

Université du Québec  
Institut national de la recherche scientifique  
Centre Eau, Terre et Environnement

# **Le Modèle Régional Canadien du Climat reproduit-il la variabilité observée des précipitations et des écoulements annuels ?**

Par  
JONATHAN JALBERT

Mémoire présenté dans le cadre du programme de maîtrise en sciences de l'eau  
pour l'obtention du grade de Maître ès sciences (M.Sc.)

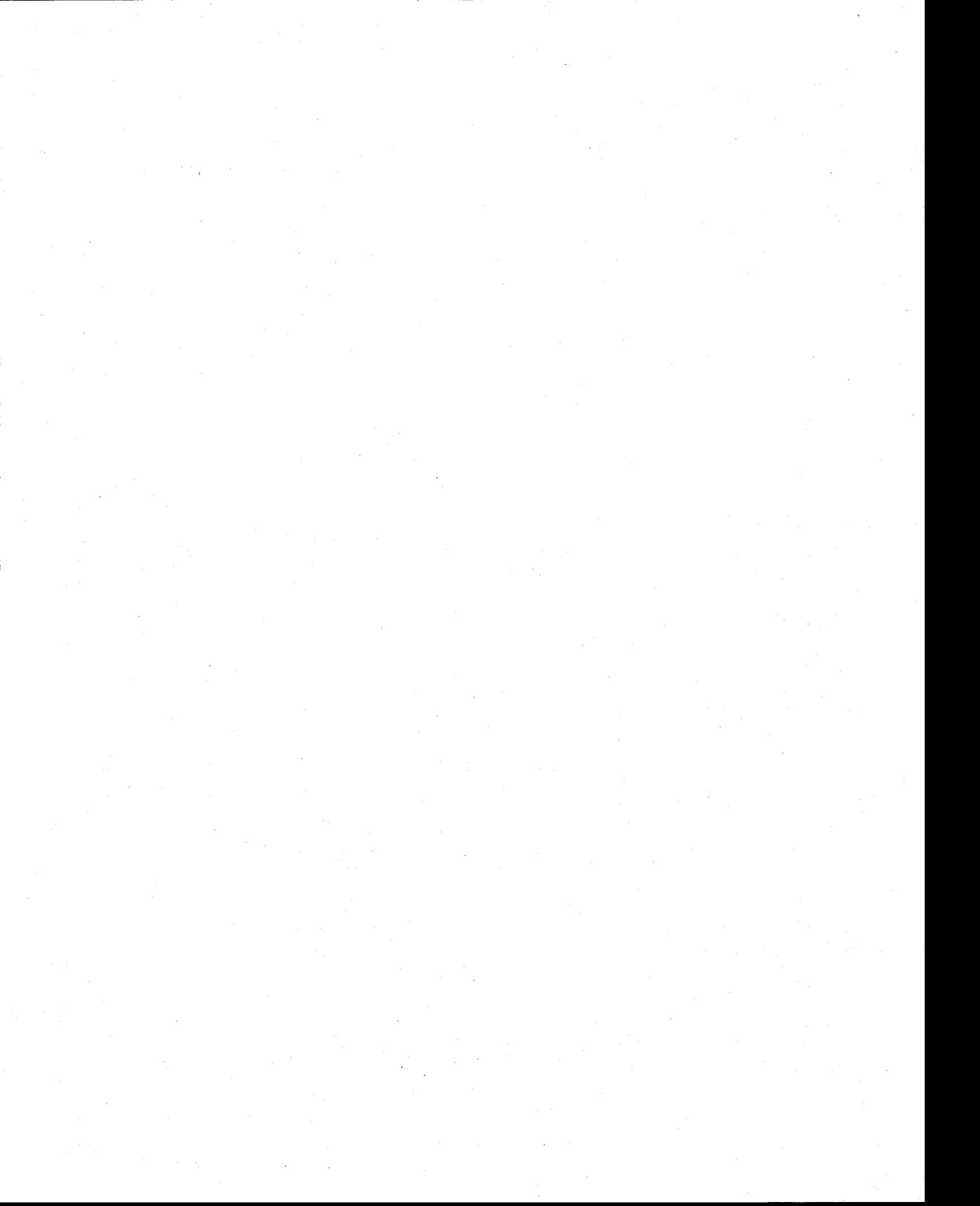
## Jury d'évaluation

|                           |                                  |
|---------------------------|----------------------------------|
| Directeur de recherche:   | Anne-Catherine Favre<br>INRS-ETE |
| Codirecteur de recherche: | Michel Slivitzky<br>INRS-ETE     |
| Examineur interne:        | Alain Mailhot<br>INRS-ETE        |
| Examineur externe:        | Debbie J. Dupuis<br>HEC Montréal |



# Résumé

Un modèle climatique, comme le Modèle Régional Canadien du Climat (MRCC), doit être apte à reproduire, au minimum, la valeur centrale et la dispersion des séries chronologiques pour estimer, par exemple, la probabilité d'occurrence des valeurs extrêmes. Ceci peut être vérifié en comparant les séries chronologiques simulées lorsque le modèle est piloté par les conditions climatiques passées et les séries observées correspondantes. Le but du projet de recherche est donc de vérifier la capacité du MRCC à reproduire la variabilité observée des précipitations et des écoulements annuels à l'échelle des régions hydrographiques du Québec. Cette vérification s'effectue en comparant les intervalles de confiance du coefficient de variation, une mesure de dispersion adimensionnelle, afin de tester la significativité des différences. À la suite de l'évaluation de la robustesse des intervalles de confiance, on peut dire que le MRCC reproduit la variabilité relative des précipitations et des écoulements annuels.



# Remerciements

Je souhaite d'abord remercier Anne-Catherine Favre, ma directrice de recherche, qui m'a accordé tout au long de la maîtrise, une confiance extraordinaire. Elle a non seulement encouragé mes initiatives, mais a aussi pris une part importante dans leurs réalisations. Je pense particulièrement à ses démarches et son appui dans la concrétisation de deux stages durant la maîtrise. Aussi, sa générosité se manifeste bien au-delà du cadre académique ce qui en fait non seulement une directrice de recherche exceptionnelle, mais une femme exceptionnelle.

Je suis reconnaissant envers mon codirecteur Michel Slivitzky qui a accepté de partager son expérience et sa vision remarquable. Également, j'ai eu beaucoup de plaisir à travailler avec Mériem Saïd, l'associée de recherche d'Anne-Catherine. Nos discussions autour d'un café, face au fleuve, dans un décor soviétique, ont été très enrichissantes. Je remercie aussi Karine Dion qui, de par son exemple, est une source de motivation et de courage.

Je souhaiterais aussi remercier les étudiants avec qui j'ai complété cette maîtrise (Audrey, Vincent, Joanie, Samuel, Hélène). Je tiens à remercier plus particulièrement Véronique Tremblay qui a su élargir ma vision scientifique au-delà des maths et de la physique. Si je suis un meilleur scientifique aujourd'hui, c'est en grande partie grâce à elle.

Je remercie Justine Lafrance, ma fiancée, qui n'a jamais cessé de m'encourager et qui a su être très patiente lors d'un de mes stages à l'étranger. En dernier lieu, je tiens à remercier ma famille qui tout au long de ma formation académique n'a jamais cessé de m'appuyer.



# Table des matières

|                                                              |           |
|--------------------------------------------------------------|-----------|
| Résumé                                                       | iii       |
| Remerciements                                                | v         |
| Table des matières                                           | viii      |
| Liste des tableaux                                           | x         |
| Table des figures                                            | xii       |
| <b>1 Introduction</b>                                        | <b>1</b>  |
| 1.1 Les changements climatiques et l'eau                     | 1         |
| 1.1.1 Les modèles climatiques globaux (MCG)                  | 3         |
| 1.1.2 Les modèles climatiques régionaux (MCR)                | 3         |
| 1.1.3 La validation des modèles climatiques                  | 4         |
| 1.2 Les objectifs du mémoire                                 | 7         |
| <b>2 Cadre théorique</b>                                     | <b>9</b>  |
| 2.1 Le coefficient de variation                              | 9         |
| 2.2 Les intervalles de confiance du coefficient de variation | 11        |
| 2.2.1 Les approches basées sur une quantité pivot            | 12        |
| 2.2.2 L'approche de Hendricks et Robey (1936)                | 15        |
| 2.2.3 L'approche de Arvesen (1972)                           | 16        |
| 2.2.4 L'approche de Wong et Wu (2002)                        | 17        |
| 2.2.5 L'approche de Bonett(2006)                             | 18        |
| <b>3 Cadre expérimental</b>                                  | <b>23</b> |
| 3.1 La description des données                               | 23        |
| 3.1.1 L'échelle d'analyse temporelle                         | 23        |
| 3.1.2 Le territoire étudié                                   | 24        |
| 3.1.3 Les écoulements observés                               | 25        |
| 3.1.4 Les précipitations réanalysées                         | 27        |
| 3.1.5 Les écoulements et précipitations simulés              | 30        |

|          |                                                                                   |           |
|----------|-----------------------------------------------------------------------------------|-----------|
| 3.2      | Le critère de comparaison choisi . . . . .                                        | 31        |
| 3.3      | Les caractéristiques des données . . . . .                                        | 32        |
| 3.4      | La méthode de Monte-Carlo . . . . .                                               | 35        |
| 3.5      | La comparaison des coefficients de variation . . . . .                            | 37        |
| 3.5.1    | Regrouper les données par bassins versants . . . . .                              | 38        |
| 3.5.2    | Vérifier la robustesse des intervalles de confiance . . . . .                     | 38        |
| 3.5.3    | Calculer les intervalles de confiance . . . . .                                   | 39        |
| 3.5.4    | Tester la significativité des différences des coefficients de variation . . . . . | 39        |
| <b>4</b> | <b>Résultats</b>                                                                  | <b>41</b> |
| 4.1      | Les résultats des simulations de Monte-Carlo . . . . .                            | 41        |
| 4.1.1    | Le cas normal ( $\gamma = 3$ ) . . . . .                                          | 41        |
| 4.1.2    | Les cas où $\gamma > 3$ . . . . .                                                 | 44        |
| 4.2      | Les intervalles de confiance des coefficients de variation . . . . .              | 45        |
| 4.2.1    | Écoulements . . . . .                                                             | 45        |
| 4.2.2    | Précipitations . . . . .                                                          | 45        |
| <b>5</b> | <b>Discussion</b>                                                                 | <b>51</b> |
| 5.1      | La robustesse des intervalles de confiance . . . . .                              | 51        |
| 5.2      | L'inférence à l'aide des intervalles de confiance . . . . .                       | 53        |
| 5.2.1    | L'inférence sur les écoulements . . . . .                                         | 53        |
| 5.2.2    | Les différences entre les réanalyses . . . . .                                    | 54        |
| 5.2.3    | L'inférence sur les précipitations . . . . .                                      | 55        |
| 5.2.4    | La différence entre les bassins versants . . . . .                                | 57        |
| 5.3      | La critique de la méthode . . . . .                                               | 58        |
| 5.3.1    | L'hypothèse de normalité . . . . .                                                | 58        |
| 5.3.2    | L'indépendance et la stationnarité . . . . .                                      | 59        |
| 5.4      | La comparaison avec d'autres approches . . . . .                                  | 61        |
| 5.4.1    | Les tests d'égalité des coefficients de variation . . . . .                       | 61        |
| 5.4.2    | Les intervalles de crédibilité bayésien . . . . .                                 | 63        |
| 5.5      | La complémentarité avec d'autres approches . . . . .                              | 64        |
| <b>6</b> | <b>Conclusion</b>                                                                 | <b>67</b> |
|          | <b>Bibliographie</b>                                                              | <b>70</b> |
|          | <b>A Notions fondamentales de statistique</b>                                     | <b>79</b> |
|          | <b>B Caractéristiques des données</b>                                             | <b>81</b> |

# Liste des tableaux

|     |                                                                                                                                                                                                                                                                                                                                    |    |
|-----|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 1.1 | Caractéristiques des bassins versants étudiés par les études de Music et Caya (2009) et Gagnon <i>et al.</i> (2009). . . . .                                                                                                                                                                                                       | 6  |
| 3.1 | Liste et caractéristiques des bassins versants du Québec à l'étude (à noter que l'acronyme MDDEP désigne le Ministère du Développement Durable, de l'Environnement et des Parcs). . . . .                                                                                                                                          | 26 |
| 3.2 | Dénombrement des valeurs extrêmes telles que définies précédemment parmi les différentes sources de données pour les écoulements et les précipitations annuels. . . . .                                                                                                                                                            | 36 |
| 4.1 | Pourcentage de couverture empirique des intervalles de confiance pour une taille d'échantillon de $n = 20$ et pour un niveau de confiance nominal de 95%. . . . .                                                                                                                                                                  | 42 |
| 4.2 | Pourcentage de couverture empirique des intervalles de confiance pour une taille d'échantillon de $n = 40$ et pour un niveau de confiance nominal de 95%. . . . .                                                                                                                                                                  | 42 |
| 4.3 | Pourcentage de couverture empirique des intervalles de confiance pour une taille d'échantillon de $n = 100$ et pour un niveau de confiance nominal de 95%. . . . .                                                                                                                                                                 | 43 |
| 4.4 | Pourcentage de couverture empirique des intervalles de confiance pour une taille d'échantillon de $n = 1000$ et pour un niveau de confiance nominal de 95%. . . . .                                                                                                                                                                | 43 |
| 5.1 | P-valeurs des tests d'égalité du coefficient de variation entre les écoulements observés et simulés. La colonne de gauche énumère les 21 bassins versants étudiés. Les acronymes <i>BM</i> , <i>FM</i> et <i>SR</i> correspondent aux tests de Bennett (1976), Feltz et Miller (1996) et de Conover (1980) respectivement. . . . . | 62 |
| B.1 | Liste des séries chronologiques exhibant une autocorrélation (tau de Kendall) significative. Le tableau indique aussi l'ordre de l'autocorrélation.                                                                                                                                                                                | 82 |

B.2 Statistiques et p-valeurs des séries annuelles de chacun des bassins versants. La p-valeur brute représente le seuil expérimental du test de normalité de D'Agostino *et al.* (1990) sur la série complète. La p-valeur représente le seuil expérimental du test lorsque les extrêmes sont retirés de la série. . . . . 83

# Table des figures

|     |                                                                                                                                                                                                                                                                                                                                  |    |
|-----|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 3.1 | Carte géographique des 21 bassins versants étudiés. . . . .                                                                                                                                                                                                                                                                      | 27 |
| 3.2 | Diagramme en boîte des débits annuels observés pour les 21 bassins versants à l'étude. . . . .                                                                                                                                                                                                                                   | 28 |
| 3.3 | Le grand domaine AMNO et le petit domaine QC (encadré blanc) utilisés dans les simulations du MRCC. La carte illustre la topographie en mètre de cette région. Tiré de Music <i>et al.</i> (2009). . . . .                                                                                                                       | 30 |
| 3.4 | P-valeurs du test de normalité pour chacune des sources de données. Les bassins en rouge sont ceux dont le test suggère de rejeter l'hypothèse de normalité au seuil de 10%. . . . .                                                                                                                                             | 33 |
| 3.5 | Droites de Henry ( <i>Q-Q plots</i> ) pour chacune des séries ne pouvant être supposées normales. . . . .                                                                                                                                                                                                                        | 34 |
| 4.1 | Pourcentage de couverture empirique de l'intervalle de Bonett (2006) pour plusieurs tailles d'échantillon et pour plusieurs valeur de coefficient d'aplatissement. . . . .                                                                                                                                                       | 44 |
| 4.2 | Intervalle de confiance de Vangel (1996) de niveau 95% des coefficients de variation des écoulements observés et simulés par le MRCC (acy) pour chacun des bassins versants étudiés. Les bandes symbolisent les intervalles de confiance tandis que les astérisques indiquent les valeurs des coefficients de variation. . . . . | 46 |
| 4.3 | Intervalle de confiance de Bonett (2006) de niveau 95% des coefficients de variation des écoulements observés et simulés par le MRCC (acy) pour chacun des bassins versants étudiés. Les bandes symbolisent les intervalles de confiance tandis que les astérisques indiquent les valeurs des coefficients de variation. . . . . | 47 |
| 4.4 | Intervalle de confiance de Vangel (1996) de niveau 95% des coefficients de variation des précipitations simulées par le MRCC (acy) et des précipitations réanalysées (CRU et WM2) pour chacun des bassins versants étudiés. . . . .                                                                                              | 48 |

|     |                                                                                                                                                                                                                                                                                                                               |    |
|-----|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 4.5 | Intervalle de confiance de Bonett (2006) de niveau 95% des coefficients de variation des précipitations simulées par le MRCC (acy) et des précipitations réanalysées (CRU et WM2) pour chacun des bassins versants étudiés. . . . .                                                                                           | 49 |
| 5.1 | Localisation des stations pluviométriques d'Environnement Canada ayant enregistré au moins une donnée durant la période 1961-1999. . . . .                                                                                                                                                                                    | 56 |
| 5.2 | Série chronologique des écoulements annuels du bassin versant de la rivière Moisie. La ligne en tirets noirs verticale représente la rupture de la série. Les lignes rouges horizontales en tirets tracent la moyenne plus l'écart type et la moyenne moins l'écart type, avant et après la rupture. . . . .                  | 59 |
| 5.3 | Intervalle de crédibilité bayésien de niveau 95% des coefficients de variations des écoulements observés et simulés par le MRCC (acy) pour chacun des bassins versants étudiés. Les bandes symbolisent les intervalles de crédibilité tandis que les astérisques indiquent les valeurs des coefficients de variation. . . . . | 64 |

# Chapitre 1

## Introduction

### 1.1 Les changements climatiques et l'eau

Les changements climatiques, qui se rapportent dans le cadre de ce travail à tous changements du climat dans le temps, qu'il soit d'origine naturelle ou anthropique, risquent de bouleverser le cycle hydrologique aux échelles mondiales et régionales. En fait, ce phénomène *«devrait accentuer les facteurs actuels de stress hydriques tels que la croissance démographique, l'évolution économique et le changement d'affectation des terres»* (GIEC, 2008). Le stress hydrique supplémentaire ainsi engendré pourra avoir des conséquences importantes sur la disponibilité de la ressource en eau et affecter des secteurs tels que l'approvisionnement en eau potable, la production alimentaire, le maintien des écosystèmes naturels, la production énergétique ou encore le transport maritime.

Prévoir les conséquences des changements climatiques du prochain siècle sur le cycle hydrologique n'est pas une tâche facile mais elle est d'une importance capitale. Dans un contexte où les impacts du réchauffement climatique sont inéluctables, il devient essentiel pour l'humanité, d'une part, d'atténuer ces changements et, d'autre part, de s'y adapter. L'adaptation aux changements climatiques, par le déploiement d'une planification stratégique, peut diminuer largement la vulnérabilité des installations et des organisations humaines à moyen et à long termes. La nécessité de se doter de plans d'intervention efficaces face aux extrêmes climatiques plus fréquents (effets très probables des changements climatiques envisagés par le Groupe d'experts intergouvernemental sur l'évolution du climat, GIEC) est devenue essentielle, même pour les sociétés industrialisées. Par exemple, la vague de chaleur de l'été 2003 sur l'Europe ainsi que l'ouragan Katrina en 2005 aux États-Unis ont témoigné de la vulnérabilité des sociétés occi-

dentales face aux extrêmes climatiques. Au Québec, le dégel du pergélisol entraînant l'évacuation de villages et l'effondrement des routes, la crise du verglas et les fortes précipitations causant le «*déluge du Saguenay*» sont des événements récents qui ont démontré la vulnérabilité du Québec face aux extrêmes climatiques.

La place de l'eau au Québec est très importante non seulement pour la consommation domestique ou industrielle, mais aussi pour la voie maritime du St-Laurent ainsi que pour la production hydroélectrique. Le Québec est producteur de plus de 43 000 MW de puissance électrique d'origine hydraulique, soit plus de 93% de la puissance électrique totale<sup>1</sup> installée sur le territoire. La priorité d'action du Québec au niveau énergétique est de développer 4 500 MW supplémentaires d'origine hydroélectrique afin d'accroître l'exportation d'énergie propre.<sup>2</sup> L'hydroélectricité est 30 fois moins polluante au point de vue des émissions de gaz à effet de serre que les centrales modernes au gaz naturel et 60 fois moins polluante que les centrales au charbon. Hydro-Québec, société d'état et principal exploitant des centrales hydrauliques, soutient qu'elle fait partie de la solution à la lutte aux changements climatiques. Une question légitime concerne l'évolution de la ressource hydraulique dans un contexte de réchauffement climatique. L'analyse fréquentielle<sup>3</sup> utilisée pour dimensionner les ouvrages hydroélectriques suppose la stationnarité des séries chronologiques. Cependant, cette méthode est-elle toujours valide dans un contexte où la variabilité du climat risque de changer au cours des prochaines années? Ainsi, est-ce qu'une crue d'occurrence une fois en cent ans aujourd'hui pourrait survenir plus fréquemment dans quelques années? Ce changement hypothétique de variabilité pourrait avoir des conséquences majeures sur la sécurité des installations.

C'est pourquoi il est important d'estimer l'évolution de la ressource en eau, non seulement pour estimer le potentiel énergétique et pour dimensionner les ouvrages, mais aussi pour se prémunir des conséquences néfastes des changements climatiques. Pour ce faire, les modèles climatiques peuvent fournir des scénarios de changement climatique.

---

<sup>1</sup><http://www.mrn.gouv.qc.ca/energie/statistiques/statistiques-production-electricite.jsp>, site consulté le 21 septembre 2009

<sup>2</sup><http://www.mrn.gouv.qc.ca/energie/strategie/strategie-hydroelectricite.jsp>, site consulté le 21 septembre 2009

<sup>3</sup> «*L'analyse fréquentielle* est une méthode statistique de prédiction consistant à étudier les événements passés, caractéristiques d'un processus donné (hydrologique ou autre), afin d'en définir les probabilités d'apparition future.» (Meylan *et al.*, 2008, chap. 1)

### 1.1.1 Les modèles climatiques globaux (MCG)

Afin d'être optimale, une planification stratégique pour lutter contre les changements climatiques, élaborée par les scientifiques, les gestionnaires et les décideurs, nécessite des scénarios climatiques, c'est-à-dire des projections probables de l'ampleur des changements climatiques. Il est impossible de concevoir une Terre miniature sous une hotte de laboratoire et de vérifier les impacts d'une augmentation des gaz à effet de serre sur le cycle hydrologique. Par conséquent, les modèles climatiques numériques ont été développés afin de prédire quantitativement ces impacts. Les modèles numériques sont désignés comme les seuls outils disponibles pour obtenir des estimations quantitatives des changements climatiques futurs (GIEC, 2001). «Un *modèle climatique* est une représentation mathématique simplifiée du système climatique de la Terre» (GIEC, 2001). Les équations décrivant le climat découlent directement des processus physiques de conservation de la quantité de mouvement, de l'énergie et de la masse. La résolution de ces équations s'effectue sur une grille tridimensionnelle globale. L'utilité des modèles climatiques réside dans leur capacité à simuler les rétroactions face à l'évolution future des agents de forçage radiatifs, ce qui est impossible par simple extrapolation des courbes observées. Par exemple, suite à une augmentation des températures, il y aura probablement plus d'évapotranspiration et donc conséquemment plus de nuages (Randal *et al.*, 2007). Or, les nuages ont un effet refroidissant en réfléchissant une partie de la radiation solaire. Les modèles climatiques peuvent ainsi simuler les rétroactions connues.

Étant donné la puissance de calcul informatique limitée, les modèles climatiques globaux possèdent souvent une résolution spatiale trop grossière pour simuler adéquatement certains phénomènes à plus petite échelle comme les précipitations (Wood *et al.*, 2004). Ce manque de précision peut être problématique dans les études d'impacts régionales. Par exemple, un modèle climatique global ne pourrait pas être utilisé dans une étude d'impact sur les orages convectifs localisés dans le temps et l'espace. Barrow *et al.* (2004) estiment qu'une amélioration de la résolution des modèles (tant spatiale que temporelle) est une condition essentielle à une meilleure simulation des processus en jeu.

### 1.1.2 Les modèles climatiques régionaux (MCR)

Deux approches différentes existent pour améliorer la résolution d'un modèle climatique global, soient la mise à l'échelle statistique et la mise à l'échelle dynamique. La mise à l'échelle statistique permet de raffiner la résolution grossière d'un modèle global

par des techniques statistiques tel que la régression et les générateurs de climat. La revue de Fowler *et al.* (2007) concernent les différentes approches de mise à l'échelle à statistique pour les besoins hydrologiques. La mise à l'échelle dynamique consiste à construire un modèle climatique régional, en quelque sorte similaire aux modèles globaux, mais uniquement sur une région donnée. En limitant la région modélisée (le domaine de simulation), les résolutions spatiale et temporelle du modèle peuvent s'accroître de façon à ce que les processus à plus petite échelle puissent être simulés. Étant donné que ces modèles ne couvrent pas tout le globe, ils doivent être pilotés, c'est-à-dire que les conditions aux frontières doivent être spécifiées. Dans le cadre d'une étude de changement climatique, un modèle climatique global peut fournir les conditions aux frontières d'un modèle régional. Une revue des différentes approches de modélisation climatique est disponible dans l'article de Laprise (2008).

### 1.1.3 La validation des modèles climatiques

Avant d'utiliser un modèle climatique dans une étude d'impact, il convient d'abord de vérifier la validité des simulations résultantes. Smith et Hulme (1998) proposent de sélectionner les modèles climatiques simulant le plus fidèlement le climat actuel. Ce critère se base sur la prémisse (non vérifiable) selon laquelle les modèles climatiques sélectionnés ont plus de chance de représenter un climat futur fiable. Afin de valider un modèle, il faut tout d'abord choisir judicieusement les critères de validation.

Le cycle hydrologique régule le climat terrestre par plusieurs interactions. Watson et the Core Writing Team (2001) affirment que l'incapacité des modèles climatiques à reproduire le cycle hydrologique, les interactions primaires ainsi que les rétroactions, constitue une des principales sources d'incertitude des projections de changement climatique. Dès lors, le cycle de l'eau prédit par un modèle devrait être validé d'une part pour son rôle capital dans la physique du climat, et, d'autre part, parce que le cycle hydrologique est primordial pour tous les écosystèmes.

Toujours est-il que la validation du modèle doit s'effectuer en comparant les variables simulées aux observations météorologiques ou climatiques du climat actuel. En fait, pour pouvoir valider complètement un modèle, il faudrait comparer toutes les variables simulées aux observations, ce qui est évidemment impossible. Les flux de surface ou les chaleurs latentes et sensibles sont les variables privilégiées pour valider le cycle hydrologique d'un modèle (Music et Caya, 2007). Toutefois, les observations de telles variables sont très rares, voire inexistantes pour certaines régions. Il faut donc adapter la méthode de validation aux données disponibles (Music et Caya, 2007). À l'échelle régionale, les observations hydrologiques disponibles sont généralement les hauteurs d'eau

précipitées, enregistrées par des stations pluviométriques, et les débits de rivières aux sites des stations hydrométriques. Ces deux variables serviront de base de comparaison entre les résultats issus d'un modèle climatique et les observations hydrologiques.

Les trois sections suivantes présentent les différentes méthodes de validation des précipitations et des écoulements<sup>4</sup> issus des modèles régionaux que l'on retrouve dans la littérature.

### La validation des précipitations simulées

La comparaison entre les précipitations simulées par le modèle climatique et les observations mesurées aux stations pluviométriques n'est pas directe. En effet, la mesure d'un pluviomètre est ponctuelle dans l'espace tandis que les sorties du modèle représentent une météorologie moyenne sur la maille de simulation. Le modèle doit donc être comparé à des précipitations moyennes sur une surface, pouvant être obtenues par diverses méthodes d'interpolation. Malgré que la comparaison entre une tuile du modèle et une station pluviométrique ponctuelle n'est pas naturelle, certaines études comme celle de Rivington *et al.* (2008) utilisent encore cette approche.

De façon générale, les précipitations simulées sont comparées aux réanalyses. Les réanalyses représentent des *pseudo observations* sur tout le territoire d'intérêt. En général, ces dernières sont obtenues par une interpolation statistique des données observées sur tout le globe ou sur une région limitée. Les variables «réanalysées» sont disponibles sur une grille où la valeur d'une tuile correspond à la moyenne du champ sur cette tuile. Cette comparaison est possible soit à l'échelle mensuelle, par exemple dans l'analyse de Dankers *et al.* (2007), ou soit à l'échelle annuelle, comme dans l'étude de Kim *et al.* (2008). Ces derniers ont également considéré la capacité du modèle régional à reproduire la variabilité interannuelle des téléconnexions, c'est-à-dire les oscillations atmosphériques, notamment la téléconnexion *El Niño*.

Les précipitations simulées et observées peuvent également être agrégées par bassins versants. Le bassin versant est la superficie d'alimentation d'un cours d'eau ou d'un lac. L'eau s'écoule vers l'exutoire du bassin versant, soit le point le plus bas. La comparaison s'effectue donc sur des séries chronologiques représentant la météorologie moyenne du bassin versant. L'étude de Music et Caya (2007) a comparé les moyennes annuelles ainsi que le cycle annuel moyen des précipitations observées et simulées sur le bassin du Mississippi.

---

<sup>4</sup>Partie des précipitations qui s'écoule à la surface du sol

TAB. 1.1 – Caractéristiques des bassins versants étudiés par les études de Music et Caya (2009) et Gagnon *et al.* (2009).

|                                      | Music et Caya (2009) |           |             | Gagnon <i>et al.</i> (2009) |           |
|--------------------------------------|----------------------|-----------|-------------|-----------------------------|-----------|
|                                      | St-Laurent           | Mackenzie | Mississippi | Châteauguay                 | Chaudière |
| superficie<br>( $km^2$ )             | 1 610 000            | 1 787 000 | 3 290 000   | 2 530                       | 6 682     |
| débit annuel<br>moyen<br>( $m^3/s$ ) | 12 600               | 9 600     | 18 400      | 38                          | 132       |

### La validation des écoulements simulés

L'écoulement d'un bassin versant possède l'avantage d'intégrer spatialement et temporellement les événements météorologiques se déroulant sur son territoire. Par conséquent, les écoulements constituent une variable de comparaison privilégiée entre les simulations et les données observées. D'une part parce que la comparaison est naturelle : l'écoulement observé intègre spatialement et temporellement les événements météorologiques à la manière d'une maille du modèle. D'autre part parce que la comparaison peut s'effectuer même si les données pluviométriques sur la région sont inexistantes. Dans cette optique, Arora *et al.* (2000) ont comparé les moyennes annuelles sur cinq ans des débits simulés à ceux observés sur le Mississippi. Music et Caya (2007) ont aussi comparé les débits du Mississippi par le biais des moyennes annuelles et du cycle annuel moyen de 1988 à 1997. Dans la même démarche, Music et Caya (2009) ont validé les débits annuels moyens ainsi que le cycle annuel moyen sur trois grands bassins versants : le Mackenzie, le St-Laurent et le Mississippi pour la période 1961 à 1999. Les caractéristiques de ces trois grands bassins versants sont résumées dans le tableau 1.1<sup>5</sup>.

Les comparaisons peuvent aussi s'effectuer sur des bassins versants de plus petite superficie. Par exemple, Music *et al.* (2009) ont comparé les débits annuels moyens sur 21 bassins versants du Québec. Frigon *et al.* (2007) ont vérifié si le modèle climatique pouvait reproduire les débits annuels moyens sur un composite de 10 bassins versants du Québec. Gagnon *et al.* (2009) ont utilisé deux bassins de très petites tailles (voir le tableau 1.1) pour valider les débits simulés, soit ceux de la rivière Chaudière et de la rivière Châteauguay.

<sup>5</sup>Source : Environnement Canada, [http://www.qc.ec.gc.ca/CSL/INF/inf027\\_f.html](http://www.qc.ec.gc.ca/CSL/INF/inf027_f.html), page consultée le 20 octobre 2009.

## Le cycle de l'eau validé par un modèle hydrologique

Afin de valider le cycle de l'eau simulé par un modèle climatique régional, il est possible d'intégrer les variables simulées représentant les conditions hydro-météorologiques dans un modèle hydrologique permettant d'obtenir les débits d'un cours d'eau en un point donné. Le principal avantage de cette approche réside dans la possibilité d'obtenir le débit d'un cours d'eau. La comparaison par rapport à un site de jaugeage est donc plus directe. Cependant, cette méthode introduit une incertitude supplémentaire lors de l'étape de calibration du modèle hydrologique.

Plusieurs études utilisant cette méthodologie ont été publiées. Akhtar *et al.* (2009) ont comparé le cycle annuel moyen des précipitations et des débits journaliers simulés par un modèle hydrologique sur la région de l'Himalaya, où les conditions hydro-météorologiques étaient simulées par un MRC. Les précipitations ont été comparées aux réanalyses et les débits aux débits observés. Dankers *et al.* (2007) ont étudié si un modèle régional était apte à reproduire le temps de retour moyen des événements extrêmes sur le Danube à six stations de jaugeage. Cao *et al.* (2007) ont utilisé un modèle hydrologique pour valider les moyennes mensuelles des débits. Dans une même optique, Kay *et al.* (2006) ont vérifié si un modèle régional pouvait reproduire la fréquence observée des crues via un modèle hydrologique sur 15 bassins versants anglais. Kleinn *et al.* (2005) ont utilisé un modèle hydrologique pour valider les débits journaliers, le cycle annuel moyen et la distribution de fréquences des précipitations journalières sur le Rhin. Hay *et al.* (2002) ont utilisé les données issues d'un MRC comme intrant à un modèle hydrologique à un pas de temps journalier sur quatre bassins versants des États-Unis continental toujours dans un but de comparaison des débits simulés et observés.

Cette méthode permet aussi de valider le cycle de l'eau simulé par un MCG car le modèle hydrologique est apte à simuler les phénomènes à petite échelle (voir notamment Ducharme *et al.*, 1998, 2003).

## 1.2 Les objectifs du mémoire

Le but de ce mémoire consiste à vérifier si le Modèle Régional Canadien du Climat (MRCC, Caya et Laprise, 1999; Laprise *et al.*, 2003; Plummer *et al.*, 2006; Music et Caya, 2007; Laprise, 2008; Music et Caya, 2009) est apte à reproduire la variabilité temporelle des précipitations et des écoulements annuels observés sur le territoire québécois. L'accent dans ce projet est mis sur la significativité statistique des différences

entre les variabilités observée et simulée des écoulements et des précipitations.

Le principal objectif consiste à comparer la variabilité des données observées à celle des données simulées par le MRCC sur la base de leur coefficient de variation. Plus spécifiquement, cette comparaison s'est effectuée en utilisant les intervalles de confiance du coefficient de variation. Pour atteindre cet objectif, il a d'abord été nécessaire de prendre en considération la non normalité (au sens statistique) de certaines séries de données. Certaines séries annuelles ne pouvaient être supposées issues d'une loi normale étant donné la présence de valeurs extrêmes. Par conséquent, le second objectif du mémoire consiste à évaluer la robustesse des intervalles de confiance en présence de valeurs extrêmes dans les séries. L'évaluation de la robustesse s'est effectuée par une étude de simulation.

La structure de ce mémoire n'est pas tout à fait conventionnelle. Le chapitre concernant la présentation des résultats et celui concernant la discussion des résultats ont été séparés. Le premier de ces chapitres, le chapitre 4, expose uniquement les résultats sans tenter de les expliquer ni d'en tirer des conclusions. La discussion des résultats est présentée dans le chapitre 5 afin de permettre une discussion globale concernant la méthode, les limites de celle-ci ainsi qu'une comparaison avec d'autres approches citées dans la littérature. Au préalable, le chapitre 2, *Cadre Théorique*, expose la théorie concernant le coefficient de variation. De plus, dans ce chapitre, le choix des intervalles de confiance retenus est justifié.

# Chapitre 2

## Cadre théorique

Ce chapitre expose les fondements mathématiques et statistiques nécessaires à la compréhension de ce mémoire. Cette section a été construite à l'aide de la revue de littérature élaborée par Saïd (2009).

### 2.1 Le coefficient de variation

Le coefficient de variation constitue une statistique de dispersion relative (adimensionnelle) dans laquelle la mesure de dispersion est divisée par une mesure de tendance centrale. Cette statistique a été introduite par Pearson (1896) dans une comparaison entre la taille des organes des hommes et celles des femmes. Selon Pearson (1896), *«lorsque l'étude se base sur les mesures absolues, c'est, évidemment, paresseux de comparer la variabilité des organes mâles plus gros directement à la variabilité des organes femelles plus petits.»*<sup>1</sup> En effet, la grandeur des mesures affecte non seulement la moyenne, mais aussi l'écart à la moyenne. Depuis Pearson, cette statistique est largement utilisée dans plusieurs domaines tels la chimie, la médecine, la finance, la biologie, etc. (voir Boyer, 2007, pour une revue détaillée des utilisations du coefficient de variation) En fait, cette mesure est justifiée lorsque l'intérêt porte sur la variabilité relative et non sur la variabilité absolue.

---

<sup>1</sup>La citation exacte de Pearson (1896) : «In dealing with the comparative variation of men and women (or, indeed, very often of the two sexes of any animal), we have constantly to bear in mind that relative size influences not only the means but the deviations from the means. When dealing with absolute measurements, it is, of course, idle to compare the variation of the larger male organ directly with the variation of the smaller female organ. The same remark applies also to the comparison of large and small built races.»

**Définition 2.1.1.** *Le coefficient de variation théorique se définit par*

$$V = \frac{\sigma}{\mu}, \quad (2.1)$$

où  $\mu$  et  $\sigma^2$  sont respectivement l'espérance et la variance de la loi de probabilité  $\mathcal{F}$  sous-jacente.

Un estimateur naturel basé sur un échantillon aléatoire de taille  $n$  du coefficient de variation peut se définir comme

$$\hat{V} = \frac{S}{\bar{X}}, \quad (2.2)$$

où les variables aléatoires  $X_1, X_2, \dots, X_n$  sont indépendantes et identiquement distribuées selon une loi de probabilité  $\mathcal{F}$  d'espérance  $\mu$  et de variance  $\sigma^2$ ,

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad (2.3)$$

représente la moyenne échantillonnale et

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \quad (2.4)$$

dénote la variance échantillonnale.

Bien que cet estimateur soit le plus naturel, il est possible d'en construire d'autres. En effet, le numérateur et le dénominateur pourraient être estimés respectivement par d'autres mesures de dispersion et de tendance centrale. Dans un cadre non paramétrique notamment, la moyenne pourrait être remplacée par la médiane et l'écart-type par l'intervalle interquartile.

Dans le cadre de ce travail, nous nous attarderons à tester le caractère significatif des différences entre des coefficients de variation. Deux démarches peuvent être mises en oeuvre pour arriver à cette fin, soit par comparaison des intervalles de confiance des coefficients de variation ou soit en appliquant un test statistique d'égalité à ces derniers. Le présent mémoire abordera la comparaison des intervalles de confiance, car ceux-ci illustrent l'incertitude sur la statistique. Les résultats concernant les tests d'égalités sont fournis à titre comparatif à la section 5.4.1 du chapitre 5.

## 2.2 Les intervalles de confiance du coefficient de variation

Il est généralement plus intéressant de fournir un intervalle de valeurs plausibles d'une statistique que d'en donner une estimation ponctuelle, notamment parce que l'intervalle illustre le degré d'incertitude de la statistique.

**Définition 2.2.1.** Notons  $a$  et  $b$  des statistiques calculées à partir d'un échantillon aléatoire i.i.d.  $X_1, X_2, \dots, X_n$ . Supposons que  $\mathbb{P}(a < b) = 1$ . Pour  $\alpha \in (0, 1)$ ,  $[a, b]$  est un intervalle de confiance de niveau  $1 - \alpha$  pour un paramètre  $\theta$  de la population si

$$\mathbb{P}(\theta \in [a, b]) = \mathbb{P}(a \leq \theta \leq b) = 1 - \alpha.$$

Casella et Berger (2002) et Saporta (2006) exposent plusieurs méthodes pour obtenir de tels intervalles.

La comparaison des intervalles de confiance permet de vérifier si la différence observée entre les statistiques est significative. En effet, si les valeurs plausibles de deux statistiques issues de deux échantillons ne se répètent pas, c'est-à-dire que leurs intervalles de confiance ne se recoupent pas, les statistiques sont significativement différentes. Dans le cas contraire, si une ou plusieurs valeurs sont plausibles à la fois pour les deux statistiques, la différence n'est pas significative dans ce cas étant donné que les deux statistiques pourraient être égales.

La construction d'un intervalle de valeurs plausibles requiert la probabilité d'occurrence des valeurs de la statistique. Autrement dit, la loi de probabilité de la statistique est nécessaire à l'élaboration d'un intervalle de confiance. Les bornes  $a$  et  $b$  de la définition 2.2.1 sont déterminées par les quantiles théoriques appropriés de la loi de probabilité correspondante.

**Remarque 2.2.1.** Il est aussi possible d'utiliser des méthodes de rééchantillonnage comme le bootstrap et le Jackknife (voir Saporta, 2006, chap. 15) afin de s'affranchir de déterminer la loi de probabilité. Dans ce contexte, les intervalles de confiance obtenus sont empiriques.

Plus particulièrement, pour le coefficient de variation, la loi de probabilité de l'estimateur défini à l'équation 2.2 peut être déterminée en supposant que la loi sous-jacente correspond à une normale  $\mathcal{N}(\mu, \sigma^2)$  de moyenne  $\mu$  et de variance  $\sigma^2$ .

**Remarque 2.2.2.** L'hypothèse de normalité introduite ici est cohérente avec le cadre d'application de ce mémoire. En effet, les séries annuelles des bassins versants sont

presque toutes normalement distribuées. Un peu plus loin, aux sections 2.2.3 et 2.2.5, deux approches ne supposant pas la normalité seront présentées.

Boyer (2007) propose de calculer la distribution de l'inverse de l'estimateur défini par l'équation 2.2. En additionnant  $\mu$  et  $-\mu$  au numérateur et en se basant sur les définitions A.0.1 et A.0.2 de l'annexe A, l'inverse du coefficient de variation s'écrit

$$\frac{1}{\hat{V}} = \frac{\bar{X}}{S} = \frac{\mu + (\bar{X} - \mu)}{(\sigma/\sqrt{n-1})\sqrt{\frac{(n-1)S^2}{\sigma^2}}} = \frac{\mu + \frac{\sigma}{\sqrt{n}}Z}{\frac{\sigma}{\sqrt{n-1}}\sqrt{U}}, \quad (2.5)$$

où  $Z$  dénote une variable aléatoire  $\mathcal{N}(0, 1)$  indépendante de  $U \sim \chi_{n-1}^2$ , distribuée selon une loi du khi-deux à  $n - 1$  degrés de liberté. En divisant par  $\sigma$  et en multipliant par  $\sqrt{n}$ , l'équation se résume à celle obtenue par Johnson et Welch (1940) que voici

$$\frac{\sqrt{n}}{\hat{V}} = \frac{\sqrt{n}}{V} \frac{1}{\sqrt{\frac{U}{n-1}}} + \frac{Z}{\sqrt{\frac{U}{n-1}}} \sim t_{n-1}(\sqrt{n}/V), \quad (2.6)$$

où  $t_{n-1}(\sqrt{n}/V)$  désigne une distribution de Student à  $n - 1$  degrés de liberté de paramètre de non centralité  $\delta = \sqrt{n}/V$  (voir définition A.0.3). Le problème majeur de cette loi de probabilité de l'estimateur  $V = S/\bar{X}$  réside dans le fait que le paramètre de non-centralité  $\delta = \sqrt{n}/V$  dépend du coefficient de variation  $V$  théorique, qui est évidemment inconnu. Cette distribution est complexe, même dans le cas où la normalité de la loi sous-jacente est supposée. Plusieurs approximations de cette distribution ont été par conséquent développées afin de construire un intervalle de confiance approximatif du coefficient de variation.

La prochaine section résume quatre approches basées sur une quantité pivot pour estimer un intervalle de confiance du coefficient de variation. Une méthode basée sur la densité conjointe de la moyenne et de la variance échantillonnales est présentée à la section 2.2.2. Une approche non paramétrique reposant sur l'estimateur *Jackknife* est présentée à la section 2.2.3. La section 2.2.4 synthétise une approche basée sur le logarithme de la vraisemblance. Finalement, une approche non paramétrique est présentée à la section 2.2.5.

## 2.2.1 Les approches basées sur une quantité pivot

Afin de construire un intervalle de confiance de niveau  $(1 - \alpha)$  sur le coefficient de variation d'un échantillon  $x_1, x_2, \dots, x_n$  de taille  $n$  issu d'une loi normale  $\mathcal{N}(\mu, \sigma^2)$ , Vangel (1996) propose l'approche suivante : soit  $Y_\nu \sim \chi_\nu^2$ , où  $\nu = (n-1)$ , soit  $W_\nu = Y_\nu/\nu$

et la variable aléatoire  $Q$  définie par

$$Q = \left( \frac{\hat{V}^2}{1 + \theta \hat{V}^2} \right) \left( \frac{1 + V^2}{V^2} \right), \quad (2.7)$$

où  $\theta = \theta(\nu, \alpha)$  est une fonction supposée connue. Notons  $q = \chi_\alpha^2(\nu)$ , le quantile d'ordre  $\alpha$  de la loi de  $W_\nu$ . Étant donné que  $W_\nu$  est connue et ne dépend pas de  $V$ ,  $Q$  représente donc un pivot approximatif (voir définition A.0.4) pour la construction d'un intervalle de confiance du coefficient de variation théorique. La démarche repose sur le fait que  $\theta$  est choisi de façon à ce que la distribution de la quantité pivot soit approximativement une loi du khi-deux à  $\nu$  degrés de liberté,

$$\mathbb{P}(Q \leq t) \approx \mathbb{P}(W_\nu \leq t). \quad (2.8)$$

Pour que cette approximation soit valide, Vangel (1996) spécifie que le coefficient de variation théorique doit être suffisamment petit afin que la probabilité d'occurrence d'observations négatives soit presque nulle, de façon à ce que l'estimation  $\hat{V}$  du coefficient de variation ne soit pas négative. En effet, le support d'un khi-deux ne peut qu'être positif.

**Remarque 2.2.3.** *Pour la majorité des auteurs ayant travaillé sur le coefficient de variation, la limite supérieure correspond conventionnellement à  $V \leq 1/3$ . Dans ce cas, la probabilité qu'une observation soit négative correspond à  $\mathbb{P}(X < 0) = 0,0013$  et la probabilité que la moyenne soit négative diminue à mesure que la taille de l'échantillon augmente  $\mathbb{P}(\bar{X} < 0) = \mathbb{P}(Z \leq -3\sqrt{n})$ .*

Le problème résiduel repose sur le choix de  $\theta$ , car celui-ci déterminera la qualité de l'approximation exprimée à l'équation 2.8. Vangel (1996) propose de développer l'erreur  $e(t) = \mathbb{P}(Q \leq t) - \mathbb{P}(W_\nu \leq t)$  en série de Taylor afin d'évaluer la qualité de l'approximation en fonction de  $\theta$ . Le terme d'erreur se représente sous la forme suivante<sup>2</sup> :

$$e(t) = t f_{W_\nu}(t) \left\{ \left[ \theta t - 1 + \frac{(1-t)\nu - 1}{\nu + 1} \right] V^2 + O(V^4) \right\}, \quad (2.9)$$

où  $f_{W_\nu}(t)$  dénote la densité de  $W_\nu$ .

Quatre choix de  $\theta$  correspondant à quatre différentes approximations de la densité de  $\hat{V}$  ont été proposés dans la littérature par trois différents auteurs. Le terme d'erreur résultant de ces choix est développé dans le but de comparer les différentes approximations.

<sup>2</sup>Pour la démarche complète, se référer à l'annexe A de l'article de Vangel (1996).

## L'approximation de McKay (1932)

L'approximation de McKay (1932) consiste à poser

$$\theta = \frac{\nu}{\nu + 1}. \quad (2.10)$$

Dans ce cas particulier, l'erreur devient

$$e_{McKay}(t) = tf_{W_\nu}(t) \left\{ \frac{-2V^2}{\nu + 1} + O(V^4) \right\}. \quad (2.11)$$

**Remarque 2.2.4.** La statistique définie par McKay (1932) diffère de celle introduite à la définition 2.1.1. McKay (1932) propose comme estimateur de  $V^* = \frac{\sigma}{|\mu|}$ , l'équation suivante

$$\hat{V}^* = \frac{S_*}{|\bar{X}|}, \quad \text{où } S_*^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2. \quad (2.12)$$

McKay (1932) a obtenu cette approximation en travaillant directement avec les densités, en n'exploitant pas la quantité pivot exposée dans cette section. Pour obtenir l'approximation de la distribution, l'auteur a fait l'hypothèse que  $V^* \leq 1/3$  ce qui rejoint directement la remarque 2.2.3. Par conséquent, la valeur absolue de la moyenne peut être négligée. En remplaçant  $V^*$  par  $\frac{n}{n-1}V$  et en posant  $\theta = \frac{\nu}{\nu+1}$ , l'approximation résultante est finalement la même que celle définie à l'équation 2.8.

## L'approximation de David (1949)

David (1949) propose de poser  $\theta = 1$ . Le terme d'erreur se réduit alors à

$$e_{David}(t) = tf_{W_\nu}(t) \left\{ \frac{(t-2)V^2}{\nu + 1} + O(V^4) \right\}. \quad (2.13)$$

## L'approximation «naïve»

L'approximation «naïve», nommée et développée par Vangel (1996), consiste à poser  $\theta = 1/t$ . Vangel (1996) a étudié cette option car le calcul de l'intervalle de confiance est simple dans ce cas. En effet, l'intervalle de confiance du coefficient de variation s'obtient en divisant les bornes de l'intervalle de confiance de l'écart-type  $\sigma$  par  $\bar{X}$ . L'approximation «naïve» conduit au terme d'erreur suivant

$$e_{naïve}(t) = tf_{W_\nu}(t) \left\{ \frac{(1-t)\nu - 1}{\nu + 1} V^2 + O(V^4) \right\}. \quad (2.14)$$

### L'approximation de Vangel (1996)

Vangel (1996) suggère de fixer

$$\theta = \frac{\nu}{\nu + 1} \left( \frac{2}{\nu t} + 1 \right) \quad (2.15)$$

afin que la composante  $O(V^2)$  du terme d'erreur soit nulle, c'est-à-dire

$$e_{Vangel}(t) = t f_{W_\nu}(t) \{O(V^4)\}. \quad (2.16)$$

### Évaluation des intervalles de confiance approximatifs basés sur le pivot 2.7

Les intervalles de confiance approximatifs de McKay (1932) et ceux proposés par Vangel (1996) sont asymptotiquement exacts, c'est-à-dire qu'ils sont exacts lorsque  $n$  tend vers l'infini. De plus, en comparant les termes d'erreurs, nous obtenons les inégalités suivantes

$$\begin{aligned} e_{David}(t) &> e_{McKay}(t) \\ e_{naïve}(t) &> e_{McKay}(t) \\ e_{McKay}(t) &> e_{Vangel}(t) \end{aligned}$$

car  $\nu > 0$  et  $t > 0$ . Vangel (1996) a confirmé par simulation que l'intervalle de confiance basé sur l'équation 2.15 est plus précis que celui basé sur l'équation 2.10.

L'intervalle «naïf» dans ce mémoire est rejeté en raison de son comportement asymptotique non exact. Celui de David (1949) est aussi proscrit car son terme d'erreur est plus grand que celui de McKay (1932) et de Vangel (1996). L'intervalle de McKay (1932) est quant à lui conservé parce qu'il est abondamment utilisé dans la littérature. Celui de Vangel (1996) est aussi étudié dans ce travail puisqu'il est le plus précis.

### 2.2.2 L'approche de Hendricks et Robey (1936)

Afin de déterminer un intervalle de confiance pour le coefficient de variation, Hendricks et Robey (1936) utilisent la densité de probabilité conjointe de la moyenne et de l'écart-type échantillonnaux,

$$f_{\bar{X}, S}(\bar{x}, s) = \frac{n^{n/2} s^{n-2}}{2^{(n-2)/2} \sqrt{\pi} \sigma^n \Gamma\left(\frac{n-1}{2}\right)} \exp\left\{-\frac{n}{2\sigma^2} [(\bar{x} - \mu)^2 + s^2]\right\}, \quad (2.17)$$

déterminée par Fisher (1925) pour une population normale. Hendricks et Robey (1936) ont estimé la densité marginale de  $\hat{V} = \frac{S}{\bar{X}}$  en intégrant en coordonnées polaires et en supposant que les valeurs négatives de  $X$  et de  $\bar{X}$  sont si faibles qu'elles peuvent être négligées. La distribution est la suivante :

$$f_{\hat{V}}(v) = \frac{2}{\sqrt{\pi}\Gamma(\frac{n-1}{2})} \frac{v^{n-2}}{(1+v^2)^{n/2}} \exp\left\{-\frac{n}{2V}\left(\frac{v^2}{1+v^2}\right)\right\} \\ \times \sum_{i=0}^{n-1} \frac{(n-1)!\Gamma(\frac{n-i}{2})}{(n-1-i)!i!} \frac{n^{i/2}}{2^{i/2}V^i} \frac{1}{(1+v^2)^{i/2}}. \quad (2.18)$$

Bien que cette dernière distribution soit complexe, la démarche pour l'obtenir est similaire à celle de la section précédente, les hypothèses émises incluent un échantillon issu d'une loi normale,  $\mathbb{P}(X < 0) \approx 0$  et  $\mathbb{P}(\bar{X} < 0) \approx 0$ .

Cet intervalle de confiance ne sera pas étudié dans le cadre de ce travail étant donné la complexité de sa distribution pour être finalement développé sous les mêmes hypothèses que celles de McKay (1932) et Vangel (1996).

### 2.2.3 L'approche de Arvesen (1972)

Pour estimer la distribution du coefficient de variation d'une population issue de n'importe quelle loi, Arvesen (1972) suggère d'utiliser la technique du *Jackknife* (voir Saporta, 2006, chap. 15). Le *Jackknife* représente une technique permettant d'obtenir des intervalles de confiance approximatifs basés seulement sur l'information contenue dans l'échantillon, sans supposer de loi de probabilité sous-jacente.

**Définition 2.2.2** (La procédure Jackknife). *Soit  $x_1, x_2, \dots, x_n$  un échantillon aléatoire issu d'une certaine loi  $\mathcal{F}$  de paramètre  $\theta$  et soit  $T_n = T_n(x_1, x_2, \dots, x_n)$  un estimateur pour  $\theta$ . On calcule  $n$  statistiques, disons  $T_n^{(i)}$ ,  $i = 1, \dots, n$ , où  $T_n^{(i)}$  est obtenue comme  $T_n$  mais en excluant l'observation  $X_i$ . L'estimateur Jackknife correspond donc*

$$\hat{\theta}_{JK} = nT_n - \frac{n-1}{n} \sum_{i=1}^n T_n^{(i)},$$

où sa variance est donnée par

$$S_{\hat{\theta}_{JK}}^2 = \frac{1}{n} \sum_{i=1}^n \frac{(T_n^{(i)} - \hat{\theta}_{JK})^2}{n-1}.$$

Selon la conjecture de Tukey (1958), il est possible de construire un intervalle de confiance pour  $\theta$  à l'aide de

$$\frac{\hat{\theta}_{JK} - \theta}{S_{\hat{\theta}_{JK}}} \sim t(n-1).$$

Arvesen (1972) se base sur la définition de McKay (1932) du coefficient de variation et de son estimateur, soient  $V^* = \frac{\sigma}{|\mu|}$  et  $\hat{V}^* = \frac{S^*}{|\bar{X}|}$ . Il développe ensuite une transformation de la statistique afin que sa distribution soit plus symétrique<sup>3</sup> :

$$\beta = \log \left\{ \left( \frac{1}{V^*} \right)^2 + \left[ 2 + \left( \frac{1}{V^*} \right)^2 \right]^{1/2} \right\}. \quad (2.19)$$

La distribution de la statistique  $\beta$  est donc estimée par *Jackknife*. Arvesen (1972) a comparé l'intervalle de confiance résultant à celui de McKay (1932) par simulation. Pour 1000 échantillons aléatoires de taille 25, Arvesen (1972) montre que son intervalle est plus précis que celui de McKay (1932) pour plusieurs valeurs du coefficient de variation lorsque les échantillons sont distribués selon une loi uniforme et selon une loi double exponentielle (ou loi de Laplace). Dans le cas où les échantillons sont issus d'une loi normale, l'intervalle de confiance de McKay (1932) demeure le plus précis.

**Remarque 2.2.5.** *Les études préliminaires sur les précipitations et les écoulements annuels du Québec semblent montrer que ces séries chronologiques peuvent être supposées issues d'une loi normale pour la plupart des cas (voir la section 3.3). Dans les autres cas, le poids des queues plus élevé de la distribution empirique suggère de rejeter l'hypothèse de normalité. Cependant, le poids des queues n'est pas assez important pour supposer que les précipitations ou les écoulements annuels proviendraient d'une loi uniforme, loi pour laquelle l'approche d'Arvesen (1972) est plus précise que celle de McKay (1932). Il est par conséquent naturel de vérifier la robustesse des intervalles de confiance paramétriques lorsque l'hypothèse de normalité est rejetée puisque dans le cas normal, ce sont ces intervalles qui sont les plus précis. Dans ce cas-ci, l'approche non paramétrique de Bonett (2006) (voir la section 2.2.5), utilisant les statistiques d'ordre, a été préférée à celle d'Arvesen (1972), utilisant une méthode de rééchantillonnage.*

## 2.2.4 L'approche de Wong et Wu (2002)

Soixante-dix ans après la publication de McKay (1932), Wong et Wu (2002) déterminent un autre intervalle de confiance approximatif du coefficient de variation  $V$ . L'intervalle proposé est basé sur la fonction log-vraisemblance  $\ell(\theta) = \ell(\psi, \lambda)$ , où  $\psi$  dénote le paramètre d'intérêt et  $\lambda$  un paramètre de nuisance. Notons  $\hat{\theta} = (\hat{\psi}, \hat{\lambda})$  l'estimateur du maximum de vraisemblance de  $\theta = (\psi, \lambda)$  et  $\hat{\theta}_\psi = (\theta, \hat{\lambda}_\psi)$  l'estimateur du maximum de vraisemblance de  $\theta$  avec  $\psi$  connu. La statistique du logarithme de la vraisemblance est donc définie par

$$r(\psi) = \text{sgn}(\hat{\psi} - \psi) \left\{ 2 \left( \ell(\hat{\theta}) - \ell(\hat{\theta}_\psi) \right) \right\}^{1/2}, \quad (2.20)$$

<sup>3</sup>Arvesen (1972) ne présuppose pas une loi sous-jacente.

avec  $r(\psi) \sim \mathcal{N}(0, 1)$ .

**Remarque 2.2.6.** *Étant donné que cette approche est basée sur la vraisemblance, la loi sous-jacente doit être connue. Pour les besoins de ce travail, seul le cas normal est retenu. Il est cependant possible d'appliquer cette méthode avec n'importe quelle loi paramétrique sous-jacente.*

Wong et Wu (2002) indiquent, en se basant sur les travaux de Pierce et Peters (1992), que la convergence de cette distribution approximative est plutôt lente, de l'ordre de  $O(n^{-1/2})$ . Ils proposent, par conséquent, d'utiliser la fonction log-vraisemblance modifiée  $r^*(\psi)$  définie par Barndorff-Nielsen (1986, 1991) dont l'ordre de convergence est de  $O(n^{-3/2})$ . Une étude par simulation a permis à Wong et Wu (2002) de conclure que la méthode basée sur la fonction de vraisemblance est moins précise que celle de Vangel (1996). L'approche basée sur la fonction de vraisemblance modifiée est tout à fait comparable à celle de Vangel (1996). Par conséquent, étant donné la forte hypothèse de normalité supposée par la démarche basée sur la vraisemblance et la précision égale de l'intervalle de confiance obtenu, l'approche de Vangel (1996) est préférée.

## 2.2.5 L'approche de Bonett(2006)

Afin de construire un intervalle de confiance du coefficient de variation, Bonett (2006) s'intéresse au cas non paramétrique, c'est-à-dire en utilisant les quantiles ou des fonctions de quantiles, sans supposer de loi de probabilité sous-jacente. Bonett (2006) propose la définition suivante du coefficient de variation quartile

$$V_q = \frac{Q_3 - Q_1}{Q_1 + Q_3}, \quad (2.21)$$

où  $Q_1$  et  $Q_3$  représentent respectivement les quantiles théoriques d'ordre 25% et 75%.

**Remarque 2.2.7.** *La définition du coefficient de variation quartile n'est pas la même que celle de la définition 2.1.1. Au numérateur, l'intervalle interquartile  $Q_3 - Q_1$  constitue une mesure de dispersion alors que le dénominateur n'est pas une mesure de tendance centrale. Naturellement, on se serait attendu à un dénominateur du type  $\frac{Q_1 + Q_3}{2}$  qui caractérise la valeur centrale. Bonett (2006) ne justifie pas le choix de sa statistique.*

Pour estimer la distribution du coefficient de variation quartile, Bonett (2006) se base sur la variable aléatoire suivante

$$W = \left( \frac{n}{n-1} \right) \ln \left( \frac{\hat{Q}_3 - \hat{Q}_1}{\hat{Q}_1 + \hat{Q}_3} \right) \approx \mathcal{N}(\mu_W, \sigma_W^2), \quad (2.22)$$

où  $\mu_W$  et  $\sigma_W^2$  sont la moyenne et la variance de  $W$  et  $\hat{Q}_1$  et  $\hat{Q}_3$  dénotent les quantiles échantillonnaires d'ordre 25% et 75%. Un intervalle de confiance de  $\mu_W$  peut donc être aisément construit en procédant de façon classique

$$\mathbb{P} \left\{ -z_{1-\alpha/2} \leq \frac{W - \mu_W}{\sigma_W} \leq z_{1-\alpha/2} \right\} \approx 1 - \alpha, \quad (2.23)$$

où  $z_\alpha$  représente le quantile d'ordre  $\alpha$  de la loi normale centrée réduite.

Bonett (2006) développe l'estimation de  $\sigma_W^2$ , la variance échantillonnale de  $W$ ,  $\sigma_W^2$ , qui se définit comme suit

$$\sigma_W^2 = \hat{\text{Var}} \left\{ \ln \left( \frac{\hat{Q}_3 - \hat{Q}_1}{\hat{Q}_1 + \hat{Q}_3} \right) \right\}. \quad (2.24)$$

**Remarque 2.2.8.** *Il est à noter que l'auteur ne présente pas dans son article les équations 2.22 et 2.23. Il ne justifie pas l'approximation 2.22 suivant laquelle la variable  $W$  est distribuée selon une loi normale de paramètres  $\mu_W$  et  $\sigma_W^2$ , qui dénotent la moyenne théorique et la variance échantillonnale. De plus, l'auteur définit  $\sigma_W^2$  à l'équation 2.24 en excluant le facteur  $\frac{n}{n-1}$  sans justification.*

La première étape dans l'estimation de  $\sigma_W^2$  consiste à décomposer l'expression 2.24 de la façon suivante

$$\begin{aligned} \sigma_W^2 &= \text{Var} \left\{ \ln (\hat{Q}_3 - \hat{Q}_1) - \ln (\hat{Q}_3 + \hat{Q}_1) \right\} \\ &= \text{Var} \left\{ \ln (\hat{Q}_3 - \hat{Q}_1) \right\} + \text{Var} \left\{ \ln (\hat{Q}_3 + \hat{Q}_1) \right\} \\ &\quad - 2\text{Cov} \left\{ \ln (\hat{Q}_3 - \hat{Q}_1), \ln (\hat{Q}_3 + \hat{Q}_1) \right\}. \end{aligned}$$

La méthode delta (voir Casella et Berger, 2002, chap.5), qui consiste à estimer  $\text{Var}\{g(X)\}$  par  $[g'\{\mathbb{E}(X)\}]^2 \text{Var}(X)$ , est ensuite appliquée comme suit

$$\begin{aligned} \sigma_W^2 &\approx \frac{\text{Var}(\hat{Q}_3 - \hat{Q}_1)}{\{\mathbb{E}(\hat{Q}_3 - \hat{Q}_1)\}^2} + \frac{\text{Var}(\hat{Q}_3 + \hat{Q}_1)}{\{\mathbb{E}(\hat{Q}_3 + \hat{Q}_1)\}^2} \\ &\quad - \frac{2\text{Cov}(\hat{Q}_3 - \hat{Q}_1, \hat{Q}_3 + \hat{Q}_1)}{\mathbb{E}(\hat{Q}_3 - \hat{Q}_1) \mathbb{E}(\hat{Q}_3 + \hat{Q}_1)}. \end{aligned}$$

En utilisant les propriétés des opérateurs variance et covariance et en supposant que  $\mathbb{E}(\hat{Q}_3 - \hat{Q}_1) = Q_3 - Q_1 \approx \hat{Q}_3 - \hat{Q}_1$  et  $\mathbb{E}(\hat{Q}_3 + \hat{Q}_1) = Q_3 + Q_1 \approx \hat{Q}_3 + \hat{Q}_1$ , l'équation

devient

$$\begin{aligned}
\sigma_W^2 &\approx \frac{\text{Var}(\hat{Q}_3) + \text{Var}(\hat{Q}_1) - 2\text{Cov}(\hat{Q}_3, \hat{Q}_1)}{(\hat{Q}_3 - \hat{Q}_1)^2} \\
&+ \frac{\text{Var}(\hat{Q}_3) + \text{Var}(\hat{Q}_1) + 2\text{Cov}(\hat{Q}_3, \hat{Q}_1)}{(\hat{Q}_3 + \hat{Q}_1)^2} \\
&- 2 \left\{ \frac{\text{Var}(\hat{Q}_3) - \text{Var}(\hat{Q}_1)}{(\hat{Q}_3 - \hat{Q}_1)(\hat{Q}_3 + \hat{Q}_1)} \right\}.
\end{aligned} \tag{2.25}$$

Il suffit donc d'estimer  $\text{Var}(\hat{Q}_1)$  et  $\text{Var}(\hat{Q}_3)$  afin d'obtenir une estimation de  $\sigma_W^2$  et de l'introduire dans l'intervalle de confiance 2.23. L'estimation de la variance de ces quantiles empiriques repose sur les résultats de Stuart et Ord (2004), à savoir

$$\text{Var}(\hat{Q}_1) = \frac{3}{16nf_1^2}, \quad \text{Var}(\hat{Q}_3) = \frac{3}{16nf_3^2} \quad \text{et} \quad \text{Cov}(\hat{Q}_1, \hat{Q}_3) = \frac{1}{16nf_1f_3}, \tag{2.26}$$

où  $f_1$  et  $f_3$  représentent des «pseudo-densités» des quantiles  $Q_1$  et  $Q_3$ . Pour estimer ces densités, Bonett (2006) suggère d'utiliser les intervalles de confiance approximatifs des quartiles proposés par Wilks (1959). L'intervalle de confiance de niveau 95% de  $Q_1$  s'écrit comme  $[Y_{(a)}, Y_{(b)}]$  et celui sur  $Q_3$ ,  $[Y_{(c)}, Y_{(d)}]$ , où  $Y_{(k)}$  désigne la  $k^e$  statistique d'ordre associée à un échantillon aléatoire  $y_1, \dots, y_n$ , et

$$a = \frac{n}{4} - 1.96 \left( \frac{3n}{16} \right)^{1/2}, \tag{2.27}$$

$$b = \frac{n}{4} + 1.96 \left( \frac{3n}{16} \right)^{1/2}, \tag{2.28}$$

$$c = n + 1 - b, \tag{2.29}$$

$$d = n + 1 - a, \tag{2.30}$$

où  $a$  et  $b$  doivent être arrondis à l'entier le plus proche, avec  $a \geq 1$ . Finalement, l'auteur introduit sans justification solide le seuil suivant

$$\alpha^* = 1 - \sum_{i=a}^{b-1} \binom{n}{i} \left( \frac{1}{4} \right)^i \left( \frac{3}{4} \right)^{n-i}, \tag{2.31}$$

permettant d'estimer les variances des estimateurs des quantiles.

En passant par les intervalles de confiance approximatifs des quantiles  $Q_1$  et  $Q_3$ , Bonett (2006) parvient à estimer la variance des estimateurs des quantiles sans connaître  $f_1$  et  $f_3$ ,

$$\text{Var}(\hat{Q}_1) \approx \frac{\{Y_{(b)} - Y_{(a)}\}^2}{4z_{1-\alpha^*/2}^2} \quad \text{et} \quad \text{Var}(\hat{Q}_3) \approx \frac{\{Y_{(d)} - Y_{(c)}\}^2}{4z_{1-\alpha^*/2}^2}. \tag{2.32}$$

Les équations 2.26 et 2.32 permettent d'estimer les «densités»  $f_1$  et  $f_3$

$$\hat{f}_1^2 = \frac{3(z_{1-\alpha^*}/2)^2}{4n \{Y_{(b)} - Y_{(a)}\}^2} \quad \text{et} \quad \hat{f}_3^2 = \frac{3(z_{1-\alpha^*}/2)^2}{4n \{Y_{(d)} - Y_{(c)}\}^2}. \quad (2.33)$$

Finalement, en remplaçant  $\text{Var}(\hat{Q}_1) \approx \frac{3}{16nf_1^2}$  et  $\text{Var}(\hat{Q}_3) \approx \frac{3}{16nf_3^2}$  dans l'équation 2.25, la variance échantillonnale  $\sigma_W^2$  de  $W$  peut être estimée par l'équation suivante

$$\sigma_W^2 \approx \frac{1}{16n} \left\{ \frac{\frac{3}{\hat{f}_1^2} + \frac{3}{\hat{f}_3^2} - \frac{2}{\hat{f}_1\hat{f}_3}}{(\hat{Q}_3 - \hat{Q}_1)^2} + \frac{\frac{3}{\hat{f}_1^2} + \frac{3}{\hat{f}_3^2} + \frac{2}{\hat{f}_1\hat{f}_3}}{(\hat{Q}_3 + \hat{Q}_1)^2} - \frac{\frac{3}{\hat{f}_3^2} - \frac{3}{\hat{f}_1^2}}{(\hat{Q}_3 - \hat{Q}_1)(\hat{Q}_3 + \hat{Q}_1)} \right\}, \quad (2.34)$$

pour être intégrée de la façon suivante dans l'intervalle de confiance de  $V_q$

$$\left\{ \left( \frac{n}{n-1} \right) \ln \left( \frac{\hat{Q}_3 - \hat{Q}_1}{\hat{Q}_1 + \hat{Q}_3} \right) \pm z_{1-\alpha/2} \sigma_W \right\}. \quad (2.35)$$

L'intervalle de confiance de Bonett (2006) sera conservé dans le cadre de ce mémoire, malgré les quelques éléments nébuleux dans la démarche puisque c'est le seul intervalle de confiance non paramétrique utilisant les statistiques d'ordre d'une mesure de dispersion relative. Cet intervalle de confiance pourra être très utile lorsque l'hypothèse de normalité sera rejetée.



# Chapitre 3

## Cadre expérimental

Ce chapitre aborde la démarche méthodologique développée dans le but de déterminer si le Modèle Régional Canadien du Climat reproduit la variabilité observée des écoulements et des précipitations annuels. Tout d'abord, la description des sources de données est exposée à la section 3.1. La section 3.2 justifie le choix du critère de comparaison. Puis, la section 3.3 est consacrée aux caractéristiques des données, notamment au fait que certaines séries ne peuvent être supposées comme une suite de réalisations aléatoires d'une loi normale. Une méthode d'évaluation de la robustesse des intervalles de confiance du coefficient de variation est décrite à la section 3.4 afin de prendre en compte la non normalité des séries. La section 3.5 propose une marche à suivre pour tester la significativité des différences entre les coefficients de variation observés et simulés par le MRCC.

### 3.1 La description des données

Cette section décrit le territoire géographique d'intérêt de ce mémoire et la manière dont les données ont été obtenues pour la période de référence 1961-1999.

#### 3.1.1 L'échelle d'analyse temporelle

Dans le cadre de ce travail, les écoulements annuels moyens et les précipitations annuelles moyennes seront analysés. Cette échelle d'analyse temporelle permet de satisfaire (en théorie) aux exigences d'échantillonnage en vue de l'inférence statistique,

c'est-à-dire l'indépendance des réalisations de la variable aléatoire. En effet, à cette échelle, il est attendu qu'une observation hydrologique annuelle n'a pas d'effet sur les observations subséquentes. Les mesures journalières des écoulements seraient quant à elles corrélées en raison du temps de réponse long des bassins versants de grande superficie. Par exemple, un événement pluvieux peut avoir des répercussions sur le débit à l'exutoire sur plus d'une journée. Les observations mensuelles comporteraient, quant à elles, des cycles induits par les différentes saisons. Les séries annuelles pourraient néanmoins être corrélées dans le temps dans la mesure où les processus climatiques générant les précipitations le sont. Cette corrélation multi-annuelle n'a pas été prise en compte dans ce travail puisqu'elle affectait qu'une minorité des séries chronologiques à l'étude (consulter le tableau B.1 de l'annexe B). Ce point est discuté plus en profondeur à la section 5.3.2.

Par ailleurs, la démarche s'inscrit dans un cadre climatique : on souhaite vérifier si un modèle climatique peut reproduire la variabilité observée de la climatologie actuelle, pas la météorologie observée. En effet, ce n'est pas le rôle d'un modèle climatique de prédire le débit observé à l'exutoire le 5 février 1985<sup>1</sup> ! Le but d'un modèle climatique consiste à produire des valeurs d'une variable qui sont représentatives par leur grandeur, leur variabilité et leur synchronisation, et ce, à n'importe lequel moment de l'année. L'étude sur une base annuelle est par conséquent pertinente dans un contexte climatique, où ce sont les tendances climatiques à long terme qui doivent être extraites.

### 3.1.2 Le territoire étudié

Le territoire québécois constitue la région d'intérêt dans le cadre de ce travail. La région à l'étude est composée de 21 bassins versants. «*Le bassin versant représente l'unité spatiale de référence pour l'hydrologie. Il est défini comme l'étendue drainée par un cours d'eau et ses affluents et limitée par une ligne de partage des eaux.*» (Musy et Higy, 2004, chap. 3) Autrement dit, le bassin versant représente une unité hydrologique fermée, c'est-à-dire qu'aucun écoulement ne provient de l'extérieur du bassin et l'excédent des précipitations s'écoule par une seule sortie, l'exutoire. La superficie d'un bassin versant est déterminée par la ligne de partage des eaux qui correspond le plus souvent à la ligne de crête des montagnes, des collines et des points hauts pour les sols non poreux. Dans le domaine de la gestion de l'eau, le bassin versant constitue l'unité d'analyse fondamentale qui s'impose d'elle-même étant donné que le bassin versant intègre toutes les activités, naturelles et humaines, se produisant sur son territoire. En ce sens, la végétation, l'occupation des terres, l'aménagement du territoire caractérisent le

---

<sup>1</sup>La date de naissance de ma fiancée

bassin versant.

Cette échelle d'analyse spatiale permet de relier presque directement les écoulements simulés sur la superficie du bassin versant au débit mesuré à l'exutoire ; «presque directement» parce que

- les différentes tuiles et fractions de tuiles du MRCC doivent être agrégées par bassin versant ;
- les débits mesurés à l'exutoire doivent être corrigés afin de retirer l'impact des constructions humaines ;
- le MRCC ne modélise pas tous les phénomènes physiques et hydrologiques.

Parfois, pour comparer directement le débit calculé par un jaugeage à un débit simulé, les variables du modèle climatique sont insérées dans un modèle hydrologique permettant d'estimer le débit simulé à la localisation du jaugeage. Toutefois, l'approche exploitant le bassin versant (plus particulièrement le débit à l'exutoire) permet de ne pas introduire un modèle hydrologique dans la comparaison entre les écoulements simulés et les débits mesurés. En effet, l'addition de tous les écoulements de surface simulés par le MRCC à l'intérieur des frontières d'un bassin versant correspond au débit simulé à l'exutoire de ce bassin. Toutefois, si l'analyse se destinait à une plus petite échelle temporelle que l'échelle annuelle, ce qui n'est pas le cas de ce travail, il faudrait considérer l'acheminement de l'eau dans le réseau hydrographique. Dans ce cas, un modèle hydrologique serait sans doute nécessaire pour comparer les débits des écoulements simulés aux débits observés concomitants.

Le tableau 3.1 résume les principales caractéristiques des bassins versants à l'étude. On constate que leur superficie varie de  $13\ 000\ km^2$  (9 points de grille du MRCC) à  $177\ 000\ km^2$  (91 points de grille du MRCC).

### 3.1.3 Les écoulements observés

Les écoulements observés des 21 bassins versants québécois proviennent de trois sources différentes, soient Hydro-Québec, le Service d'Hydrométrie du Ministère du Développement Durable, de l'Environnement et des Parcs (MDDEP) et de l'aluminerie Rio Tinto Alcan inc. Ces organismes ont corrigé les écoulements observés pour estimer les écoulements naturels à l'exutoire en enlevant les impacts des infrastructures humaines (par exemple les barrages hydroélectriques). Cette manipulation, bien que sujette aux erreurs, est nécessaire afin de comparer les écoulements simulés à ceux observés car les ouvrages de rétention d'eau ne sont pas modélisés dans le Modèle Régional Canadien du Climat. Haché *et al.* (1996) ont validé ces écoulements naturels historiques dans la

TAB. 3.1 – Liste et caractéristiques des bassins versants du Québec à l'étude (à noter que l'acronyme MDDEP désigne le Ministère du Développement Durable, de l'Environnement et des Parcs).

| bassin versant<br>et acronyme      | source des<br>données | superficie<br>(km <sup>2</sup> ) | nombre de points<br>de grille du MRCC | nombre<br>d'observations |
|------------------------------------|-----------------------|----------------------------------|---------------------------------------|--------------------------|
| Rivière Arnaud (ARN)               | MDDEP                 | 26900                            | 14                                    | 20                       |
| Rivière à la Baleine (BAL)         | Hydro-Québec          | 29000                            | 17                                    | 36                       |
| Rivière Bell (BEL)                 | MDDEP                 | 22200                            | 15                                    | 36                       |
| Bersimis-Outardes-Manic (BOM)      | Hydro-Québec          | 87000                            | 47                                    | 39                       |
| Réservoir Caniapiscou (CAN)        | Hydro-Québec          | 37870                            | 23                                    | 39                       |
| Réservoir Churchill Falls (CHU)    | Hydro-Québec          | 69300                            | 34                                    | 39                       |
| Rivière aux Feuilles (FEU)         | MDDEP                 | 41700                            | 22                                    | 23                       |
| Rivière Georges (GEO)              | MDDEP                 | 24200                            | 11                                    | 30                       |
| Grande Rivière de la Baleine (GRB) | Hydro-Québec          | 36300                            | 18                                    | 39                       |
| La Grande Rivière (LGR)            | MDDEP                 | 177000                           | 91                                    | 39                       |
| Réservoir Manic5 (MAN)             | Hydro-Québec          | 29240                            | 17                                    | 39                       |
| Rivière aux Mélèzes (MEL)          | MDDEP                 | 42700                            | 22                                    | 29                       |
| Rivière Moisie (MOI)               | MDDEP                 | 19000                            | 12                                    | 33                       |
| Rivière Natashquan (NAT)           | MDDEP                 | 15600                            | 9                                     | 37                       |
| Rivière Caniapiscou (Pyrite) (PYR) | MDDEP                 | 48500                            | 24                                    | 17                       |
| Rivière des Outaouais (RDO)        | Hydro-Québec          | 143000                           | 80                                    | 31                       |
| Rivière Romaine (ROM)              | MDDEP                 | 13000                            | 9                                     | 39                       |
| Rivière Rupert (RUP)               | MDDEP                 | 40900                            | 22                                    | 37                       |
| Lac Saint-Jean (SAG)               | Rio Tinto Alcan Inc.  | 73000                            | 43                                    | 39                       |
| Rivière Saint-Maurice (STM)        | Hydro-Québec          | 47200                            | 28                                    | 28                       |
| Rivière Waswanipi (WAS)            | MDDEP                 | 31900                            | 16                                    | 32                       |

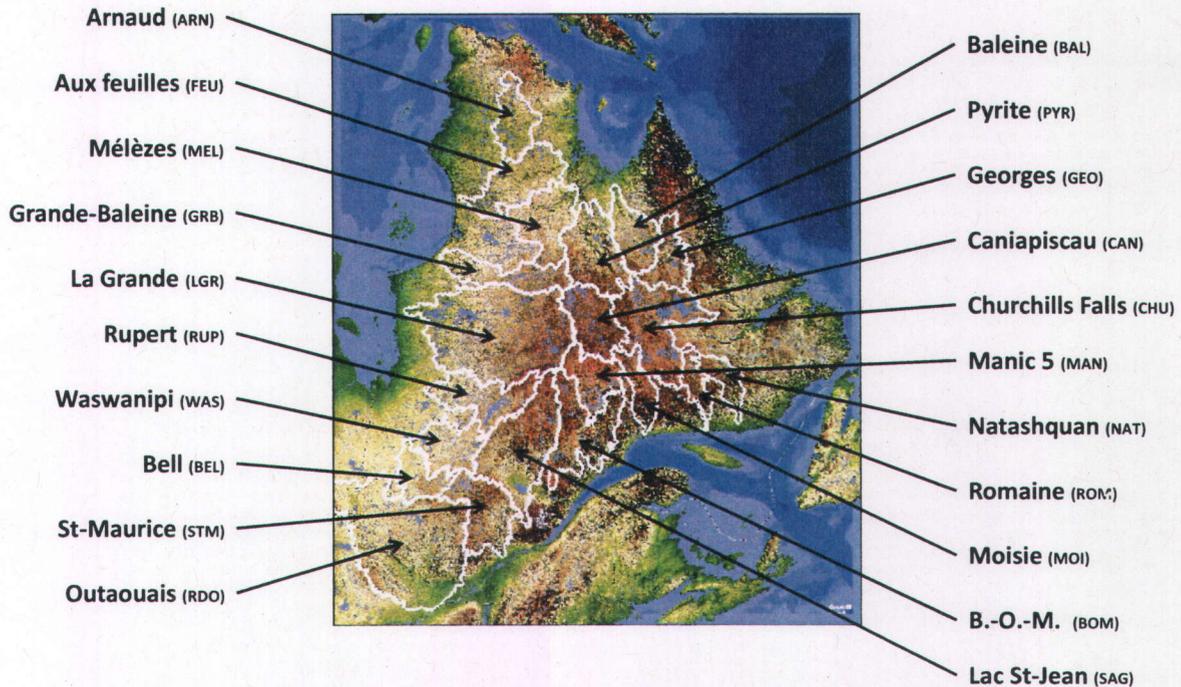


FIG. 3.1 – Carte géographique des 21 bassins versants étudiés.

première phase de la Chaire industrielle en hydrologie statistique. Pour estimer le débit à l'exutoire d'un bassin versant, plusieurs sites jaugés à l'intérieur de ses frontières sont divisés par la surface du sous-bassin versant (Roads *et al.*, 2003). Cette approche néglige néanmoins les écoulements souterrains en ne considérant que les écoulements en rivière. Par ailleurs, les séries chronologiques des écoulements ne sont pas toutes complètes sur la période d'analyse (1961 à 1999) pour certains des bassins versants étudiés (consulter le tableau 3.1). La figure 3.2 illustre les débits annuels observés des 21 bassins versants à l'étude sous la forme de diagramme en boîte en parallèle.

### 3.1.4 Les précipitations réanalysées

#### La réanalyse CRU

La réanalyse du *Climatic Research Unit* (CRU, Mitchell et Jones, 2005) comprend une base de données de six variables climatiques mensuelles interpolées sur une grille polaire stéréographique (coordonnées en degrés de latitude et longitude) de résolution 0,5° couvrant la totalité de la surface terrestre à l'exception de l'Antarctique. Les données observées aux stations météorologiques sont d'abord analysées afin de vérifier leur homogénéité. Cette vérification s'effectue sur une station en particulier en construisant

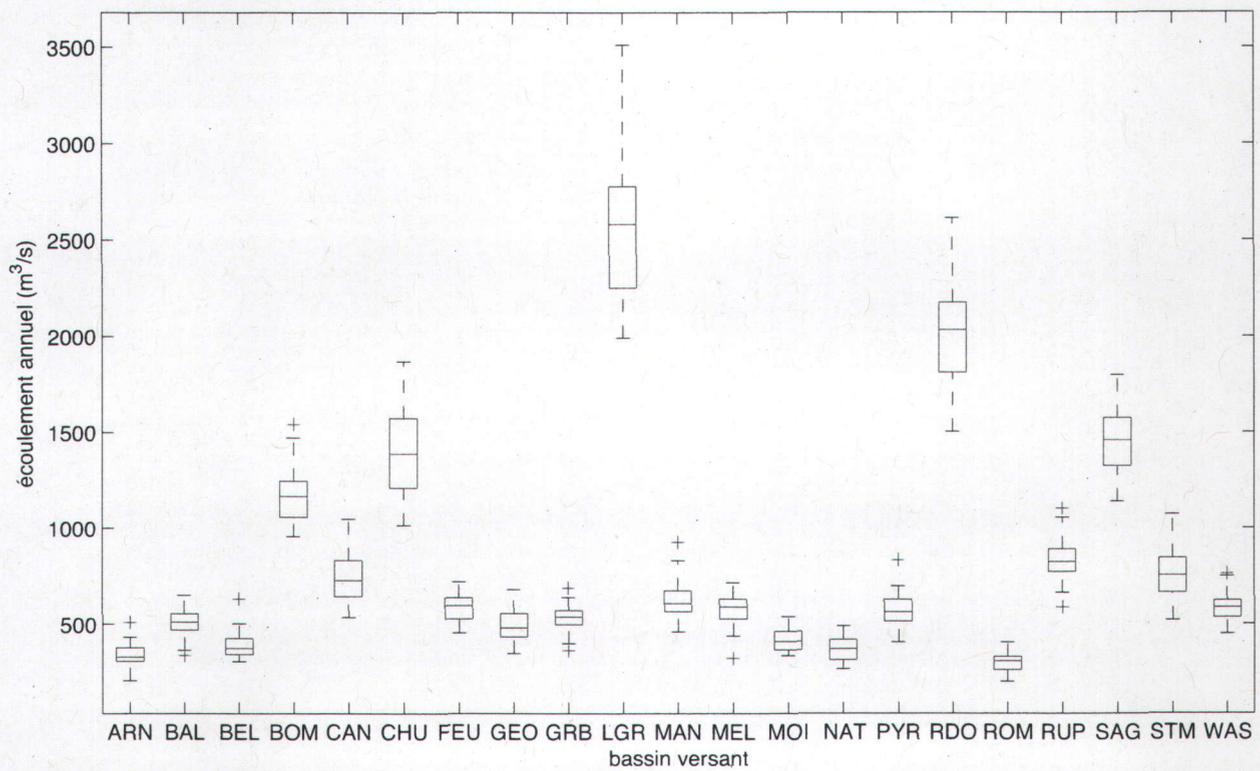


FIG. 3.2 – Diagramme en boîte des débits annuels observés pour les 21 bassins versants à l'étude.

une série climatique de référence à l'aide des stations voisines. Par la suite, l'homogénéité est vérifiée par une analyse de corrélation appliquée aux séries des premières différences de la station et de la série de référence. Le processus d'homogénéisation est itératif car l'homogénéité des stations servant à la construction de la série de référence doit aussi être vérifiée. L'interpolation des données mensuelles observées des précipitations s'effectue sur les anomalies relatives des données, c'est-à-dire la différence relative entre les valeurs observées et les normales de la période 1961-1990. Les observations extrêmes sont retirées de la série et elles sont par la suite traitées comme des valeurs manquantes. Dans les régions où les stations sont rares voire inexistantes, des stations fictives ont été ajoutées dont les observations ont une anomalie relative nulle de façon à ce que l'information des vraies stations ne soit pas extrapolée au-delà d'une certaine distance.

### La réanalyse WM2

Cette réanalyse utilise une interpolation spatiale par l'inverse de la distance pour obtenir des prédictions des températures et des précipitations moyennes mensuelles sur une grille latitude/longitude mondiale ayant une résolution de  $0,5^\circ$  (Willmott et Matsuura, 1995). Les observations d'une vingtaine de stations météorologiques ont été utilisées afin d'obtenir la prédiction par interpolation au centre d'une tuile de la grille. Pour les températures moyennes mensuelles, l'effet de l'altitude des stations a été retiré avant d'appliquer l'algorithme d'interpolation. L'effet de l'altitude est par la suite ajouté aux prédictions. Les précipitations ont été interpolées en utilisant une technique développée par Willmott et Robeson (1995), le *Climatologically Aided Interpolation (CAI)*. Les précipitations enregistrées aux stations ne sont pas interpolées directement. Auparavant, une valeur moyenne des précipitations au site de la station est retirée. Cette moyenne est obtenue soit par interpolation des stations voisines, soit en utilisant directement la moyenne des stations voisines si la densité de station est grande. Par la suite, l'algorithme d'interpolation est appliqué sur les différences entre les précipitations enregistrées et les moyennes. Finalement, la moyenne est ajoutée aux données interpolées.<sup>2</sup>

---

<sup>2</sup>Cette sous-section a été rédigée entre autres à l'aide du document explicatif de Cort J. Willmott et Kenji Matsuura, disponible en ligne : [http://climate.geog.udel.edu/~climate/html\\_pages/README.ghcn\\_clim2.html](http://climate.geog.udel.edu/~climate/html_pages/README.ghcn_clim2.html) ; page consultée le 21 octobre 2009.

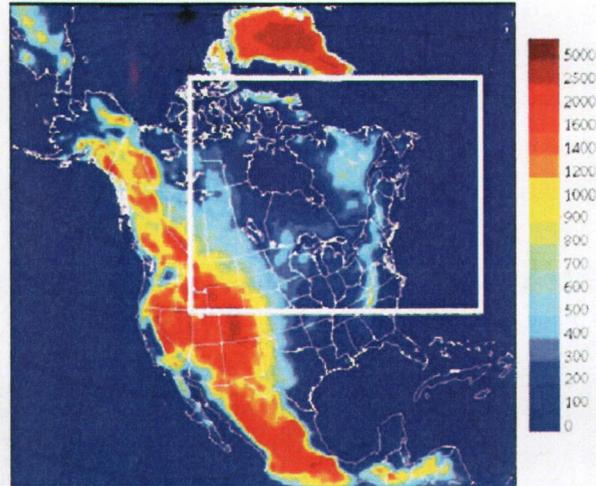


FIG. 3.3 – Le grand domaine AMNO et le petit domaine QC (encadré blanc) utilisés dans les simulations du MRCC. La carte illustre la topographie en mètre de cette région. Tiré de Music *et al.* (2009).

### 3.1.5 Les écoulements et précipitations simulés

Le Modèle Régional Canadien du Climat (MRCC, Caya et Laprise, 1999; Laprise *et al.*, 2003; Plummer *et al.*, 2006; Music et Caya, 2007; Laprise, 2008; Music et Caya, 2009) constitue un modèle climatique à haute résolution à aire limitée piloté par des modèles climatiques globaux (planétaires) ou des réanalyses à plus faibles résolutions. Un modèle régional comme le MRCC permet d'améliorer la faible résolution spatiale d'un modèle global sur une région limitée en modélisant et en intégrant les détails à plus petite échelle. Le MRCC possède une grille horizontale uniforme de projection polaire stéréographique de résolution de 45 km à 60°N. Les données du MRCC disponibles au Québec peuvent être fournies pour deux domaines de simulation : un grand domaine (AMNO) couvrant la totalité de l'Amérique du Nord et un plus petit domaine (QC) centré sur le Québec couvrant les prairies canadiennes et le Nord-Est des États-Unis (figure 3.3). Étant donné que le modèle exécute une simulation sur un domaine limité, soit AMNO ou QC, les conditions limites latérales et inférieures doivent être spécifiées : le modèle régional est piloté ou emboîté dans un modèle global. La procédure de pilotage consiste, entre autres, à spécifier aux limites latérales du modèle régional les valeurs de la pression, de la température, de la vapeur d'eau et du vent horizontal du pilote. Dans une zone tampon périphérique à la frontière du domaine, les champs atmosphériques du modèle sont progressivement mélangés à ceux du pilote jusqu'à atteindre une zone libre où les données du pilote n'influencent plus directement celles simulées.

La version 4.1.1 du MRCC, celle considérée dans la présente étude, inclut le schéma

de surface multicouche *Canadian Land Surface Scheme* (CLASS 2.7, Verseghy, 1991; Verseghy *et al.*, 1993). Ce schéma de surface détermine les échanges d'énergie et d'eau entre le sol et l'atmosphère. Les variables hydrologiques issues du modèle sont calculées par bilan atmosphérique et terrestre. Plus particulièrement, les écoulements sont obtenus par la mise en adéquation du bilan atmosphérique et du bilan hydrique, sans recourir à un modèle hydrologique. Les équations du bilan sont disponibles dans l'article de Music et Caya (2007). Au niveau du sol, le schéma de surface est divisé en trois couches : une couche superficielle de 10 cm, une couche intermédiaire racinaire de la végétation de 25 cm et une couche profonde de 3,75 m. L'hiver, le couvert de neige est traité de façon distincte thermiquement et hydrologiquement, comme une quatrième couche de sol d'épaisseur variable. CLASS résout les équations de conservation de l'énergie et de conservation de la masse d'eau pour ces couches de sol, entre autres pour calculer l'écoulement de chacune des mailles du modèle. Dans CLASS, l'écoulement total est composé de l'écoulement de surface et de l'écoulement souterrain. L'écoulement de surface est généré si la capacité d'infiltration du sol est surpassée. La capacité d'infiltration varie en fonction du type de végétation modélisée dans CLASS. Cependant, la physique et la rétention d'eau des lacs ne sont pas modélisées dans le schéma de surface CLASS, ce qui aurait pu être problématique si l'analyse s'effectuait sur des données à l'échelle infra-annuelle. En effet, les lacs occupent une place importante dans les bassins versants québécois, soit de 5% à 30 % de la surface des bassins versants étudiés.

La simulation acy générée par le MRCC est celle qui sera étudiée dans le cadre de ce projet. Cette simulation pilotée par la réanalyse planétaire ERA40 (Uppala *et al.*, 2005) sur le petit domaine de simulation QC (112 x 88 points de grilles) sera comparée aux observations par bassins versants pour la période s'échelonnant entre 1961 et 1999. La série chronologique d'un bassin versant est obtenue par l'agrégation de tous les points de grille du MRCC situés à l'intérieur de sa frontière.

### 3.2 Le critère de comparaison choisi

Afin de déterminer si les séries étudiées ont la même variabilité, le coefficient de variation, a été retenu puisqu'il est adimensionnel. Dans le contexte particulier des sciences de l'eau, plusieurs caractéristiques influencent la quantité d'eau sortante à l'exutoire d'un bassin versant, par exemple

- la taille du bassin versant,
- la quantité de pluie tombée sur le bassin,
- la quantité d'eau évapotranspirée,

- etc.

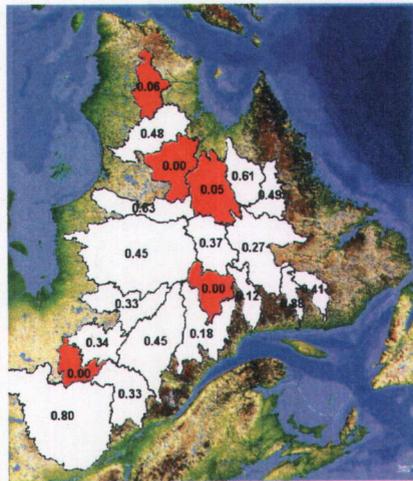
Chaque bassin versant présente des caractéristiques (taille, végétation, type de sol, occupation du territoire, etc.) qui déterminent les écoulements. De plus, les bassins sont soumis à des régimes de précipitations inhomogènes, c'est-à-dire que les pluies peuvent être plus abondantes à des altitudes et des latitudes plus élevées. Le coefficient de variation permet de comparer la variabilité relative de deux bassins versants ayant des caractéristiques de tailles, de végétations et de régimes de précipitations différentes. L'utilité de celui-ci dans le contexte de ce mémoire est de déterminer si la variabilité relative simulée par le MRCC corrobore la variabilité observée même si la modélisation imparfaite des caractéristiques des bassins versants conduit à une série simulée biaisée. Autrement dit, le coefficient de variation pourra déterminer si le MRCC reproduit bien la variabilité relative même si les séries simulées sont biaisées. Le coefficient de variation peut aussi servir à comparer la variabilité relative de deux bassins versants d'une même région de tailles différentes.

### 3.3 Les caractéristiques des données

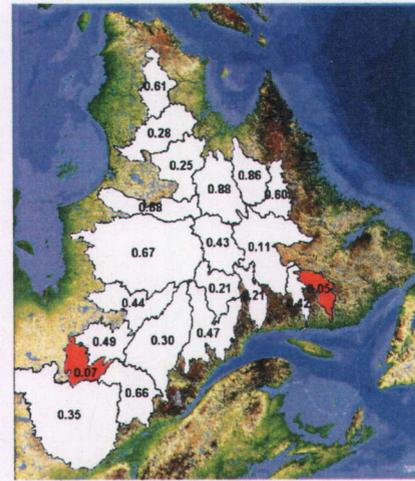
Pour tester la normalité des données utilisées, un test basé sur la symétrie et l'aplatissement de la loi normale (D'Agostino *et al.*, 1990) a été appliqué. Les résultats de ce test, illustrés à la figure 3.4, montrent que les séries annuelles des écoulements et des précipitations de certains bassins versants ne sont pas distribuées selon des lois normales au seuil de 0, 10. Pour les écoulements observés et simulés, ce ne sont pas exactement les mêmes bassins versants qui ne passent pas le test de normalité. Le même comportement se retrouve aussi pour les précipitations simulées et réanalysées.

**Remarque 3.3.1.** *Les séries chronologiques des deux sources de pseudo-observations des précipitations, les réanalyses CRU et WM2, qui ne sont pas distribuées selon une loi normale, ne sont pas les mêmes. Pourtant, ces séries sont censées représenter la réalité, c'est-à-dire la météorologie observée. Elles devraient donc être pratiquement semblables. Toutefois, les causes de cette divergence des réanalyses seront discutées à la section 5.2.2.*

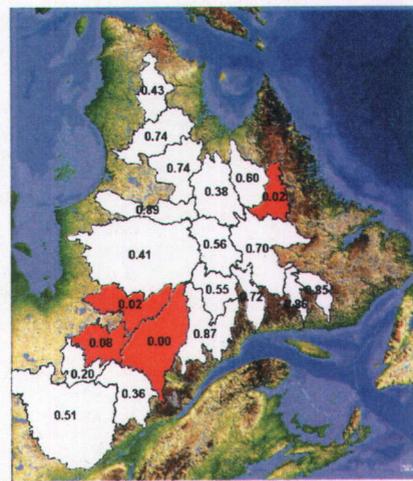
Une étude plus approfondie a permis de déterminer la cause de la non normalité des séries annuelles des écoulements et des précipitations. En fait, les séries qui ne sont pas distribuées selon une loi normale possèdent une, deux, rarement trois et exceptionnellement quatre valeurs extrêmes qui contaminent les séries. Le tableau 3.2 compile le nombre de valeurs extrêmes des différentes sources de données pour les écoulements et



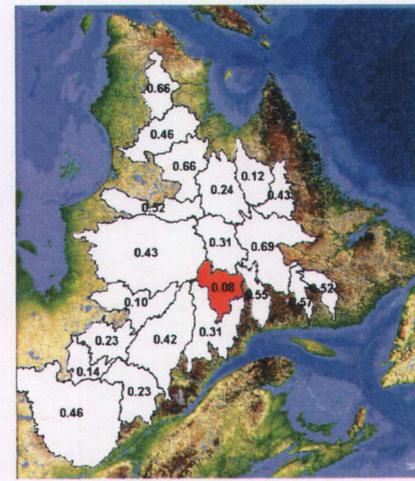
(a) écoulements observés



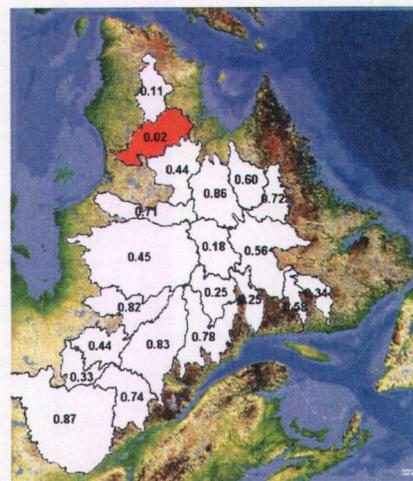
(b) écoulements simulés



(c) précipitations CRU

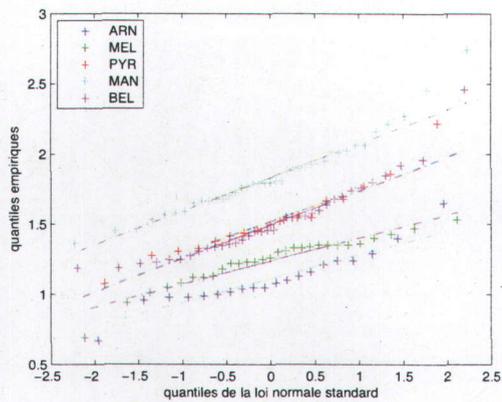


(d) précipitations WM2

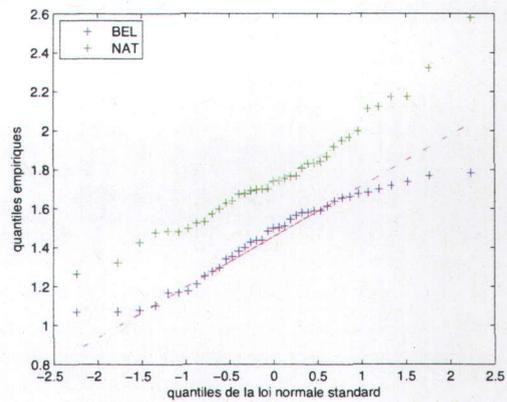


(e) précipitations simulées

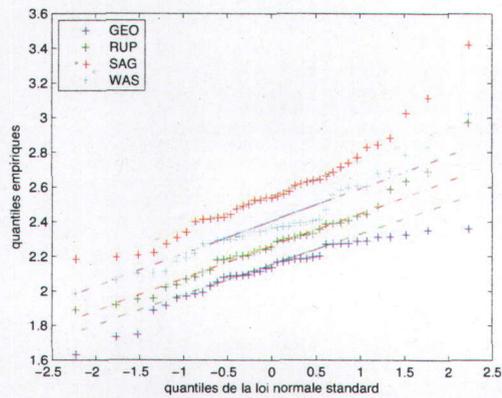
FIG. 3.4 – P-valeurs du test de normalité pour chacune des sources de données. Les bassins en rouge sont ceux dont le test suggère de rejeter l'hypothèse de normalité au seuil de 10%.



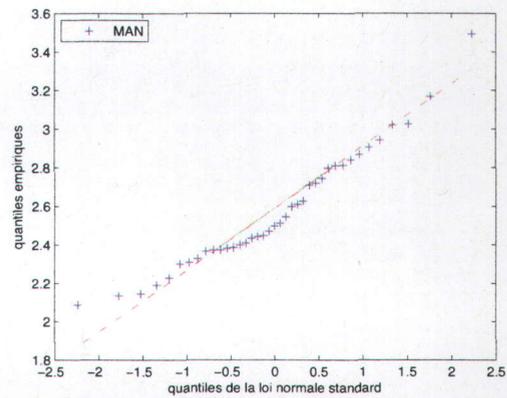
(a) écoulements observés



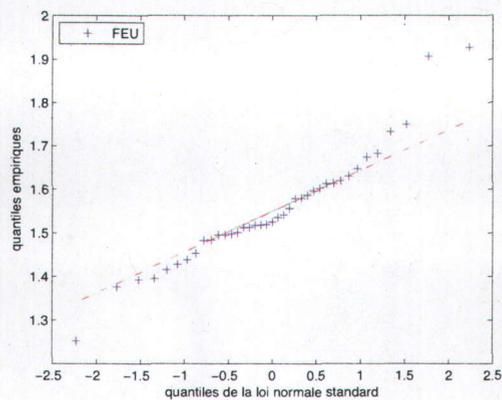
(b) écoulements simulés



(c) précipitations CRU



(d) précipitations WM2



(e) précipitations simulées

FIG. 3.5 – Droites de Henry ( $Q-Q$  plots) pour chacune des séries ne pouvant être supposées normales.

les précipitations ainsi que le nombre d'observations des séries des écoulements observés. Les valeurs extrêmes sont définies en utilisant la définition des valeurs aberrantes d'un diagramme en boîte, c'est-à-dire que  $x_i$  est une valeur extrême si

- $x_i < Q_1 - 1,5(Q_3 - Q_1)$  ;
- $x_i > Q_3 + 1,5(Q_3 - Q_1)$  ;

où  $Q_1$  et  $Q_3$  sont les quantiles empiriques d'ordre 25% et 75%. Si les extrêmes étaient retirés, ces séries pourraient être supposées issues d'une loi normale selon le test de D'Agostino *et al.* (1990). Le tableau B.2 de l'annexe B résume plusieurs statistiques des sources de données : le coefficient de variation  $V$ , l'asymétrie (skewness) et l'aplatissement (kurtosis) pour tous les bassins versants et pour toutes les sources de données. De plus, ces tableaux résumés également les p-valeurs brutes d'un test de normalité basé sur l'asymétrie et l'aplatissement de la loi normale (D'Agostino *et al.*, 1990) et les p-valeurs des séries où les valeurs extrêmes ont été retirées. On constate que lorsque les valeurs extrêmes sont retirées, les séries peuvent toutes être assumées issues de lois normales.

Pour confirmer l'hypothèse selon laquelle les séries ne sont pas distribuées normalement, les droites de Henry (*Q-Q plots*) ont été tracées à la figure 3.5 pour chacune des séries ne pouvant être supposée normale. Cette dernière figure montre clairement que le rejet de l'hypothèse de normalité est causé par la présence de valeurs extrêmes (excepté pour la série des écoulements simulés du bassin de la rivière Bell). Les écoulements moyens suivent généralement très bien une droite.

Compte tenu du fait que les données ne peuvent pas être toutes supposées normales en raison de la présence de quelques valeurs extrêmes, les intervalles de confiance retenus sont-ils applicables à ce genre de séries ? Pour vérifier la robustesse des intervalles de confiance aux valeurs extrêmes, une étude par simulation de Monte-Carlo peut être effectuée.

### 3.4 La méthode de Monte-Carlo

Les simulations de Monte-Carlo (voir Saporta, 2006, chap. 15) dans le contexte de ce travail permettent d'étudier les propriétés des intervalles de confiance lorsque des séries normales sont contaminées par des extrêmes. Les étapes méthodologiques d'une étude par simulation se résument ainsi :

1. Générer un très grand nombre d'échantillons aléatoires issus d'une loi connue.
2. Effectuer les calculs statistiques pour chacun des échantillons aléatoires.

TAB. 3.2 – Dénombrement des valeurs extrêmes telles que définies précédemment parmi les différentes sources de données pour les écoulements et les précipitations annuels.

| 1961<br>1999 | écoulements observés          |                          |                      | mrcc éc.             | mrcc acy pcp         | wm2 pcp              | cru pcp              |
|--------------|-------------------------------|--------------------------|----------------------|----------------------|----------------------|----------------------|----------------------|
|              | plus longue<br>série continue | nombre<br>d'observations | nombre<br>d'extrêmes | nombre<br>d'extrêmes | nombre<br>d'extrêmes | nombre<br>d'extrêmes | nombre<br>d'extrêmes |
| ARN          | 1963-1982 (20)                | 20                       | 1                    | 2                    | 3                    | 1                    | 0                    |
| BAL          | 1963-1998 (36)                | 36                       | 1                    | 1                    | 0                    | 1                    | 2                    |
| BEL          | 1963-1996 (34)                | 36                       | 1                    | 0                    | 0                    | 0                    | 0                    |
| BOM          | 1961-1999 (39)                | 39                       | 1                    | 0                    | 0                    | 0                    | 0                    |
| CAN          | 1961-1999 (39)                | 39                       | 0                    | 0                    | 0                    | 0                    | 4                    |
| CHU          | 1961-1999 (39)                | 39                       | 0                    | 2                    | 0                    | 0                    | 1                    |
| FEU          | 1963-1982 (20)                | 23                       | 0                    | 2                    | 3                    | 0                    | 0                    |
| GEO          | 1972-1987 (16)                | 30                       | 0                    | 2                    | 0                    | 0                    | 1                    |
| GRB          | 1961-1999 (39)                | 39                       | 2                    | 0                    | 0                    | 0                    | 0                    |
| LGR          | 1961-1999 (39)                | 39                       | 0                    | 0                    | 0                    | 0                    | 2                    |
| MAN          | 1961-1999 (39)                | 39                       | 1                    | 0                    | 0                    | 1                    | 1                    |
| MEL          | 1963-1981 (19)                | 29                       | 1                    | 2                    | 4                    | 0                    | 0                    |
| MOI          | 1968-1999 (32)                | 33                       | 0                    | 1                    | 0                    | 0                    | 0                    |
| NAT          | 1963-1999 (37)                | 37                       | 0                    | 1                    | 1                    | 0                    | 0                    |
| PYR          | 1982-1998 (17)                | 17                       | 1                    | 0                    | 4                    | 2                    | 1                    |
| RDO          | 1969-1999 (31)                | 31                       | 0                    | 0                    | 0                    | 0                    | 0                    |
| ROM          | 1961-1999 (39)                | 39                       | 0                    | 0                    | 0                    | 0                    | 0                    |
| RUP          | 1961-1997 (37)                | 38                       | 2                    | 0                    | 0                    | 0                    | 1                    |
| SAG          | 1961-1999 (39)                | 39                       | 0                    | 0                    | 0                    | 0                    | 2                    |
| STM          | 1966-1993 (31)                | 28                       | 0                    | 0                    | 1                    | 0                    | 1                    |
| WAS          | 1968-1999 (32)                | 32                       | 1                    | 1                    | 2                    | 0                    | 1                    |

### 3. Comparer les propriétés échantillonnales empiriques à celles théoriques.

L'étape 1 consiste à générer des échantillons aléatoires issus d'une loi connue. Dans le présent cas, la robustesse des intervalles de confiance doit être testée lorsque les données s'écartent de la normalité à cause de la présence de valeurs extrêmes. Une loi de probabilité théorique d'un tel comportement doit d'abord être identifiée. Tukey (1960) propose un modèle permettant d'augmenter le poids des queues d'une distribution normale  $\mathcal{N}(\mu, \sigma^2)$ . Ce modèle s'exprime sous la forme d'un mélange de lois entre la distribution normale et une distribution normale contaminante de même moyenne mais de variance  $(b\sigma)^2$ , pouvant s'écrire

$$M \sim \{1 - \lambda\}\mathcal{N}(\mu, \sigma^2) + \lambda\mathcal{N}(\mu, (b\sigma)^2) \quad (3.1)$$

où  $0 \leq \lambda \leq 1$  est la pondération de la normale contaminante.

**Remarque 3.4.1.** *Le poids des queues d'une distribution révèle en fait la probabilité d'observer des données éloignées des valeurs centrales. La dispersion d'un échantillon issu d'une distribution possédant des queues lourdes est plus grande que la dispersion d'un échantillon provenant d'une loi aux queues légères. Autrement dit, la probabilité d'observer des valeurs extrêmes issues d'une distribution aux queues lourdes est supérieure à celle d'une distribution aux queues légères. Une statistique permet de décrire le poids des queues par rapport à la valeur centrale. Cette statistique se nomme le coefficient d'aplatissement (aussi appelée le kurtosis). Celle-ci fut introduite par Karl Pearson (1857-1936) dans une série d'articles parue entre 1894 et 1905 (Fiori et Zenga, 2009).*

Le poids des queues du mélange  $M$  peut être déterminé par le coefficient d'aplatissement du mélange. Le kurtosis d'une distribution se définit comme le rapport entre le quatrième moment centré et la variance au carré :

$$\gamma = \frac{\mu_4}{\sigma^4},$$

où  $\mu_4$  est le quatrième moment centré défini par  $\mu_4 = \mathbb{E}(X^4)$ . Le coefficient d'aplatissement de la loi normale est de 3.

Les paramètres du mélange de lois  $M$  sont ajustés de façon à reproduire les différentes caractéristiques  $V_M$  et  $\gamma_M$  fixées. La fonction génératrice des moments du mélange s'exprime comme une pondération des fonctions génératrices des moments de chacune des composantes du mélange :

$$M_M(t) = (1 - \lambda)e^{t\mu + \frac{t^2\sigma^2}{2}} + \lambda e^{t\mu + \frac{t^2b^2\sigma^2}{2}}. \quad (3.2)$$

Le moment centré d'ordre  $m$  s'obtient en dérivant  $m$  fois la fonction génératrice des moments et en l'évaluant au point  $t = 0$ . Par conséquent, le coefficient de variation du mélange s'écrit comme

$$V_M = \frac{\sqrt{1 + \lambda(b^2 - 1)}\sigma}{\mu} = \sqrt{1 + \lambda(b^2 - 1)} \times V,$$

et le kurtosis

$$\gamma_M = \frac{3\{1 + \lambda(b^4 - 1)\}}{\{1 + \lambda(b^2 - 1)\}^2}. \quad (3.3)$$

Donc, en fixant  $V_M$  et  $\gamma_M$ , les paramètres  $\lambda$  et  $b$  du mélange peuvent être déterminés. Les valeurs  $\mu$  et  $\sigma$  peuvent être fixées à n'importe quelles valeurs sans perte de généralité, car la loi normale est invariante par translation linéaire. Cependant, le coefficient de variation doit demeurer inférieur ou égal à  $V = \sigma/\mu \leq 1/3$  pour que la probabilité d'observer des valeurs négatives soit négligeable (voir la remarque 2.2.3).

L'étape 2 consiste à calculer les intervalles de confiance sur tous les échantillons aléatoires générés provenant de la loi  $M$ . Étant donné que  $M$  dénote une loi théorique de paramètres connus, le coefficient de variation théorique est également connu. L'étape 3 consiste à vérifier si le pourcentage de couverture empirique des intervalles de confiance coïncide avec la valeur du seuil nominal fixé de ces intervalles.

### 3.5 La comparaison des coefficients de variation

Cette section expose la démarche permettant de répondre à la question centrale de ce mémoire, à savoir si le MRCC reproduit la variabilité des écoulements et des

précipitations observée pour la période (1961-1999). La démarche peut être divisée en quatre étapes, chacune de ces étapes constitue une sous-section dans ce qui suit.

### 3.5.1 Regrouper les données par bassins versants

La première étape consiste à agréger toutes les données à l'échelle du bassin versant. Pour les écoulements observés, il suffit de normaliser le débit d'une station de jaugeage par la superficie de son sous-bassin versant et de multiplier par la taille du bassin versant (Roads *et al.*, 2003) pour obtenir une estimation du débit total à l'exutoire. Les autres sources de données sont disponibles sous forme de grille (les écoulements et précipitations simulés par le MRCC et les précipitations réanalysées). Dans ce cas, une série pour un bassin versant particulier est obtenue en agrégeant de façon proportionnelle à la superficie les points de grilles à l'intérieur du bassin versant. De cette façon, une seule série chronologique est obtenue pour chacun des bassins versants et chacune des sources de données.

### 3.5.2 Vérifier la robustesse des intervalles de confiance

Par la suite, il convient de vérifier la robustesse des intervalles de confiance retenus (consulter la section 2, *Cadre Théorique*) dans le contexte où les séries chronologiques ne sont pas toutes normales (voir section 3.3). La robustesse est testée par une méthode de simulation de Monte-Carlo (voir la section 3.4) dans un cadre plus large que les données elles-mêmes. En effet, la robustesse des intervalles de confiance a été vérifiée pour des séries plus courtes et plus longues que les échantillons étudiés ( $20 \leq n \leq 1000$ ), pour des queues plus lourdes et plus légères ( $3 \leq \gamma_M \leq 6$ ) et pour des coefficients de variation plus petits et plus grands ( $1/10 \leq V_M \leq 1/3$ ). Dans le cas présent, le poids des queues de la distribution est mesuré par le coefficient d'aplatissement. L'augmentation du coefficient d'aplatissement permet de déterminer le comportement du pourcentage de couverture empirique des intervalles de confiance au fur et à mesure que le poids des queues de la loi normale initiale augmente. En ajustant les paramètres  $\lambda$  et  $b$  du modèle générateur d'extrêmes (eq. 3.1), 10 000 échantillons de coefficients de variation et de kurtosis différents ont été générés pour une taille d'échantillon donnée. Les intervalles de confiance à 95% du coefficient de variation ont été calculés pour toutes les simulations et comparés aux valeurs théoriques. Dans le cas où l'échantillon proviendrait d'une loi normale, 95% des 10 000 intervalles de confiance incluraient la valeur théorique du coefficient de variation.

Pour des raisons pratiques et sans perte de généralité, les simulations sont effectuées à partir d'une loi normale de moyenne  $\mu = 1,5$  et d'écart-type  $\sigma = 0,12$ , c'est-à-dire  $\mathcal{N}(1,5, (0,12)^2)$ . Dans ce cas-ci, le coefficient de variation théorique est  $V = 0,08$ . Ces valeurs pour la moyenne et l'écart-type ont été déterminées par les valeurs des écoulements annuels observés mais d'autres valeurs auraient très bien pu être retenues sans perte de généralité.

### 3.5.3 Calculer les intervalles de confiance

Les coefficients de variation ainsi que leurs intervalles de confiance doivent être calculés pour chacun des bassins versants et pour toutes les sources de données afin de tester la significativité des différences. Dans le cadre de ce travail, trois intervalles de confiance ont été retenus (consulter le chapitre 2, *Cadre Théorique*), soit deux intervalles supposant la normalité (McKay, 1932; Vangel, 1996) et un intervalle non paramétrique (Bonett, 2006).

### 3.5.4 Tester la significativité des différences des coefficients de variation

À la lumière des résultats émanant de l'étude par simulation de Monte-Carlo, le caractère significatif des différences entre les coefficients de variation des séries simulées et observées pourra être testé. Les intervalles de confiance pour chacun des bassins versants et de chacune des séries à l'étude pourront être comparés. Si les intervalles de confiance se recoupent, alors la différence n'est pas significative. Dans le cas contraire, la différence est significative.



# Chapitre 4

## Résultats

Ce chapitre présente d'abord les résultats des simulations de Monte-Carlo pour déterminer la robustesse des intervalles de confiance du coefficient de variation par rapport à la violation de l'hypothèse de normalité (voir le chapitre précédent). Par la suite, ces intervalles sont illustrés pour les écoulements et les précipitations. La comparaison des intervalles entre les observations et les simulations par bassins versants permettra de déterminer s'il existe une différence significative de variabilité relative. La discussion de ces résultats se trouve au chapitre suivant.

### 4.1 Les résultats des simulations de Monte-Carlo

Les pourcentages de couverture empirique des intervalles de confiance du coefficient de variation de McKay (1932), Vangel (1996) et Bonett (2006) de 10 000 simulations de Monte-Carlo sont résumés dans les tableaux 4.1, 4.2, 4.3 et 4.4 pour des tailles d'échantillon de 20, 40, 100 et 1000.

#### 4.1.1 Le cas normal ( $\gamma = 3$ )

Dans le cas où les échantillons sont normaux, c'est-à-dire lorsque le coefficient d'aplatissement  $\gamma$  est égal à 3, les niveaux de confiance empiriques des intervalles paramétriques (McKay, 1932; Vangel, 1996) sont pratiquement égaux au seuil nominal de 95% et ce, indépendamment de la valeur du coefficient de variation et de la taille de l'échantillon (voir Tableaux 4.1 à 4.4). En effet, les intervalles de confiance paramétriques

TAB. 4.1 – Pourcentage de couverture empirique des intervalles de confiance pour une taille d'échantillon de  $n = 20$  et pour un niveau de confiance nominal de 95%.

| I.C.   | kurtosis | Coefficient de variation |      |      |      |      |      |
|--------|----------|--------------------------|------|------|------|------|------|
|        |          | 0,1                      | 0,15 | 0,2  | 0,25 | 0,3  | 1/3  |
| Mckay  | 3        | 0,96                     | 0,95 | 0,95 | 0,95 | 0,95 | 0,95 |
|        | 4        | 0,90                     | 0,90 | 0,90 | 0,90 | 0,90 | 0,90 |
|        | 5        | 0,87                     | 0,84 | 0,85 | 0,85 | 0,85 | 0,86 |
|        | 6        | 0,84                     | 0,80 | 0,80 | 0,79 | 0,80 | 0,80 |
| Vangel | 3        | 0,96                     | 0,95 | 0,96 | 0,95 | 0,95 | 0,95 |
|        | 4        | 0,90                     | 0,90 | 0,90 | 0,90 | 0,90 | 0,90 |
|        | 5        | 0,87                     | 0,84 | 0,85 | 0,85 | 0,85 | 0,86 |
|        | 6        | 0,84                     | 0,80 | 0,80 | 0,79 | 0,80 | 0,81 |
| Bonett | 3        | 0,99                     | 0,99 | 0,99 | 0,99 | 0,99 | 0,99 |
|        | 4        | 1,00                     | 1,00 | 0,99 | 0,99 | 0,98 | 0,98 |
|        | 5        | 1,00                     | 0,99 | 0,99 | 0,98 | 0,98 | 0,97 |
|        | 6        | 1,00                     | 1,00 | 0,99 | 0,99 | 0,98 | 0,98 |

TAB. 4.2 – Pourcentage de couverture empirique des intervalles de confiance pour une taille d'échantillon de  $n = 40$  et pour un niveau de confiance nominal de 95%.

| I.C.   | kurtosis | Coefficient de variation |      |      |      |      |      |
|--------|----------|--------------------------|------|------|------|------|------|
|        |          | 0,1                      | 0,15 | 0,2  | 0,25 | 0,3  | 1/3  |
| Mckay  | 3        | 0,95                     | 0,95 | 0,95 | 0,95 | 0,95 | 0,95 |
|        | 4        | 0,90                     | 0,89 | 0,90 | 0,90 | 0,90 | 0,91 |
|        | 5        | 0,85                     | 0,85 | 0,84 | 0,85 | 0,85 | 0,86 |
|        | 6        | 0,82                     | 0,79 | 0,80 | 0,80 | 0,80 | 0,80 |
| Vangel | 3        | 0,95                     | 0,95 | 0,95 | 0,95 | 0,95 | 0,95 |
|        | 4        | 0,90                     | 0,89 | 0,90 | 0,90 | 0,90 | 0,91 |
|        | 5        | 0,85                     | 0,85 | 0,84 | 0,85 | 0,85 | 0,86 |
|        | 6        | 0,82                     | 0,79 | 0,80 | 0,80 | 0,81 | 0,80 |
| Bonett | 3        | 0,96                     | 0,96 | 0,96 | 0,96 | 0,95 | 0,95 |
|        | 4        | 0,96                     | 0,96 | 0,96 | 0,96 | 0,95 | 0,95 |
|        | 5        | 0,96                     | 0,97 | 0,97 | 0,97 | 0,96 | 0,95 |
|        | 6        | 0,96                     | 0,97 | 0,98 | 0,97 | 0,97 | 0,96 |

TAB. 4.3 – Pourcentage de couverture empirique des intervalles de confiance pour une taille d'échantillon de  $n = 100$  et pour un niveau de confiance nominal de 95%.

| I.C.   | kurtosis | Coefficient de variation |      |      |      |      |      |
|--------|----------|--------------------------|------|------|------|------|------|
|        |          | 0,1                      | 0,15 | 0,2  | 0,25 | 0,3  | 1/3  |
| McKay  | 3        | 0,95                     | 0,95 | 0,95 | 0,95 | 0,95 | 0,95 |
|        | 4        | 0,89                     | 0,89 | 0,90 | 0,90 | 0,90 | 0,90 |
|        | 5        | 0,85                     | 0,84 | 0,84 | 0,85 | 0,85 | 0,86 |
|        | 6        | 0,79                     | 0,79 | 0,80 | 0,80 | 0,81 | 0,81 |
| Vangel | 3        | 0,95                     | 0,95 | 0,95 | 0,95 | 0,95 | 0,95 |
|        | 4        | 0,89                     | 0,89 | 0,90 | 0,90 | 0,90 | 0,90 |
|        | 5        | 0,85                     | 0,84 | 0,84 | 0,85 | 0,86 | 0,86 |
|        | 6        | 0,79                     | 0,79 | 0,80 | 0,80 | 0,81 | 0,81 |
| Bonett | 3        | 0,95                     | 0,94 | 0,94 | 0,94 | 0,94 | 0,94 |
|        | 4        | 0,94                     | 0,95 | 0,95 | 0,95 | 0,94 | 0,94 |
|        | 5        | 0,95                     | 0,95 | 0,96 | 0,96 | 0,96 | 0,96 |
|        | 6        | 0,95                     | 0,95 | 0,96 | 0,96 | 0,97 | 0,97 |

TAB. 4.4 – Pourcentage de couverture empirique des intervalles de confiance pour une taille d'échantillon de  $n = 1000$  et pour un niveau de confiance nominal de 95%.

| I.C.   | kurtosis | Coefficient de variation |      |      |      |      |      |
|--------|----------|--------------------------|------|------|------|------|------|
|        |          | 0,1                      | 0,15 | 0,2  | 0,25 | 0,3  | 1/3  |
| McKay  | 3        | 0,95                     | 0,95 | 0,95 | 0,95 | 0,95 | 0,95 |
|        | 4        | 0,89                     | 0,89 | 0,89 | 0,90 | 0,90 | 0,90 |
|        | 5        | 0,84                     | 0,84 | 0,84 | 0,85 | 0,85 | 0,86 |
|        | 6        | 0,79                     | 0,79 | 0,80 | 0,80 | 0,81 | 0,81 |
| Vangel | 3        | 0,95                     | 0,95 | 0,95 | 0,95 | 0,95 | 0,95 |
|        | 4        | 0,89                     | 0,89 | 0,89 | 0,90 | 0,90 | 0,90 |
|        | 5        | 0,84                     | 0,84 | 0,84 | 0,85 | 0,85 | 0,86 |
|        | 6        | 0,79                     | 0,79 | 0,80 | 0,80 | 0,81 | 0,81 |
| Bonett | 3        | 0,95                     | 0,95 | 0,95 | 0,95 | 0,95 | 0,95 |
|        | 4        | 0,95                     | 0,95 | 0,95 | 0,95 | 0,95 | 0,95 |
|        | 5        | 0,96                     | 0,95 | 0,95 | 0,95 | 0,95 | 0,96 |
|        | 6        | 0,95                     | 0,95 | 0,95 | 0,95 | 0,96 | 0,95 |

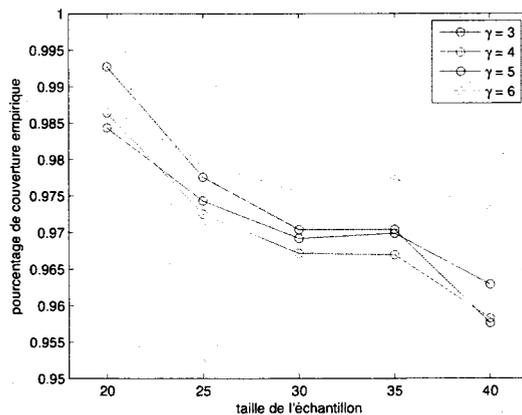


FIG. 4.1 – Pourcentage de couverture empirique de l'intervalle de Bonett (2006) pour plusieurs tailles d'échantillon et pour plusieurs valeur de coefficient d'aplatissement.

demeurent précis pour un coefficient de variation élevé et une petite taille d'échantillon puisque l'hypothèse de normalité est vérifiée. Toutefois, le coefficient de variation doit être inférieur à  $1/3$  pour que ces intervalles demeurent valides.

Contrairement aux intervalles paramétriques, l'intervalle de confiance non paramétrique du coefficient de variation quartile de Bonett (2006) est très sensible à la taille de l'échantillon. D'ailleurs, la figure 4.1 illustre les pourcentages de couverture empiriques de l'intervalle de Bonett (2006) pour plusieurs tailles d'échantillons entre 20 et 40. Cet intervalle surestime le niveau de confiance nominal fixé à 95% pour les échantillons de tailles 20 et 40 issus d'une loi normale. Néanmoins, pour une taille d'échantillon donnée, cet intervalle demeure insensible aux valeurs du coefficient de variation.

#### 4.1.2 Les cas où $\gamma > 3$

Les pourcentages de couverture empiriques des trois intervalles de confiance demeurent stables par rapport aux différentes valeurs du coefficient de variation. Par contre, les trois intervalles de confiance sont vulnérables aux excès d'aplatissement de l'échantillon. L'impact d'un coefficient d'aplatissement plus élevé est différent selon la nature paramétrique ou non paramétrique de l'intervalle. Pour les intervalles paramétriques, le seuil nominal est largement surestimé à mesure que le coefficient d'aplatissement augmente. L'opposé se produit dans le cas non paramétrique. L'intervalle de Bonett (2006), quant à lui, surestime le niveau de confiance nominal lorsque le kurtosis augmente.

Les seuils empiriques des intervalles de confiance paramétriques demeurent à peu près constants, pour un kurtosis fixé, suivant la taille de l'échantillon. L'intervalle de Bonett (2006), quant à lui, surestime d'autant plus le niveau de confiance nominal que la taille d'échantillon est petite. Afin que le niveau de confiance nominal de l'intervalle de confiance non paramétrique soit de 95% pour des kurtosis de 3 à 6, la taille d'échantillon doit être d'au moins 100.

## 4.2 Les intervalles de confiance des coefficients de variation

Les pourcentages de couverture empirique des intervalles de confiance de McKay (1932) et de Vangel (1996) sont pratiquement identiques (voir la section précédente). Par conséquent, l'intervalle de confiance de Vangel (1996) sera le seul intervalle paramétrique présenté dans cette section d'abord parce qu'il est le plus récent et aussi parce qu'il est théoriquement le plus précis (voir la section 2.2.1).

### 4.2.1 Écoulements

Les intervalles de confiance de Vangel (1996) et de Bonett (2006) du coefficient de variation des écoulements observés et simulés par le MRCC pour chacun des bassins versants sont illustrés respectivement aux figures 4.2 et 4.3.

Les intervalles de confiance des coefficients de variation se recoupent pour tous les bassins versants étudiés, à la fois dans les cas paramétriques et non paramétriques. Par conséquent, aucune des différences entre les coefficients de variation des écoulements observés et simulés n'est significative.

### 4.2.2 Précipitations

Les intervalles de confiance des coefficients de variation des précipitations annuelles simulées par le MRCC sont comparés aux intervalles de confiance des précipitations réanalysées, soient les réanalyses CRU (Mitchell et Jones, 2005) et WM2 (Willmott et Matsuura, 1995; Willmott et Robeson, 1995), pour chacun des bassins versants aux figures 4.4 et 4.5.

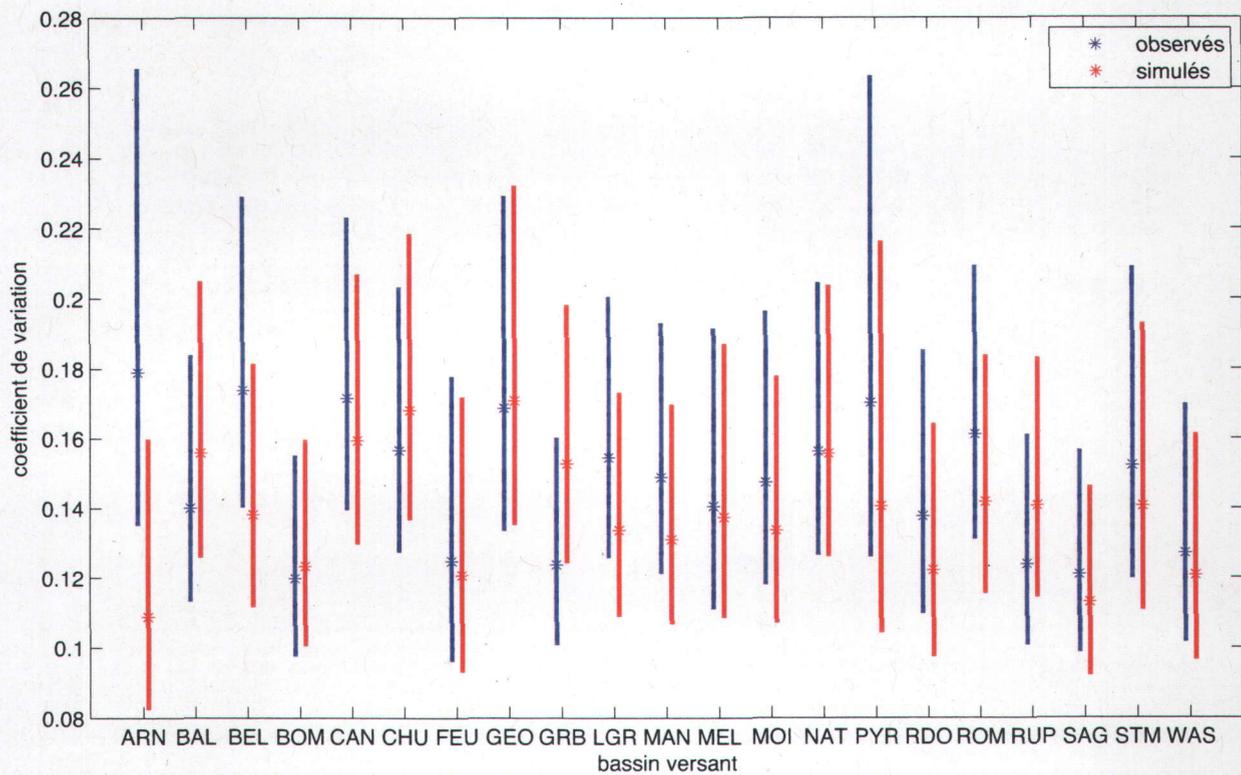


FIG. 4.2 – Intervalle de confiance de Vangel (1996) de niveau 95% des coefficients de variation des écoulements observés et simulés par le MRCC (acy) pour chacun des bassins versants étudiés. Les bandes symbolisent les intervalles de confiance tandis que les astérisques indiquent les valeurs des coefficients de variation.

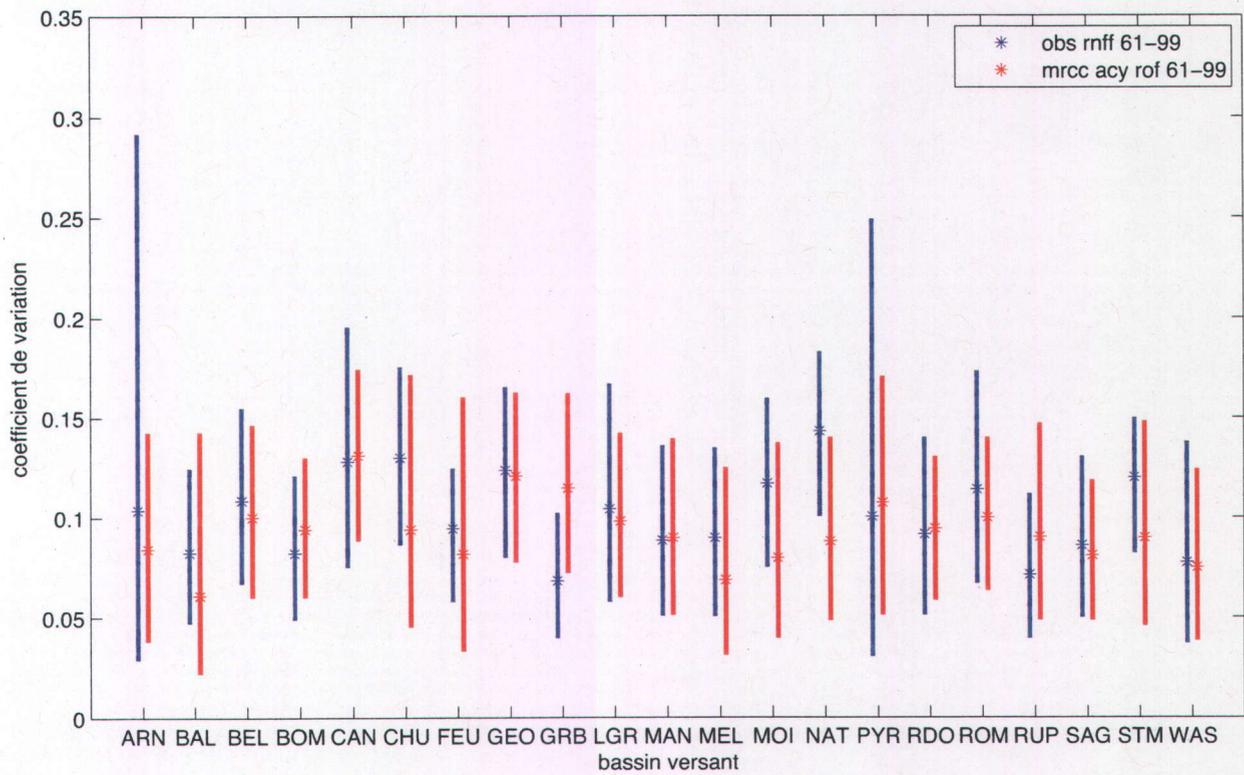


FIG. 4.3 – Intervalle de confiance de Bonett (2006) de niveau 95% des coefficients de variation des écoulements observés et simulés par le MRCC (acy) pour chacun des bassins versants étudiés. Les bandes symbolisent les intervalles de confiance tandis que les astérisques indiquent les valeurs des coefficients de variation.

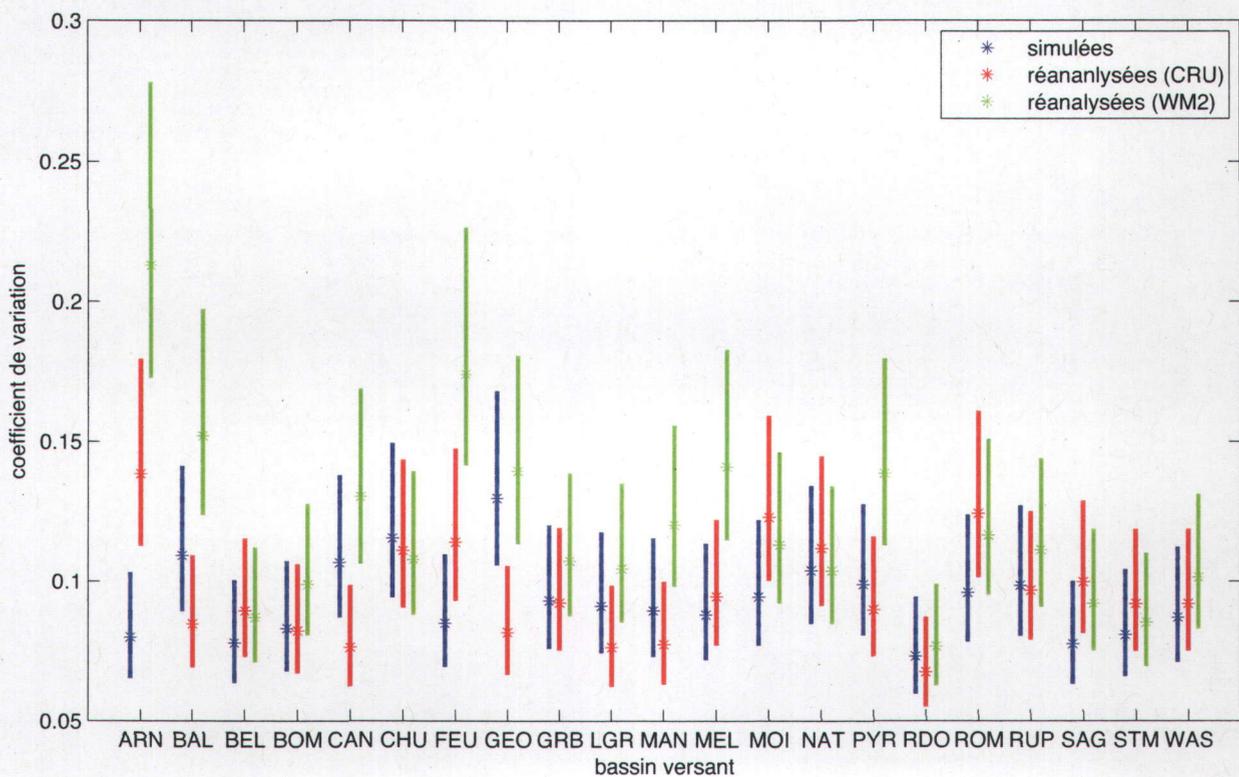


FIG. 4.4 – Intervalle de confiance de Vangel (1996) de niveau 95% des coefficients de variation des précipitations simulées par le MRCC (acy) et des précipitations réanalysées (CRU et WM2) pour chacun des bassins versants étudiés.

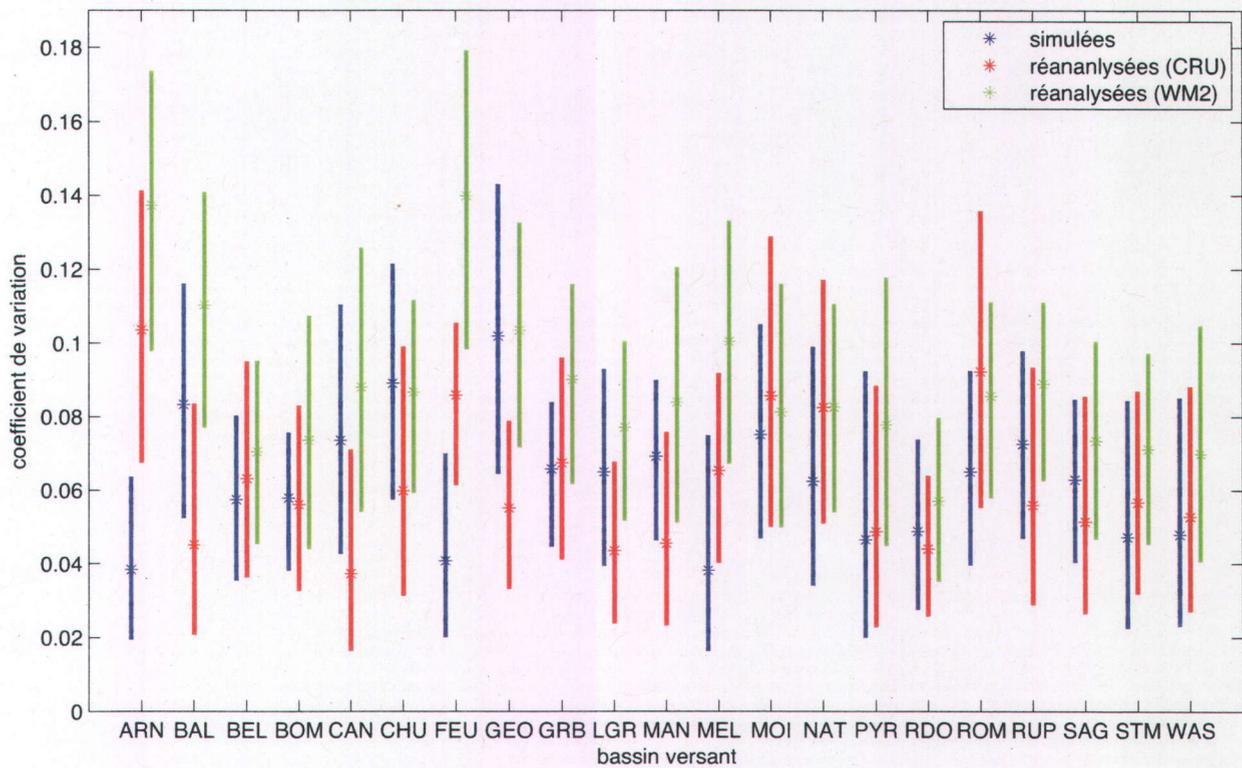


FIG. 4.5 – Intervalle de confiance de Bonett (2006) de niveau 95% des coefficients de variation des précipitations simulées par le MRCC (acy) et des précipitations réanalysées (CRU et WM2) pour chacun des bassins versants étudiés.

La différence entre les coefficients de variation des précipitations simulées et ré-analysées du bassin de la rivière Arnaud est significative pour les deux intervalles de confiance étudiés. Le même comportement se reproduit entre les précipitations simulées et la réanalyse WM2 du bassin de la rivière Aux Feuilles. Pour le bassin versant de la rivière Georges, la différence entre les précipitations simulées et la réanalyse CRU semble significative dans le cas des intervalles de confiance de Vangel (1996) mais cette différence est non significative dans le cas des intervalles de Bonett (2006). Par ailleurs, une constatation surprenante émane de ces deux dernières figures : les réanalyses sont parfois significativement différentes.

# Chapitre 5

## Discussion

Ce chapitre présente, dans un premier temps, une discussion sur la robustesse des intervalles de confiance étudiés dans ce mémoire dans le but de déterminer s'ils peuvent contribuer à l'inférence statistique dans les cas où la normalité des séries est rejetée. Par la suite, la discussion s'étend à l'inférence elle-même, c'est-à-dire à la significativité ou non des différences entre les coefficients de variation des séries observées et simulées pour chacun des bassins versants.

Plus particulièrement, les limites de la méthode, notamment les hypothèses de normalité et d'indépendance, sont discutées à la section 5.3. La section 5.4 présente, dans le but de comparer les résultats émanant de ce mémoire, les conclusions de deux autres approches ayant le même but, c'est-à-dire vérifier si le MRCC peut reproduire la variabilité des écoulements observés : une approche par tests d'égalité et une autre par inférence bayésienne. Finalement, les résultats de ce mémoire sont confrontés à ceux provenant d'autres méthodes complémentaires issues de la littérature (section 5.5).

### 5.1 La robustesse des intervalles de confiance

La robustesse des intervalles de confiance a été vérifiée par des simulations de Monte-Carlo. Les résultats de ces simulations sont résumés aux tableaux 4.1, 4.2, 4.3 et 4.4 de la section 4.1. Les conclusions suivantes ont émané de cette étude :

- dans le cas normal (un coefficient d'aplatissement de 3), les seuils observés des intervalles de confiance paramétriques (McKay, 1932; Vangel, 1996) sont égaux au seuil nominal de 5% tandis que les seuils observés de l'intervalle non paramétrique

- de Bonett (2006) diminuent à mesure que la taille de l'échantillon décroît ;
- le pourcentage de couverture empirique de tous les intervalles de confiance demeure constant pour un kurtosis fixé et pour un coefficient de variation inférieur à 1/3 ;
  - les intervalles de confiance paramétriques deviennent trop courts à mesure que l'aplatissement augmente et ce, peu importe la taille des échantillons ;
  - dans tous les cas de figure, l'intervalle non paramétrique surestime, d'autant plus le que le coefficient d'aplatissement est élevé, le niveau de confiance nominal pour des tailles d'échantillons inférieures à 100.

**Remarque 5.1.1.** *Le rétrécissement des intervalles de confiance paramétriques est confirmé théoriquement par les travaux d'Arvesen (1972) qui démontrent que le niveau de confiance réel de l'intervalle de confiance de McKay (1932) est inférieur lorsqu'il y a un excès d'aplatissement, c'est-à-dire lorsque  $\gamma > 3$ .*

Les intervalles de confiance ne respectent pas le seuil nominal fixé dans certaines conditions particulières d'aplatissement et de taille d'échantillon, mais ils peuvent néanmoins être utiles à des fins d'inférence statistique. En se référant aux caractéristiques des données, compilées à l'annexe B, les kurtosis des bassins versants sont inclus dans l'intervalle [1, 97; 5, 58]. Pour les intervalles paramétriques de McKay (1932) et de Vangel (1996), un excès d'aplatissement (un kurtosis supérieur à 3) se traduit par une diminution drastique du pourcentage de couverture des intervalles de confiance. Autrement dit, les intervalles deviennent beaucoup trop courts. Si les intervalles se recoupent, alors on peut affirmer que les différences sont non significatives. Il est cependant plus difficile de juger de la situation opposée, c'est-à-dire de déterminer si la différence est significative lorsque les intervalles de confiance trop courts ne se recoupent pas. Par conséquent, il est possible de conclure sur la non significativité des différences en utilisant des intervalles trop courts, mais la question de la significativité demeure ambiguë.

Contrairement aux intervalles paramétriques, l'intervalle de confiance non paramétrique de Bonett (2006) semble très robuste aux excès d'aplatissement. En effet, pour une taille d'échantillon donnée, le pourcentage de couverture empirique des intervalles demeure constant pour des kurtosis de 3 à 6 (consulter les tableaux 4.1 à 4.4). Cependant, cet intervalle de confiance non paramétrique est biaisé, c'est-à-dire qu'il est trop long pour des tailles d'échantillon inférieures à 100. Pour les données simulées issues du MRCC et celles provenant des réanalyses, les tailles des échantillons sont de 39, soit des séries débutant en 1961 et se terminant en 1999. D'autre part, les séries des écoulements observés ne sont pas toutes complètes (voir le tableau 3.2). Ces petites tailles d'échantillon ont pour effet d'allonger l'intervalle de confiance (voir le tableau 4.2 ainsi que la figure 4.1). En fait, plus la taille d'échantillon est petite, plus le biais est

important. Pour des tailles d'échantillon supérieures à 35, le biais semble raisonnable ( $\text{biais} \approx 0,01$ ) excepté pour les séries caractérisées par un coefficient d'aplatissement égal à 6 ( $\text{biais} \approx 0,025$ ). Par exemple, pour la série des écoulements observés de la rivière Pyrite (17 observations), l'intervalle de confiance non paramétrique risque d'être exagérément long. Il est néanmoins possible de vérifier si les coefficients de variation sont égaux avec ces intervalles de confiance. Si les intervalles ne se recoupent pas, la différence est nécessairement significative puisque la probabilité que la vraie valeur du coefficient de variation soit incluse dans l'intervalle est plus grande que le niveau nominal. Sinon, il est difficile de conclure puisque les intervalles sont trop longs. Les résultats sont concluants pour une différence significative mais ne le sont pas pour une différence non significative. Ce cas est complémentaire à celui des intervalles paramétriques.

Le comportement complémentaire des intervalles de confiance paramétriques et celui non paramétrique se révèle très utile pour l'inférence statistique appliquée dans le cadre particulier de ce mémoire. En effet, si les intervalles paramétriques se recoupent, la différence est assurément non significative au seuil fixé car ces intervalles surestiment le seuil nominal. Si les intervalles non paramétriques ne se recoupent pas, la différence est assurément significative puisque ces intervalles sous-estiment le seuil nominal fixé. Il subsiste néanmoins une situation ambiguë dans le cas où les intervalles paramétriques ne se recoupent pas et les non paramétriques se recoupent.

## 5.2 L'inférence à l'aide des intervalles de confiance

Pour tester si les coefficients de variation sont significativement différents, les intervalles de confiance sont comparés bassin par bassin. Étant donné que les intervalles de confiance n'ont pas la même robustesse par rapport à l'aplatissement et la taille de l'échantillon, la discussion se divise la plupart du temps en deux parties distinctes : une partie traitant des intervalles paramétriques et l'autre dévolue à l'intervalle non paramétrique. Dans cette section, la procédure d'inférence est appliquée à la comparaison des variabilités relatives des écoulements observés et simulés, des sources de réanalyses, des précipitations observées et simulées ainsi qu'aux différences entre bassins versants.

### 5.2.1 L'inférence sur les écoulements

Comme il a déjà été mentionné précédemment, les intervalles paramétriques sont sensibles aux excès d'aplatissement, c'est-à-dire aux kurtosis supérieurs à 3. Le cas

échéant, les intervalles de confiance deviennent trop courts. Dans le contexte particulier de l'application, les kurtosis des écoulements observés sont inclus dans l'intervalle [1,97 ; 5,58] tandis que ceux des écoulements simulés par le MRCC sont inclus dans [1,96 ; 4,08]. Nonobstant le fait que les intervalles de confiance paramétriques du coefficient de variation soient trop courts pour des excès d'aplatissement, les intervalles de la figure 4.2 se recourent pour tous les bassins versants étudiés. Par conséquent, il n'existe aucune différence significative de variabilité relative entre les écoulements annuels observés et ceux simulés par le MRCC pour chacun des 21 bassins versants à l'étude.

Les tailles échantillonnales des écoulements observés varient de 17 à 39 tandis que celles des simulations sont égales à 39. Ces petites tailles nuisent d'autant plus à la précision de l'intervalle de confiance de Bonett (2006) que la taille d'échantillon est petite. Dès lors, les intervalles deviennent trop longs. Les intervalles de confiance non paramétriques calculés par la méthodologie de Bonett (2006) se recourent pour tous les écoulements observés et simulés des bassins versants étudiés. Toutefois, ces intervalles sont trop longs, voire exagérément longs dans les cas des bassins de la rivière Arnaud et celui de la rivière Pyrite qui ne comprennent respectivement que 20 et 17 données observées. Cet intervalle de confiance n'est pas très utile pour des petites tailles échantillonnales parce qu'il n'est pas assez discriminant. Il est possible qu'il existe une ou plusieurs différences significatives, mais qu'elles ne soient pas détectables puisque les intervalles surestiment le niveau de confiance nominal de 95%. Toutefois, cette supposition n'est que très peu réaliste considérant les résultats des intervalles de confiance paramétriques discutés au paragraphe précédent.

Par conséquent, la conclusion retenue consiste à affirmer qu'il n'existe pas de différence significative entre les coefficients de variation des écoulements observés et simulés pour chacun des bassins versants étudiés car tous les intervalles de confiance paramétriques se recourent.

## 5.2.2 Les différences entre les réanalyses

La figure 4.4 illustre les intervalles de confiance paramétriques des coefficients de variation des précipitations simulées et réanalysées. Les intervalles de confiance des deux sources de réanalyse ne se recourent pas pour les bassins versants des rivières Baleine, Georges et Caniapiscau. Ces différences sont d'autant plus troublantes que les réanalyses sont supposées dépendre la météorologie observée. Il est évidemment impossible qu'il existe plus d'une valeur d'écoulement annuel observé pour un même bassin versant. Par conséquent, il ne devrait pas exister de différence significative entre les deux sources de réanalyse, à moins qu'une de ces réanalyses soit erronée. Cette dissemblance

peut être expliquée par le manque d'observations pluviométriques disponibles dans le Nord du Québec. Les algorithmes d'interpolation des différents organismes fournissant des réanalyses ne gèrent pas tous de la même façon le manque de données à proximité des points à interpoler. Plus particulièrement, dans le cas de la réanalyse CRU, des stations fictives ayant des anomalies relatives nulles sont ajoutées sur le territoire afin que les observations ne soient pas extrapolées à des distances trop éloignées (voir la section 3.1.4). Quant à la réanalyse WM2, l'algorithme permet d'extrapoler les observations jusqu'à de très grandes distances. Cette différence au niveau de l'algorithme d'interpolation peut expliquer la divergence des réanalyses dans les régions où les observations pluviométriques sont rares. Par exemple, pour les bassins versants nordiques, la réanalyse WM2 est beaucoup plus humide que CRU. Quoi qu'il en soit, aucune de ces réanalyses ne peut être considérée meilleure que l'autre tout simplement parce qu'il n'existe pas de données disponibles pour les comparer.

**Remarque 5.2.1.** *La figure 5.1 présente une carte localisant les stations pluviométriques d'Environnement Canada ayant enregistré au moins une donnée durant la période 1961-1999. Cette figure illustre notamment la très faible densité d'observations pluviométriques dans le Nord du Québec. L'interpolation réalisée par les organismes fournissant les réanalyses utilisent vraisemblablement ces données. D'autres organismes comme Hydro-Québec, Environnement Québec et la Sopfeu possèdent aussi des stations pluviométriques, mais leurs périodes de mesure sont généralement beaucoup moins longues que celles d'Environnement Canada.*

Le manque de données pluviométriques au nord du bassin de la Grande Rivière rend très difficile, voire impossible, une comparaison des valeurs simulées des précipitations à celles réanalysées. Les informations les plus fiables pour ces bassins nordiques<sup>1</sup> demeurent les écoulements puisque ces derniers intègrent tous les événements hydro-météorologiques en un point, le débit à l'exutoire (qui possède tout de même une part d'incertitude).

### 5.2.3 L'inférence sur les précipitations

Les précipitations issues du MRCC et des réanalyses CRU et WM2 sont caractérisées par des kurtosis entre [2,05; 4,72] et des coefficients de variation  $V \leq 1/3$ . Les intervalles paramétriques sont trop courts pour  $\gamma > 3$ . En excluant les bassins nordiques où les observations pluviométriques sont rares, tous les intervalles de confiance des précipitations simulées recourent les intervalles des deux réanalyses. Donc, les précipitations

<sup>1</sup>Les acronymes de ces bassins sont ARN, BAL, CAN, CHU, FEU, GEO, GRB, MEL et PYR

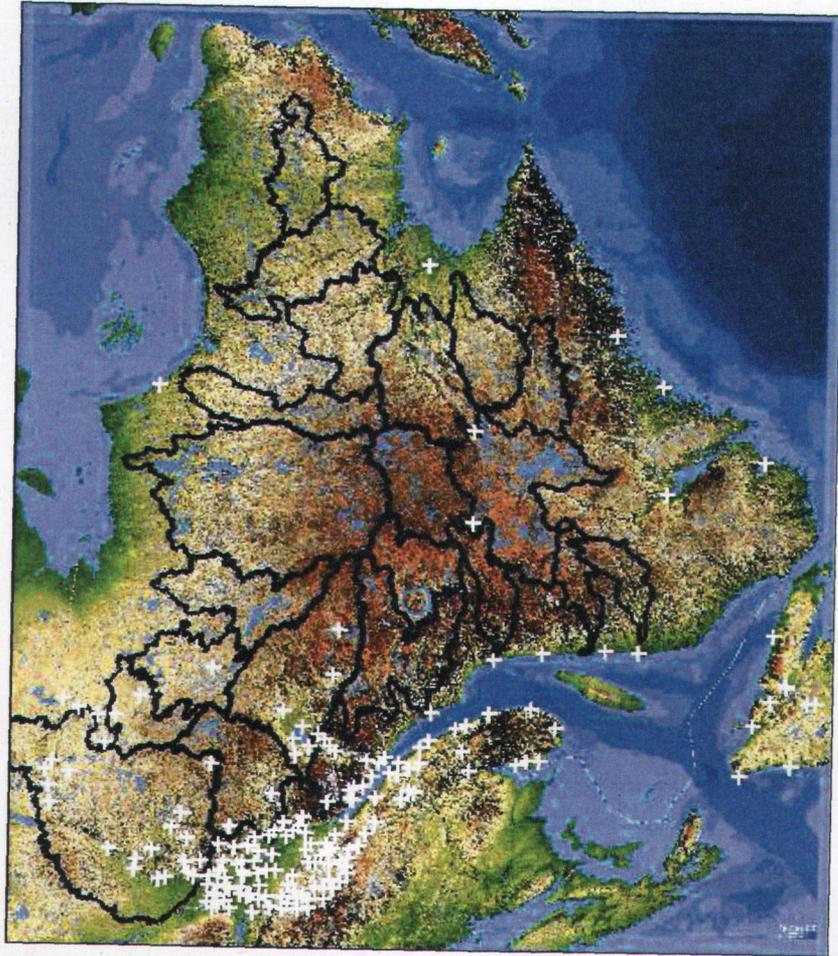


FIG. 5.1 – Localisation des stations pluviométriques d'Environnement Canada ayant enregistré au moins une donnée durant la période 1961-1999.

simulées par le MRCC possèdent la même variabilité relative que les précipitations ré-analysées pour ces bassins. Dans le cas des bassins nordiques, les intervalles de confiance du MRCC recourent au moins une des deux réanalyses sauf pour le bassin de la rivière Arnaud. Il est cependant difficile de conclure étant donné que les réanalyses sont plus ou moins crédibles à ces latitudes.

La figure 4.5 montre que les intervalles de confiance non paramétriques des précipitations simulées et réanalysées se recourent pour tous les bassins versants non nordiques. Cependant, l'intervalle non paramétrique de Bonett (2006) est trop long pour une taille d'échantillon de 39. On ne peut donc pas identifier des différences significatives à la lumière de ces intervalles.

Par conséquent, les précipitations simulées par le MRCC possèdent la même variabilité relative que les observations pour les bassins non nordiques car tous les intervalles de confiance paramétriques se recourent. Dans le cas des bassins nordiques, par exemple le bassin de la rivière Arnaud, les conclusions demeurent incertaines puisqu'il n'existe pas suffisamment d'observations pour comparer adéquatement les simulations.

#### 5.2.4 La différence entre les bassins versants

Chacun des intervalles de confiance des figures 4.2 à 4.5 recoupe tous les intervalles des autres bassins versants quels que soient les variables étudiées et la nature paramétrique ou non paramétrique des intervalles (en excluant les bassins nordiques pour les précipitations). Par conséquent, la variabilité relative des écoulements et des précipitations n'est pas significativement différente à l'échelle du Québec. Le coefficient de variation standardise la variabilité par la moyenne. Par conséquent, tout ce qui influence la variable d'intérêt est pris en compte. Plus particulièrement, en standardisant la variabilité des écoulements annuels par le coefficient de variation, on s'assure de prendre en compte l'hétérogénéité des moyennes des précipitations sur le territoire du Québec ainsi que la taille des bassins versants, deux variables qui régulent les écoulements. Les valeurs moyennes des précipitations peuvent être altérées par la latitude et l'altitude. Ces résultats démontrent que la variabilité relative des précipitations annuelles est la même pour les 21 bassins versants étudiés. Autrement, en standardisant les écoulements uniquement par la taille des bassins, la différence des précipitations moyennes à l'échelle du territoire québécois n'est pas prise en compte. Le cas échéant, la variabilité relative des écoulements ne serait pas la même pour tout le Québec. Néanmoins, la standardisation des écoulements par la taille du bassin versant permet de mettre l'accent sur la différence des lames d'eau de chacun des bassins soumis à un régime de précipitations différent.

## 5.3 La critique de la méthode

La méthodologie sur laquelle ce mémoire est fondé possède quelques limites. Ces limites s'expriment notamment par les hypothèses initiales de normalité et d'indépendance des données utilisées.

### 5.3.1 L'hypothèse de normalité

L'hypothèse de normalité de toutes les séries de données a été vérifiée par le test de normalité de D'Agostino *et al.* (1990) basé sur la symétrie et l'aplatissement de la loi normale. À la section 3.3, il a été mentionné que certaines séries ne pouvaient pas être supposées issues d'une variable normale au seuil de 10%. Il a aussi été déduit que la non normalité était causée par la présence de valeurs extrêmes dans les séries et que si ces extrêmes étaient retirés, les séries pourraient provenir d'une loi normale (consulter les p-valeurs de tests de normalité avec et sans les valeurs extrêmes exposées à l'annexe B). La méthodologie employée dans ce mémoire pour tester la robustesse des intervalles de confiance a été d'effectuer des simulations de Monte-Carlo. Un mélange de lois normales a été utilisé pour générer les échantillons contaminés par la présence de valeurs extrêmes. La robustesse des intervalles de confiance n'a pas été testée uniquement pour les caractéristiques des données. Ces intervalles ont été testés pour une gamme élargie de coefficients de variation, d'excès d'aplatissement et de tailles d'échantillon dont les caractéristiques des données sont incluses dans cette gamme. L'utilisation d'une autre distribution statistique qu'un mélange de lois aurait pu être utile pour modéliser l'asymétrie de certaines séries, ce que le mélange de normales de même moyenne ne pouvait accomplir. Cependant, l'hypothèse de base de la non normalité était que les données sont issues de lois normales contaminées par les extrêmes. La présence d'une valeur extrême n'influe pas seulement sur le kurtosis échantillonnal mais également sur l'asymétrie. L'asymétrie des échantillons a été supposée provenant aussi des valeurs extrêmes.

Par ailleurs, la robustesse des intervalles de confiance n'a pas été testée pour des valeurs de kurtosis plus petites que 3. De telles valeurs d'aplatissement ne peuvent pas être modélisées par un mélange de lois normales. Néanmoins, pour toutes les sources de données utilisées, une faible valeur de kurtosis n'entraîne généralement pas le rejet de l'hypothèse de normalité, c'est-à-dire que les faibles valeurs de kurtosis ne sont pas suffisamment petites pour s'écarter significativement de la normalité, à part la surprenante série des écoulements simulés de la rivière Bell. Par conséquent, les intervalles de confiance sont supposés adéquats également pour de petits kurtosis.

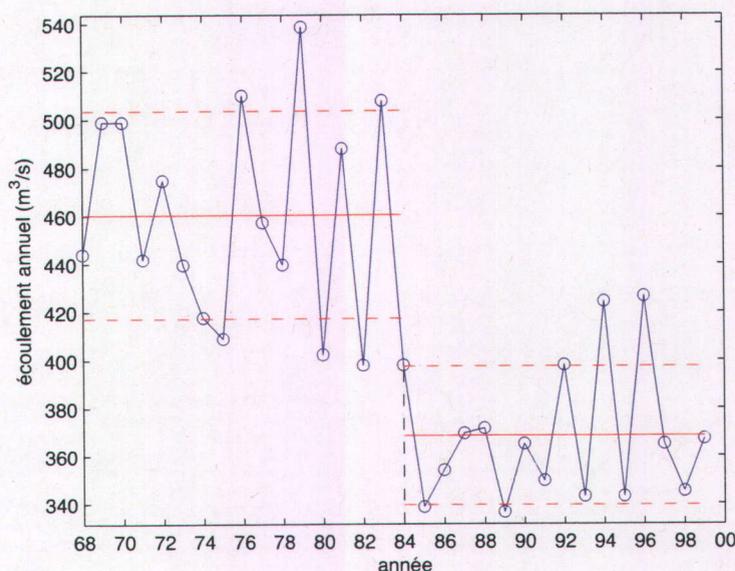


FIG. 5.2 – Série chronologique des écoulements annuels du bassin versant de la rivière Moisie. La ligne en tirets noirs verticale représente la rupture de la série. Les lignes rouges horizontales en tirets tracent la moyenne plus l'écart type et la moyenne moins l'écart type, avant et après la rupture.

Une autre façon de tester l'égalité des coefficients de variation sans effectuer de simulations aurait été d'ajuster des lois statistiques possédant des queues plus lourdes aux données, par exemple la loi gamma, Gumbel ou Pearson type III. À partir de ces lois, la distribution du coefficient de variation aurait pu être déterminée pour construire un intervalle de confiance. Notons que pour la loi gamma, le coefficient de variation ne dépend que de l'inverse du paramètre de précision. Cependant, la distribution du coefficient de variation n'étant pas simple dans le cas où la normalité de la loi est supposée, elle se complexifierait davantage dans le cas non normal. Par ailleurs, des méthodes de simulations bayésiennes auraient pu être appliquées pour calculer les intervalles de crédibilité du coefficient de variation pour n'importe quelle loi. Dans le cadre de ce mémoire, cette méthode n'a pas été retenue car la robustesse des intervalles de confiance était suffisante pour ne pas complexifier la méthodologie par une approche bayésienne.

### 5.3.2 L'indépendance et la stationnarité

Dans le cadre de ce mémoire, il a été mentionné à la section 3.1 que la taille des bassins versants ainsi que leur capacité de rétention n'étaient pas suffisantes pour que les écoulements soient autocorrélés dans le temps. Autrement dit, une observation particu-

lière de la série n'est pas influencée par l'observation de l'année précédente et n'influence pas celle de l'année subséquente. Cette supposition est justifiée par l'absence d'autocorrélation significative pour la majorité des séries chronologiques à l'étude (voir le tableau B.1 de l'annexe B). Toutefois, les processus météorologiques générant les précipitations, et par la suite les écoulements, peuvent être autocorrélés dans le temps. Par exemple, la téléconnexion climatique *El Niño* influence le régime des précipitations de façon cyclique. Ces processus climatologiques corrélés dans le temps pourraient expliquer l'autocorrélation observée dans certaines des séries chronologiques à l'étude (voir l'annexe B).

Par ailleurs, les processus climatologiques non stationnaires dans le temps pourraient être responsables de la présence de rupture dans plusieurs séries chronologiques des écoulements annuels observés. Ces ruptures significatives se manifestent surtout dans les bassins versants de la Côte-Nord. Plus particulièrement, Dion *et al.* (2009) montrent qu'il existe une rupture significative dans les séries des écoulements observés pour quatre bassins versants, soient les bassins des rivières Moisie, Natashquan et Romaine et le bassin de la rivière Baleine. Ces ruptures sur les écoulements observés ont d'ailleurs été confirmées dans plusieurs autres publications (Perreault *et al.*, 2000a,b; Jandhyala *et al.*, 2009). La figure 5.2 illustre les écoulements annuels du bassin versant de la rivière Moisie, située sur la Côte-Nord. De plus, l'étude de Dion *et al.* (2009) affirme que «*les ruptures observées sont en outre remarquablement bien reproduites par le MRCC sur le domaine QC*». Lorsqu'une série exhibe une ou plusieurs ruptures, elle n'est plus stationnaire puisque la loi de probabilité change au cours du temps.

Les intervalles de confiance du coefficient de variation retenus, et plus généralement tous les intervalles présentés dans ce travail, supposent notamment la stationnarité de la série. Les ruptures exhibées par quatre bassins versants ont été ignorées dans le cadre de ce mémoire. Une façon de vérifier la robustesse des intervalles de confiance pour des séries présentant des ruptures aurait été de modéliser ce comportement avec un modèle probabiliste de persistance. Des échantillons aléatoires auraient pu être générés à partir de ce modèle puis introduits dans les simulations de Monte-Carlo. Cependant, plusieurs questions découlent lors de la mise en oeuvre d'un modèle de rupture : comment modéliser la rupture, avec quelles lois, etc. ? Une approche alternative qui s'adapte à la présence de rupture est présentée à la section 5.4.2.

## 5.4 La comparaison avec d'autres approches

Les différences entre les coefficients de variations des écoulements observés et simulés ont aussi été abordées par le biais de deux autres approches différentes dans le cadre de cette maîtrise. Toutefois, ces autres démarches n'ont pas été incluses car celles-ci ne constituaient pas le coeur du mémoire. L'une d'elles consistait à tester l'égalité des coefficients de variation en utilisant trois tests statistiques d'égalité tandis que l'autre exploitait des intervalles de crédibilité bayésiens.

### 5.4.1 Les tests d'égalité des coefficients de variation

Outre les intervalles de confiance, plusieurs tests d'égalité des coefficients de variation ont été publiés. Fung et Tsang (1998), Forkman (2006) et Boyer (2007) présentent des revues de littérature ainsi que des comparaisons par simulation de certains de ces tests issus de la littérature. De ces revues de littérature, trois tests ont été retenus afin de vérifier l'égalité des coefficients de variation, soient deux tests paramétriques supposant la normalité de la loi de probabilité sous-jacente ainsi qu'un test non paramétrique ne faisant aucune hypothèse sur la distribution. Deux tests paramétriques ont été retenus car ils se basaient sur deux statistiques complètement différentes. Les tests paramétriques de Bennett (1976) modifié et de Feltz et Miller (1996) sont basés respectivement sur l'approximation de McKay (1932) et sur le développement en série de Taylor. Outre les tests paramétriques, le test de la somme des rangs au carré de Conover (1980), permettant de tester l'égalité des variances, a été adapté au coefficient de variation en procédant par un changement de variable. La méthodologie basée sur les tests d'égalité des coefficients de variation est tout à fait similaire à celle fondée sur les intervalles de confiance. Les tests ont d'abord fait l'objet d'une étude par simulation de Monte-Carlo identique à la méthodologie décrite à la section 3.4 pour déterminer si ceux-ci étaient robustes à la non normalité des séries de données. À l'instar des intervalles de confiance, les tests paramétriques surestimaient le seuil nominal de 5%. Par ailleurs, le test non paramétrique de Conover (1980) n'était pas suffisamment puissant pour tester efficacement l'égalité des coefficients de variation. Les p-valeurs de ces tests d'égalité du coefficient de variation entre les écoulements observés et simulés sont résumées au tableau 5.1. Les résultats des tests sont tout à fait cohérents avec la méthodologie des intervalles de confiance : il n'y a pas de différence significative entre les écoulements observés et simulés. Pour de plus amples informations sur la démarche, consulter Saïd *et al.* (2009).

TAB. 5.1 – P-valeurs des tests d'égalité du coefficient de variation entre les écoulements observés et simulés. La colonne de gauche énumère les 21 bassins versants étudiés. Les acronymes *BM*, *FM* et *SR* correspondent aux tests de Bennett (1976), Feltz et Miller (1996) et de Conover (1980) respectivement.

| bv  | BM   | FM   | SR   |
|-----|------|------|------|
| ARN | 0,17 | 0,17 | 0,64 |
| BAL | 0,24 | 0,24 | 0,58 |
| BEL | 0,33 | 0,33 | 0,93 |
| BOM | 0,86 | 0,86 | 0,63 |
| CAN | 0,66 | 0,66 | 0,61 |
| CHU | 0,67 | 0,67 | 0,58 |
| FEU | 0,52 | 0,52 | 0,65 |
| GEO | 0,58 | 0,58 | 0,91 |
| GRB | 0,21 | 0,21 | 0,11 |
| LGR | 0,38 | 0,38 | 0,47 |
| MAN | 0,44 | 0,44 | 0,94 |
| MEL | 0,60 | 0,60 | 0,88 |
| MOI | 0,76 | 0,77 | 0,34 |
| NAT | 0,97 | 0,97 | 0,43 |
| PYR | 0,86 | 0,86 | 0,58 |
| RDO | 0,54 | 0,54 | 0,51 |
| ROM | 0,44 | 0,44 | 0,42 |
| RUP | 0,40 | 0,40 | 0,20 |
| SAG | 0,68 | 0,68 | 0,61 |
| STM | 0,59 | 0,59 | 0,87 |
| WAS | 0,39 | 0,40 | 0,23 |

### 5.4.2 Les intervalles de crédibilité bayésien

L'égalité des coefficients de variation des écoulements a également été testée par une approche bayésienne<sup>2</sup> développée par Évin *et al.* (2009). Le principal avantage de cette méthode réside dans le fait qu'elle ne suppose pas une seule distribution pour les données. En effet, cette dernière approche ajuste de façon bayésienne neuf modèles probabilistes précédemment définis. Ces neuf modèles sont issus de trois familles de lois différentes : les familles normale, gamma et Gumbel. Pour chacune de ces familles, trois modèles ont été considérés, la loi simple, les mélanges de deux lois de la même famille où toutes les données sont indépendantes et le mélange de deux lois de même famille avec persistance markovienne. L'ajustement de ces modèles sur les données s'effectue par l'échantillonnage de Gibbs à l'aide des distributions conditionnelles complètes *a posteriori* de chacun des paramètres des modèles. Dans le cas du mélange de lois Gumbel avec persistance markovienne, il était impossible d'obtenir une expression analytique pour la distribution conditionnelle complète *a posteriori* pour le paramètre de précision. Par conséquent, l'algorithme de Metropolis-Hastings a été appliqué pour ajuster ce modèle sur les données. Le choix du modèle s'est effectué en comparant les facteurs de Bayes de ces modèles. Le facteur de Bayes représente en quelque sorte la probabilité que les observations soient issues du modèle considéré. Par conséquent, un modèle avec rupture peut être choisi si son facteur de Bayes est supérieur par rapport aux autres modèles. Les intervalles de crédibilité sur le coefficient de variation sont construits à l'aide des 30 000 itérations des chaînes de Markov Monte-Carlo. Les intervalles de crédibilité constituent des intervalles construits à l'aide des quantiles échantillonnaires des distributions empiriques obtenues par chaîne de Markov Monte-Carlo sur le coefficient de variation. Les modèles probabilistes choisis pour un bassin versant peuvent être différents pour les observations et les simulations. Les intervalles de crédibilité du coefficient de variation sont construits avec les paramètres du modèle sélectionné. Dans le cas où les modèles choisis sont différents pour les observations et les simulations, la comparaison des coefficients de variation demeure néanmoins possible. Les résultats de cette approche, illustrés à la figure 5.3, sont tout à fait cohérents avec ceux de la section 4.2 : tous les intervalles de crédibilité se recoupent. Donc, il n'existe pas de différence de variabilité relative entre les écoulements annuels observés et simulés. Les détails de la méthodologie sont disponibles dans l'article de Évin *et al.* (2009).

La méthodologie de cette approche est complètement différente de celle du mémoire. En effet, la méthodologie du mémoire tirait profit de la robustesse des intervalles de confiance à la violation de l'hypothèse de normalité tandis que cette dernière approche ajuste le modèle probabiliste le plus adéquat sur les données pour effectuer l'inférence.

---

<sup>2</sup>dans le cadre d'un stage à l'IREQ supervisé par James Merleau et Luc Perreault

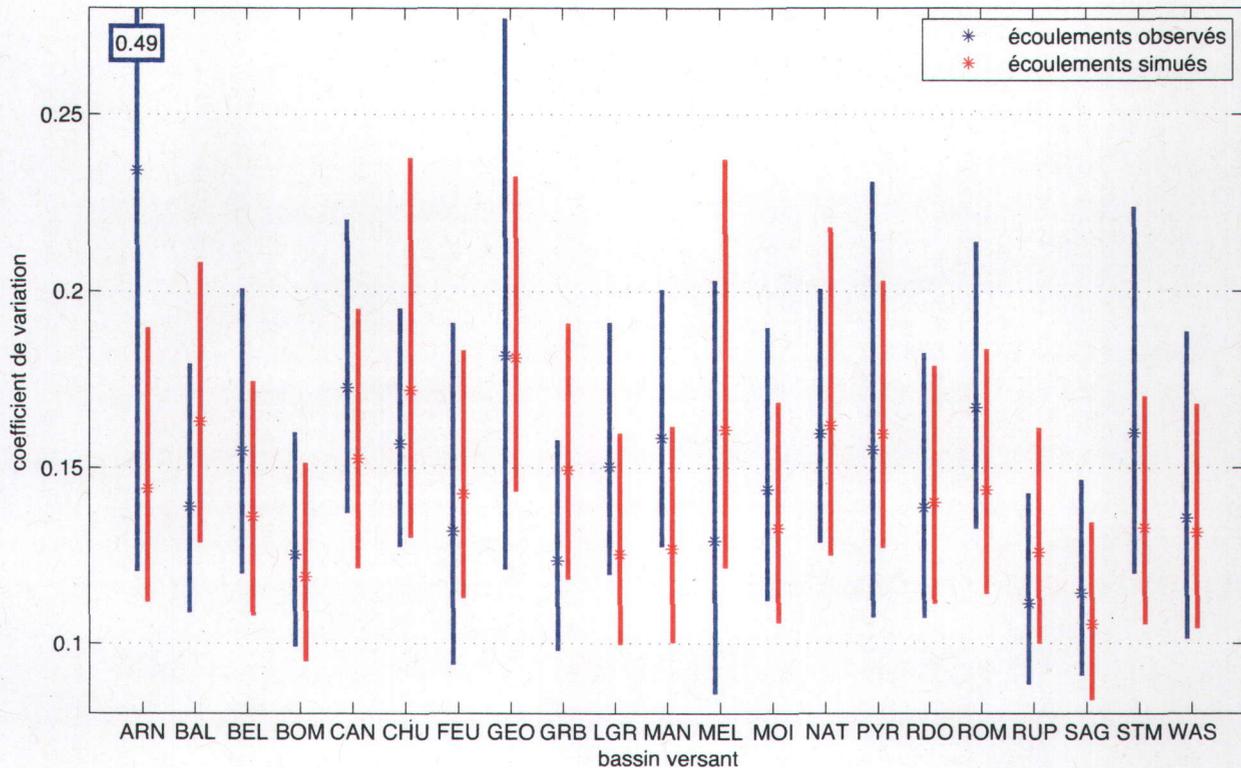


FIG. 5.3 – Intervalle de crédibilité bayésien de niveau 95% des coefficients de variations des écoulements observés et simulés par le MRCC (acy) pour chacun des bassins versants étudiés. Les bandes symbolisent les intervalles de crédibilité tandis que les astérisques indiquent les valeurs des coefficients de variation.

Même si les concepts et les méthodologies de ces deux méthodes sont différents, les deux approches mènent à la conclusion qu'il n'y a pas de différence significative des coefficients de variation entre les observations et les données simulées par le MRCC.

## 5.5 La complémentarité avec d'autres approches

Les résultats de ce mémoire complètent bien certains travaux antérieurs sur la capacité du Modèle Régional Canadien du Climat à reproduire l'hydrologie observée. Les travaux de Music *et al.* (2009) ont montré que la simulation acy du MRCC reproduisait sans biais significatif les moyennes annuelles des écoulements observés pour 16 des 21 bassins versants de la figure 3.1 pour les années 1961 à 1999. Selon cette étude, le modèle semble biaisé pour les bassins des rivières Georges, Moisie et Natashquan ainsi

que pour les réservoirs Bermis-Manic-Outardes et Manic<sup>3</sup>. Malgré que les écoulements annuels de ces bassins soient biaisés, les résultats de ce mémoire montrent que les variabilités relatives simulées sur ces bassins ne sont pas significativement différentes de celles observées, dans la mesure des limites mentionnées de la méthode.

D'autres travaux ont été publiés concernant la capacité du Modèle Régional Canadien du Climat à reproduire les valeurs moyennes des écoulements et des précipitations observés, mais les tailles des bassins versants sont trop différentes pour une comparaison directe avec l'approche de ce mémoire. Par exemple, l'approche de Music et Caya (2007) et Music et Caya (2009) s'est effectuée sur trois grands bassins versants Nord-Américain (Mississippi, Mackenzie et le St-Laurent) et celle de Gagnon *et al.* (2009) où la taille des bassins était très petite (bassins de la Chaudière et de la Châteauguay). Dans le même ordre d'idées, Frigon *et al.* (2007) ont montré que le modèle était apte à reproduire les écoulements observés d'un composite de 10 des 21 bassins versants de la figure 3.1.

---

<sup>3</sup>Les bassins versants biaisés ne sont pas spécifiquement identifiés. Cette dernière liste de bassins est celle que nous pensons être la plus probable d'être biaisée en se basant sur la figure 4 de Music *et al.* (2009)



# Chapitre 6

## Conclusion

Avant de conclure, rappelons d'abord le but du projet de recherche qui consistait à déterminer si le MRCC reproduisait la variabilité annuelle observée des précipitations et des écoulements à l'échelle de 21 bassins versants québécois. Pour ce faire, le coefficient de variation a été la statistique retenue notamment parce que celle-ci possède la propriété d'adimensionnalité. La méthodologie utilisée explicitement dans ce travail de recherche fut l'inférence statistique par intervalle de confiance, à la fois paramétrique et non paramétrique, appliquée au coefficient de variation. La significativité d'une différence de coefficient de variation était décrétée si les deux intervalles associés à chacune des statistiques ne se recoupaient pas. Dans le cadre de ce travail, trois intervalles issus de la littérature ont été retenus, soient deux intervalles paramétriques supposant la normalité de la loi de probabilité sous-jacente ainsi qu'un intervalle non paramétrique. Ces intervalles ont d'abord été testés empiriquement par une méthode de simulation de Monte-Carlo afin d'évaluer le comportement de ces intervalles lorsque l'hypothèse de normalité n'est pas respectée. Les intervalles paramétriques sous-estimaient le niveau de confiance nominal fixé à mesure que l'aplatissement de la distribution augmentait. Quant à l'intervalle non paramétrique, il surestimait le niveau de confiance nominal lorsque la taille de l'échantillon était inférieure à 100. Les comportements complémentaires de ces intervalles sous diverses conditions, notamment la taille des échantillons et le kurtosis, ont permis de prendre une décision concernant l'égalité des coefficients de variation, même lorsque les distributions sous-jacentes ne pouvaient être supposées normales.

Pour les écoulements annuels, tous les intervalles de confiance des coefficients de variation des écoulements observés et simulés par le MRCC se recourent. Par conséquent, la variabilité relative des écoulements annuels de la simulation acy du MRCC n'est pas significativement différente de la variabilité relative des écoulements observés.

La comparaison des précipitations moyennes annuelles est plus difficile en raison du manque de données pluviométriques dans le Nord du Québec. En effet, pour les bassins versants nordiques, les données pluviométriques sont très rares voire inexistantes, ce qui est à l'origine des différences significatives entre les sources de réanalyses. Dans ce contexte où les réanalyses sont différentes, le niveau de confiance de la comparaison du modèle climatique aux données de précipitations réanalysées n'est pas très grande. Pour les bassins versants du Sud du Québec, la densité d'observations pluviométriques est suffisante pour comparer adéquatement les résultats du modèle aux précipitations réanalysées. La variabilité relative des précipitations générées par la simulation acy du MRCC n'est pas significativement différente des précipitations annuelles observées pour le sud du Québec. Autrement dit, les résultats obtenus en comparant les coefficients de variation permettent de conclure que le MRCC arrive à reproduire la variabilité relative des écoulements observés des 21 bassins versants et à reproduire la variabilité relative des précipitations du Sud du Québec.

Les limites de la méthode proposée concernent essentiellement les hypothèses suivantes :

- les données doivent être indépendantes,
- la distribution sous-jacente des données doit être normale pour les intervalles de confiance paramétriques.

L'indépendance des données annuelles a été justifiée par le fait que les 21 bassins versants étudiés ne possédaient pas une assez grande capacité de rétention pour que les précipitations d'une année influencent les écoulements des années subséquentes. Cependant, les processus générant les précipitations, moteur de l'écoulement, ne sont probablement pas indépendants. Par exemple, les téléconnexions climatiques comme *El Niño* pourrait induire un cycle multi-annuel des précipitations pour certaines régions, et conséquemment, les écoulements de cette région hydrographique. Les précipitations et, par le fait même les écoulements, pourraient être autocorrélés d'une année à l'autre. Toutefois, cette corrélation ne semble pas problématique pour la méthodologie de ce mémoire puisque les coefficients de variation estimés ne sont pas différents de ceux estimés avec une approche de persistance markovienne qui tient compte de l'autocorrélation (voir la section 5.4.2).

Par ailleurs, l'hypothèse de normalité pourrait elle aussi être mise en doute. En effet, pour plusieurs bassins versants, la présence de valeurs extrêmes dans les séries contamine la normalité de ces séries. Dans ce cas, à défaut d'effectuer des simulations de Monte-Carlo, il serait sans aucun doute possible d'ajuster d'autres lois de probabilité à queues plus lourdes. À partir de ces lois, la distribution du coefficient de variation pourrait être déterminée pour l'élaboration d'un intervalle de confiance. Remarquons d'ailleurs que le coefficient de variation de la loi gamma ne dépend que de l'inverse du

paramètre de précision. Cependant, l'inférence sur le coefficient de variation étant déjà assez complexe dans le cas normal, elle se complexifierait davantage pour d'autres lois.

L'approche de comparaison de la variabilité relative par le biais du coefficient de variation développée dans ce mémoire pourrait être régionalisée en fonction des comportements régionaux des bassins versants. En effet, plusieurs bassins versants pourraient être regroupés en raison de leur corrélation spatiale (Goussanou, 2008). Dans ce contexte, le coefficient de variation pourrait être utilisé pour tester si les variabilités relatives des groupes de bassins versants sont significativement différentes dépendamment de la région où ils se trouvent. La régionalisation des statistiques permet d'accroître le niveau de confiance de l'inférence associée puisqu'en regroupant les bassins versants, une plus grande quantité d'information se retrouve disponible pour estimer les statistiques d'intérêts. La prise en compte de la corrélation spatiale permettrait de tracer un portrait géographique des variabilités relatives à l'échelle du territoire québécois. Néanmoins, il y aurait un gain pour l'inférence uniquement si la corrélation pouvait être modélisée, ce qui est loin d'être trivial.

Par ailleurs, cette approche pourrait aussi prendre en compte les ruptures exhibées dans certaines des séries à l'étude. Dans ce travail, l'hypothèse de stationnarité des observations a été émise mais ce ne sont pas toutes les séries qui sont stationnaires. Par exemple, les bassins de la Côte-Nord admettent une rupture de la moyenne dans les années 1980. Les modèles utilisés devraient alors être assez souples pour représenter ce comportement. Par exemple, un modèle de mélange de lois avec persistance markovienne pourrait être utilisé pour modéliser les ruptures dans les séries. La prise en compte de la corrélation spatiale des bassins versants pourrait dans ce cas accroître la crédibilité d'une potentielle rupture dans les séries d'un groupe de bassins versants d'une région donnée.

D'autre part, la méthode pourrait être généralisée pour des échelles temporelles inférieures qu'à celle annuelle. Cependant, cette diminution d'échelle temporelle nécessiterait la prise en compte de la dépendance temporelle des séries de données.

Finalement, ce travail de recherche a démontré que le Modèle Régional Canadien du Climat reproduit la variabilité relative observée des écoulements et des précipitations annuels. Ce travail constitue par conséquent une validation supplémentaire du modèle en climat actuel, ce qui selon Smith et Hulme (1998) augmente la crédibilité des prédictions de climat futur issues de ce modèle.



# Bibliographie

- AKHTAR, M., AHMAD, N. et BOOIJ, M. J. (2009). Use of regional climate model simulations as input for hydrological models for the Hindukush-Karakorum-Himalaya region. *Hydrology and Earth System Sciences*, 13(7):1075–1089.
- ARORA, V. K., CHIEW, F. H. S. et GRAYSON, R. B. (2000). The use of river runoff to test CSIRO9 land surface scheme in the Amazon and Mississippi River Basins. *International Journal of Climatology*, 20(10):1077–1096.
- ARVESEN, J. N. (1972). Approximate confidence intervals for coefficients of variation. *Mimeograph Series number 305, Department of Statistics, Division of Mathematical Sciences, Purdue University, USA*.
- BARNDORFF-NIELSEN, O. (1986). Inference on full or partial parameters based on the standardized signed log likelihood ratio. *Biometrika*, 73:307–322.
- BARNDORFF-NIELSEN, O. (1991). Modified signed log likelihood ratio. *Biometrika*, 78:557–563.
- BARROW, E., MAXWELL, B. et GACHON, P. (2004). *La variabilité et le changement climatique au Canada : le passé, le présent et le futur*. DSCA Séries d'évaluations scientifique N°2, Service météorologique du Canada, Environnement Canada, Toronto, Ontario.
- BENNETT, B. (1976). On an approximate test for homogeneity of coefficients of variation. *Experientia Supplementum*, 22:169–171.
- BONETT, D. G. (2006). Confidence interval for a coefficient of quartile variation. *Computational Statistics & Data Analysis*, 50(11):2953–2957.
- BOYER, J. G. (2007). *Topics Involving the Gamma Distribution : the Normal Coefficient of Variation and Conditional Monte Carlo*. Thèse de doctorat, North Carolina State University. 203 pages.
- CAO, L. J., DONG, W. J., XU, Y. L., ZHANG, Y. et SPARROW, M. (2007). Validating the runoff from the PRECIS model using a large-scale routing model. *Advances in Atmospheric Sciences*, 24:855–862.

- CASELLA, G. et BERGER, R. (2002). *Statistical Inference*. Duxbury Press, Pacific Grove, CA, USA, seconde édition. 660 pages.
- CAYA, D. et LAPRISE, R. (1999). A semi-implicit semi-Lagrangian regional climate model : The Canadian RCM. *Monthly Weather Review*, 127(3):341–362.
- CONOVER, W. (1980). *Practical Nonparametric Statistics*. John Wiley & Sons, New-York, seconde édition. 511 pages.
- D'AGOSTINO, R. B., BELANGER, A. et D'AGOSTINO JR, R. B. (1990). A suggestion for using powerful and informative tests of normality. *The American Statistician*, 44(4):316–321.
- DANKERS, R., CHRISTENSEN, O. B., FEYEN, L., KALAS, M. et de ROO, A. (2007). Evaluation of very high-resolution climate model data for simulating flood hazards in the Upper Danube Basin. *Journal of Hydrology*, 347(3-4):319–331.
- DAVID, F. N. (1949). Note on the application of Fisher's k-statistics. *Biometrika*, 36:383–393.
- DION, K., SLIVITZKY, M., SAÏD, M. et FAVRE, A.-C. (2009). Analyses des ruptures et des tendances dans les écoulements annuels observés et simulés par le MRCC en climat actuel (1961-1999) pour les 21 bassins versants de la péninsule Québec/Labrador. Rapport technique, Institut national de la recherche scientifique, Centre Eau, Terre et Environnement. Québec (Qc), Canada, 55 pages.
- DUCHARNE, A., GOLAZ, C., LEBLOIS, E., LAVAL, K., POLCHER, J., LEDOUX, E. et de MARSILY, G. (2003). Development of a high resolution runoff routing model, calibration and application to assess runoff from the LMD GCM. *Journal of Hydrology*, 280(1-4):207–228.
- DUCHARNE, A., LAVAL, K. et POLCHER, J. (1998). Sensitivity of the hydrological cycle to the parametrization of soil hydrology in a GCM. *Climate Dynamics*, 14(5):307–327.
- ÉVIN, G., MERLEAU, J. et PERREAULT, L. (2009). Mixtures of normal, gamma, and Gumbel distributions. *To be submitted to the Journal of the Royal Statistical Society : Series B (Statistical Methodology)*.
- FELTZ, C. J. et MILLER, G. E. (1996). An asymptotic test for the equality of coefficients of variation from  $k$  populations. *Statistics in Medicine*, 15(6):647–658.
- FIORI, A. M. et ZENGA, M. (2009). Karl Pearson and the origin of kurtosis. *International Statistical Review*, 77(1):40–50.
- FISHER, R. (1925). Applications of Student's distribution. *Metron*, 5(3):90–104.

- FORKMAN, J. (2006). Statistical inference for the coefficient of variation in normally distributed data. Research Report 2006 :2, Centre of Biostochastics, Swedish University of Agricultural Sciences. 29 pages.
- FOWLER, H. J., BLENKINSOP, S. et TEBALDI, C. (2007). Linking climate change modelling to impacts studies : recent advances in downscaling techniques for hydrological modelling. *International Journal of Climatology*, 27(12):1547–1578.
- FRIGON, A., SLIVITZKY, M., CAYA, D. et ROY, R. (2007). Construction de projections hydro-climatiques et leurs incertitudes associées à partir de simulations issues de modèles régionaux de climat : application à la gestion des eaux des réservoirs hydroélectriques du Québec. *La Houille Blanche*, pages 97–103.
- FUNG, W. K. et TSANG, T. S. (1998). A simulation study comparing tests for the equality of coefficients of variation. *Statistics in Medicine*, 17(17):2003–2014.
- GAGNON, P., KONAN, B., ROUSSEAU, A. N. et SLIVITZKY, M. (2009). Hydrometeorological validation of a Canadian Regional Climate Model simulation within the Chaudière and Châteauguay watersheds (Quebec, Canada). *Canadian Journal of Civil Engineering*, 36(2):253–266.
- GIEC (2001). Bilan 2001 des changements climatiques : Les éléments scientifiques. Rapport technique, Rapport du Groupe de travail I du GIEC. Genève, Suisse, 97 pages.
- GIEC (2008). Bilan 2007 des changements climatiques. Rapport technique, Contribution des Groupes de travail I, II et III au quatrième Rapport d'évaluation du Groupe d'experts intergouvernemental sur l'évolution du climat [Équipe de rédaction principale, Pachauri, R.K. et Reisinger, A.]. Genève, Suisse, 103 pages.
- GOUSSANOU, A. (2008). Modélisation de la relation entre les écoulements issus du Modèle régional canadien du climat (MRCC) et les observations en tenant compte de la dépendance spatiale. Mémoire de maîtrise, Université Laval, Faculté des sciences et de génie. 80 pages.
- HACHÉ, M., PERREAULT, L. et BOBÉE, B. (1996). Rapport final du Projet C1 - Validation des apports naturels historiques. Rapport technique No R-478, INRS-Eau. 222 pages.
- HAY, L. E., CLARK, M. P., WILBY, R. L., GUTOWSKI, W. J., LEAVESLEY, G. H., PAN, Z., ARRITT, R. W. et TAKLE, E. S. (2002). Use of regional climate model output for hydrologic simulations. *Journal of Hydrometeorology*, 3(5):571–590.
- HENDRICKS, W. A. et ROBEY, K. W. (1936). The sampling distribution of the coefficient of variation. *The Annals of Mathematical Statistics*, 7(3):129–132.

- JANDHYALA, V. K., LIU, P. et FOTOPOULOS, S. B. (2009). River stream flows in the northern Québec Labrador region : A multivariate change point analysis via maximum likelihood. *Water Resources Research*, 45(W02408).
- JOHNSON, N. et WELCH, B. (1940). Applications of the non-central t-distribution. *Biometrika*, 31(3-4):362–389.
- KAY, A. L., REYNARD, N. S. et JONES, R. G. (2006). RCM rainfall for UK flood frequency estimation. I. Method and validation. *Journal of Hydrology*, 318(1-4):151–162.
- KENDALL, M. (1938). A new measure of rank correlation. *Biometrika*, 30(1-2):81–93.
- KIM, J., JUNG, H. S., MECHOSO, C. R. et KANG, H. S. (2008). Validation of a multidecadal RCM hindcast over East Asia. *Global and Planetary Change*, 61(3-4):225–241.
- KLEINN, J., FREI, C., GURTZ, J., LUTHI, D., VIDALE, P. L. et SCHAR, C. (2005). Hydrologic simulations in the Rhine basin driven by a regional climate model. *Journal of Geophysical Research-Atmospheres*, 110(D4).
- LAPRISE, R. (2008). Regional climate modelling. *Journal of Computational Physics*, 227(7):3641–3666.
- LAPRISE, R., CAYA, D., FRIGON, A. et PAQUIN, D. (2003). Current and perturbed climate as simulated by the second-generation Canadian Regional Climate Model (CRCM-II) over northwestern North America. *Climate Dynamics*, 21(5-6):405–421.
- MCKAY, A. T. (1932). Distribution of the coefficient of variation and the extended "t" distribution. *Journal of the Royal Statistical Society*, 95(4):695–698.
- MEYLAN, P., FAVRE, A.-C. et MUSY, A. (2008). *Hydrologie fréquentielle : Une science prédictive*. Presses polytechniques et universitaires romandes, Lausanne, Suisse, première édition. 184 pages.
- MITCHELL, T. D. et JONES, P. D. (2005). An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *International Journal of Climatology*, 25(6):693–712.
- MUSIC, B. et CAYA, D. (2007). Evaluation of the hydrological cycle over the Mississippi River basin as simulated by the Canadian regional climate model (CRCM). *Journal of Hydrometeorology*, 8:969–988.
- MUSIC, B. et CAYA, D. (2009). Investigation of the sensitivity of water cycle components simulated by the Canadian Regional Climate Model to the land surface parameterization, the lateral boundary data, and the internal variability. *Journal of Hydrometeorology*, 10(1):3–21.

- MUSIC, B., FRIGON, A., SLIVITZKY, M. et MUSY, A. (2009). Runoff modeling within the Canadian Regional Climate Model (CRCM) : analysis over the Quebec/Labrador watersheds. *soumis aux IAHS publications*, (333):183–194.
- MUSY, A. et HIGY, C. (2004). *Hydrologie : 1. Une science de la nature*. Presses polytechniques et universitaires romandes, Lausanne, Suisse. 314 pages.
- PEARSON, K. (1896). Mathematical contributions to the theory of evolution. III. Regression, heredity, and panmixia. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 187:253–318.
- PERREAULT, L., BERNIER, J., BOBÉE, B. et PARENT, E. (2000a). Bayesian change-point analysis in hydrometeorological time series. Part 1. The normal model revisited. *Journal of Hydrology*, 235(3-4):221–241.
- PERREAULT, L., BERNIER, J., BOBÉE, B. et PARENT, E. (2000b). Bayesian change-point analysis in hydrometeorological time series. Part 2. Comparison of change-point models and forecasting. *Journal of Hydrology*, 235(3-4):242–263.
- PIERCE, D. et PETERS, D. (1992). Practical use of higher-order asymptotics for multiparameter exponential-families. *Journal of The Royal Statistical Society Series B-Methodological*, 54(3):701–737.
- PLUMMER, D. A., CAYA, D., FRIGON, A., COTE, H., GIGUERE, M., PAQUIN, D., BINER, S., HARVEY, R. et DE ELIA, R. (2006). Climate and climate change over North America as simulated by the Canadian RCM. *Journal of Climate*, 19(13):3112–3132.
- RANDALL, D., WOOD, R., BONY, S., COLMAN, R., FICHEFET, T., FYFE, J., KATSOV, V., PITMAN, A., SHUKLA, J., SRINIVASAN, J., SUMI, R. S. A. et TAYLOR, K. (2007). Contribution du Groupe de travail I au quatrième Rapport d'évaluation du Groupe d'experts intergouvernemental sur l'évolution du climat [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M.Tignor and H.L. Miller (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. 996 pages.
- RIVINGTON, M., MILLER, D., MATTHEWS, K. B., RUSSELL, G., BELLOCCHI, G. et BUCHAN, K. (2008). Evaluating regional climate model estimates against site-specific observed data in the UK. *Climatic Change*, 88(2):157–185.
- ROADS, J., LAWFORD, R., BAINTO, E., BERBERY, E., CHEN, S., FEKETE, B., GALLO, K., GRUNDSTEIN, A., HIGGINS, W., KANAMITSU, M., KRAJEWSKI, W., LAKSHMI, V., LEATHERS, D., LETTENMAIER, D., LUO, L., MAURER, E., MEYERS, T., MILLER, D., MITCHELL, K., MOTE, T., PINKER, R., REICHLER, T., ROBINSON, D., ROBOCK,

- A., SMITH, J., SRINIVASAN, G., VERDIN, K., VINNIKOV, K., HAAR, T. V., VOROSMARTY, C., WILLIAMS, S. et YAROSH, E. (2003). GCIP water and energy budget synthesis (WEBS). *Journal of Geophysical Research-Atmospheres*, 108(D16).
- SAÏD, M. (2009). Coefficient de variation : estimation, inférence... et plus si affinités. Rapport technique, Institut national de la recherche scientifique, Centre Eau, Terre et Environnement. Québec (Qc), Canada, 42 pages.
- SAÏD, M., JALBERT, J. et FAVRE, A.-C. (2009). Coefficient de variation pour l'écoulement annuel simulé par le MRCC en climat actuel (1961–1999) pour les 21 bassins versants de la péninsule Québec/Labrador. Rapport technique, Institut national de la recherche scientifique, Centre Eau, Terre et Environnement. Québec (Qc), Canada, 45 pages.
- SAPORTA, G. (2006). *Probabilités, analyse des données et statistique*. Editions TECHNIP, 2e édition révisée et augmentée édition. 622 pages.
- SMITH, J. et HULME, M. (1998). Climate change scenarios Chapter 3 in, Handbook on methods of climate change impacts assessment and adaptation strategies (ed. Feenstra, J., Burton, I. & Smith, JB & Tol, RSJ) UNEP/IES, Version 2.0. *Amsterdam. Holland*.
- STUART, A. et ORD, J. (2004). *Kendall's advances theory of statistics, Volume I : Distribution theory*. Edward Arnold, sixième édition. London, UK.
- TUKEY, J. W. (1958). Bias and confidence in not quite large samples. *The Annals of Mathematical Statistics*, 29(2):614. Résumé d'article.
- TUKEY, J. W. (1960). A survey of sampling from contaminated distributions. *Contributions to Probability and Statistics*, pages 448–485.
- UPPALA, S. M., KÅLLBERG, P. W., SIMMONS, A. J., ANDRAE, U., BECHTOLD, V. D. C., FIORINO, M., GIBSON, J. K., HASELER, J., HERNANDEZ, A., KELLY, G. A., LI, X., ONOGI, K., SAARINEN, S., SOKKA, N., ALLAN, R. P., ANDERSSON, E., ARPE, K., BALMASEDA, M. A., BELJAARS, A. C. M., BERG, L. V. D., BIDLOT, J., BORMANN, N., CAIRES, S., CHEVALLIER, F., DETHOF, A., DRAGOSAVAC, M., FISHER, M., FUENTES, M., HAGEMANN, S., HÓLM, E., HOSKINS, B. J., ISAKSEN, L., JANSSEN, P. A. E. M., JENNE, R., MCNALLY, A. P., MAHFOUF, J.-F., MORCRETTE, J.-J., RAYNER, N. A., SAUNDERS, R. W., SIMON, P., STERL, A., TRENBERTH, K. E., UNTCH, A., VASILJEVIC, D., VITERBO, P. et WOOLLEN, J. (2005). The ERA-40 reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 131(612):2961–3012.
- VANGEL, M. G. (1996). Confidence intervals for a normal coefficient of variation. *American Statistician*, 50(1):21–26.

- VERSEGHY, D. L. (1991). CLASS-A Canadian Land Surface Scheme for GCMs. I. Soil model. *International Journal of Climatology*, 11(2):111–133.
- VERSEGHY, D. L., MCFARLANE, N. A. et LAZARE, M. (1993). CLASS - A Canadian Land-Surface Scheme for GCMs . II. Vegetation model and coupled runs. *International Journal of Climatology*, 13(4):347–370.
- WATSON, R. et the CORE WRITING TEAM (2001). *Climate Change 2001 : Synthesis Report*. Cambridge University Press. 398 pages.
- WILKS, S. (1959). *Nonparametric Statistical Inference*. Probability and Statistics : The Harald Cramér Volume. 331–354.
- WILLMOTT, C. J. et MATSUURA, K. (1995). Smart interpolation of annually averaged air-temperature in the United-States. *Journal of Applied Meteorology*, 34(12):2577–2586.
- WILLMOTT, C. J. et ROBESON, S. M. (1995). Climatologically Aided Interpolation (CAI) of terrestrial air-temperature. *International Journal of Climatology*, 15(2):221–229.
- WONG, A. C. M. et WU, J. (2002). Small sample asymptotic inference for the coefficient of variation : normal and nonnormal models. *Journal of Statistical Planning and Inference*, 104(1):73–82.
- WOOD, A. W., LEUNG, L. R., SRIDHAR, V. et LETTENMAIER, D. P. (2004). Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. *Climatic Change*, 62(1-3):189–216.



# Annexe A

## Notions fondamentales de statistique

**Définition A.0.1.** Si  $X_i \sim i.i.d.\mathcal{N}(\mu, \sigma^2)$ ,  $\forall i = 1, \dots, n$ , alors

$$Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1).$$

**Définition A.0.2.** Si  $X_i \sim i.i.d.\mathcal{N}(\mu, \sigma^2)$ ,  $\forall i = 1, \dots, n$ , alors

$$T = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim t(n-1),$$

où  $t(\nu)$  désigne la distribution de Student à  $\nu$  degrés de liberté.

**Définition A.0.3.** Si  $Z \sim \mathcal{N}(0, 1)$  et  $U \sim \chi_\nu^2$ , alors

$$\frac{Z + \delta}{\sqrt{U/\nu}} \sim t^\delta(\nu),$$

où  $t^\delta(\nu)$  est la loi de Student décentrée de paramètre de non centralité  $\delta$  dont la distribution est donné par

$$f(t) = \frac{1}{\Gamma(\frac{\nu}{2})\sqrt{\nu\pi}2^{(\nu-1)/2}} \int_0^\infty x^\nu e^{-\{x^2 + (tx\nu^{-1/2} - \delta)^2\}/2} dx, \quad t \in \mathbb{R}.$$

**Définition A.0.4.** Une variable aléatoire  $Q(X_1, \dots, X_n, \theta)$  est une quantité pivot (ou un pivot) si la distribution de  $Q(X_1, \dots, X_n, \theta)$  est indépendante de tous les paramètres, c'est-à-dire que pour  $X \sim \mathcal{F}(x|\theta)$ ,  $Q(X_1, \dots, X_n, \theta)$  suit la même distribution peu importe les valeurs de  $\theta$ . Définition tirée de Casella et Berger (2002, chap.9).



# Annexe B

## Caractéristiques des données

Le tableau B.1 répertorie les séries chronologique ayant une autocorrélation significative. L'autocorrélation a été calculé avec le tau de Kendall (Kendall, 1938).

Le tableau B.2 résume plusieurs statistiques des sources de données : le coefficient de variation  $V$ , l'asymétrie (skewness) et l'aplatissement (kurtosis) pour tous les bassins versants et pour toutes les sources de données. De plus, ces tableaux compilent également les p-valeurs brutes d'un test de normalité basé sur l'asymétrie et l'aplatissement de la loi normale (D'Agostino *et al.*, 1990) et les p-valeurs des séries où les valeurs extrêmes ont été retirées. On constate que lorsque les valeurs extrêmes sont retirées, les séries peuvent toutes être assumées tirées de lois normales.

TAB. B.1 – Liste des séries chronologiques exhibant une autocorrélation (tau de Kendall) significative. Le tableau indique aussi l'ordre de l'autocorrélation.

|                | source              | bassin versant        | ordres    |
|----------------|---------------------|-----------------------|-----------|
| écoulements    | observés            | Rivière à la Baleine  | 4         |
|                | observés            | Rivière Moisie        | 2-4-5-7-9 |
|                | observés            | Rivière Natashquan    | 1         |
|                | simulés (MRCC acy)  | Rivière à la Baleine  | 2         |
|                | simulés (MRCC acy)  | Rivière des Outaouais | 7         |
| précipitations | réanalysées (CRU)   | Rivière Arnaud        | 1         |
|                | réanalysées (CRU)   | Rivière aux Feuilles  | 7         |
|                | réanalysées (CRU)   | La Grande Rivière     | 3         |
|                | réanalysées (CRU)   | Rivière Saint-Maurice | 2-7       |
|                | réanalysées (WM2)   | Rivière Arnaud        | 2         |
|                | réanalysées (WM2)   | Rivière Saint-Maurice | 7         |
|                | simulées (MRCC acy) | Rivière des Outaouais | 7         |

TAB. B.2 – Statistiques et p-valeurs des séries annuelles de chacun des bassins versants. La p-valeur brute représente le seuil expérimental du test de normalité de D'Agostino *et al.* (1990) sur la série complète. La p-valeur représente le seuil expérimental du test lorsque les extrêmes sont retirés de la série.

(a) écoulements observés

| bv  | V    | skewness | kurtosis | p-valeur brute | p-valeur |
|-----|------|----------|----------|----------------|----------|
| ARN | 0,18 | 0,62     | 4,67     | 0,06           | 0,19     |
| BAL | 0,14 | -0,34    | 2,89     | 0,61           | 0,91     |
| BEL | 0,17 | 1,34     | 5,58     | 0,00           | 0,33     |
| BOM | 0,12 | 0,65     | 3,03     | 0,18           | 0,47     |
| CAN | 0,17 | 0,46     | 2,46     | 0,37           | 0,37     |
| CHU | 0,16 | 0,26     | 2,13     | 0,27           | 0,27     |
| FEU | 0,12 | 0,07     | 2,02     | 0,48           | 0,48     |
| GEO | 0,17 | 0,47     | 2,66     | 0,49           | 0,49     |
| GRB | 0,12 | -0,13    | 3,30     | 0,63           | 0,96     |
| LGR | 0,15 | 0,39     | 2,44     | 0,45           | 0,45     |
| MAN | 0,15 | 1,00     | 4,62     | 0,00           | 0,37     |
| MEL | 0,14 | -1,11    | 4,71     | 0,00           | 0,65     |
| MOI | 0,15 | 0,39     | 1,97     | 0,12           | 0,12     |
| NAT | 0,16 | 0,18     | 2,18     | 0,41           | 0,41     |
| PYR | 0,17 | 0,91     | 4,19     | 0,05           | 0,92     |
| RDO | 0,14 | 0,17     | 2,41     | 0,80           | 0,80     |
| ROM | 0,16 | 0,02     | 2,47     | 0,88           | 0,88     |
| RUP | 0,12 | 0,28     | 3,65     | 0,33           | 0,54     |
| SAG | 0,12 | 0,06     | 2,19     | 0,45           | 0,45     |
| STM | 0,15 | 0,45     | 3,38     | 0,33           | 0,33     |
| WAS | 0,13 | 0,52     | 3,06     | 0,34           | 0,49     |

(b) écoulements simulés

| bv  | V    | skewness | kurtosis | p-valeur brute | p-valeur |
|-----|------|----------|----------|----------------|----------|
| ARN | 0,14 | -0,19    | 3,27     | 0,61           | 0,95     |
| BAL | 0,17 | -0,04    | 3,05     | 0,86           | 0,67     |
| BEL | 0,15 | -0,35    | 1,96     | 0,07           | 0,07     |
| BOM | 0,12 | 0,20     | 2,25     | 0,47           | 0,47     |
| CAN | 0,16 | 0,43     | 2,48     | 0,43           | 0,43     |
| CHU | 0,17 | 0,65     | 3,50     | 0,11           | 0,91     |
| FEU | 0,14 | 0,10     | 3,98     | 0,28           | 0,53     |
| GEO | 0,19 | 0,15     | 3,34     | 0,60           | 0,87     |
| GRB | 0,15 | 0,02     | 2,33     | 0,68           | 0,68     |
| LGR | 0,13 | 0,23     | 2,43     | 0,67           | 0,67     |
| MAN | 0,13 | 0,64     | 2,75     | 0,21           | 0,21     |
| MEL | 0,15 | 0,04     | 4,08     | 0,25           | 0,78     |
| MOI | 0,14 | 0,64     | 2,85     | 0,21           | 0,32     |
| NAT | 0,16 | 0,75     | 3,71     | 0,05           | 0,62     |
| PYR | 0,16 | 0,14     | 2,57     | 0,88           | 0,88     |
| RDO | 0,15 | -0,40    | 2,32     | 0,35           | 0,35     |
| ROM | 0,14 | 0,45     | 2,91     | 0,42           | 0,42     |
| RUP | 0,14 | -0,45    | 2,72     | 0,44           | 0,44     |
| SAG | 0,11 | -0,54    | 2,51     | 0,30           | 0,30     |
| STM | 0,14 | 0,01     | 2,31     | 0,66           | 0,66     |
| WAS | 0,15 | -0,41    | 2,85     | 0,49           | 0,49     |

(c) précipitation CRU

| bv  | V    | skewness | kurtosis | p-valeur brute | p-valeur |
|-----|------|----------|----------|----------------|----------|
| ARN | 0,14 | 0,45     | 2,86     | 0,43           | 0,43     |
| BAL | 0,08 | -0,11    | 3,38     | 0,60           | 0,57     |
| BEL | 0,09 | 0,65     | 2,70     | 0,20           | 0,20     |
| BOM | 0,08 | -0,13    | 2,53     | 0,87           | 0,87     |
| CAN | 0,08 | -0,17    | 3,36     | 0,56           | 0,52     |
| CHU | 0,11 | -0,21    | 3,09     | 0,70           | 0,94     |
| FEU | 0,11 | 0,02     | 2,36     | 0,74           | 0,74     |
| GEO | 0,08 | -0,94    | 3,60     | 0,02           | 0,11     |
| GRB | 0,09 | -0,16    | 2,79     | 0,89           | 0,89     |
| LGR | 0,08 | 0,14     | 3,65     | 0,41           | 0,13     |
| MAN | 0,08 | -0,36    | 2,95     | 0,55           | 0,91     |
| MEL | 0,09 | 0,05     | 2,37     | 0,74           | 0,74     |
| MOI | 0,12 | 0,02     | 2,35     | 0,72           | 0,72     |
| NAT | 0,11 | 0,16     | 2,56     | 0,85           | 0,85     |
| PYR | 0,09 | 0,36     | 3,35     | 0,38           | 0,70     |
| RDO | 0,07 | -0,19    | 2,27     | 0,51           | 0,51     |
| ROM | 0,12 | 0,07     | 2,47     | 0,86           | 0,86     |
| RUP | 0,10 | 0,75     | 4,28     | 0,02           | 0,93     |
| SAG | 0,10 | 1,03     | 4,72     | 0,00           | 0,86     |
| STM | 0,09 | 0,50     | 2,61     | 0,36           | 0,47     |
| WAS | 0,09 | 0,71     | 3,55     | 0,08           | 0,56     |

(d) précipitations WM2

| bv  | V    | skewness | kurtosis | p-valeur brute | p-valeur |
|-----|------|----------|----------|----------------|----------|
| ARN | 0,21 | 0,07     | 3,31     | 0,66           | 0,55     |
| BAL | 0,15 | -0,60    | 3,62     | 0,12           | 0,43     |
| BEL | 0,09 | 0,37     | 2,06     | 0,14           | 0,14     |
| BOM | 0,10 | 0,47     | 2,36     | 0,31           | 0,31     |
| CAN | 0,13 | 0,52     | 2,97     | 0,31           | 0,31     |
| CHU | 0,11 | 0,25     | 2,48     | 0,69           | 0,69     |
| FEU | 0,17 | -0,16    | 2,23     | 0,46           | 0,46     |
| GEO | 0,14 | -0,46    | 2,76     | 0,43           | 0,43     |
| GRB | 0,11 | 0,10     | 2,24     | 0,52           | 0,52     |
| LGR | 0,10 | 0,40     | 2,43     | 0,43           | 0,43     |
| MAN | 0,12 | 0,75     | 3,37     | 0,08           | 0,35     |
| MEL | 0,14 | -0,27    | 3,00     | 0,66           | 0,66     |
| MOI | 0,11 | 0,15     | 2,27     | 0,55           | 0,55     |
| NAT | 0,10 | 0,03     | 2,23     | 0,52           | 0,52     |
| PYR | 0,14 | -0,15    | 4,05     | 0,24           | 0,71     |
| RDO | 0,08 | -0,08    | 2,20     | 0,46           | 0,46     |
| ROM | 0,12 | 0,35     | 2,55     | 0,57           | 0,57     |
| RUP | 0,11 | 0,66     | 3,51     | 0,10           | 0,10     |
| SAG | 0,09 | 0,45     | 2,54     | 0,42           | 0,42     |
| STM | 0,09 | -0,21    | 2,08     | 0,23           | 0,23     |
| WAS | 0,10 | 0,62     | 2,76     | 0,23           | 0,23     |

(e) précipitations simulées

| bv  | V    | skewness | kurtosis | p-valeur<br>brute | p-valeur |
|-----|------|----------|----------|-------------------|----------|
| ARN | 0,08 | 0,27     | 4,44     | 0,11              | 0,74     |
| BAL | 0,11 | -0,25    | 2,39     | 0,60              | 0,60     |
| BEL | 0,08 | -0,53    | 2,60     | 0,33              | 0,33     |
| BOM | 0,08 | 0,24     | 2,71     | 0,78              | 0,78     |
| CAN | 0,11 | 0,23     | 2,05     | 0,18              | 0,18     |
| CHU | 0,12 | 0,07     | 2,26     | 0,56              | 0,56     |
| FEU | 0,08 | 0,76     | 4,43     | 0,02              | 0,73     |
| GEO | 0,13 | -0,02    | 2,35     | 0,72              | 0,72     |
| GRB | 0,09 | 0,25     | 2,53     | 0,71              | 0,71     |
| LGR | 0,09 | 0,42     | 2,52     | 0,45              | 0,45     |
| MAN | 0,09 | 0,28     | 2,12     | 0,25              | 0,25     |
| MEL | 0,09 | 0,20     | 3,54     | 0,44              | 0,86     |
| MOI | 0,09 | 0,57     | 2,48     | 0,25              | 0,25     |
| NAT | 0,10 | 0,51     | 2,92     | 0,34              | 0,66     |
| PYR | 0,10 | 0,17     | 2,84     | 0,86              | 0,97     |
| RDO | 0,07 | -0,10    | 2,50     | 0,87              | 0,87     |
| ROM | 0,10 | 0,36     | 2,64     | 0,58              | 0,58     |
| RUP | 0,10 | -0,01    | 2,42     | 0,82              | 0,82     |
| SAG | 0,08 | -0,18    | 2,90     | 0,83              | 0,83     |
| STM | 0,08 | 0,26     | 2,80     | 0,74              | 0,91     |
| WAS | 0,09 | -0,33    | 3,29     | 0,44              | 0,81     |

