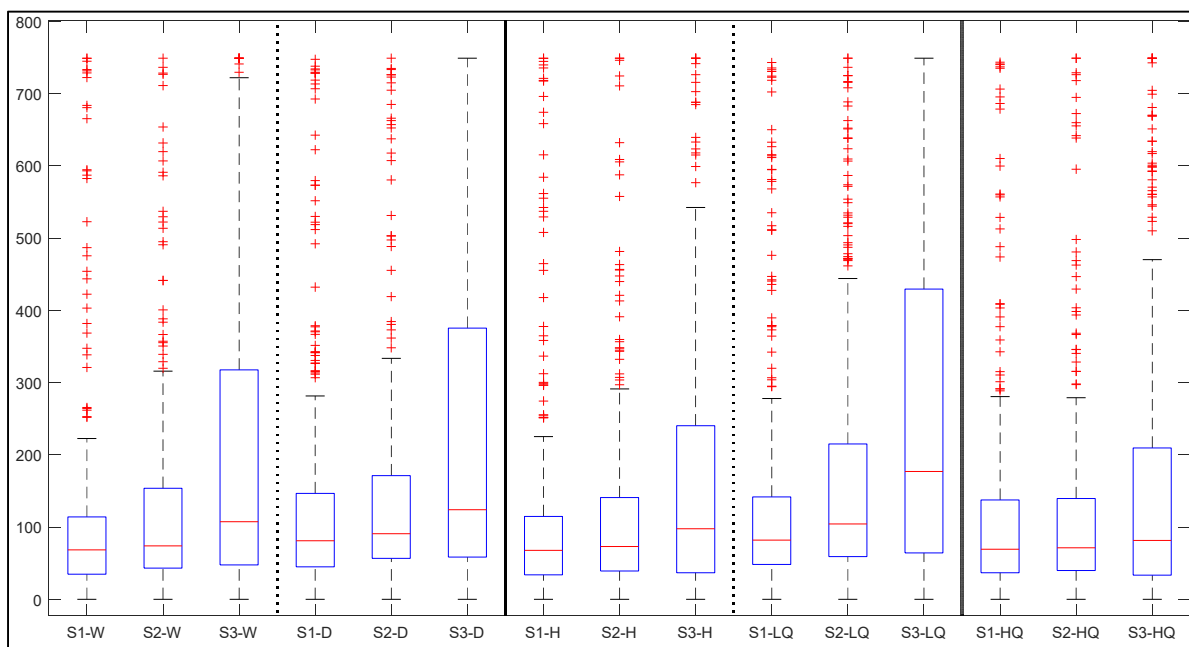


Figure 4.1 Stabilité du paramètre X1 pour trois sous-groupes de bassins versants

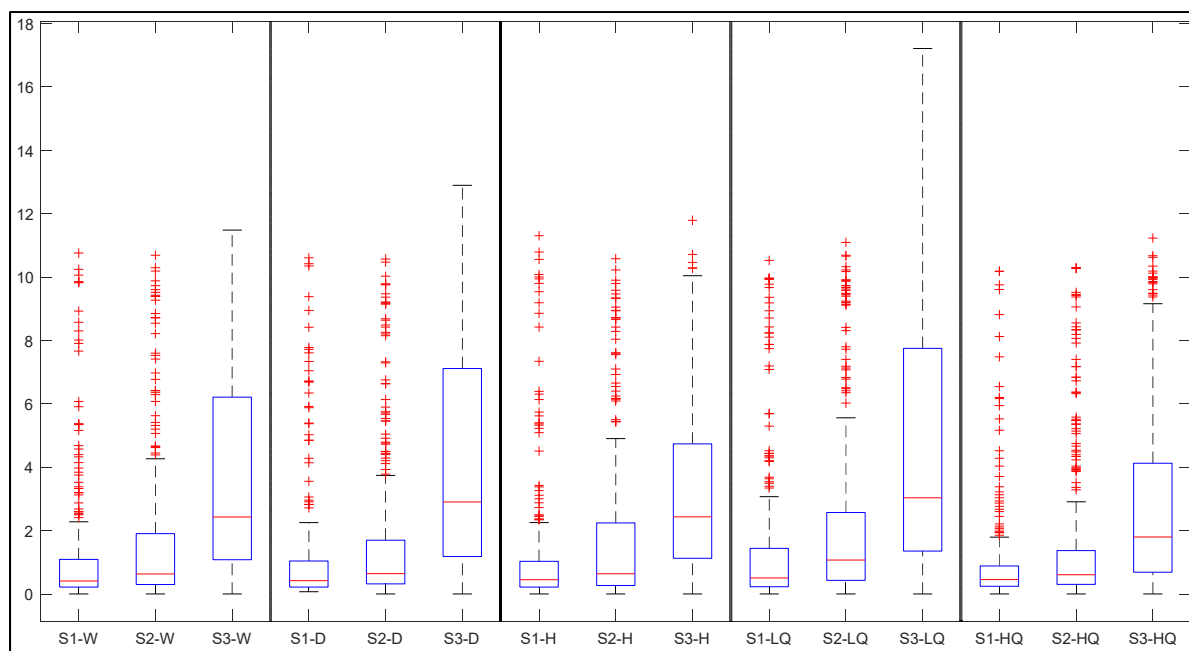


Figure 4.2 Stabilité du paramètre X2 pour trois sous-groupes de bassins versants

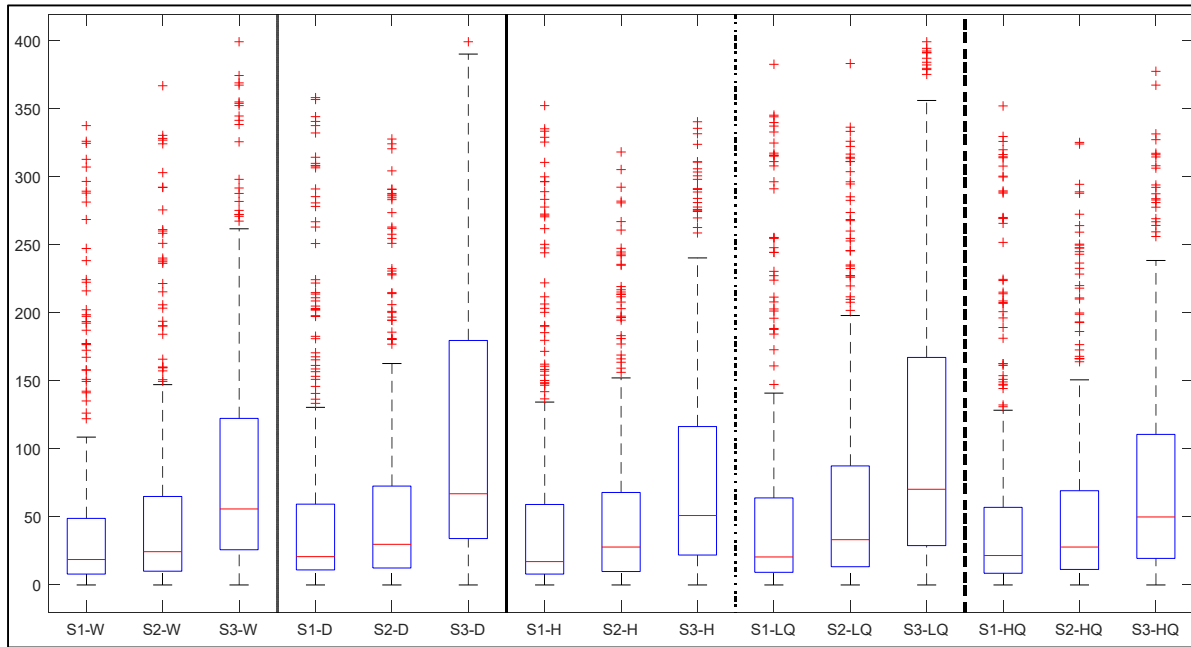


Figure 4.3 Stabilité du paramètre X3 pour trois sous-groupes de bassins versants

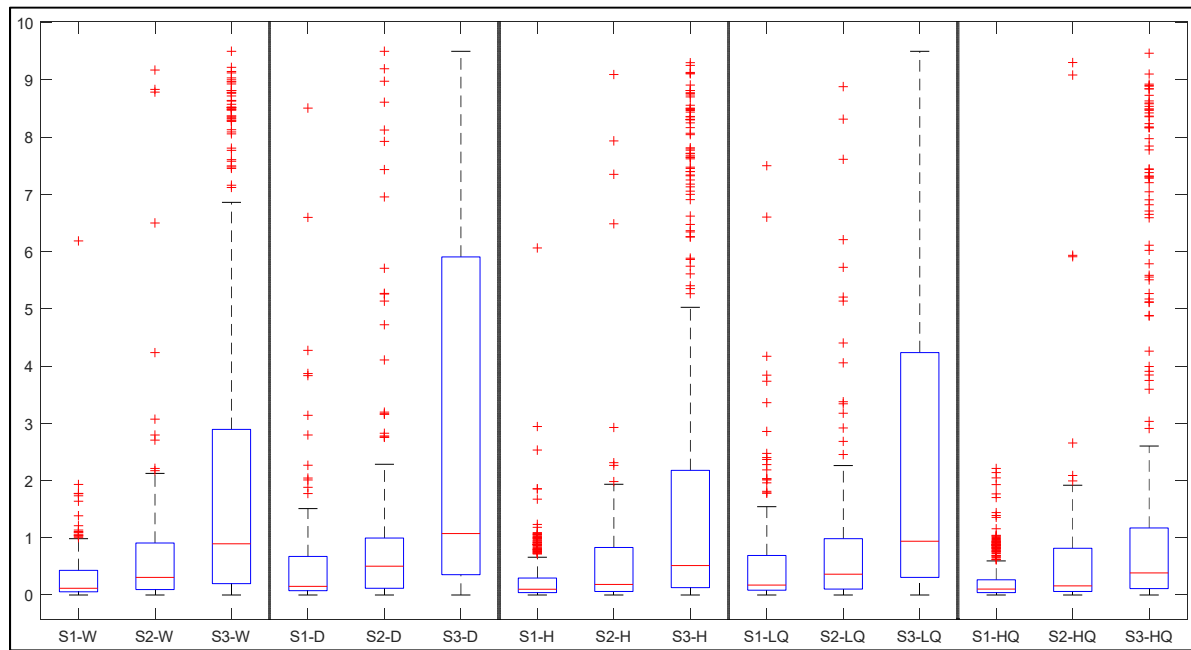


Figure 4.4 Stabilité du paramètre X4 pour trois sous-groupes de bassins versants

On remarque une large fourchette des valeurs des paramètres du groupe S3, quelle que soit la stratégie de calibration utilisée. Plus du quart de ces valeurs se trouvent bien au-delà des écarts maximums des deux autres sous-groupes. Les groupes S1 et S2 se démarquent par une stabilité plus importante, malgré que la médiane des paramètres du S1 demeure la plus basse pour l'ensemble des stratégies ainsi que pour tous les paramètres.

4.4.4 Bassins sous influence nivale

Axées sur l'influence que la précipitation neigeuse peut exercer sur la stabilité des paramètres optimaux de CemaNeige, la Figure 4.5 et la Figure 4.6 illustrent un certain lien entre la chute de neige annuelle moyenne, pour les 602 bassins où celle-ci excède 50 mm EEN, et la tendance évolutive des paramètres X5 et X6. L'axe vertical correspond à la pente de Sen mesurée pour chaque bassin, alors que l'axe horizontal montre les différentes valeurs de chute de neige annuelle moyenne. Les bassins représentés par un point orange sont ceux qui comportent une tendance significative selon un niveau de confiance alpha de 0,05, alors que les points bleus servent à montrer les résultats de tous les autres bassins influencés par la neige.

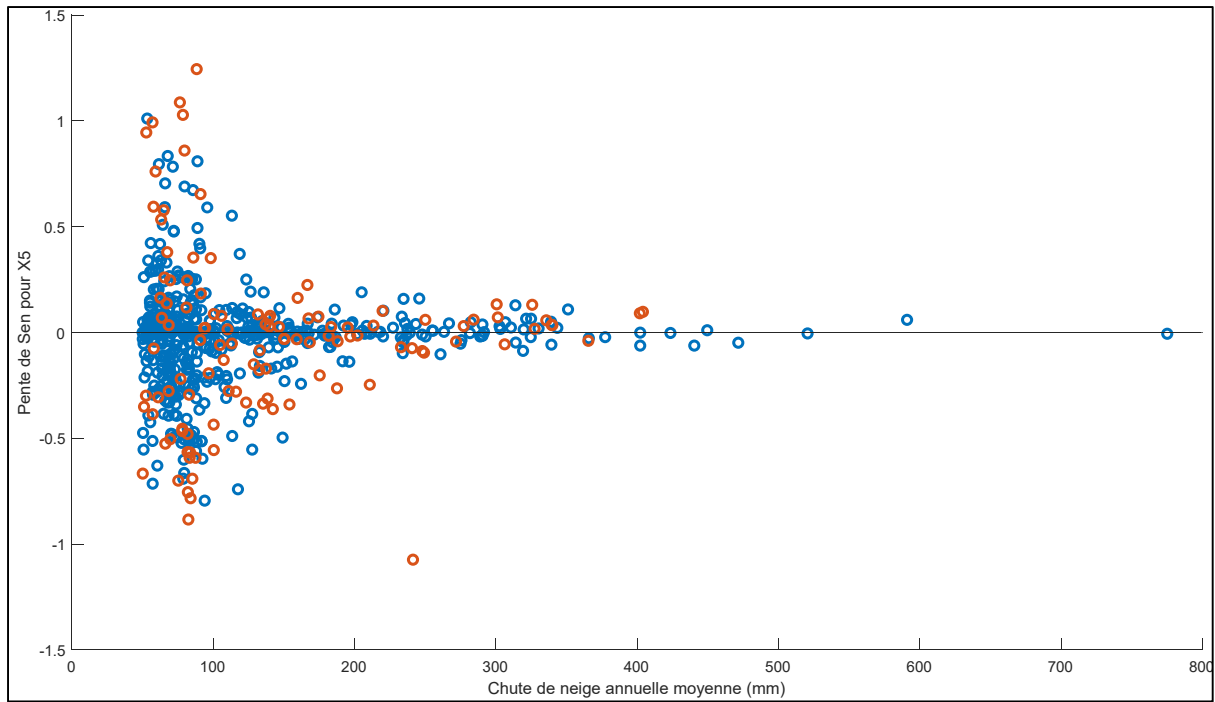


Figure 4.5 Évolution du paramètre X5 par rapport à la chute de neige annuelle moyenne (stratégie des années chaudes) pour les bassins où la pente de Sen est significative (orange) et non significative (bleu)

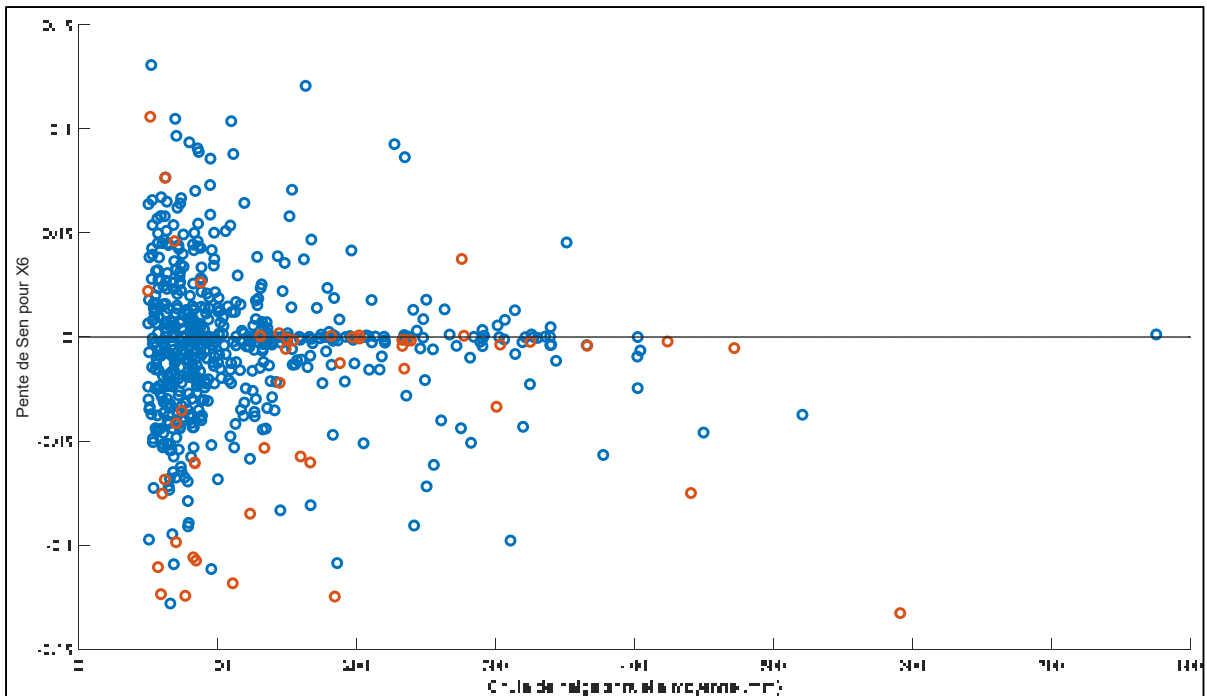


Figure 4.6 Évolution du paramètre X6 par rapport à la chute de neige annuelle moyenne (stratégie des années chaudes) pour les bassins où la pente de Sen est significative (orange) et non significative (bleu)

Au total, cet ensemble contient 602 individus. On dénote qu'un patron se dessine pour les deux paramètres analysés, bien que la majorité des bassins choisis reçoivent moins de 200 mm EEN de neige en moyenne annuelle. Globalement, les pentes de Sen ascendantes et descendantes semblent également réparties pour cet échantillon.

La stratégie des années chaudes comporte, tel que mentionné dans l'article, un classement décroissant d'années de calibration basé sur la température moyenne annuelle, incrémentées de 5 ans. Cette évolution progressive des périodes de calibration rend pertinente l'analyse des paramètres 5 et 6 dans de telles conditions. L'évolution du manteau nival dépend directement de la température au sein d'un régime hydrologique, bien que les précipitations liquides peuvent également avoir une influence importante en ce qui concerne l'énergie thermique présente.

Tenant compte des résultats obtenus précédemment, une investigation des tendances (significative ou non) à l'échelle régionale a démontré qu'aucune conclusion ne peut être établie à ce sujet actuellement. En effet, parmi le nombre relativement faible de bassins présentant une tendance, aucun patron régional particulier ne se dessine lorsque ceux-ci sont placés sur une carte. Cette observation concerne toutes les stratégies de calibration et tous les paramètres.

4.5 Discussion

Les résultats énoncés précédemment permettent de tracer un portrait sur quatre sujets distincts : la présence de tendances statistiquement significatives au sein des valeurs de quelques paramètres, la correspondance entre la direction de ces tendances et le type de conditions expérimentées en calibration, la stabilité des paramètres X1 à X4 et, finalement, le lien entre la chute de neige annuelle moyenne et l'évolution des paramètres X5 et X6 pour la stratégie des années chaudes.

Tout d'abord, on peut voir que les paramètres X2 et X3 se démarquent particulièrement quant au nombre de bassins versants analysés qui présentent une tendance significative, allant jusqu'à 40% des 921 bassins pour certaines stratégies. À l'intérieur du modèle, le paramètre X2 fait office de facteur de pondération dans les échanges d'eaux souterraines, pouvant être à la fois positif ou négatif selon la direction du transfert. Les volumes échangés sont calculés à partir de la formule du facteur F (Équation 4.1). Tel que démontré à l'ANNEXE II, les eaux souterraines (F) s'additionnent au volume du réservoir de routage ainsi qu'au débit provenant du second hydrogramme unitaire ($HU2$, plus allongé). Une augmentation du paramètre X2 pourrait signifier que des échanges plus importants s'effectuent entre ces sections du modèle.

$$F = X_2 \times (R/X_3)^{3.5} \quad (4.1)$$

Lorsque le réservoir de routage R atteint sa capacité maximale, donnée par X3 (mm) pour le jour suivant, la valeur maximale que peut prendre F est de X2. Et puisque ces deux paramètres

font partie de la même équation, il est possible que pour les stratégies où le paramètre X2 subit une augmentation, le paramètre X3 subisse également une hausse compensatoire, simplement pour permettre d'équilibrer la valeur de F (voir les stratégies 'LQ' et 'HQ'). En effet, lorsque la valeur de X2 gagne en importance, le paramètre X3 doit aussi croître pour que le résultat de l'équation reste inchangé. Toutefois, lorsque les pentes de ces paramètres sont de signes opposés, la valeur de F suivra la même direction que X2 (hausse ou baisse). Dans ce cas, il faut voir les stratégies 'D' et 'H', où F subit majoritairement une baisse et une hausse, plus le nombre d'années de calibration est élevé. Cette portion de la structure de GR4J donne des indications sur les quantités d'eaux souterraines échangées, sans toutefois permettre de tirer une conclusion qui relie directement le climat de calibration à la valeur des paramètres X2 et X3.

Que ce soit pour les tendances statistiquement significatives ou pour les tendances non significatives (Tableau 4.3 et Tableau 4.4), le contraste des pentes observées pour les paramètres X1, X2 et X3 suit le contraste des climats de calibration. Par exemple, le signe de la pente expérimentée pour la stratégie des années humides s'oppose à celui des années sèches, et vice-versa. Cette observation va de pair avec les travaux de Coron et al. (2011), démontrant une relation réelle entre le climat des périodes de calibration et la valeur de certains paramètres. Toutes les stratégies subissent ce contraste, malgré que ce dernier soit davantage prépondérant pour les pentes significatives.

Selon (Li et al., 2015), un modèle à performances élevées en calibration offre une stabilité de paramètres plus importante qu'un modèle à faible performance dans des conditions climatiques changeantes. Les Figure 4.1, Figure 4.2, Figure 4.3 et Figure 4.4 corroborent cette affirmation puisque le premier sous-groupe (S1), pour lequel les performances sont les plus élevées, démontre clairement une stabilité accrue pour les paramètres X1 à X4. En fait, plus l'ensemble de bassins versants est modélisé efficacement, plus la stabilité de ces quatre paramètres semble importante. Toujours selon la même référence, un modèle plus robuste devrait théoriquement être priorisé lors d'études d'impacts des changements climatiques.

En plus de démontré une grande stabilité des paramètres, le premier sous-groupe (S1) comporte un nombre de tendances statistiquement significatives ($\alpha = 0,05$) comparable à ceux observés pour la totalité des bassins, tel que démontré au Tableau 4.5. La concordance entre la direction de la pente (évolution du paramètre) et la stratégie de calibration utilisée est toutefois plus marquée pour S1 que pour tous les bassins analysés (voir Tableau 4.6), bien que ces résultats présentent encore trop de bruit pour être indicateurs d'une relation claire.

Tableau 4.5 Nombre de bassins (%) présentant une tendance significative au sein des 307 membres du sous-groupe 1 (S1)

| | W | D | H | LQ | HQ |
|----|-----|-----|-----|-----|-----|
| X1 | 20% | 26% | 29% | 27% | 33% |
| X2 | 33% | 32% | 42% | 49% | 45% |
| X3 | 20% | 29% | 40% | 27% | 42% |
| X4 | 27% | 30% | 32% | 30% | 32% |
| X5 | 13% | 13% | 11% | 14% | 11% |
| X6 | 6% | 8% | 5% | 3% | 7% |

Tableau 4.6 Nombre de bassins présentant une tendance significative ascendante ou descendante pour chaque paramètre au sein des membres du sous-groupe 1 (S1)

| | W | | D | | H | | LQ | | HQ | |
|----|-----------|-----------|-----------|-----------|-----------|------------|------------|-----------|-----------|------------|
| | As. | Des. | As. | Des. | As. | Des. | As. | Des. | As. | Des. |
| X1 | 29 | 32 | 20 | 60 | 70 | 18 | 13 | 71 | 84 | 17 |
| X2 | 74 | 28 | 39 | 58 | 82 | 48 | 139 | 12 | 8 | 129 |
| X3 | 32 | 28 | 80 | 8 | 6 | 117 | 74 | 8 | 4 | 124 |
| X4 | 46 | 38 | 26 | 65 | 78 | 20 | 45 | 48 | 74 | 24 |
| X5 | 11 | 30 | 17 | 23 | 13 | 20 | 12 | 32 | 14 | 20 |
| X6 | 3 | 15 | 7 | 18 | 7 | 7 | 5 | 5 | 5 | 17 |

Du côté des bassins davantage influencés par un couvert de neige, on perçoit une stabilité (pente inférieure à ± 0.3) du paramètre X5 pour les bassins où la chute annuelle moyenne de neige surpasse les 200 mm EEN, lorsque la stratégie 'W' est employée (Figure 4.5). Ce paramètre a pour objectif de faire le pont entre la température moyenne journalière et l'énergie thermique contenue dans le manteau (neige accumulée). Une grande stabilité de ce paramètre pour des bassins où la neige est un facteur important peut signifier que le paramètre X5 perd en partie son influence pour une région recevant une grande quantité de neige. En d'autres

termes, le paramètre X5 fluctue peu lorsque la quantité de neige est importante et que davantage d'années de calibration aux températures moyennes plus basses sont utilisées. Toutefois, il faut mentionner que l'analyse de ce phénomène ne s'appuie que sur quelques bassins où la tendance apparaît comme statistiquement significative. De plus, un patron semblable se dessine pour le paramètre X6, à la Figure 4.6, mais les résultats ne laissent pas entrevoir une stabilité aussi évidente que pour le paramètre X5.

CONCLUSION

À partir d'un ensemble de 921 bassins versants nord-américains, cinq stratégies de calibration de modèles hydrologiques ont été testées, en plus de comparer les répercussions de différents nombres d'années de calibration (5 à 40 ans) sur les performances en validation. L'analyse de ces stratégies fut réalisée à partir du modèle pluie-débit GR4J-Cemaneige. Un total de 46 années de données hydrométéorologiques furent utilisées pour chaque bassin versant. Dans le but de cibler des anomalies climatiques les plus grandes possibles, les périodes de validation étaient composées des cinq années aux températures les plus élevées, aux précipitations annuelles les plus élevées/faibles, ainsi qu'aux débits les plus élevés/faibles. Les 41 autres années disponibles furent utilisées pour la calibration, sous forme d'échantillons allant de 5 à 40 ans (par multiples de 5). Une étude de la stabilité des paramètres fut également réalisée à partir des valeurs des paramètres optimaux obtenus suite aux processus de calibration. Cette analyse consistait à vérifier la présence de tendances (significatives ou non) au sein des huit valeurs optimales d'un paramètre pour chaque stratégie de calibration, en plus de vérifier si les paramètres du modèle GR4J étaient plus stables pour des bassins aux performances supérieures. En dernier lieu, une recherche fut menée dans le but d'investiguer les relations entre la chute de neige annuelle moyenne et la stabilité des deux paramètres de CemaNeige à travers la stratégie des années chaudes.

Les résultats montrent des divergences entre les différents bassins pour les cinq stratégies de calibration testées. Toutefois, des conclusions globales peuvent être tirées. Les résultats démontrent que d'écartier quelques-unes des années les plus froides de la période de calibration est une pratique fiable lorsque la période de validation est constituée des années les plus chaudes. Dans le cas des quatre autres stratégies de calibration, où la période de validation est fortement influencée par les précipitations, retirer un bon nombre d'années aux anomalies moindres est la solution la plus efficace. Une approche de calibration basée sur toutes les années disponibles (40 ans dans ce cas-ci) apparaît comme étant robuste dans la plupart des cas, bien que non-optimale. À l'opposé, l'utilisation d'un très petit nombre d'années (5 à 10 ans) n'est pas une approche recommandable.

Les évolutions que subissent les paramètres X1, X2 et X3 à mesure que des années de calibration sont ajoutées sont généralement de directions contraires pour des climats de calibration contrastés. Par exemple, le paramètre X2 subit plus souvent une diminution pour la stratégie des années sèches, alors qu'il subit majoritairement une augmentation pour la stratégie des années humides. Peu importe la stratégie de calibration utilisée, une forte efficacité en calibration sur un bassin versant est garante d'une stabilité accrue des quatre paramètres de GR4J. Les valeurs des paramètres de CemaNeige semblent plus stables pour les bassins où la précipitation neigeuse annuelle moyenne excède 200 mm et que l'échantillon d'années de calibration contient davantage d'années froides.

Dans l'ensemble, cette recherche a permis d'améliorer les connaissances sur le choix d'une stratégie de calibration pour tenter d'atténuer les incertitudes liées aux conditions hydroclimatiques futures. Les conclusions obtenues à ce sujet démontrent que le choix des années de calibration est important, bien que la variabilité à travers les différents bassins versants étudiés est trop grande pour donner des recommandations strictes. L'utilisation de modèles hydrologiques pour des études d'impacts des changements climatiques requiert une meilleure compréhension de la transposabilité temporelle des paramètres optimaux. Pour améliorer cette compréhension, davantage de recherches doivent étudier l'évolutions des valeurs de paramètres à travers différentes périodes de calibration et ce, pour différents modèles hydrologiques.

ANNEXE I

ÉVOLUTION DES VALEURS DE PARAMÈTRES (X1 À X6)

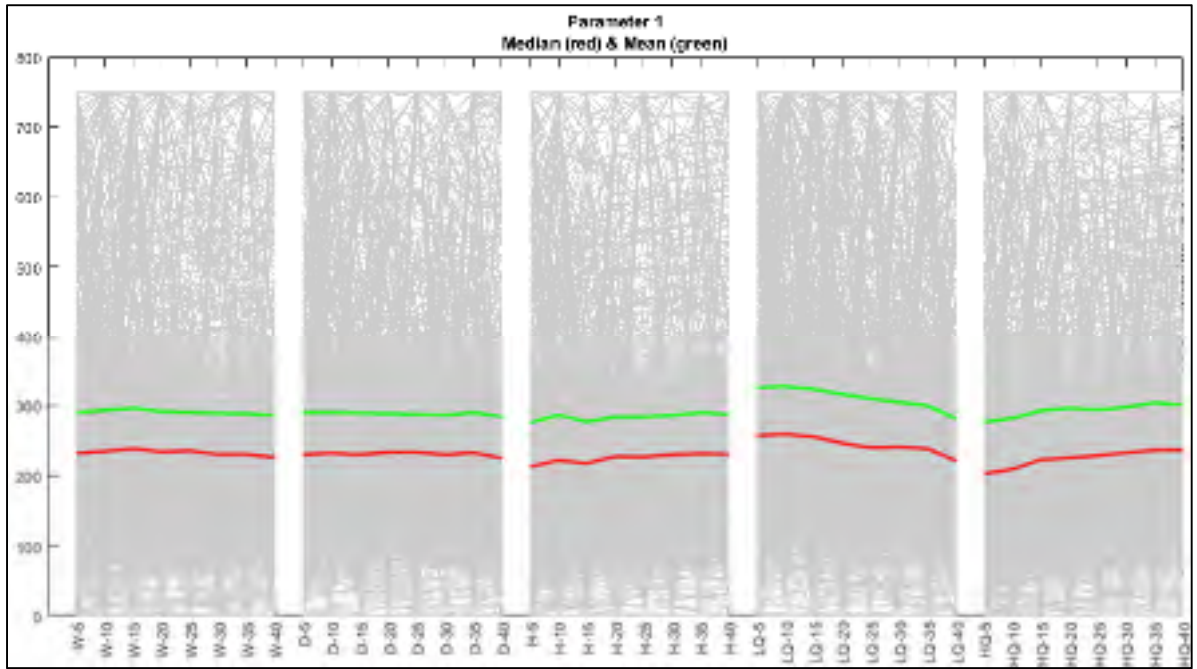


Figure-A I-1 Valeurs du paramètre X1 pour toutes les stratégies (testées sur 921 bassins)

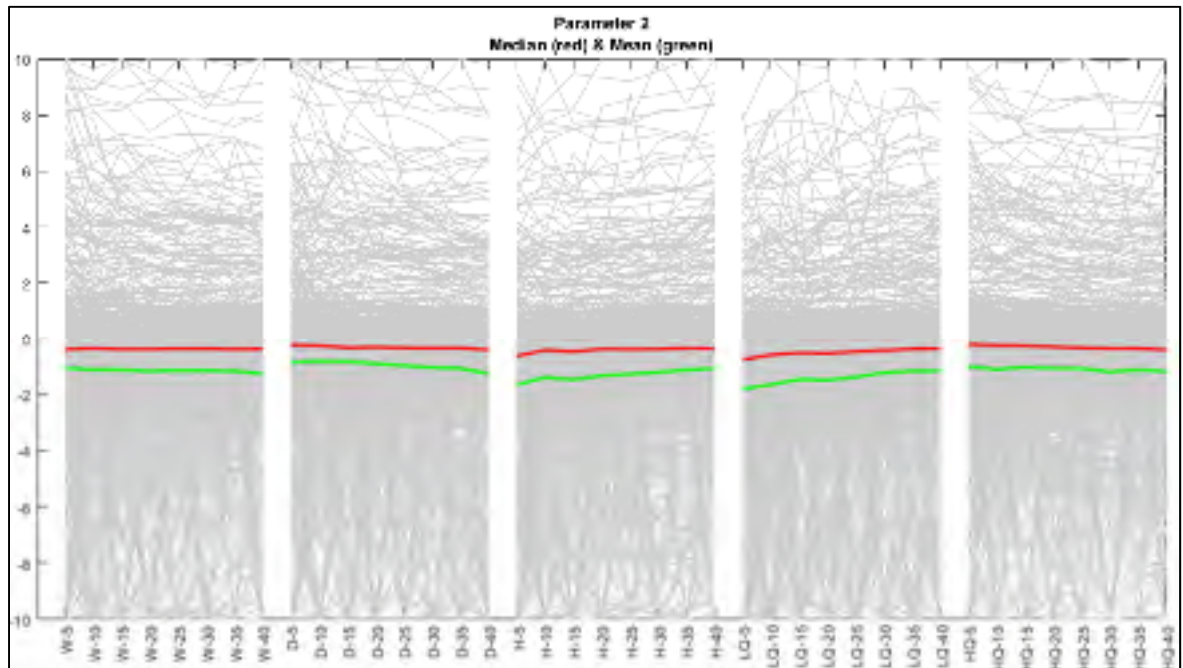


Figure-A I-2 Valeurs du paramètre X2 pour toutes les stratégies (testées sur 921 bassins)

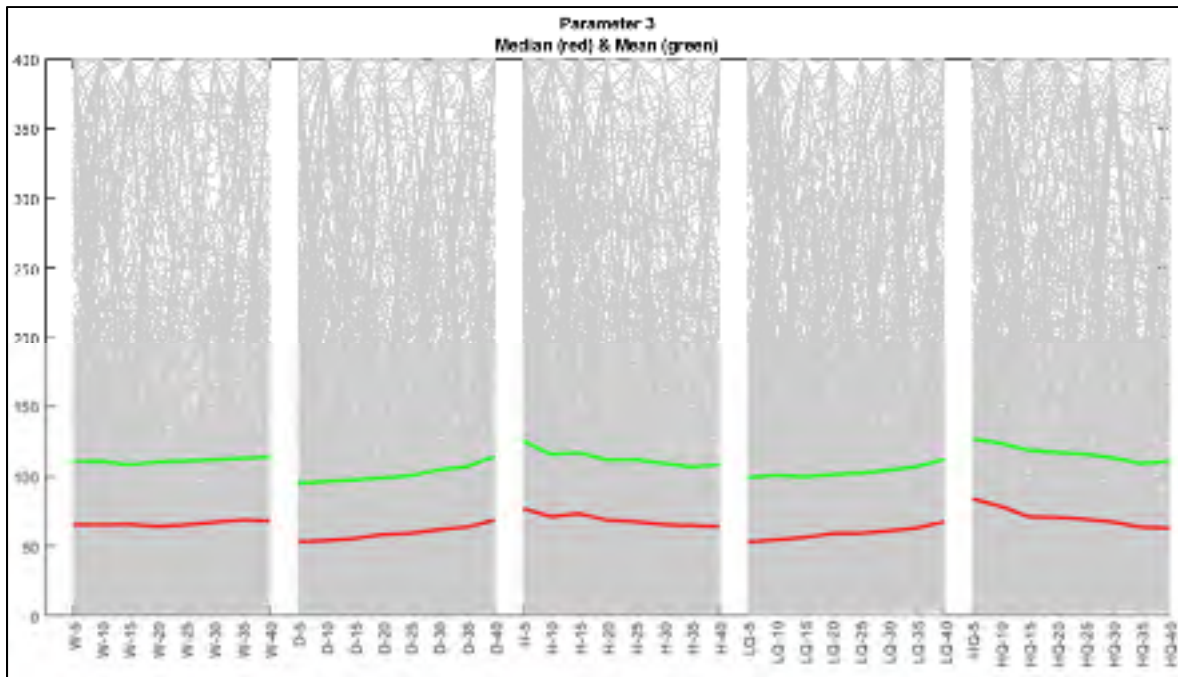


Figure-A I-3 Valeurs du paramètre X3 pour toutes les stratégies (testées sur 921 bassins)

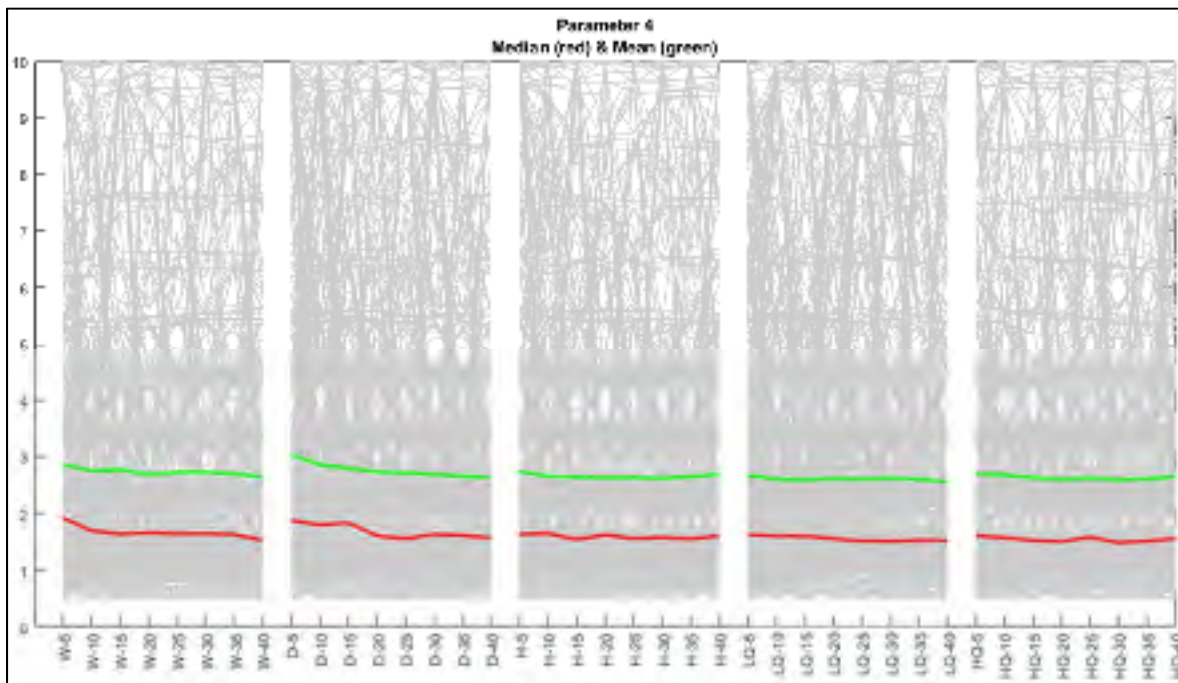


Figure-A I-4 Valeurs du paramètre X4 pour toutes les stratégies (testées sur 921 bassins)

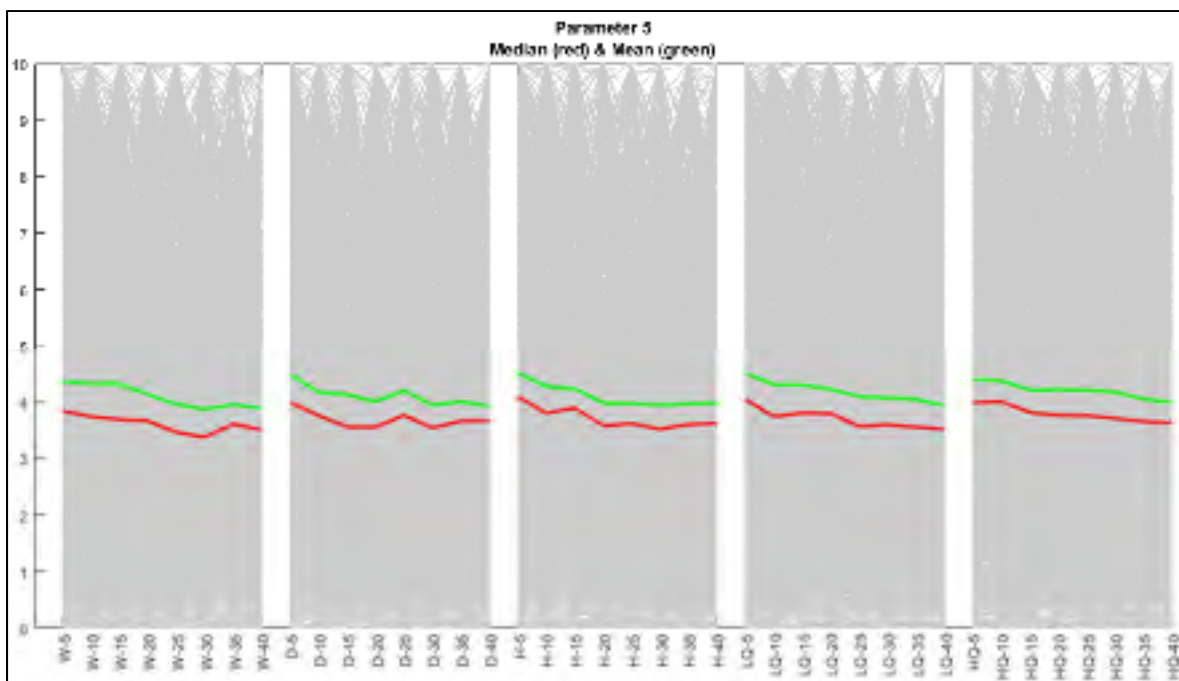


Figure-A I-5 Valeurs du paramètre X5 pour toutes les stratégies (testées sur 921 bassins)

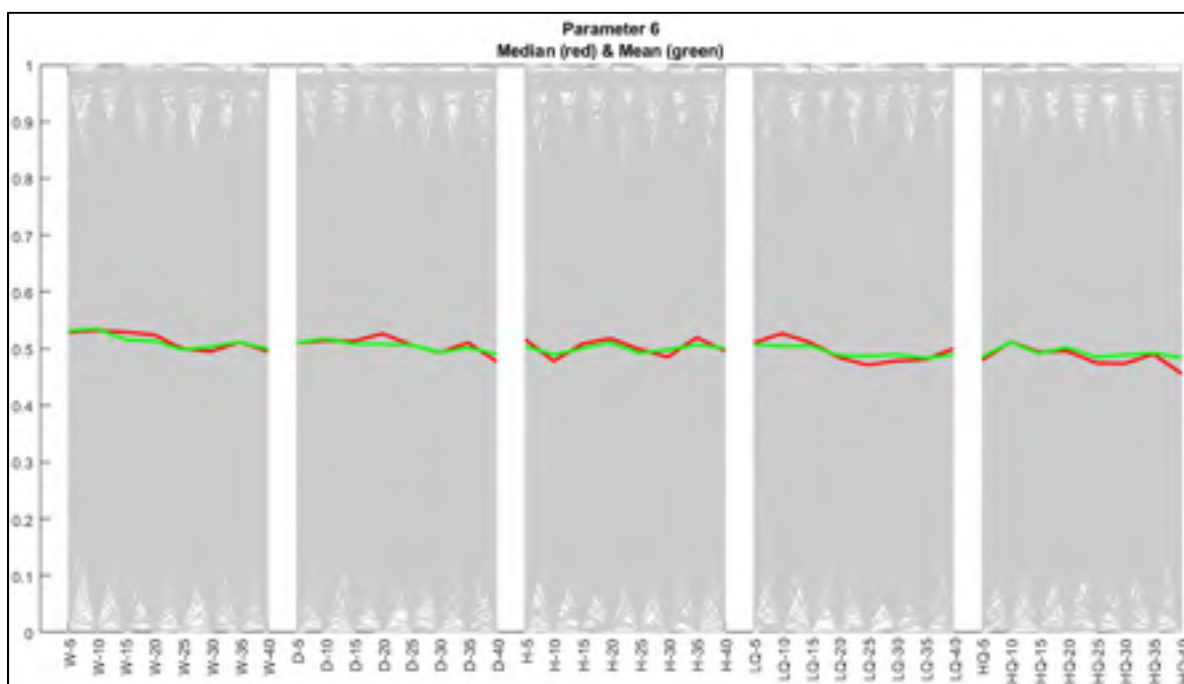


Figure-A I-6 Valeurs du paramètre X6 pour toutes les stratégies (testées sur 921 bassins)

ANNEXE II

STRUCTURE DU MODÈLE PLUIE-DÉBIT GR4J-CemaNeige

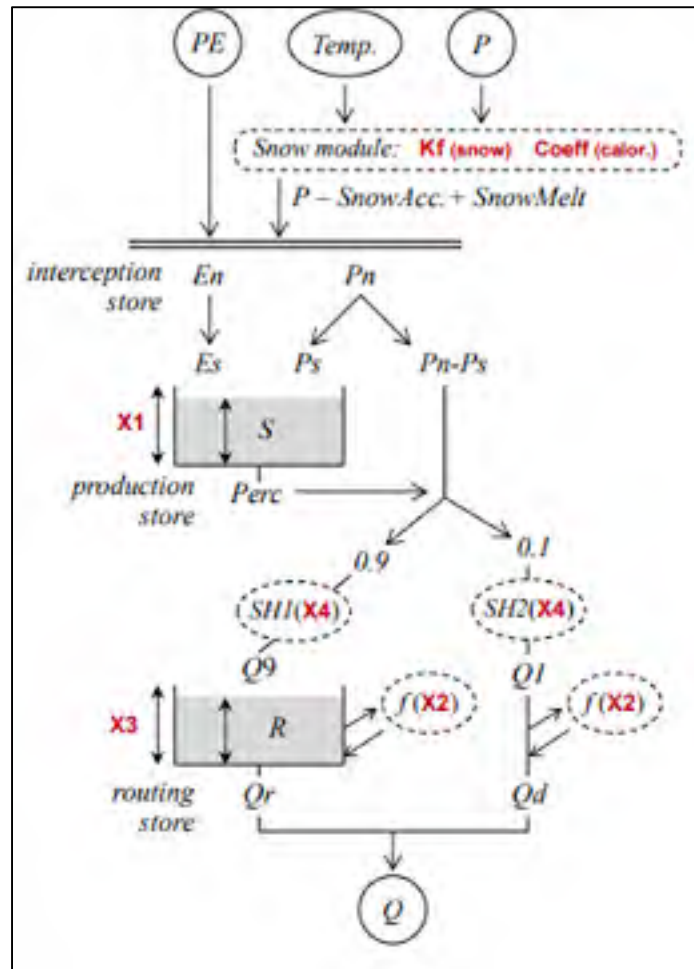


Figure-A II-1 Schéma de production du modèle GR4J-CemaNeige
Tirée de Coron et al. (2014)

ANNEXE III

COORDONNÉES DES BASSINS INFLUENCÉS PAR LA NEIGE



Figure-C III-1 Emplacements géographiques des bassins versants influencés par une chute de neige annuelle moyenne supérieure ou égale à 50 mm EEN

LISTE DE RÉFÉRENCES BIBLIOGRAPHIQUES

- Anderson, E. A. (2002). Calibration of conceptual hydrologic models for use in river forecasting. *Office of Hydrologic Development, US National Weather Service, Silver Spring, MD.*
- Andréassian, V., Perrin, C., & Michel, C. (2004). Impact of imperfect potential evapotranspiration knowledge on the efficiency and parameters of watershed models. *Journal of Hydrology*, 286(1-4), 19-35.
- Arsenault, R., Brissette, F., Chen, J., Guo, Q., & Dallaire, G. (2020). NAC2H: The North American Climate Change and Hydroclimatology Data Set. *Water Resources Research*, 56(8), e2020WR027097.
- 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), 1-12.
- 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.
- Beven, K., & Freer, J. (2001). Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *Journal of Hydrology*, 249(1-4), 11-29.
- 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.
- Burkey, J. (2006). A non-parametric monotonic trend test computing Mann-Kendall Tau, Tau-b, and Sen's slope written in Mathworks-MATLAB implemented using matrix rotations. *King County, Department of Natural Resources and Parks, Science and Technical Services Section: Seattle, WA.*
- Conover, W. (1981). Nonparametric methods. *Practical Nonparametric Statistics*, 2, 3.
- Coron, L., Andréassian, V., Bourqui, M., Perrin, C., & Hendrickx, F. (2011). Pathologies of hydrological models used in changing climatic conditions: a review. Dans *Symposium J-H02 at IUGG 2011* (pp. 6).

- Coron, L., Andréassian, V., Perrin, C., Bourqui, M., & Hendrickx, F. (2014). On the lack of robustness of hydrologic models regarding water balance simulation: a diagnostic approach applied to three models of increasing complexity on 20 mountainous catchments.
- Coron, L., Andreassian, 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).
- Coron, L., Thirel, G., Delaigue, O., Perrin, C., & Andréassian, V. (2017). The suite of lumped GR hydrological models in an R package. *Environmental Modelling & Software*, 94, 166-171.
- Deser, C., Knutti, R., Solomon, S., & Phillips, A. S. (2012). Communication of the role of natural variability in future North American climate. *Nature Climate Change*, 2(11), 775-779.
- Folland, C., Karl, T., Christy, J., Clarke, R., Gruza, G., Jouzel, J., . . . Wang, S. (2001). Observed climate variability and change. *Climate change*, 2001, 99.
- Gharari, S., Hrachowitz, M., Fenicia, F., & Savenije, H. (2013). An approach to identify time consistent model parameters: sub-period calibration. *Hydrology and Earth System Sciences*, 17(1), 149.
- Giuntoli, I., Villarini, G., Prudhomme, C., & Hannah, D. M. (2018). Uncertainties in projected runoff over the conterminous United States. *Climatic change*, 150(3-4), 149-162.
- Goosse, H., Renssen, H., Timmermann, A., & Bradley, R. S. (2005). Internal and forced climate variability during the last millennium: a model-data comparison using ensemble simulations. *Quaternary Science Reviews*, 24(12-13), 1345-1360.
- 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(1-2), 80-91.
- Hansen, N., & Ostermeier, A. (1996). Adapting arbitrary normal mutation distributions in evolution strategies: The covariance matrix adaptation. Dans *Proceedings of IEEE international conference on evolutionary computation* (pp. 312-317). IEEE.
- Hansen, N., & Ostermeier, A. (2001). Completely derandomized self-adaptation in evolution strategies. *Evolutionary computation*, 9(2), 159-195.
- Hawkins, E., & Sutton, R. (2012). Time of emergence of climate signals. *Geophysical Research Letters*, 39(1).

- 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), 1-22.
- Hollander, M., Wolfe, D. A., & Chicken, E. (2013). *Nonparametric statistical methods* (Vol. 751). John Wiley & Sons.
- IPCC. (2013). *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
- Kay, A., Davies, H., Bell, V., & Jones, R. (2009). Comparison of uncertainty sources for climate change impacts: flood frequency in England. *Climatic change*, 92(1-2), 41-63.
- Kirchner, J. W. (2006). Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. *Water Resources Research*, 42(3).
- Klemeš, V. (1986). Operational testing of hydrological simulation models. *Hydrological sciences journal*, 31(1), 13-24.
- 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.
- Le Lay, M., Galle, S., Saulnier, G.-M., & Braud, I. (2007). Exploring the relationship between hydroclimatic stationarity and rainfall-runoff model parameter stability: A case study in West Africa. *Water Resources Research*, 43(7).
- Lehner, F., Deser, C., & Terray, L. (2017). Toward a new estimate of “time of emergence” of anthropogenic warming: Insights from dynamical adjustment and a large initial-condition model ensemble. *Journal of Climate*, 30(19), 7739-7756.
- Li, H., Beldring, S., & Xu, C.-Y. (2015). Stability of model performance and parameter values on two catchments facing changes in climatic conditions. *Hydrological sciences journal*, 60(7-8), 1317-1330.
- Lorenz, E. N. (1995). *The essence of chaos*. University of Washington press.
- 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.

- Martel, J.-L., Mailhot, A., Brissette, F., & Caya, D. (2018). Role of natural climate variability in the detection of anthropogenic climate change signal for mean and extreme precipitation at local and regional scales. *Journal of Climate*, 31(11), 4241-4263.
- Mayewski, P. A., Rohling, E. E., Stager, J. C., Karlén, W., Maasch, K. A., Meeker, L. D., . . . Holmgren, K. (2004). Holocene climate variability. *Quaternary research*, 62(3), 243-255.
- Mendoza, P. A., Clark, M. P., Mizukami, N., Newman, A. J., Barlage, M., Gutmann, E. D., . . . Arnold, J. R. (2015). Effects of hydrologic model choice and calibration on the portrayal of climate change impacts. *Journal of Hydrometeorology*, 16(2), 762-780.
- Merz, R., Parajka, J., & Blöschl, G. (2011). Time stability of catchment model parameters: Implications for climate impact analyses. *Water Resources Research*, 47(2).
- Motavita, D., 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.
- 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.
- Oudin, L., Andréassian, V., Perrin, C., Michel, C., & Le Moine, N. (2008). Spatial proximity, physical similarity, regression and ungaged catchments: A comparison of regionalization approaches based on 913 French catchments. *Water Resources Research*, 44(3).
- 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(1-4), 290-306.
- Pachauri, R. K., Allen, M. R., Barros, V. R., Broome, J., Cramer, W., Christ, R., . . . Dasgupta, P. (2014). *Climate change 2014: synthesis report. Contribution of Working Groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change*. Ipcc.
- Perrin, C., Michel, C., & Andréassian, V. (2003). Improvement of a parsimonious model for streamflow simulation. *Journal of Hydrology*, 279(1-4), 275-289.
- 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.

- 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(3-4), 626-636.
- Prudhomme, C., & Davies, H. (2009). Assessing uncertainties in climate change impact analyses on the river flow regimes in the UK. Part 1: baseline climate. *Climatic change*, 93(1-2), 177-195.
- Razavi, S., & Tolson, B. A. (2013). An efficient framework for hydrologic model calibration on long data periods. *Water Resources Research*, 49(12), 8418-8431.
- Riboust, P., Thirel, G., Le Moine, N., & Ribstein, P. (2019). Revisiting a simple degree-day model for integrating satellite data: implementation of SWE-SCA hystereses. *Journal of hydrology and hydromechanics*, 67(1), 70-81.
- Seiller, G., Hajji, I., & Anctil, F. (2015). Improving the temporal transposability of lumped hydrological models on twenty diversified US watersheds. *Journal of Hydrology: Regional Studies*, 3, 379-399.
- Tarek, M., Brissette, F. P., & Arsenault, R. (2019). Evaluation of the ERA5 reanalysis as a potential reference dataset for hydrological modeling over North-America. *Hydrology and Earth System Sciences Discussions*, 1-35.
- Thirel, G., Andréassian, V., & Perrin, C. (2015). On the need to test hydrological models under changing conditions: Taylor & Francis.
- Valéry, A. (2010). *Modélisation précipitations débit sous influence nivale: Elaboration d'un module neige et évaluation sur 380 bassins versants* (Doctorat Hydrobiologie, Institut des Sciences et Industries du Vivant et de l'Environnement AgroParisTech).
- Valéry, A., Andréassian, V., & Perrin, C. (2014a). 'As simple as possible but not simpler': What is useful in a temperature-based snow-accounting routine? Part 1—Comparison of six snow accounting routines on 380 catchments. *Journal of Hydrology*, 517, 1166-1175.
- Valéry, A., Andréassian, V., & Perrin, C. (2014b). '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.
- Vansteenkiste, T., Tavakoli, M., Ntegeka, V., De Smedt, F., Batelaan, O., Pereira, F., & Willems, P. (2014). Intercomparison of hydrological model structures and calibration approaches in climate scenario impact projections. *Journal of Hydrology*, 519, 743-755.

- Vaze, J., Post, D., Chiew, F., Perraud, J.-M., Viney, N., & Teng, J. (2010). Climate non-stationarity–validity of calibrated rainfall–runoff models for use in climate change studies. *Journal of Hydrology*, 394(3-4), 447-457.
- Vogel, C., & o'Brien, K. (2017). *Coping with climate variability*. Routledge.
- 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.
- Yue, S., Pilon, P., & Cavadias, G. (2002). Power of the Mann–Kendall and Spearman's rho tests for detecting monotonic trends in hydrological series. *Journal of Hydrology*, 259(1-4), 254-271.
- 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.