

Université du Québec
INRS-ETE

HOMOGÉNÉISATION DES SÉRIES DE
PRÉCIPITATIONS : IDENTIFICATION DES TECHNIQUES
LES PLUS PROMETTEUSES ET NOUVEAUX
DÉVELOPPEMENTS

Par
Claudie Beaulieu

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

Jury d'évaluation :

Examinatrice externe	Sophie Dabo-Niang Université Lille-3
Examineur externe	Paul J. Pilon Commission Mixte Internationale
Examinatrice interne	Monique Bernier INRS-ETE
Co-directeur de recherche	Ousmane Seidou Université d'Ottawa
Directeur de recherche	Taha B.M.J. Ouarda INRS-ETE

Thèse présentée le 23 mars 2009

Résumé

Le terme *homogénéisation* désigne la correction de séries de données climatiques contaminées par des ruptures artificielles dues à des modifications dans les réseaux d'observations (déplacement de station, changement d'instrument de mesure, changement dans l'environnement immédiat d'une station, changement d'observateur, etc.). Ces ruptures artificielles sont présentes dans la plupart des enregistrements climatiques et peuvent interférer avec les variations réelles du climat. La détection et la correction de ces ruptures est nécessaire pour construire des bases de données climatiques qui serviront ultérieurement à analyser le signal climatique et suivre son évolution dans le temps. Omettre de corriger ces ruptures artificielles implique de risquer que les projections futures basées sur les observations passées soient biaisées.

Dans cette thèse, deux objectifs généraux sont visés : déterminer les techniques les plus prometteuses pour l'homogénéisation des précipitations totales dans la province de Québec et ses alentours et proposer de nouvelles méthodes plus flexibles et/ou plus performantes que celles qui existent. La flexibilité additionnelle recherchée inclut entre autres la détection simultanée de plusieurs sauts, la relaxation de l'hypothèse de normalité sous-jacente à la plupart des méthodes existantes et l'inclusion de l'information additionnelle provenant des métadonnées et du jugement de l'expert dans l'analyse possible avec l'approche bayésienne.

Différentes techniques ont été comparées dans une étude de simulation de Monte Carlo afin d'identifier celles qui semblent prometteuses pour homogénéiser les séries de précipitations totales annuelles de la province de Québec, au Canada, et ses alentours, compte-tenu de leurs caractéristiques spatiales et temporelles. Cette comparaison a été effectuée sur des séries synthétiques homogènes et inhomogènes. Les séries synthétiques ont été générées à partir des caractéristiques statistiques des enregistrements d'un ensemble de stations situées dans la province de Québec et ses

alentours. Parmi les techniques classiques, celles basées sur le test du rapport de vraisemblance ont donné les meilleurs résultats.

Trois méthodes bayésiennes de détection de rupture ont été adaptées et appliquées au problème. La première est une technique de régression bayésienne qui a une formulation plus générale que les autres modèles de régression utilisés auparavant pour vérifier l'homogénéité d'une série. La deuxième est une autre approche régressive bayésienne qui permet de détecter un nombre inconnu de sauts et leurs positions simultanément. La troisième méthode bayésienne est un test d'homogénéité permettant de détecter un changement dans les paramètres d'une distribution appartenant à la famille exponentielle. Dans le cas où les observations sont normales, le test permet de vérifier la présence d'un changement de moyenne et de déterminer sa position. Le cas normal de ce test a été utilisé. Ces trois techniques ont été comparées aux précédentes sur les mêmes séries synthétiques. La méthode bayésienne qui permet de détecter un nombre inconnu de sauts et celle du test d'homogénéité bayésien se sont montrées très prometteuses pour détecter des inhomogénéités dans des séries de précipitations totales annuelles.

Une technique plus générale qui permet de détecter des inhomogénéités dans des séries de précipitations totales a été proposée. Le cas gamma du test permettant de détecter un changement dans les paramètres d'une distribution appartenant à la famille exponentielle a été repris. L'avantage est que la distribution gamma est plus flexible que la normale. Elle permet de représenter différents niveaux d'asymétrie et en même temps, d'approximer la distribution normale. Un cas d'étude a été réalisé et montre que cette technique est appropriée pour détecter des inhomogénéités dans les précipitations à différents pas de temps. Enfin, cette technique de détection de changement a une formulation générale qui peut s'étendre à plusieurs variables climatiques ayant une distribution qui appartient à la famille exponentielle (loi normale, gamma, Poisson, ...).

Avant propos

Cette thèse présente les travaux de recherche menés au cours de mes études doctorales. La structure de la présente thèse suit la structure standard des thèses par articles de l'INRS-ETE. La première partie de la thèse comporte une synthèse générale des travaux effectués. Cette synthèse a pour objectif de survoler les principaux résultats obtenus au cours de la thèse et de discuter de leur portée. La deuxième partie de la thèse contient cinq articles publiés, soumis ou sur le point d'être soumis à des revues avec comité de lecture. Les résultats non inclus dans les articles sont présentés en Annexe.

Articles et contribution de chaque auteur

- [1] **Beaulieu, C.**, Ouarda, T.B.M.J., Seidou, O. (2007) Synthèse des techniques d'homogénéisation des séries climatiques et analyse d'applicabilité aux séries de précipitations. *Hydrological Sciences-Journal-des Sciences Hydrologiques* **52** (1), 18-37.
- [2] **Beaulieu, C.**, Seidou, O., Ouarda, T.B.M.J., Zhang, X., Boulet, G., Yagouti, A. (2008) Intercomparison of homogenization techniques for precipitation data. *Water Resources Research* **44** (2), W02425.
- [3] **Beaulieu, C.**, O. Seidou, T.B.M.J. Ouarda, Zhang, X. (2009a) Intercomparison of homogenization techniques for precipitation data continued: the comparison of 2 recent bayesian changepoint models. Accepté dans *Water Resources Research*.
- [4] **Beaulieu, C.**, Ouarda, T.B.M.J., Seidou, O. (2009b) A Bayesian Normal Homogeneity Test for the detection of artificial discontinuities in climatic series. En révision à *International Journal of Climatology*.
- [5] **Beaulieu, C.**, Ouarda, T.B.M.J., Seidou, O. (2009c) A Bayesian Gamma Homogeneity Test for the detection of artificial discontinuities in precipitation series. À soumettre.

Dans le premier article, C. Beaulieu a effectué la revue de littérature sur les techniques d'homogénéisation des données climatiques et a rédigé l'article. T.B.M.J. Ouarda et O. Seidou ont suggéré des articles à inclure dans la revue de littérature et ont révisé l'article. L'idée originale provient de T.B.M.J. Ouarda.

Dans le deuxième article, le choix des méthodes à inclure dans la comparaison et du modèle de génération des séries synthétiques a été fait conjointement par C. Beaulieu, T.B.M.J. Ouarda, G. Boulet et A. Yagouti. La génération des séries synthétiques, l'étude de simulation et la rédaction de l'article ont été réalisés par C. Beaulieu. O. Seidou a programmé une des méthodes qui fait partie de l'étude de comparaison. O. Seidou, T.B.M.J. Ouarda et X. Zhang ont révisé la version finale du manuscrit. L'idée originale provient de T.B.M.J. Ouarda et G. Boulet.

Dans le troisième article, C. Beaulieu a réalisé l'étude de simulation, l'application au cas réel et a rédigé l'article. O. Seidou a aidé dans la programmation des codes qui ont servi à l'étude de simulation. T.B.M.J. Ouarda et O. Seidou ont fourni leurs commentaires durant l'exécution du travail. T.B.M.J. Ouarda, O. Seidou et X. Zhang ont révisé le manuscrit. L'idée originale provient de T.B.M.J. Ouarda.

Les quatrième et cinquième articles sont l'idée et le travail de C. Beaulieu. T.B.M.J. Ouarda et O. Seidou ont fourni leurs commentaires durant l'exécution du travail. T.B.M.J. Ouarda et O. Seidou ont révisé les deux manuscrits.

Remerciements

Cette thèse est le fruit de plusieurs années de travail durant lesquelles il m'a été donné de recevoir un grand support de plusieurs personnes. Je voudrais tout d'abord remercier mon superviseur Taha Ouarda et mon co-superviseur Ousmane Seidou sans qui ce travail n'aurait pu être accompli. Je dois dire que j'ai été extrêmement privilégiée puisqu'ils m'ont tous deux supervisé de manière très complémentaire et attentive. Je désire donc remercier Taha, mon directeur et mon mentor, pour avoir dirigé mon travail de thèse et pour m'avoir aidé à me préparer pour ma carrière. Tu m'as poussé à dépasser mes limites, graduellement, toujours avec un parfait dosage. Je suis consciente que si je suis plus forte et plus confiante aujourd'hui, tu es en grande partie responsable. Je tiens également à te remercier pour ta confiance, la plupart du temps plus grande que ma propre confiance, et pour m'avoir donné des conditions de travail et des opportunités exceptionnelles. Je remercie également mon co-directeur, Ousmane, pour sa patience, sa disponibilité, sa compréhension, ses encouragements et son calme si apaisant. Mon expérience au doctorat a été tellement enrichie par ta présence... J'espère sincèrement, Taha et Ousmane, pouvoir un jour vous rendre la pareille.

J'aimerais également remercier les membres de mon comité, Sophie Dabo-Niang, Paul Pilon et Monique Bernier, pour avoir accepté et pris le temps d'évaluer ma thèse, et pour avoir fait des commentaires qui ont grandement amélioré sa qualité.

Je suis également redevable à plusieurs collègues de travail : à André St-Hilaire, pour son aide, sa disponibilité et ses x lettres de recommandation (j'ai arrêté de compter), à Salaheddine El-Adlouni et Fateh Chebana, mes encyclopédies de statistique, à Karem Chokmani et Erwan Gloaguen, pour leurs précieux conseils, à Bernard Bobée, le pionnier de l'équipe d'hydrologie statistique, à Xuebin Zhang, Gilles Boulet, Abderrahmane Yagouti et Lucie Vincent pour leur collaboration à mes travaux

et à Zeljka Ristic-Rudolf pour son support. À mes futurs collègues de Princeton pour leur accueil chaleureux.

J'aimerais aussi remercier Christian Genest, qui m'a encadré au début de ma maîtrise et qui s'est retiré de mon projet lorsque j'ai fait un passage au doctorat. Je lui dois en partie la décision d'avoir fait une thèse et lui en serai toujours reconnaissante.

J'aimerais remercier le Fonds de Recherche sur la Nature et les Technologies qui m'a accordé une bourse de doctorat et qui m'a permis d'avoir une vie confortable tout au long de mes études doctorales.

Merci à mon amoureux, Vladimir, pour sa patience et sa compréhension, à toute ma famille, et à mes amis de l'INRS et d'ailleurs, qui ont toujours été là pour me remonter le moral dans les périodes plus difficiles et toujours présents pour célébrer mes réussites.

Enfin, j'aimerais remercier spécialement ma mère, pour son amour inconditionnel, sa confiance en moi, et le support moral et financier qu'elle m'a accordé tout au long de mes études. C'est à ma mère que je dédie cette thèse.

Table des matières

<i>Résumé</i>	<i>iii</i>
<i>Avant propos</i>	<i>v</i>
<i>Articles et contribution de chaque auteur</i>	<i>vii</i>
<i>Remerciements</i>	<i>ix</i>
<i>Table des matières</i>	<i>xi</i>
<i>Liste des tableaux</i>	<i>xiii</i>
<i>Liste des figures</i>	<i>xv</i>
<i>PARTIE 1 : SYNTHÈSE</i>	<i>1</i>
<i>1. Introduction</i>	<i>1</i>
1.1. Contexte	<i>1</i>
1.2. Problématique	<i>2</i>
1.3. Objectifs	<i>5</i>
1.4. Méthodologie.....	<i>6</i>
1.5. Organisation de la synthèse.....	<i>7</i>
<i>2. Revue de littérature</i>	<i>9</i>
2.1. Principales causes d'inhomogénéités	<i>9</i>
2.2. Homogénéisation	<i>11</i>
2.2.1. Métadonnées	<i>12</i>
2.2.2. Séries voisines, de référence et de comparaison.....	<i>14</i>
2.2.3. Détection de rupture	<i>16</i>
2.2.4. Ajustement	<i>24</i>
2.3. Applications faites aux séries de précipitations	<i>24</i>
2.4. Études comparatives	<i>26</i>
2.4.1. Études comparatives indépendantes	<i>26</i>
2.4.2. Effort global de comparaison	<i>29</i>
<i>3. Comparaison, adaptation et développement de méthodes d'homogénéisation</i>	<i>31</i>
3.1. Comparaison de méthodes d'homogénéisation des précipitations totales annuelles	<i>31</i>
3.1.1. Séries synthétiques.....	<i>31</i>
3.1.2. Première sélection des méthodes	<i>32</i>
3.1.3. Nouvelles techniques adaptées.....	<i>36</i>
3.1.4. Résultats.....	<i>37</i>
3.1.5. Pistes d'amélioration.....	<i>40</i>
3.2. Homogénéisation des précipitations totales avec des distributions asymétriques	<i>43</i>

3.3. Importance des hypothèses de base	45
3.3.1. Homogénéité des voisines	45
3.3.2. Indépendance	46
3.3.3. Distribution.....	47
3.3.4. Non-stationnarité	48
4. Conclusions et perspectives de recherche	49
4.1. Conclusions.....	49
4.2. Perspectives de recherche	52
4.2.1. Applications avec des distributions a priori informatives	52
4.2.2. Vérification de la performance de la technique BGHT.....	53
4.2.3. Sauts multiples.....	53
4.2.4. Autres types d'inhomogénéités	54
4.2.5. Homogénéisation des séries de vitesse du vent.....	54
4.2.6. Détection de changement dans différentes variables climatiques	55
4.2.7. Précipitations totales journalières	55
5. Références.....	57
<i>PARTIE 2 : ARTICLES.....</i>	<i>67</i>
<i>6. Article 1: Synthèse des techniques d'homogénéisation des séries climatiques et analyse d'applicabilité aux séries de précipitations.....</i>	<i>69</i>
<i>7. Article 2: Intercomparison of homogenization techniques for precipitation data.....</i>	<i>111</i>
<i>8. Article 3: Intercomparison of homogenization techniques for precipitation data continued: the comparison of two recent Bayesian changepoint models.....</i>	<i>169</i>
<i>9. Article 4: A Bayesian Normal Homogeneity Test for the detection of artificial discontinuities in climatic series</i>	<i>221</i>
<i>10. Article 5: A Bayesian Gamma Homogeneity Test for the detection of artificial discontinuities in precipitation series</i>	<i>269</i>
<i>ANNEXES.....</i>	<i>309</i>
<i>11. Annexe A : Résultats non publiés.....</i>	<i>311</i>
<i>12. Annexe B : Pseudonymes utilisés pour désigner les techniques d'homogénéisation....</i>	<i>317</i>
<i>13. Annexe C : Notation.....</i>	<i>319</i>

Liste des tableaux

TABLEAU 1. EFFET DES DIFFÉRENTES MODIFICATIONS DES CONDITIONS DE MESURE SUR DIFFÉRENTES VARIABLES CLIMATIQUES TELLES QUE LA TEMPÉRATURE MOYENNE (TMOY), LA TEMPÉRATURE MINIMALE (TMIN) ET MAXIMALE (TMAX), LES PRÉCIPITATIONS TOTALES (PTOT), LA PRESSION (PRES), L'HUMIDITÉ (HUM), LA DIRECTION DU VENT (VDIR) ET LA FORCE DU VENT (VFOR). TIRÉ DE HEINO (1997) ET DE MESTRE (2000).....	10
TABLEAU 2. PSEUDONYME ET RÉFÉRENCE DES TECHNIQUES D'HOMOGÉNÉISATION CONSIDÉRÉES LORS DE LA PREMIÈRE SÉLECTION.....	35
TABLEAU A1 : POURCENTAGE DE REJET DE L'HYPOTHÈSE D'HOMOGÉNÉITÉ SUR LES SÉRIES AVEC UN SAUT DE VARIANCE.....	313
TABLEAU A2: DIFFÉRENCE ENTRE LES POURCENTAGES DE REJET DE L'HYPOTHÈSE D'HOMOGÉNÉITÉ SUR LES SÉRIES AVEC UN SAUT DE VARIANCE ET SUR LES SÉRIES HOMOGENES.....	313
TABLEAU A3 : NOMBRE DE CAS (%) POUR LESQUELS UN SAUT EST DÉTECTÉ À L'INTÉRIEUR DE LA TENDANCE.....	314
TABLEAU A4 : NOMBRE DE CAS (%) POUR LESQUELS DEUX SAUTS OU PLUS SONT DÉTECTÉS À L'INTÉRIEUR DE LA TENDANCE.....	314

Liste des figures

FIGURE 1. PHOTOGRAPHIES DE LA STATION MONT JOLI A) PRISE AVANT 1967 B) PRISE APRÈS 1967. SOURCE : LUCIE VINCENT, ENVIRONNEMENT CANADA.	13
FIGURE 2. MÉTADONNÉES DE LA STATION QUÉBEC.	13
FIGURE 3. EXEMPLE ILLUSTRANT LA SÉPARATION DU SIGNAL CLIMATIQUE ET DES RUPTURES ARTIFICIELLES AVEC UNE SÉRIE DE COMPARAISON A) SÉRIE DE BASE AVEC TENDANCE, B) SÉRIES VOISINES AVEC TENDANCE C) SÉRIE DE COMPARAISON DE A) ET B), D) SÉRIE DE BASE AVEC TENDANCE ET SAUT, E) SÉRIES VOISINES AVEC TENDANCE ET F) SÉRIE DE COMPARAISON DE D) ET E).	16
FIGURE 4. CARTE DES STATIONS UTILISÉES POUR LA GÉNÉRATION DES SÉRIES SYNTHÉTIQUES.	32
FIGURE A1 : CRITÈRE DE PERFORMANCE MOYEN OBTENU SUR LES SÉRIES AVEC DEUX SAUTS EN FONCTION DE LA DISTANCE ENTRE DEUX SAUTS CONSÉCUTIFS.	312
FIGURE A2 : CRITÈRE DE PERFORMANCE MOYEN OBTENU SUR LES SÉRIES AVEC TROIS SAUTS EN FONCTION DE L'ORDRE DES SIGNES DES TROIS SAUTS CONSÉCUTIFS.	312

PARTIE 1 : SYNTHÈSE

1. Introduction

1.1. Contexte

La plupart des séries climatiques contiennent des ruptures artificielles introduites par des modifications des conditions de mesure (déplacement d'une station, changement d'observateur, remplacement d'un instrument de mesure, modification de l'environnement immédiat d'une station, etc.). Ces ruptures artificielles peuvent être du même ordre de grandeur que le signal climatique réel. Par conséquent, l'analyse de la variation temporelle (tendance, cycle, etc.) des séries climatiques peut être biaisée par leur présence. Donc, leur détection et leur correction est importante avant de prendre des décisions basées sur les analyses de ces séries climatiques. En effet, des séries de données homogènes assurent que les changements et variations identifiés dans ces dernières sont induits par le climat (Karl *et al.*, 1993). Des techniques statistiques peuvent être utilisées pour détecter les inhomogénéités, en utilisant la théorie classique de détection de rupture (Zwiers et Von Storch, 2004).

En climatologie, une rupture artificielle, causée par des modifications des conditions de mesure, se désigne par le terme 'inhomogénéité'. Une série est dite 'homogène' lorsque les variations qu'elle contient proviennent uniquement des fluctuations du climat (Conrad, 1948). Dans le cas contraire, on dit que la série est 'inhomogène'. Le processus de détection et de correction des ruptures d'origine non climatique se nomme 'homogénéisation'. En statistique, cette terminologie possède un sens un peu différent. Tout au long de cette thèse, la terminologie propre à la climatologie sera employée.

Plusieurs techniques d'homogénéisation sont présentées dans la littérature. Des revues de littérature sur l'homogénéisation des données sont présentées par Peterson *et al.* (1998), Aguilar *et al.* (2003) et Beaulieu *et al.* (2007). La plupart du temps, la séparation du signal climatique réel et des ruptures artificielles de la série de base (celle dont on étudie l'homogénéité) se fait à l'aide de séries voisines ou d'une série de référence (fonction d'une ou de plusieurs séries voisines) climatiquement similaires à la série de base. Les séries voisines doivent être homogènes, sinon des inhomogénéités dans l'une d'elles pourraient être attribuées à la série de base. On émet l'hypothèse implicite que les stations voisines sont affectées par les mêmes fluctuations du climat que la série de base et on recherche des ruptures dans la série de base qui ne se produisent pas dans les séries voisines. Ceci permet d'éliminer l'essentiel du signal climatique et de faire ressortir uniquement les ruptures artificielles.

Pour expliquer la cause d'une inhomogénéité, évaluer la probabilité qu'une série soit inhomogène et/ou identifier les positions de discontinuités potentielles, les métadonnées peuvent être consultées. Ces dernières sont des informations historiques sur les conditions de mesure d'une station. Elles sont constituées des enregistrements de la station, des annuaires météorologiques, des fiches d'inspection, des photographies de la station et de son environnement, d'une entrevue avec la personne responsable d'une station, etc. Pour une liste détaillée avec des exemples, le lecteur peut se référer à Peterson *et al.* (1998) et à Aguilar *et al.* (2003).

1.2. Problématique

Plusieurs variables climatiques sont homogénéisées et présentées dans la littérature dont la température (Moberg et Alexandersson, 1997; Vincent *et al.*, 2001; Li *et al.*, 2004; Yagouti *et al.*, 2006), les précipitations (Hanssen-Bauer et Forland, 1994), la pression (Young, 1993), la vitesse du vent (Thomas *et al.*, 2005), les données prises par radiosonde (Lanzante et Klein, 2003; Haimberger, 2007),

etc. Des séries de précipitations totales fiables sont très importantes pour différentes applications en hydrologie et en climatologie. Elles sont utilisées comme entrant dans les modèles hydrologiques et dans les modèles climatiques qui servent à faire des projections futures du climat. De plus, des séries de précipitations totales homogènes sont essentielles pour étudier les variations du climat. Cependant, la grande variabilité spatiale des précipitations fait en sorte que ce paramètre est difficile à homogénéiser. Lorsque le réseau d'observation est peu dense et que les stations voisines sont éloignées, cela peut devenir très problématique. La performance des différentes techniques d'homogénéisation dépend des caractéristiques des données. Plusieurs études comparatives ont été réalisées afin de déterminer les méthodes d'homogénéisation les plus prometteuses pour la température (Easterling et Peterson, 1992; Ducré-Robitaille *et al.*, 2003; DeGaetano, 2006), et la pression de surface (Slonosky *et al.*, 1999). Une étude a été réalisée par Lubes-Niel *et al.* (1998) pour comparer différents tests d'homogénéité (au sens statistique) sur des séries de précipitations totales annuelles. Toutefois, les techniques les plus prometteuses pour détecter des inhomogénéités (au sens climatologique) dans les séries de précipitations totales annuelles n'ont pas été identifiées.

La plupart des techniques d'homogénéisation reposent sur plusieurs hypothèses : la normalité des données, l'indépendance, l'homogénéité des séries voisines et l'homogénéité des variances. En général, l'hypothèse de normalité permet de faciliter l'analyse statistique des observations. Dans le cas des données annuelles, cette hypothèse est souvent vérifiée, mais cela dépend des caractéristiques des données. Dans le cas des données mensuelles, cette hypothèse devient moins facile à respecter, particulièrement pour les précipitations (Von Storch et Zwiers, 1999; Wilks, 2006). Pour l'indépendance, formuler cette hypothèse ne pose pas de problème dans le cas des données annuelles. Par contre, plus le pas de temps diminue, plus cette hypothèse devient contestable (Lund *et al.*, 2007). La plupart des procédures comparent la série de base avec des séries voisines (ou une série de référence) supposées homogènes. Or, il est difficile de garantir que les séries voisines sont homogènes. En fait, des séries voisines homogènes sont souvent difficiles à obtenir, voir même

impossible. Il existe plusieurs approches pour minimiser les inhomogénéités potentielles et former une série de référence homogène (Peterson *et al.*, 1998). Les inhomogénéités provenant des séries voisines risquent tout de même d'être associées à la série de base (Menne et Williams, 2005). Enfin, les mathématiques de la plupart des méthodes reposent sur l'homogénéité des variances. Si la variance dans la série de base ou dans une des séries voisines n'est pas constante, cela pourrait biaiser la détection de ruptures. En résumé, les postulats sur lesquels reposent la plupart des techniques d'homogénéisation ne sont pas toujours vérifiés. Des méthodes plus flexibles seraient avantageuses.

En général, les techniques d'homogénéisation ont été développées pour détecter un seul saut. Cependant, la présence de plusieurs inhomogénéités est envisageable, particulièrement sur de longues séries de données. Pour remédier à ce problème, les techniques d'homogénéisation sont souvent appliquées en utilisant une procédure de segmentation afin de détecter des sauts multiples. Toutefois, une application séquentielle peut affecter la puissance des techniques puisqu'elles sont utilisées sur des segments de plus en plus petits.

Plusieurs techniques d'homogénéisation se basent sur l'information tirée des métadonnées (on ne teste et/ou corrige que les dates pour lesquelles des changements sont rapportés). Cependant, les métadonnées sont souvent incomplètes, inexistantes, erronées ou remplies d'informations non pertinentes. De ce fait, retrouver la cause d'une inhomogénéité peut devenir très délicat, particulièrement pour des données anciennes. Il se peut alors qu'une inhomogénéité détectée soit impossible à valider par les métadonnées. À l'inverse, un changement documenté n'a pas nécessairement un effet significatif sur les données. Cela justifie l'utilisation de techniques d'homogénéisation qui comparent la série de base avec une série de référence et qui peuvent détecter des ruptures artificielles non documentées. Par ailleurs, lorsque la série de base est corrigée sans connaître la cause d'une inhomogénéité, une rupture artificielle risque d'être introduite dans la série. Dans cette thèse, on s'intéressera aux méthodes bayésiennes qui ont l'avantage de pouvoir combiner l'information provenant des métadonnées et de l'analyse de rupture réalisée en utilisant des séries

voisines. En effet, l'approche bayésienne permet d'intégrer de l'information disponible sur les dates de changement qui ne serait pas utilisable en appliquant une technique classique. L'analyse bayésienne permet de mettre à profit de l'information vague ou incomplète dans l'analyse qui ne pourrait être utilisée autrement. Si l'information utilisée est erronée, il y a un risque que l'analyse soit affectée, mais le fait de combiner la loi a priori avec la vraisemblance contenue dans les observations peut faire en sorte de minimiser le risque d'obtenir des résultats biaisés. Enfin, si aucune information n'est disponible, l'analyse sera basée uniquement sur les observations.

1.3. Objectifs

Les objectifs de la thèse de doctorat sont de:

- Déterminer les techniques les plus prometteuses pour l'homogénéisation des précipitations totales annuelles dans la province de Québec et ses alentours.
- Développer des nouvelles méthodes d'homogénéisation moins restrictives, qui permettront notamment :
 - la détection simultanée de plusieurs sauts;
 - d'inclure l'information provenant des métadonnées et du jugement de l'expert dans l'analyse;
 - de relaxer l'hypothèse de normalité faite lors de l'homogénéisation d'une série.
- Évaluer la capacité de méthodes bayésiennes récemment développées à détecter des inhomogénéités et comparer leur performance à celle des méthodes classiques.

1.4. Méthodologie

Pour répondre au premier objectif de la thèse, une revue de littérature des méthodes d'homogénéisation des séries climatiques ainsi qu'une réflexion sur l'applicabilité des différentes techniques à des séries de précipitations totales annuelles a été réalisée et constitue le premier article de cette thèse (Beaulieu *et al.*, 2007). Par la suite, les méthodes qui semblaient les plus prometteuses ont été identifiées sur la base de leur potentiel d'applicabilité aux séries de précipitations totales annuelles de la province de Québec et ses alentours. Ces dernières ont été comparées en les appliquant sur plusieurs ensembles de séries synthétiques de précipitations totales annuelles homogènes, avec un saut, avec des sauts multiples, avec une tendance et avec un saut de variance. Les séries synthétiques ont été générées à partir d'un ensemble de stations situées au Québec et ses alentours avec des séries de données de qualité (peu de données manquantes, longues séries d'observations, etc.). Ces résultats sont présentés dans Beaulieu *et al.* (2008).

Le deuxième objectif de cette thèse est de proposer de nouvelles méthodes qui permettent d'intégrer différentes sources d'information dans l'analyse par une distribution a priori informative (la connaissance de l'expert, les métadonnées, etc.), de détecter plusieurs sauts simultanément et de relaxer l'hypothèse de normalité. Les approches bayésiennes sont les plus appropriées pour remplir la première condition et ont reçu une grande attention dans cette thèse. Les bases théoriques sont développées dans cette thèse, mais les exemples d'applications qui incluent la connaissance de l'expert sortent du cadre de la thèse et feront partie de travaux futurs. Au total, trois méthodes bayésiennes sont abordées. Les articles présentés dans Beaulieu *et al.* (2009a, 2009b, 2009c) sont tous en lien avec cet objectif.

Dans Beaulieu *et al.* (2009a), deux méthodes bayésiennes qui permettent de détecter une rupture dans une relation de régression linéaire ont été appliquées aux mêmes séries synthétiques que dans Beaulieu *et al.* (2008). La première permet de tester la présence et de détecter la position d'un changement de paramètre dans une relation de régression linéaire (Seidou *et al.*, 2007). La seconde

permet d'inférer un nombre inconnu de sauts et leurs positions dans une relation de régression linéaire (Seidou et Ouarda, 2007). Ces deux techniques étaient en développement lors de l'étude comparative de Beaulieu et al (2008).

Une méthode bayésienne qui permet de détecter un changement dans une distribution appartenant à la famille exponentielle est proposée. Le développement de cette méthode est présenté dans Lee (1998). Plusieurs variables climatiques ont des distributions appartenant à la famille exponentielle. L'intérêt de cette méthode est sa formulation générale et flexible qui permet d'inclure différentes sources d'information dans l'analyse. Le cas normal de cette méthode a été comparé aux autres techniques précédentes dans Beaulieu *et al.* (2009b). De plus, une étude de sensibilité aux différents paramètres a priori y est présentée. Puisque les précipitations totales saisonnières ou mensuelles et les rapports base-voisines ont souvent une distribution asymétrique, le cas gamma est proposé. Un cas d'étude est présenté et démontre que cette nouvelle méthode peut être utilisée à la fois pour détecter des inhomogénéités dans les précipitations totales annuelles, saisonnières et mensuelles (Beaulieu *et al.*, 2009c).

1.5. Organisation de la synthèse

La synthèse de cette thèse est organisée comme suit : le chapitre 2 présente une revue de littérature sur l'homogénéisation des séries climatiques en général. Le chapitre 3 présente une discussion sur les résultats obtenus dans les cinq publications scientifiques de cette thèse. Enfin, la conclusion et des perspectives de recherches futures sont présentées dans le chapitre 4.

2. Revue de littérature

Ce chapitre présente une revue de littérature générale sur l'homogénéisation des données climatiques. Tout d'abord, les différentes causes d'inhomogénéités sont présentées. Ensuite, les différentes approches pour aborder le problème sont présentées. Par la suite, les principaux résultats de quelques études comparatives de plusieurs techniques d'homogénéisation sont présentés. Enfin, des applications sur des séries réelles de précipitations sont présentées.

2.1. Principales causes d'inhomogénéités

Les principales causes d'inhomogénéités dans les séries climatologiques dépendent du paramètre qui est mesuré (tableau 1). Par exemple, des changements des heures d'observations ou de méthodes de calcul peuvent causer une inhomogénéité dans les séries de températures moyennes et d'humidité, mais n'affectent pas les précipitations ou la pression. Les ruptures dans les séries de précipitations peuvent être dues à des changements d'instrumentation, de la hauteur de l'instrument, de l'environnement immédiat de la station, de l'exposition de la station et des relocations. Les déplacements et changements d'exposition sont susceptibles d'introduire des ruptures très importantes dans les séries de précipitations. Daly *et al.* (2007) expliquent comment différents observateurs peuvent introduire des biais dans les séries de précipitations. De manière générale, la principale cause de rupture dans la plupart des séries climatologiques est probablement le déplacement de station. De plus, le déplacement d'une station est souvent accompagné d'un changement d'instrumentation, d'observateur et d'environnement (Mestre, 2000).

Tableau 1. Effet des différentes modifications des conditions de mesure sur différentes variables climatiques telles que la température moyenne (Tmoy), la température minimale (Tmin) et maximale (Tmax), les précipitations totales (Ptot), la pression (Pres), l'humidité (Hum), la direction du vent (Vdir) et la force du vent (Vfor). Tiré de Heino (1997) et de Mestre (2000).

Type de changement	Variables						
	Tmoy	Tmin, Tmax	Ptot	Pres	Hum	Vdir	Vfor
Instrumentation	-	+	+	-	++	-	++
Hauteur instrument	+	+	+	+++	+	+	++
Exposition	+	++	+++	-	+	++	++
Heures observations	+	+	-	-	+	-	+
Méthodes de calcul	++	-	-	-	+	-	+
Déplacement	++	+++	+++	-	+	++	+++
Environnement	+	+	+	-	+	+	++
Observateurs	-	-	-	-	-	-	+

Pas de problèmes majeurs : -, quelques inhomogénéités : +, ruptures importantes : ++ et ruptures très importantes : +++.

Les séries de précipitations sont particulièrement susceptibles de contenir des inhomogénéités et des biais de grande amplitude (Easterling *et al.*, 1996). En général, les mesures de précipitations sont biaisées et tendent à sous-estimer les vraies précipitations (Groisman et Legates, 1994). Les mesures de précipitations sont surtout affectées par le vent et par l'effet Venturi qui est produit au dessus du cône du pluviomètre. Donc, les changements qui font en sorte que l'exposition de la station au vent est différente sont les plus susceptibles de perturber les données. La capture des précipitations diminue avec le vent qui lui, augmente avec l'altitude, ce qui a pour effet de biaiser les mesures. Les modifications de la hauteur du capteur au dessus du sol, de la forme du capteur et de l'auvent du capteur peuvent induire une rupture dans les données, puisque le vent perturbe la mesure d'une manière différente (Groisman et Legates, 1994; Yang *et al.*, 1999). En effet, même les précipitations mesurées par deux capteurs différents placés côte-à-côte sont souvent différentes (Sevruk, 1996). Des changements mineurs tels qu'un changement de la hauteur du capteur au dessus du sol de 1 mètre peuvent introduire des inhomogénéités de très grande amplitude (Karl *et al.*, 1993). Des changements dans l'environnement de la station comme la croissance de végétation, la construction ou la démolition d'un édifice et l'urbanisation peuvent affecter les mesures de

précipitations puisque ces changements affectent la circulation du vent (Groisman et Legates, 1994). Comme la vitesse du vent et l'exposition de la station au vent augmentent avec l'altitude, les déplacements de stations avec changement d'élévation peuvent introduire des inhomogénéités dans la série.

Puisque le vent influence grandement les mesures de précipitations, il semblerait logique que la détection d'inhomogénéités se fasse relativement aux fluctuations du vent. Cependant, aux pas de temps mensuels, saisonniers et annuels, on peut faire l'hypothèse que l'effet introduit par une rupture est constant et ignorer l'effet du vent (Mestre, 2000).

2.2. Homogénéisation

Plusieurs approches sont utilisées pour l'homogénéisation des données climatiques. Certains auteurs utilisent des méthodes directes basées sur les métadonnées, des comparaisons d'instruments côte à côte ou des études statistiques de changements d'instruments (Peterson *et al.*, 1998). Ces approches ne consistent pas à détecter un changement avec une méthode statistique, elles permettent plutôt de corriger directement la série de données. Elles sont avantageuses et devraient être utilisées lorsque les informations sur le changement sont précises. Toutefois, elles sont inapplicables lorsque les informations sont vagues, incomplètes ou inexistantes. Puisque c'est souvent le cas, des méthodes d'homogénéisation dites 'indirectes', qui utilisent des séries voisines pour détecter des discontinuités, ont été développées. Des revues de littérature des différentes approches sont présentées dans Peterson *et al.* (1998), Aguilar *et al.* (2003) et Beaulieu *et al.* (2007). Dans cette thèse, on se concentre sur les méthodes indirectes. En effet, la plupart du temps, les métadonnées sont incomplètes et l'utilisation des méthodes indirectes est nécessaire pour vérifier l'homogénéité d'une série. Dans cette section, les étapes typiques à suivre pour l'homogénéisation d'une série climatique sont expliquées, à

savoir : l'analyse des métadonnées, la sélection des séries voisines, la détection de rupture et la correction des données.

2.2.1. Métadonnées

Des métadonnées fiables et complètes sont importantes pour s'assurer que les séries de données ont été recueillies dans des conditions constantes et par le fait même, que les conclusions qui découleront de leur analyse seront valides (Aguilar *et al.*, 2003). Dans le cas où les conditions d'observations ont changé avec le temps, les métadonnées permettent de retracer l'histoire de la station de mesure et ainsi, de corriger les ruptures éventuelles induites artificiellement par les changements de conditions d'observations ou encore, de nuancer les résultats d'analyses effectuées sur ces données.

Les métadonnées sont composées des enregistrements ou des photographies d'une station, de ses annuaires météorologiques, des rapports d'inspection ou encore des entrevues avec les personnes qui en sont responsables (Peterson *et al.*, 1998). La figure 1 présente un exemple de métadonnées : des photographies de la station Mont Joli, Québec, Canada. Ces photographies indiquent que la station a été déplacée. Étant donné la nature du déplacement (la station est passée du toit noir à l'herbe verte), il est possible que les données aient été affectées, particulièrement les températures. La figure 2 présente un autre exemple de métadonnées : un rapport sur les différents déplacements de la station de la ville de Québec, Québec, Canada. Cette dernière a été déplacée plusieurs fois et il est possible que certains déplacements aient introduit des ruptures artificielles dans les données. Donc, les métadonnées peuvent contenir différentes informations pertinentes sur la station.

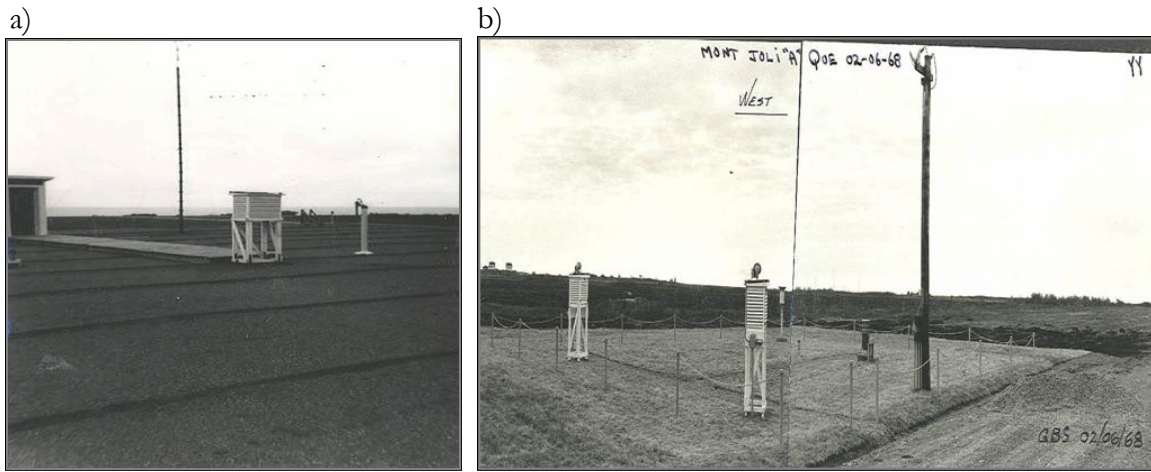


Figure 1. Photographies de la station Mont Joli a) prise avant 1967 b) prise après 1967.
Source : Lucie Vincent, Environnement Canada.

Period_Start	Period_End	SIR_No	STN_ID	Station_Nam	CLMT_ID	Date_of_Change	Lat_Degree	Lat_Minute	Lon_Degree	Lon_Minute	Station_Elevation	Instrument_Elevation	Cistern_Elevation	Comments
06-oct-43	06-oct-43	1	5251	Québec A	7016294	06-oct-43					245'		253'	
06-juil-55	09-juil-55	1	5251	Québec A	7016294	06-juil-55	46	48	72		23 245'		253'	
27-sept-58	30-sept-58	1	5251	Québec A	7016294	27-sept-58	46	48	71		23 245'		252'	The station was relocated in the
27-janv-64	11-mars-66	3	5251	Québec A	7016294	27-janv-64	46	48	71		23 245'		252'	The station was relocated in the
25-janv-67	30-janv-67	1	5251	Québec A	7016294	25-janv-67	46	48	71		23 245'		252'	Instrument site was relocated to
07-févr-68	27-févr-76	5	5251	Québec A	7016294	07-févr-68	46	48	71		23 245'	245'	252'	
18-mars-77	03-févr-84	6	5251	Québec A	7016294	18-mars-77	46	48	71		24 239'(72.8m)	239'(72.8m)	252'(76.8m)	Location and elevation change
27-févr-85	10-oct-86	3	5251	Québec A	7016294	27-févr-85	46	48	71		24 72.8m	70.0m	76.8m	No reference is made in SIR on
04-mars-91	06-mai-94	3	5251	Québec A	7016294	04-mars-91	46	47	71		23 74.1m	70.0m	76.8m	Reason unknown for
15-janv-96	19-janv-96	1	5251	Québec A	7016294	15-janv-96	46	47	71		23 74.1m	70.0m	76.8m	Station was probably

Figure 2. Métadonnées de la station Québec.
Source : Lucie Vincent, Environnement Canada.

Puisque les inhomogénéités sont dues à des changements dans les conditions d'observations, il y a toujours une cause reliée à une inhomogénéité. De ce fait, certaines méthodes d'homogénéisation n'acceptent que les ruptures artificielles qui sont documentées dans les métadonnées. Certains auteurs affirment même que les métadonnées constituent le seul outil permettant d'homogénéiser les données (Auer *et al.*, 2004). Cette approche est pertinente lorsque les métadonnées sont complètes et exemptes d'erreur. Malheureusement, elles sont souvent incomplètes et imparfaites, d'où la nécessité d'utiliser des méthodes statistiques qui permettent de détecter des ruptures (Peterson *et al.*, 1998). Par exemple, Hanssen-Bauer et Forland (1994) ont détecté des inhomogénéités de fortes amplitudes sans en identifier les causes malgré des métadonnées en

principe de bonne qualité. Dans ce cas, les métadonnées peuvent toutefois fournir de l'aide pour appuyer les résultats de certains tests d'homogénéité.

2.2.2. Séries voisines, de référence et de comparaison

L'amplitude des inhomogénéités présentes dans les séries climatiques peut être du même ordre de grandeur que celle des vraies fluctuations du climat (Aguilar *et al.*, 2003). De ce fait, la série à homogénéiser (série de base), est souvent comparée avec des séries voisines (qui appartiennent à la même région climatique). Les séries voisines doivent être homogènes et exposées aux mêmes fluctuations du climat que la série de base. L'utilisation de séries voisines n'est pas efficace lorsque toutes les stations dans la même région ont subi le même changement (Staudt *et al.*, 2007). L'utilisation de plusieurs séries voisines permet de minimiser le risque qu'elles soient toutes affectées par la même inhomogénéité.

Certains auteurs utilisent plutôt une série de référence (fonction d'une ou de plusieurs séries voisines) pour comparer avec la série de base. Une technique pour créer une série de référence homogène qui contient le même signal climatique que la série de base a été suggérée par Peterson et Easterling (1994). Ils utilisent les séries différenciées pour calculer le coefficient de corrélation entre les séries voisines et la série de base, pour éviter qu'une inhomogénéité affecte le coefficient de corrélation. Pour créer la série de référence, des poids sont attribués aux séries voisines en utilisant les coefficients de corrélation obtenus avec les séries différenciées.

Pour éliminer l'essentiel du signal climatique et faire ressortir les ruptures artificielles, une série de comparaison entre la série de base et la série de référence ou les séries voisines est produite. L'analyse de cette série permettra de décider si la station de base est homogène ou non. Par exemple, la série de comparaison peut être formée de différences ou de rapports (selon la variable étudiée) entre la série de base et les voisines. La figure 3 présente un exemple qui illustre l'utilité des séries

voisines et de base avec des séries créées artificiellement. La série de base (Fig. 3a) et les séries voisines (Fig. 3b) contiennent toutes une tendance linéaire. Or, la série de comparaison formée des rapports entre la série de base et la moyenne des trois séries voisines est stationnaire (Fig. 3c). Si un test de détection de rupture était appliqué à la série de base, un saut serait probablement détecté du à la présence de la tendance. Par contre, si le test de détection de rupture est appliqué à la série de comparaison, la série de base devrait être considérée homogène. Si la série de base contient à la fois un saut abrupt et une tendance (Fig. 3d), et que les séries voisines contiennent la même tendance que la série de base (Fig. 3e), la série de comparaison (Fig. 3f) permettra de mettre en évidence la rupture de la série de base qui ne se produit pas dans les séries voisines. Il est possible aussi que le saut (Fig. 3d) soit quand même détecté sans utiliser les séries voisines. Par contre, la tendance risque de biaiser le facteur de correction. Un modèle de régression qui représente la série de base en fonction des séries voisines ou de références peut également être utilisé pour isoler les ruptures artificielles.

Les précipitations ont une forte variabilité spatio-temporelle. De ce fait, il peut s'avérer difficile de trouver des stations voisines qui appartiennent à la même région climatique. Si les stations voisines sont trop éloignées ou d'altitude trop différente, les inhomogénéités dans la série de base risquent d'être masquées par une trop grande variabilité. La qualité de l'homogénéisation des précipitations dépend de la densité du réseau d'observations et de la topographie de la zone d'étude. De plus, le degré d'importance de ces divergences dépend du pas de temps de données à homogénéiser (Beaulieu *et al.*, 2007).

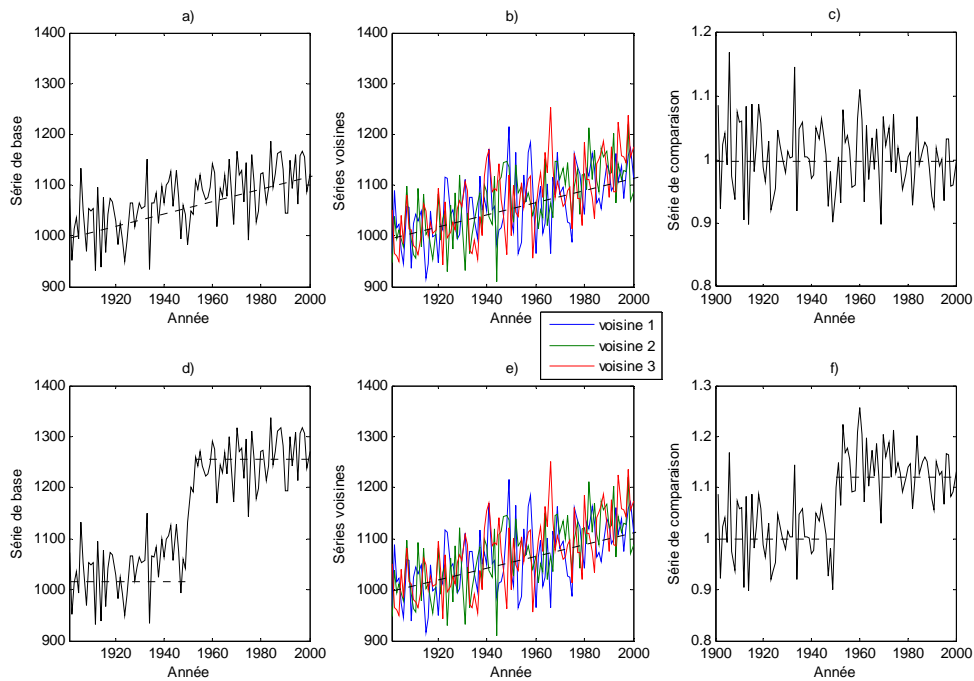


Figure 3. Exemple illustrant la séparation du signal climatique et des ruptures artificielles avec une série de comparaison a) série de base avec tendance, b) séries voisines avec tendance c) série de comparaison de a) et b), d) série de base avec tendance et saut, e) séries voisines avec tendance et f) série de comparaison de d) et e).

2.2.3. Détection de rupture

La détection de rupture peut se faire de différentes façons. La plupart du temps, elle se fait en appliquant une méthode statistique de détection de rupture sur la série de comparaison ou en ajustant un modèle de régression. Les sauts détectés sont comparés aux métadonnées et ensuite, on décide quels sauts seront identifiés comme étant des inhomogénéités. Lorsque des séries voisines ou de référence ne sont pas disponibles, la détection peut se faire directement dans la série de base.

Dans cette section, les différentes approches pour homogénéiser les données aux pas de temps annuel et mensuel sont résumées. Les détails sont présentés dans Beaulieu *et al.* (2007). Les méthodes développées pour homogénéiser les données journalières ne sont pas présentées dans cette thèse. Toutefois, le lecteur peut se référer notamment à Menne et Duchon (2001), Della-Marta et Wanner (2006) et Brandsma et Konnen (2006).

Les techniques d'homogénéisation sont classifiées en cinq groupes principaux : les méthodes graphiques, les méthodes qui sont basées sur l'hypothèse de normalité (de la série de base, de la série de comparaison ou des résidus de régression), les méthodes basées sur une distribution non normale, les méthodes non paramétriques et les méthodes hybrides. Dans le premier groupe, les méthodes permettent de détecter graphiquement une inhomogénéité. Dans les autres groupes, des techniques statistiques sont utilisées pour la détection de rupture. Les techniques d'homogénéisation sont regroupées selon le modèle de rupture. Comme la littérature statistique sur la détection de rupture est abondante, le nombre de méthodes présentées dans cette section est restreint à celles qui ont été développées spécifiquement pour faire de l'homogénéisation. Pour une revue de littérature des différentes techniques statistiques de détection de rupture, le lecteur peut se référer à Chen et Gupta (2000, 2001). Plusieurs des méthodes citées dans cette section, le sont également dans les sections subséquentes de cette thèse. Pour alléger la présentation, la première fois qu'une technique est citée, un pseudonyme lui est attribué. Par la suite, le pseudonyme est employé pour désigner la méthode. La liste complète des pseudonymes avec leurs références respectives est également présentée en Annexe B.

2.2.3.1. Méthodes graphiques

Les méthodes présentées dans cette section consistent à détecter graphiquement une inhomogénéité. La détection peut se faire sur des observations cumulées, des différences, des rapports ou encore des déviations cumulées entre la série de base et les séries voisines. Ceci a pour but d'accentuer les ruptures pour qu'elles soient facilement détectables visuellement.

L'analyse des doubles accumulations (GAC2), proposée par Kohler (1949), consiste à faire un graphique des valeurs cumulées de la série de base en fonction des valeurs cumulées de la série de référence. La droite de régression est superposée au graphique. Une rupture de la pente indique un changement dans la série de base ou dans la série de référence. La technique des ellipses de Bois

(Bois, 1986) et celle du vecteur des indices annuels (Brunet-Moret, 1979) sont des extensions de l'analyse des doubles accumulations.

L'analyse graphique des déviations cumulées (GDCU) a été proposée par Craddock (1979) pour homogénéiser des séries de précipitations totales annuelles. Un changement de pente sur le graphique indique une inhomogénéité. Pour éviter qu'un saut dans la série de référence soit identifié, Craddock (1979) répète le processus avec plusieurs séries de référence et compare les résultats. Si le même saut revient à chaque fois, il est attribué à la série de base. Jones *et al.* (1986) proposent d'examiner les séries de différences entre la série de base et ses voisines (GDIF). Rhoades et Salinger (1993) se basent sur le même principe en examinant les sommes cumulatives parallèles (GSCP).

Le filtre de Kolmogorov-Zurbenko (GKZU) a été développé pour des données non-stationnaires et affectées par une forte saisonnalité (Zurbenko *et al.*, 1996). Les régions affectées par une discontinuité se remarquent par une hausse de variabilité ou un changement de pente de la série filtrée.

2.2.3.2. Méthodes basées sur l'hypothèse de normalité

2.2.3.2.1. Un changement de moyenne

Plusieurs méthodes cherchent à détecter une rupture de la moyenne dans la série de comparaison. Le modèle peut s'exprimer ainsi :

$$q_i = \begin{cases} \mu + e_i & i = 1, \dots, p \\ \mu + \delta + e_i & i = p + 1, \dots, n \end{cases} \quad (1)$$

où q_i est la valeur de la série de comparaison au temps i , μ est la moyenne, p est la position du saut et δ son amplitude. Les résidus du modèle sont représentés par e_i . Pour détecter la position du saut, vérifier sa significativité et/ou estimer son amplitude, plusieurs techniques peuvent être employées.

Karl et Williams (1987) proposent d'identifier les positions de rupture potentielles en consultant les métadonnées, et d'appliquer un test d'égalité des moyennes local (Student ou Wilcoxon) pour vérifier si le changement est significatif (PMETA).

Lorsque la position du saut n'est pas connue, plusieurs tests de détection peuvent être utilisés. Le test basé sur le rapport des vraisemblances de Hawkins (1977) et adapté à l'homogénéisation par Alexandersson (1986) est le plus souvent utilisé (SNHT). L'emplacement du saut le plus probable est identifié en calculant une suite de statistiques, chacune associée à une position dans la série. La valeur maximale de cette suite correspond à la position qui sera testée. Cette dernière est comparée à une valeur critique obtenue par simulation (Alexandersson, 1986 ; Khaliq et Ouarda, 2007). La recherche de la valeur maximale d'une statistique pour toutes les positions successives dans la série est proposée aussi par Buishand (1982, 1984) et par Jaruskova (1996). Ces deux techniques ont les pseudonymes respectifs de BUIT et JARU. Par ailleurs, Gullett *et al.* (1990) utilisent une fenêtre mobile avec une taille fixe de 10 observations, et calculent la statistique de Student pour comparer les cinq observations qui précèdent et qui suivent chaque position dans la série (STUS). Les statistiques de Student significatives correspondent aux positions des inhomogénéités. Wang *et al.* (2007) ont développé un test de Student pénalisé pour diminuer les fausses détections dans les extrémités de la série (PSTU). Une méthode bayésienne univariée (BAUN) peut aussi résoudre ce problème (Lee et Heghinian, 1977 ; Ouarda *et al.*, 1999 ; Perreault *et al.*, 1999).

Certaines méthodes permettent de détecter une rupture dans la série de base qui ne se produit pas dans la série de référence. Le modèle est tel que :

$$y_i = \begin{cases} \mu + \beta x_i + e_i & i = 1, \dots, p \\ \mu + \delta + \beta x_i + e_i & i = p + 1, \dots, n \end{cases} \quad (2)$$

où y_i est la valeur de la série de base au temps i , x_i est la valeur de la série de référence au temps i et β est le coefficient de la série de référence.

Pour détecter la position du saut, le test bivarié (BIVT) peut être utilisé. Ce test a été développé par Maronna et Yohai (1978) et adapté à l'homogénéisation par Potter (1981). La série de base et la série de référence suivent une distribution normale bivariée. Cette technique est basée sur un test de rapport des vraisemblances. Les approches bayésiennes développées par Asselin *et al.* (1999) et par Rasmussen (2001) permettent également de résoudre ce problème. Ces deux techniques ont les pseudonymes respectifs BABI et BAYE.

2.2.3.2.2. Un changement de moyenne et/ou de tendance

Les modèles de régression linéaire permettent à la fois la détection de saut et de tendance. Vincent (1998) a proposé l'ajustement de différents modèles de régression successifs pour détecter un saut abrupt et/ou une tendance (REGM). Par exemple, le modèle suivant est utilisé pour détecter un saut :

$$y_i = \begin{cases} \tau + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + e_i & i = 1, \dots, p \\ \tau + \delta + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + e_i & i = p+1, \dots, n \end{cases} \quad (3)$$

où y_i est la valeur de la série de base au temps i , τ est l'ordonnée à l'origine, x_{ij} est la valeur de la série voisine j au temps i . Il y a n observations et k séries voisines. L'emplacement du saut, p , est déterminé en ajustant le modèle pour toutes les positions possibles et en sélectionnant celui avec la plus petite somme des carrés résiduelle. Le choix du modèle est effectué avec le test de Fisher. L'estimation de l'amplitude du saut est donnée par δ et son seuil significatif se calcule avec la statistique de Student. L'approche bayésienne de Seidou *et al.* (2007) a une formulation générale qui permet également la détection d'un saut et/ou d'un changement de tendance (BARE).

La régression à deux phases (REG2) permet également de détecter un changement de moyenne et/ou de tendance dans une série. Le modèle qui représente une série inhomogène est :

$$y_i = \begin{cases} \tau_1 + \lambda_1 i + e_i & i = 1, \dots, p \\ \tau_2 + \lambda_2 i + e_i & i = p+1, \dots, n \end{cases} \quad (4)$$

où y_i est la valeur de la série de base au temps i , τ_1 et τ_2 sont les ordonnées à l'origine avant et après le changement, λ_1 et λ_2 sont les tendances avant et après le changement et p est la position du changement. Les résidus du modèle sont représentés par e_i . L'emplacement du saut est déterminé selon les moindres carrés. Le choix entre ce modèle et celui qui représente une série homogène se fait avec le test de Fisher. La régression à deux-phases a été introduite par Hinkley (1969). Solow (1987) a proposé quelques ajustements à la méthode. Solow (1988) présente une approche bayésienne pour résoudre ce problème. Easterling et Peterson (1995) proposent d'ajuster le modèle de régression à deux-phases sur une série de comparaison. Lund et Reeves (2002) suggèrent d'utiliser une statistique de Fisher révisée. Wang (2003) propose un modèle dans lequel les pentes avant et après la rupture sont égales. Lund *et al.* (2007) reprennent ce modèle en apportant une extension pour le cas où les données sont autocorrélées et périodiques.

Enfin, un test basé sur le rapport des vraisemblances qui permet la détection d'une tendance a été développé par Alexandersson et Moberg (1997).

2.2.3.2.3. Plusieurs sauts de moyenne

En présence de sauts multiples, le problème peut se diviser en deux parties : le choix de modèle et la combinaison des ruptures.

a) Choix de modèle

Le choix du meilleur modèle pour représenter les positions des sauts et leurs amplitudes peut se faire avec des critères basés sur la vraisemblance pénalisée comme ceux développés par Akaike (1973), Schwarz (1978) ou Caussinus et Lyazrhi (1997). Les critères d'Akaike (AIC) et de Schwarz (SIC) sont communément utilisés en choix de modèle et pour sélectionner les variables explicatives en régression multiple. Le critère de Caussinus et Lyazrhi (CIC) a été conçu spécifiquement pour les modèles à rupture. La technique développée par Caussinus et Mestre (2004) est basée sur ce critère pour la sélection de modèle combinée à un algorithme de programmation dynamique pour les combinaisons de ruptures (CAME). Cette technique permet d'homogénéiser un ensemble de données. Les séries sont

tour à tour considérées comme étant la série de base et toutes les autres sont les séries voisines. Ensuite, des séries de comparaison sont formées et lorsque le même changement se produit dans toutes les séries, il est attribué à la série de base. La méthode MASH, développée par Szentimrey (1997, 1999) permet également de détecter des sauts multiples. La méthode est basée sur le principe du rapport de vraisemblances. L'hypothèse que les séries voisines sont homogènes n'est pas nécessaire.

Gullett *et al.* (1991) proposent d'ajuster sept modèles de régression qui comprennent de 0 à six inhomogénéités et de choisir le meilleur modèle avec un test de Fisher (REMU). La méthode bayésienne (BAMS) de Fearnhead (2006) permet d'inférer le nombre de sauts et leurs positions. Elle a été adaptée à la régression linéaire multiple par Seidou et Ouarda (2007).

Lorsque des métadonnées complètes sont disponibles, il est aussi possible d'appliquer des tests localement (e.g. test de Student ou de Scheffé).

b) Combinatoire

En pratique, les techniques développées pour détecter au plus un saut sont souvent appliquées de manière itérative pour détecter des sauts multiples. Cela consiste à détecter une rupture, la conserver en mémoire, en détecter une seconde, et ainsi de suite. La procédure de segmentation des séries hydrométéorologiques (SEGH) développée par Hubert *et al.* (1989) fonctionne de cette façon. L'écart quadratique entre la série et la segmentation considérée est calculé et la segmentation optimale au sens des moindres carrés est retenue. Le test des contrastes de Scheffé est appliqué pour vérifier si la segmentation est significative. Kehagias *et al.* (2006) proposent une généralisation de cette méthode. Pour des données normales, la technique de programmation dynamique de Hawkins (2001) est équivalente à la recherche de toutes les combinaisons de ruptures.

2.2.3.3. Méthodes paramétriques pour des données non normales

Thompson (1984) a développé une méthode d'homogénéisation des précipitations totales mensuelles qui n'est pas basée sur la normalité. Avec cette méthode, la fréquence des précipitations est caractérisée par un processus de Poisson et son amplitude suit une distribution exponentielle

(POIS). Le test d'homogénéité est basé sur le rapport des vraisemblances. Il compare les estimateurs du maximum de vraisemblance du modèle dans chaque segment. Cette technique ne permet pas de positionner les ruptures, seuls les sauts documentés dans les métadonnées sont testés.

2.2.3.4. Méthodes non paramétriques

Plusieurs tests non paramétriques peuvent être utilisés pour détecter des inhomogénéités. Le test de Pettitt (1979) peut être appliqué pour détecter la position de la rupture (PTTI). Si la position de la rupture est connue des métadonnées par exemple, le test de Mann-Whitney ou de Wilcoxon peut être appliqué localement (Karl et Williams, 1987). Allen et DeGaetano (2000) utilisent une approche semblable, mais sur des séries de températures extrêmes qui dépassent un certain seuil (NMETA). Par ailleurs, Lanzante (1996) suggère une application successive du test de Mann-Whitney ou de Wilcoxon (WILS) pour détecter plusieurs sauts.

2.2.3.5. Méthodes hybrides

Des méthodes hybrides qui combinent des tests paramétriques et non paramétriques ont également été proposées dans la littérature. L'idée est que ces approches hybrides permettent de réduire le nombre de fausses alarmes et d'éviter de confondre les sauts abrupts et les tendances. Tayanc *et al.* (1998) proposent d'utiliser GDIF et d'appliquer la méthode d'analyse de la variance de Kruskal et Wallis (1952) (KWAL) et le test d'indépendance de Wald et Wolfowitz (1943) (WWOL). Boroneant et Tomozeiu (1999) proposent d'appliquer successivement SNHT, PTTI et le test de tendance de Mann-Kendall (Mann, 1945 ; Kendall, 1975) (KEND). Wijngaard *et al.* (2003) proposent d'appliquer successivement SNHT, PTTI, BUIT et le test du rapport de Von Neumann (1941) (VNEU). Par la suite, les stations sont classifiées selon le nombre de tests significatifs.

La théorie des ondelettes (ONDE) a également été utilisée en homogénéisation pour ajuster des séries de températures anciennes (Datsenko *et al.*, 2002). La décomposition en base d'ondelettes permet d'ajuster les séries à différents pas de temps.

2.2.4. Ajustement

La dernière étape d'une analyse d'homogénéité consiste à ajuster les séries de données pour les rendre homogènes. Généralement, les séries de données sont corrigées par rapport aux conditions du segment le plus récent de la série. Ceci a pour but de s'assurer que les futures observations seront aussi homogènes, tant et aussi longtemps qu'il n'y a pas de nouveau changement dans les conditions de mesure de la station (Aguilar *et al.*, 2003). Les facteurs d'ajustement sont souvent obtenus en calculant une moyenne dans chacun des segments de la série de comparaison séparés par une inhomogénéité. Les segments affectés par une inhomogénéité sont corrigés avec ces facteurs. Pour les inhomogénéités de type tendance, la tendance de la série de comparaison est calculée et la série de base peut être redressée en fonction de cette dernière.

2.3. Applications faites aux séries de précipitations

Plusieurs méthodes d'homogénéisation ont été développées dans le but d'homogénéiser des séries de précipitations totales annuelles. Ces techniques incluent notamment GAC2, GDCU, BIVT, BUIT et SNHT. D'autres techniques comme JARU, MASH et CAME ont été développées pour plusieurs variables climatiques, incluant les précipitations. Cette section présente les différentes applications faites aux précipitations totales annuelles ou mensuelles.

La technique SNHT est sans aucun doute la plus utilisée. Hanssen-Bauer et Forland (1994) ont homogénéisé un ensemble de séries de précipitations totales annuelles en Norvège avec SNHT. La même technique a été utilisée en France (Lamarque et Jourdain, 1994), en Suède et en Finlande (Peterson *et al.*, 1998). En Italie, SNHT et GDCU ont été utilisés pour des précipitations totales mensuelles (Buffoni *et al.*, 1999). Dans la région d'Emilia-Romagna en Italie, l'homogénéité des précipitations totales estivales a été testée avec SNHT (Tomozeiu *et al.*, 2000). La même méthode a été appliquée sur des séries de précipitations totales mensuelles en Ibérie, en France et en Afrique du

Nord (Gonzalez-Rouco *et al.*, 2001), en Finlande (Tuomenvirta, 2001), en Allemagne (Muller-Westermeier, 2004) et en Espagne (Gonzalez Hidalgo *et al.*, 2004). En Turquie, l'homogénéité des séries de précipitations totales mensuelles a été testée avec SNHT et PTTI par Karabork *et al.* (2007) et avec la méthode SNHT uniquement dans Göktürk *et al.* (2008). Au Canada, les différents instruments qui ont été utilisés pour mesurer les précipitations sont présentés dans Metcalfe *et al.* (1997) et dans Devine et Mékis (2008). Mékis et Hogg (1999) ont réhabilité des séries de précipitations journalières au Canada pour des biais introduits par des changements d'instruments documentés dans les métadonnées. Leur approche est directe et ne nécessite pas l'utilisation de séries voisines pour comparer avec la série de base. Des comparaisons d'instruments côte-à-côte ou des études statistiques des différents instruments ont été utilisées par ces auteurs. Différents ajustements (vent, pertes par évaporation, pertes par mouillage, etc.) ont été apportés selon le type d'instrument utilisé. Par la suite, SNHT et REGM ont été appliquées pour homogénéiser des séries de précipitations totales annuelles et mensuelles pour des déplacements documentés dans les métadonnées (Mékis et Vincent, 2004).

D'autres techniques sont utilisées, selon l'approche adoptée par pays. Aux États-Unis, la méthode PMETA est appliquée pour tester si les changements documentés dans les métadonnées ont introduits des inhomogénéités dans la série. En Nouvelle-Zélande et dans les îles du pacifique, GSCP est utilisé. La méthode REG2, version de Wang (2003), a été utilisée pour homogénéiser des indices de températures et de précipitations extrêmes en Amérique Centrale et du Sud (Aguilar *et al.*, 2005). La méthode hybride de Wijngaard *et al.* (2003) a été utilisée pour détecter des inhomogénéités dans des séries de précipitations en Europe. Ces auteurs appliquent la méthode sur le nombre de jours humides par année. De plus, cette approche a été employée pour tester l'homogénéité des séries de pluie en Éthiopie (Conway *et al.*, 2004). Chez Météo-France, la méthode CAME est utilisée (Moisselin et Schneider, 2002; Moisselin *et al.*, 2002). En Suisse, les précipitations totales annuelles ont été homogénéisées avec le logiciel THOMAS qui combine 12 méthodes d'homogénéisation

différentes (Begert *et al.*, 2005). Le logiciel HOCLIS qui permet d'appliquer les méthodes SNHT, MASH, et GDCU a été utilisé pour homogénéiser des séries de précipitations totales mensuelles en Europe par Auer *et al.* (2005) et en Italie par Brunetti *et al.* (2006).

2.4. Études comparatives

Plusieurs techniques d'homogénéisation sont présentées dans la littérature. Le choix de la technique à utiliser dépend de plusieurs facteurs tels que la variable à homogénéiser, la variabilité spatiale et temporelle des données, les caractéristiques statistiques des séries, les métadonnées et la densité du réseau d'observations (Aguilar *et al.*, 2003). Étant donné le grand nombre de techniques disponibles, le choix de la technique à utiliser peut être difficile. De ce fait, plusieurs auteurs ont fait des études comparatives de plusieurs techniques d'homogénéisation, pour identifier les plus performantes selon différentes situations. Les résultats de ces études dépendent de l'effectif et des caractéristiques des séries synthétiques générées et doivent donc être interprétés avec prudence. Néanmoins, ces études permettent l'identification des techniques qui semblent les plus prometteuses dans un contexte particulier. Dans la littérature, on retrouve plusieurs études comparatives indépendantes. Dernièrement, une étude de comparaison globale a été lancée pour que toutes les méthodes d'homogénéisation disponibles dans la littérature soient comparées sur une base commune.

2.4.1. Études comparatives indépendantes

L'étude comparative la plus récente est présentée dans Reeves *et al.* (2007) sur des séries générées à partir d'une distribution normale standardisée. Les méthodes SNHT, WILS, deux versions de REG2 (Lund et Reeves, 2002 ; Wang, 2003), REGM, AIC, SIC et une nouvelle procédure proposée dans le cadre de l'étude sont comparées. La plupart des techniques comparées a été légèrement modifiée.

Une des conclusions de cette étude est qu'AIC engendre un trop grand nombre de fausses détections. Les autres techniques donnent des pourcentages de fausses détections acceptables. Lorsque la série de base contient un saut, SNHT et la nouvelle technique développée donnent les meilleurs résultats. En présence d'une tendance, REG2 et la procédure développée par Reeves *et al.* (2007) sont les plus performantes.

Plusieurs études ont été effectuées avec des séries de températures synthétiques. Ces études sont basées sur des séries synthétiques qui ont été générées avec des modèles autorégressifs et une structure de corrélation entre la série de base et les séries voisines. Easterling et Peterson (1992) ont comparé plusieurs méthodes sur des séries synthétiques de température avec un et plusieurs sauts. Les méthodes SNHT, BIVT, REG2, GAC2, GSCP et un modèle de régression univarié ont été comparées dans ce travail. Ducré-Robitaille *et al.* (2003) ont comparé huit techniques d'homogénéisation différentes sur des séries synthétiques de température. Des séries homogènes, avec un saut et avec plusieurs sauts ont été générées. Les techniques qui ont été utilisées sont SNHT, REGM, REG2 (version d'Easterling et Peterson, 1995), STUS, WILS et BAUN. Menne et Williams (2005) ont comparé SNHT, REGM et REG2 (Lund et Reeves, 2002) en utilisant différentes approches pour créer la série de référence. Enfin, DeGaetano (2006) a comparé sept techniques d'homogénéisation sur des séries synthétiques de température : SNHT, BIVT, REGM, REG2 (Lund et Reeves, 2002), BAUN, PMETA et NMETA. Les séries de référence générées ont des corrélations et des variances différentes. Des séries de base homogènes, avec un saut, avec plusieurs sauts et avec une tendance ont été générées.

Dans toutes ces études, la plupart des méthodes sont capables de détecter des sauts de grande amplitude. Dans tous les cas, SNHT se situe parmi les méthodes les plus performantes pour détecter un ou plusieurs sauts et ce, même pour les sauts ayant une faible amplitude. Menne et Williams (2005) ont montré aussi que cette technique est très sensible aux inhomogénéités dans les séries voisines. La méthode BIVT donne des résultats semblables à ceux obtenus avec SNHT dans

les deux études qui incluent cette méthode. En présence d'une tendance, ces deux techniques ne sont pas robustes. Par ailleurs, les deux modèles de régression sont les plus appropriés en présence d'une tendance. De plus, REG2 et REGM sont également parmi les plus performantes pour détecter des sauts multiples dans Ducré-Robitaille *et al.* (2003). Par ailleurs, REGM n'a pas bien performé en présence de sauts multiples dans DeGaetano (2006). La méthode BAUN a bien performé pour détecter une discontinuité, mais moins en présence de sauts multiples. Les méthodes basées sur les métadonnées PMETA et NMETA ont performé autant que SNHT et BIVT dans l'étude de DeGaetano (2006). Menne et Williams (2005) ont montré que l'application de deux tests successifs permet de diminuer le risque de fausse détection sans affecter la puissance des tests.

Une étude comparative de différentes techniques de détection de rupture a été réalisée par Lubes-Niel *et al.* (1998) sur des séries synthétiques de précipitations totales annuelles. Différents ensembles de séries synthétiques ont été générés : séries normales indépendantes, séries lognormales indépendantes, séries gamma indépendantes, séries autocorrélées et séries avec une tendance. Les méthodes comparées sont : KEND, PTTI, BAUN, PSSH et BUIT. Ces techniques ne vérifient pas toutes les mêmes hypothèses, mais le but du travail était d'étudier leur robustesse dans différentes conditions.

La méthode bayésienne BAUN a un fort pourcentage de sauts faussement détectés sur les séries homogènes. Les méthodes basées sur l'hypothèse d'indépendance (PTTI, BUIT et BAUN) des observations ne sont pas robustes en présence d'une tendance ou d'autocorrélation. Les méthodes basées sur l'hypothèse de normalité (BUIT, BAUN et PSSH) sont robustes lorsque cette hypothèse est violée. Les techniques basées sur l'hypothèse que la variance est constante sont robustes lorsque les séries contiennent des ruptures de variance et que la moyenne est constante.

Slonosky *et al.* (1999) ont comparé des méthodes d'homogénéisation sur des séries de pression moyenne en surface en Europe. La méthode graphique GDIF, la méthode qui permet de détecter des sauts multiples CAME et SNHT ont été comparées. La méthode GDIF a donné les

meilleurs résultats. La méthode SNHT s'est montrée performante lorsque de bonnes séries de référence peuvent être construites. Enfin, la méthode CAME nécessite un grand nombre de séries voisines (au moins six) pour être en mesure de fournir de bons facteurs de correction.

2.4.2. Effort global de comparaison

Plusieurs techniques d'homogénéisation ont été développées et peu d'études comparatives ont été réalisées pour identifier les techniques les plus appropriées. Il est donc nécessaire de faciliter cette comparaison de méthodes en établissant des standards de base communs et en promouvant les techniques les plus efficaces. Au cours de l'année 2008, plusieurs scientifiques européens qui travaillent sur l'homogénéisation des données ont entrepris une action commune, COST – HOME, qui vise à développer une procédure générale pour homogénéiser les données climatiques et environnementales : <http://www.homogenisation.org/>. Une procédure générale sera dérivée des techniques statistiques les plus appropriées pour détecter et corriger des inhomogénéités dans des séries climatiques à différents pas de temps et des standards seront établis. L'action consiste à recenser les différentes méthodes d'homogénéisation des séries climatiques et les comparer sur une base commune. Plusieurs groupes de travail ont été désignés dans cette action commune. Les différentes tâches des groupes de travail consistent à faire la liste des différentes méthodes d'homogénéisation des données, établir des standards de comparaison avec des séries réelles et synthétiques, comparer les différentes techniques de détection et les différentes techniques de correction et émettre des recommandations pour l'homogénéisation des données annuelles, mensuelles et journalières. Enfin, un logiciel permettant d'appliquer la nouvelle procédure sera développé. La création de ce groupe d'action vient confirmer le besoin de réaliser des études comparatives afin de déterminer les techniques les plus prometteuses pour détecter des inhomogénéités dans les séries de précipitations totales annuelles, saisonnières et mensuelles.

3. Comparaison, adaptation et développement de méthodes d'homogénéisation

Ce chapitre présente une discussion générale sur les idées et les résultats des articles présentés dans la deuxième partie de cette thèse. Tout au long de la discussion, l'article dont il est question est cité. De plus, certains résultats qui n'ont pas été inclus dans les articles sont discutés ici. Cette discussion est organisée comme suit : les principaux résultats liés aux méthodes d'homogénéisation des précipitations totales annuelles sont discutés. Par la suite, l'homogénéisation des séries de précipitations totales en présence d'asymétrie est discutée. Enfin, l'importance des différentes hypothèses de base des méthodes d'homogénéisation est discutée.

3.1. Comparaison de méthodes d'homogénéisation des précipitations totales annuelles

3.1.1. Séries synthétiques

Des séries synthétiques de précipitations totales annuelles ont été générées pour constituer une base de comparaison des différentes méthodes d'homogénéisation. Ces dernières représentent les caractéristiques moyennes d'un ensemble de stations situées dans la province de Québec et ses alentours. Le choix des stations a été fait selon la qualité des séries de données (longues séries d'observations, peu de données manquantes, etc.). La localisation des stations utilisées est présentée dans la figure 4. Les caractéristiques moyennes des séries de cet ensemble sont : une moyenne de 1089 mm, un écart-type de 142 mm et une autocorrélation de 0.02. La corrélation entre la série de base et ses voisines a été fixée à 0.55 pour représenter la corrélation moyenne entre les stations qui ont moins de 300 km de distance. Cette distance a été fixée de manière arbitraire, en supposant que

des stations voisines situées à plus de 300 km de distance ne seront pas assez corrélées avec la station de base pour constituer de bonnes séries voisines. Dans Groisman et Easterling (1994), le rayon de corrélation des précipitations totales a été estimé pour différentes régions en Amérique du Nord. Dans la région qui correspond à la zone d'étude, le rayon de corrélation estimé est de 350 km. L'insertion des sauts de moyenne, des tendances et des sauts de variance dans les séries synthétiques de base est expliquée en détails dans Beaulieu *et al.* (2008).

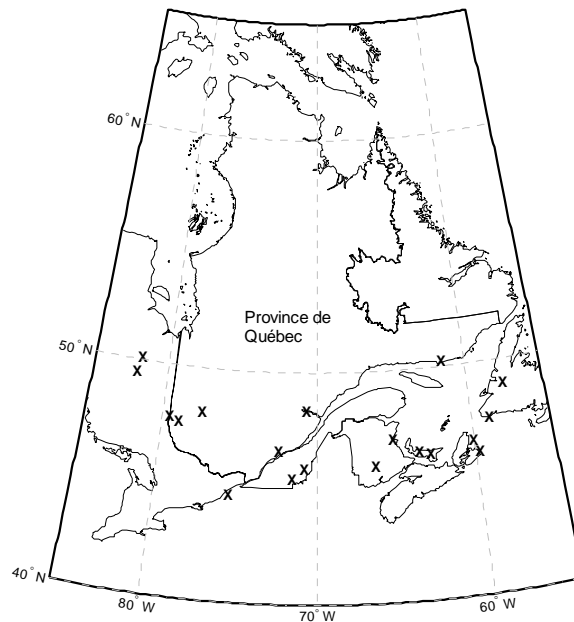


Figure 4. Carte des stations utilisées pour la génération des séries synthétiques.

3.1.2. Première sélection des méthodes

Plusieurs techniques peuvent être utilisées pour détecter des inhomogénéités dans des séries de précipitations totales annuelles. Une revue de littérature des méthodes d'homogénéisation des données climatiques a été réalisée dans Beaulieu *et al.* (2007). Des méthodes d'homogénéisation qui semblaient prometteuses pour les précipitations totales annuelles de la province de Québec et ses alentours ont été identifiées pour faire partie d'une étude comparative (Beaulieu *et al.*, 2008). La

sélection des techniques a été faite en fonction de plusieurs critères. Premièrement, les techniques sélectionnées ne devaient pas être basées sur les métadonnées. Selon le type de gestion du réseau d'observations, les métadonnées peuvent être inaccessibles, incomplètes ou inexistantes, ce qui peut rendre l'utilisation de ces techniques inappropriée. Deuxièmement, les techniques sélectionnées devaient permettre l'utilisation des séries voisines pour comparer avec la série de base. En effet, des changements naturels (e.g. tendance) pourraient être présents dans les séries de précipitations totales annuelles, mais ne doivent pas être détectés en tant qu'inhomogénéités. Les méthodes d'homogénéisation permettant d'isoler les changements naturels des changements artificiels sont basées sur une comparaison avec des séries voisines ou une série de référence. Troisièmement, en se basant sur les résultats des études comparatives précédentes, les techniques qui ont bien performées auparavant, ont été sélectionnées : REGM et SNHT. Parmi celles qui n'ont pas bien performé dans les études comparatives antérieures, certaines techniques ont tout de même été utilisées avec une modification dans le but de les améliorer. Des techniques d'homogénéisation qui répondent aux différents critères et qui n'ont pas été comparées auparavant ont été sélectionnées : JARU et BIVT. La méthode CAME permet de détecter plusieurs sauts et ne nécessite pas de série de référence homogène. Plusieurs séries sont homogénéisées simultanément en étant tour à tour série de base et série voisine. Cette technique n'a pas été sélectionnée pour l'étude comparative puisqu'elle est plus performante lorsqu'un grand nombre de séries voisines est utilisé (Slonosky *et al.*, 1999). Au Canada, les réseaux d'observations sont peu denses et des séries voisines avec de longues séries de données de qualité peuvent être difficiles à obtenir. Donc, l'utilisation de la méthode CAME pourrait être problématique étant donné les caractéristiques des réseaux d'observations de la région d'étude. La méthode MASH fonctionne selon le même principe que CAME et n'a pas été incluse dans l'étude comparative.

La sélection des méthodes s'est aussi faite en fonction de l'étude comparative de Ducre-Robitaille *et al.* (2003). La méthode BAUN a été utilisée dans cette étude comparative. Cette méthode

a bien performé pour détecter un saut, mais s'est montrée moins performante en présence de sauts multiples. De ce fait, une autre méthode bayésienne, BAYE, a été sélectionnée pour l'étude comparative. Dans Ducré-Robitaille *et al.* (2003), STUS n'a pas donné de bons résultats. Cependant, ils ont suggérés d'augmenter la taille de la fenêtre mobile du test pour augmenter sa performance. Cette technique a été incluse dans Beaulieu *et al.* (2008) et cette suggestion a été suivie. De plus, comme le test est appliqué plusieurs fois sur la même série d'observations, le risque que l'erreur de type I se produise est augmenté. Par conséquent, de nouveaux seuils significatifs ont été obtenus par simulation pour faire en sorte que l'erreur de type I soit réellement de 5%. Dans Ducré-Robitaille *et al.* (2003), les seuils significatifs ont été obtenus par simulation également. La méthode WILS n'a pas bien performé dans Ducré-Robitaille *et al.* (2003). Il s'agit d'une application successive du test de Wilcoxon, selon le même principe que STUS. Cependant, Ducré-Robitaille *et al.* (2003) n'ont pas ajusté les seuils critiques comme avec STUS. Par conséquent, dans Beaulieu *et al.* (2008), les seuils critiques utilisés avec WILS ont été ajustés. Enfin, plusieurs versions de REG2 ont été présentées dans la littérature de climatologie (Solow, 1987; Easterling et Peterson, 1995; Lund et Reeves, 2002; Wang, 2003; Lund *et al.*, 2007). Ducré-Robitaille *et al.* (2003) ont comparé la version présentée dans Easterling et Peterson (1995), et ont obtenu un fort pourcentage de fausse détection. Ces derniers suggèrent l'utilisation de la version révisée de Lund et Reeves (2002) pour améliorer la méthode. De ce fait, cette version a été comparée dans Beaulieu *et al.* (2008). Cependant, les versions de Wang (2003) et Lund *et al.* (2007) n'ont pas été testées.

Une autre méthode non paramétrique, PTTI, aurait pu faire partie de l'étude comparative. Cependant, Lubes-Niel *et al.* (1998) ont montré que la puissance de PTTI est faible. Donc, ce test n'a pas été inclus dans l'étude comparative. Lubes-Niel *et al.* (1998) ont aussi obtenu une faible puissance pour PSSH et cette procédure n'a pas été incluse dans l'étude comparative. Le test BUIT répond également aux critères utilisés pour la sélection des méthodes, mais ne fait pas partie de l'étude de simulation.

Enfin, la méthode PSTU ne fait pas partie de l'étude comparative puisqu'elle a été publiée après que l'article de Beaulieu *et al.* (2008) ait été soumis, mais pourrait être appropriée pour détecter des inhomogénéités dans les précipitations totales annuelles. Enfin, d'autres techniques de détection de ruptures présentées dans la littérature statistique pourraient s'avérer prometteuses. La liste des techniques mentionnées dans cette section ainsi que leurs pseudonymes et références est présentée dans le tableau 2.

Tableau 2. Pseudonyme et référence des techniques d'homogénéisation considérées lors de la première sélection.

Pseudonyme	Méthode	Référence(s)
BAUN	Approche bayésienne univariée pour un saut	-Lee et Heghinian (1977) -Ouarda <i>et al.</i> (1999) -Perreault <i>et al.</i> (1999)
BAYE*	Méthode bayésienne univariée pour la détection d'un saut	-Rasmussen (2001)
BIVT*	Approche bivariée	-Maronna & Yohai (1978) -Potter (1981)
BUIT	Test sur les déviations cumulées	-Buishand (1982, 1984)
CAME	Méthode de Caussinus et Mestre	-Mestre (2000) -Caussinus et Mestre (2004)
JARU*	Méthode de Jaruskova	-Jaruskova (1996)
PSSH	Procédure de segmentation de série hydrométéorologique	-Hubert <i>et al.</i> (1989) -Kehagias <i>et al.</i> (2006)
PSTU	Test de Student pénalisé	-Wang <i>et al.</i> (2007)
PTTI	Test de Pettitt	-Pettitt (1979) -Boroneant et Tomozeiu (1999) -Wijngaard (2003)
REG2*	Régression à deux phases	-Solow (1987) -Easterling et Peterson (1995) -Lund et Reeves (2002) -Wang (2003)
REGM*	Régression multiple	-Vincent (1998)
SNHT*	Test d'homogénéité standard normal (standard normal homogeneity test)	-Alexandersson (1986) -Alexandersson et Moberg (1997) -Khaliq et Ouarda (2007)
STUS*	Test de Student séquentiel	-Gullett <i>et al.</i> (1990) -Ducré-Robitaille <i>et al.</i> (2003)
WILS*	Test de Wilcoxon séquentiel	-Lanzante (1996) -Ducré-Robitaille <i>et al.</i> (2003)

* Technique comparée dans Beaulieu et al (2008)

3.1.3. Nouvelles techniques adaptées

3.1.3.1. Régression linéaire bayésienne

Dans Beaulieu *et al.* (2009a), deux méthodes bayésiennes permettant de détecter un changement dans un modèle de régression linéaire multiple ont été adaptées pour détecter des inhomogénéités dans des séries de précipitations totales annuelles. La première qui a été adaptée, BARE, a été développée par Seidou *et al.* (2007) avec une formulation plus générale que celle de BAYE qui a été comparée dans Beaulieu *et al.* (2008). L'autre technique proposée a été développée par Seidou et Ouarda (2007) pour détecter des changements multiples dans une régression linéaire multiple (BAMS). Cette technique permet d'obtenir de manière analytique les distributions a posteriori du nombre de sauts et de leurs positions. Ces deux techniques ont été comparées avec les huit précédentes (Beaulieu *et al.*, 2008), listées dans le tableau 2, sur les mêmes séries synthétiques pour évaluer leur performance à détecter des inhomogénéités dans des séries de précipitations totales annuelles.

3.1.3.2. Test d'homogénéité bayésien

Dans Beaulieu *et al.* (2009b), un test d'homogénéité bayésien qui permet de détecter un changement dans les paramètres d'une distribution appartenant à la famille exponentielle a été adapté pour faire de l'homogénéisation. Entre autres, les distributions normales et gamma font partie de cette famille. Ce test a été développé par Lee (1998). Le cas normal (BNHT) a été comparé avec les méthodes précédentes en utilisant les mêmes séries synthétiques que dans Beaulieu *et al.* (2008, 2009a). Cette approche est beaucoup plus simple que les autres méthodes bayésiennes comparées précédemment et les distributions a posteriori sont analytiques. La technique a été modifiée pour permettre l'utilisation d'autres distributions a priori dans lesquelles l'information provenant des métadonnées peut être incluse. Cette technique est générale et ne se restreint pas uniquement à la

distribution normale comme toutes les méthodes paramétriques comparées précédemment. La distribution gamma appartient aussi à la famille exponentielle et sera abordée dans la section 3.2.

3.1.4. Résultats

3.1.4.1. Performance générale

Les différentes techniques comparées dans Beaulieu *et al.* (2008, 2009a, 2009b) ont toutes leurs forces et faiblesses. Cependant, plusieurs techniques ont donné une bonne performance générale avec de faibles taux de fausse détection sur les séries homogènes et des pourcentages de détection élevés sur les séries avec un ou plusieurs sauts. Ces techniques sont : BAMS, BNHT, BIVT, SNHT et JARU. Pour détecter un saut, la méthode BNHT a donné les meilleurs résultats avec un pourcentage de sauts positionnés plus élevé que les autres méthodes. Pour détecter plusieurs sauts, les méthodes BAMS et BNHT sont les plus performantes. La méthode WILS n'a pas été identifiée parmi les plus performantes. Néanmoins, cette technique donne des résultats légèrement en dessous des techniques identifiées comme étant les plus performantes. Certaines techniques semblent moins prometteuses parce qu'elles ont un taux de détection plus faible : BARE, STUS, REGM et REG2. La technique BAYE est moins prometteuse parce qu'elle mène à trop de fausses détections. Cette technique ne devrait pas être utilisée pour faire de l'homogénéisation due à un trop grand risque de fausse détection.

Dans Beaulieu *et al.* (2008), des séries synthétiques de différentes longueurs (60 et 100 ans) ont été utilisées comme base de comparaison des méthodes. Dans les autres articles de cette thèse dans lesquels des méthodes ont été appliquées sur des séries synthétiques de précipitations totales annuelles, les séries de 60 ans n'ont pas été utilisées. En effet, l'ordre de performance des techniques demeure proportionnel selon qu'elles soient appliquées sur les séries de 60 ou 100 ans. En général, le pourcentage de détection augmente légèrement avec la longueur de la série.

Dans Beaulieu *et al.* (2008), l'effet de la distance entre les sauts consécutifs et du signe des sauts (positif ou négatif) sur la performance des méthodes a été étudié. Dans Beaulieu *et al.* (2009a, 2009b), ces effets n'ont pas été analysés, mais les résultats sont présentés en Annexe A. Pour BAMS et BNHT, les résultats sont semblables à ceux présentés dans Beaulieu *et al.* (2008). En effet, deux sauts consécutifs très rapprochés ou très éloignés semblent plus difficiles à détecter que deux sauts séparés par une distance plus moyenne. Des sauts ayant des signes alternés semblent aussi plus faciles à détecter. Ces observations ne sont pas vérifiées avec la méthode BARE.

Dans Beaulieu *et al.* (2008), les méthodes d'homogénéisation ont été appliquées sur des séries qui contiennent un saut de variance et une tendance. Encore une fois, ces résultats n'ont pas été présentés dans les articles de Beaulieu *et al.* (2009a, 2009b) afin d'alléger les articles puisqu'un exemple de cas pratique y a été inclus. Toutefois, les résultats obtenus avec BAMS, BARE et BNHT sur les séries synthétiques avec un saut de variance sont présentés en Annexe A. La plupart des techniques font l'hypothèse que la variance est constante tout au long de la série et le non-respect de cette hypothèse peut augmenter le risque de fausse détection. Cependant, BAMS ne fait pas l'hypothèse que la variance est constante et permet de détecter un saut de variance, ce qui explique que le nombre de sauts détectés est beaucoup plus grand.

Plusieurs techniques d'homogénéisation comparées permettent de détecter un changement abrupt de la moyenne, en supposant qu'elle est constante dans chaque segment. Or, il est possible d'observer des inhomogénéités de type tendance (croissance de la végétation, urbanisation, etc.). La réponse des différentes techniques à la présence d'une tendance a été étudiée dans Beaulieu *et al.* (2008). La plupart des techniques tend à détecter un ou plusieurs sauts à l'intérieur de la tendance, selon l'amplitude de cette dernière. Pour les méthodes BAMS, BARE et BNHT, les résultats sont présentés en Annexe A. En présence d'une tendance, BARE risque de détecter un saut à l'intérieur de la tendance, surtout si elle a une forte amplitude. Par contre, BAMS et BNHT risquent de détecter un ou plusieurs sauts à l'intérieur de la tendance, selon son amplitude. Les modèles de régression

peuvent tenir compte d'une tendance et sont donc avantageux dans ce cas. Lors de l'application d'une méthode d'homogénéisation, une attention particulière devrait être portée pour éviter d'interpréter une tendance comme un ou plusieurs sauts abrupts puisque le cas échéant, la correction ne sera pas adéquate.

Une étude de sensibilité au choix des paramètres a priori a été réalisée avec la méthode BNHT (Beaulieu *et al.*, 2009b). Cette étude était possible puisque cette technique a un temps d'exécution très rapide. Les autres méthodes bayésiennes présentées dans cette thèse sont plus longues à exécuter, c'est pourquoi des études de sensibilité n'ont pas été réalisées avec ces dernières. La méthode BNHT a été appliquée plusieurs fois sur les différents ensembles de séries synthétiques avec différentes probabilités de non-changement a priori. Sur les séries homogènes, pour éviter un trop grand nombre de fausses détections, la probabilité de non-changement a priori doit être supérieure ou égale à 0.5. Sur les séries avec un saut, une faible probabilité de non-changement a priori permet de détecter des sauts de faible amplitude, qui ne seraient pas détectés autrement. De plus, même avec une probabilité de non-changement a priori très élevée, les sauts de forte amplitude sont détectés dans pratiquement 100% des cas.

Les méthodes bayésiennes BAMS, BARE et BNHT ont été appliquées à des séries synthétiques homogènes indépendantes et avec des autocorrélations positives (coefficients de 0.2 et 0.4) afin d'étudier l'effet de l'autocorrélation sur le risque de fausse détection. Un nombre total de 5000 séries homogènes ayant une autocorrélation de 0.2 et de 5000 séries homogènes ayant une autocorrélation de 0.4 a été généré. Les résultats sont présentés dans l'Annexe A. Avec les trois techniques, le risque de fausse détection est augmenté en présence d'autocorrélation.

3.1.5. Pistes d'amélioration

3.1.5.1. Étude comparative

Les séries synthétiques utilisées comme base de comparaison des différentes méthodes d'homogénéisation dans Beaulieu *et al.* (2008, 2009a, 2009b) ont des caractéristiques fixes. La première piste d'amélioration de ce travail de comparaison serait d'appliquer les méthodes d'homogénéisation sur des séries synthétiques ayant des caractéristiques plus variables. En effet, la performance des différentes techniques pourrait varier sur des séries ayant des caractéristiques différentes comme la corrélation base-voisines, la variance des séries voisines et l'autocorrélation. Il serait intéressant de quantifier l'effet de ces différentes caractéristiques. En effet, cela permettrait d'élargir le domaine de validité des résultats de comparaison des différentes méthodes présentées dans cette thèse.

DeGaetano (2006) a appliqué plusieurs méthodes d'homogénéisation sur des séries de températures moyennes annuelles avec des corrélations différentes (entre 0.84 et 0.94) entre la série de base et ses voisines. Cet auteur n'a pas identifié de relation claire entre la performance des différentes techniques et la corrélation entre la série de base et ses voisines. Cependant, l'ordre de grandeur des corrélations base-voisines testées est très restreint. L'effet de la corrélation entre la série de base et les séries voisines nécessite plus d'approfondissement, particulièrement pour les précipitations qui ont une forte variabilité spatiale. On peut s'attendre à ce que les inhomogénéités soient plus facilement détectables lorsque les séries voisines sont fortement corrélées avec la série de base. Similairement, il se peut que la variance dans les séries voisines influence la performance des techniques. Une grande variance pourrait avoir pour effet de masquer les inhomogénéités dans la série de base. Cet aspect nécessite également plus d'approfondissement. L'autocorrélation peut également influencer la détection d'inhomogénéités. Par exemple, en détection de tendance, une autocorrélation positive augmente le risque d'erreur de type I (Yue *et al.*, 2002). DeGaetano (2006) a

observé le même phénomène avec différentes méthodes d'homogénéisation. La performance des techniques comparées pour détecter des inhomogénéités dans les précipitations totales annuelles pourrait varier avec des séries ayant une autocorrélation différente.

3.1.5.2. Méthodes

La plupart des méthodes d'homogénéisation détectent plus facilement des sauts artificiels dans les extrémités. Pour éviter ce problème, les sauts détectés dans les extrémités ont été ignorés. Cependant, ce problème pourrait être adressé de manière plus sophistiquée. Wang *et al.* (2007) ont corrigé ce problème en introduisant une pénalité dans la fonction de vraisemblance. Les méthodes comparées pourraient être améliorées en utilisant la même approche.

Dans Reeves *et al.* (2007), une modification est suggérée pour améliorer la méthode SNHT. Ils proposent d'utiliser la variance combinée puisque la variance est biaisée en présence d'une inhomogénéité. Khaliq et Ouarda (2007) ont amélioré les valeurs critiques de la méthode SNHT. Ces deux modifications pourraient améliorer les résultats obtenus avec cette technique, mais n'ont pas été considérées dans ce travail.

Les modèles de régression REG2 et REGM pourraient aussi éventuellement donner de meilleurs résultats. La méthode REG2 est désavantagée dans ce travail puisqu'elle permet de détecter à la fois un changement de la moyenne et un changement de tendance, mais des séries synthétiques avec changement de tendance n'ont pas été générées. Donc, plus de paramètres que nécessaire ont été estimés. La version proposée par Wang (2003) a un paramètre en moins à estimer et aurait peut-être mieux performé dans cette étude. La méthode REGM a un faible taux de détection. L'application successive du test d'autocorrélation des résidus, du test de Fisher pour le choix de modèle et du test de Student pour vérifier si l'amplitude du saut est significative peut avoir pénalisé la détection.

Les seuils critiques utilisés avec les techniques WILS et STUS ont été obtenus par simulation puisque ces deux statistiques sont calculées successivement pour toutes les positions possibles de

saut dans la série. Ceci a permis d'obtenir un seuil critique global de 5% pour des longueurs de séries de 60 ou 100 ans. Pour la détection de sauts multiples, ces techniques ont pu être désavantagées puisqu'elles ont été appliquées sur des segments plus petits que 60 ou 100 observations. Des valeurs critiques pour plusieurs tailles de séries pourraient être obtenues grâce à une étude de simulation similaire à celle réalisée par Khaliq et Ouarda (2007).

Les trois méthodes bayésiennes comparées sur les séries synthétiques de précipitations totales annuelles ont toutes une formulation différente. Leur performance dépend fortement des distributions a priori utilisées. En utilisant des distributions a priori différentes, les résultats pourraient être améliorés.

3.1.5.3. Cas d'application

Un exemple d'application sur les précipitations totales annuelles à Québec est présenté dans Beaulieu *et al.* (2009a; 2009b). Dans Beaulieu *et al.* (2009a), des lois a priori non informatives ont été utilisées pour les méthodes bayésiennes (BAMS et BARE) puisque les résultats du cas d'application sont comparés avec des méthodes classiques. Dans Beaulieu *et al.* (2009b), différentes lois a priori et différentes probabilités de non-changement a priori ont été utilisées. Dans cet exemple, l'information permettant d'ajuster une loi a priori informative provient des métadonnées. En effet, deux changements documentés ont été identifiés comme étant susceptibles d'introduire une inhomogénéité dans les données. De ce fait, des lois a priori uniformes qui permettent de restreindre l'intervalle de détection autour des dates pour lesquelles des changements sont documentés ont été utilisées. Des lois a priori triangulaires qui permettent de donner une plus forte probabilité de changement aux dates pour lesquelles des changements sont documentés ont aussi été utilisées. Peu importe la loi a priori et la probabilité de non-changement a priori utilisées, le même changement est détecté dans tous les cas. Comme la rupture est de forte amplitude, on ne peut pas voir l'effet du choix de la loi a priori sur le résultat de l'analyse. Une piste d'amélioration pour les cas d'application

serait de choisir des cas d'études pour lesquels la rupture a une plus faible amplitude et qui permettraient d'observer l'impact de l'utilisation de lois a priori informatives.

Dans cet exemple, l'information provenant des métadonnées a été intégrée dans l'analyse. Cependant, d'autres types d'information peuvent être inclus dans l'analyse en utilisant l'approche bayésienne. En effet, la loi a priori sur la moyenne avant et après la rupture pourrait être construite à partir de segments de séries ou de séries voisines non utilisées dans l'analyse. De plus, un expert hydrologue, climatologue, qui possède une expérience en homogénéisation des précipitations peut être en mesure de construire une loi a priori sur l'amplitude de la rupture sachant un certain type de modification des conditions de mesure.

3.2. Homogénéisation des précipitations totales avec des distributions asymétriques

Au pas de temps annuel, la distribution des séries de précipitations totales peut être considérée normale (Von Storch et Zwiers, 1999). À des pas de temps inférieurs, la distribution des précipitations est asymétrique et est mieux représentée par la distribution gamma (Wilks, 2006). Cette distribution est aussi simple que la distribution normale, mais beaucoup plus flexible puisqu'elle peut représenter différents niveaux d'asymétrie.

Les différentes méthodes d'homogénéisation sont souvent appliquées sur une série de rapports entre la série de base et les séries voisines. Dans ce cas, c'est la série de rapports qui doit respecter une distribution donnée. Les rapports de précipitations totales mensuelles ou saisonnières peuvent être asymétriques et sont mieux représentés par la distribution gamma. De plus, des rapports nuls ou négatifs sont impossibles, mais peuvent avoir une probabilité d'occurrence non nulle avec une distribution normale. La distribution gamma est bornée par zéro, donc semble plus logique pour ajuster des rapports de précipitations totales.

Une nouvelle approche pour détecter des inhomogénéités dans les précipitations totales annuelles, saisonnières ou mensuelles est présentée dans Beaulieu *et al.* (2009c). Cette méthode a été proposée par Lee (1998) et permet de détecter un changement dans le paramètre d'échelle d'une série qui a une distribution gamma. Un cas d'étude est présenté et montre que les rapports de précipitations à des pas de temps saisonniers ou mensuels sont mieux représentés par la distribution gamma que par la normale. Au pas de temps annuel, les deux distributions sont appropriées. Donc, la méthode proposée est plus générale et permet de détecter des inhomogénéités à des pas de temps variés. De plus, la diminution du pas de temps permet d'appliquer la technique d'homogénéisation sur des séries de données plus longues (12 fois plus longues dans le cas des séries mensuelles). Par le fait même, la puissance des techniques d'homogénéisation peut être augmentée. Par contre, en réduisant le pas de temps, la présence d'autocorrélation dans la série est plus probable, il serait donc nécessaire de quantifier l'effet de l'autocorrélation sur le risque de fausse détection et sur la puissance de cette technique. Enfin, la présence de saisonnalité pourrait interférer avec la détection de changement. En supposant que la saisonnalité est régionale, on fait l'hypothèse que celle-ci sera retranchée en produisant la série de rapports.

En pratique, les méthodes d'homogénéisation sont souvent appliquées sur des séries annuelles, mois par mois, et les ajustements diffèrent pour chaque mois. Les facteurs d'ajustements mois par mois sont calculés avec peu d'observations et sont avantageux lorsque les séries de données sont longues et de qualité (Staudt *et al.*, 2007). Les facteurs d'ajustements devraient dépendre du mois ou de la saison seulement si la saisonnalité est très prononcée (Staudt *et al.*, 2007).

Cette nouvelle approche fait l'hypothèse que le paramètre de forme est constant tout au long de la série et supérieur à 1 pour une forme de distribution réaliste. Le paramètre de forme peut être estimé par maximum de vraisemblance ou par la méthode des moments. Supposons que le paramètre de forme est représenté par α et que le paramètre d'échelle est représenté par η . La moyenne (μ), la variance (σ^2) et le coefficient d'asymétrie (γ) peuvent s'exprimer en fonction des

paramètres de forme et d'échelle et sont données par : $\mu = \alpha\eta$, $\sigma^2 = \alpha\eta^2$ et $\gamma = 2/\sqrt{\alpha}$. Lorsque le paramètre de forme est très élevé, le coefficient d'asymétrie tend vers zéro et la normalité est approximée.

3.3. Importance des hypothèses de base

3.3.1. Homogénéité des voisines

Dans cette thèse, l'approche utilisée pour détecter des inhomogénéités consiste à comparer la série de base avec des séries voisines. L'hypothèse que les séries voisines sont homogènes est très importante, puisque des inhomogénéités dans ces dernières pourraient mener à des fausses détections dans la série de base. La formulation de la série de référence ou de la série de comparaison varie selon les différentes techniques d'homogénéisation. Lorsque les séries voisines sont homogènes, le choix de la formulation de la série de référence ou de comparaison a peu d'effet (Menne et Williams, 2005). Lorsque des inhomogénéités sont présentes dans les séries voisines, cela affecte plus les techniques qui ont les meilleurs taux de détection (Menne et Williams, 2005). L'effet des inhomogénéités dans les séries voisines sur la performance des techniques comparées n'a pas été étudié dans cette thèse. Par contre, il paraît normal d'envisager que les techniques qui se sont montrées les plus sensibles à la présence d'inhomogénéités soient les plus affectées. On pourrait également s'attendre à ce que les fausses détections dues à des inhomogénéités dans les séries voisines dépendent de l'amplitude de ces dernières. Lorsque les inhomogénéités des séries voisines ne sont pas assez grandes pour entraîner des fausses détections, les facteurs de correction peuvent néanmoins être biaisés. L'hypothèse d'homogénéité des séries voisines est une faiblesse pour toutes les méthodes comparées et proposées dans cette thèse.

3.3.2. Indépendance

Les méthodes comparées et proposées dans cette thèse sont basées sur l'hypothèse d'indépendance des données. Cette hypothèse est importante puisque la présence d'autocorrélation peut faire en sorte d'augmenter le nombre de fausses détections (Lund *et al.*, 2007). La plupart du temps, les précipitations totales annuelles sont indépendantes. De ce fait, l'effet de l'autocorrélation sur le taux de détection des techniques n'a pas été quantifié avec toutes les techniques comparées dans cette thèse. Cependant, en se basant sur les résultats présentés en Annexe A sur des séries autocorrélées et ceux de DeGaetano (2006) et de Lund *et al.* (2007), on peut s'attendre à ce que la présence d'autocorrélation puisse engendrer des fausses détections et diminuer la puissance de détection. La présence d'autocorrélation dans les séries ayant un pas de temps inférieur (e.g. mensuel ou journalier) est plus probable que dans les séries annuelles.

D'autre part, puisque les méthodes comparées et proposées dans cette thèse sont basées sur l'hypothèse d'indépendance des données, il est important de vérifier cette hypothèse lorsqu'une méthode d'homogénéisation est appliquée. Une attention particulière doit être accordée à cette étape, puisque la présence d'une inhomogénéité peut faire en sorte que cette hypothèse soit rejetée (Perreault *et al.*, 1999; Beaulieu *et al.*, 2009b, 2009c). En effet, certains auteurs proposent même d'appliquer plusieurs tests successifs, dont un test d'indépendance, pour vérifier l'homogénéité d'une série. Tayanc *et al.* (1998) proposent d'appliquer le test d'indépendance de Wald et Wolfowitz (WWOL) et Wijngaard *et al.* (2003) proposent d'appliquer VNEU, le rapport de Von Neumann (1941). Le test WWOL consiste à transformer la série en une suite de signes positifs ou négatifs, selon que l'observation soit inférieure ou supérieure à la médiane de la série. En présence d'un saut, les observations avec le même signe risquent de se retrouver en forte concentration dans le même segment et l'hypothèse d'indépendance risque d'être rejetée. La statistique de Von Neumann calcule un rapport des différences entre les observations consécutives dans la série sur la variance. La présence d'une inhomogénéité dans la série risque de gonfler la variance et ainsi diminuer la valeur de

la statistique, faisant ainsi en sorte que l'hypothèse d'indépendance soit rejetée. Donc, ces deux tests d'indépendance peuvent indiquer la présence d'une inhomogénéité dans la série. Dans Beaulieu *et al.* (2009b, 2009c), l'hypothèse d'indépendance est vérifiée dans chaque segment respectif, après la détection du saut.

3.3.3. Distribution

La majorité des méthodes comparées dans Beaulieu *et al.* (2008, 2009a, 2009b) est basée sur l'hypothèse de normalité, sauf la méthode non paramétrique WILS. On peut s'attendre à ce que cette dernière soit robuste à toute distribution. Par contre, la robustesse des techniques basées sur la normalité n'a pas été vérifiée. De la même façon, la robustesse de BGHT à des données qui ne sont pas distribuées selon la distribution gamma n'a pas été vérifiée. On peut s'attendre à ce que les méthodes basées sur la normalité soient plus efficaces pour des données annuelles. Par contre, BGHT devrait être plus efficace pour des données saisonnières ou mensuelles.

D'autre part, les tests d'adéquation à une distribution quelconque peuvent être affectés en présence d'une inhomogénéité (Perreault *et al.*, 1999; Beaulieu *et al.*, 2009b, 2009c). Par exemple, la statistique de Chi-deux, communément utilisé pour vérifier l'ajustement d'une distribution, calcule l'écart entre les valeurs observées et espérées selon une distribution quelconque. Une inhomogénéité pourrait avoir pour effet d'agrandir cet écart et ainsi, augmenter la valeur de la statistique de Chi-deux. Une statistique de Chi-deux élevée fera en sorte que l'hypothèse nulle sera plus facilement rejetée. De plus, l'histogramme peut avoir une forme bimodale due à la présence d'une rupture dans la série. Il peut alors s'avérer difficile de reconnaître la forme de la distribution de la série de données. La distribution peut être vérifiée dans chaque segment après la détection du saut.

3.3.4. Non-stationnarité

3.3.4.1. Saut de variance

La plupart des méthodes étudiées dans cette thèse est basée sur l'hypothèse que la variance de la série est constante, sauf WILS et BAMS. La technique WILS est non paramétrique, alors un changement de variance ne devrait pas l'affecter. La méthode BAMS permet de détecter un changement dans les paramètres d'une régression linéaire multiple, incluant la variance. Dans Beaulieu *et al.* (2008), la robustesse des différentes méthodes à un saut de variance a été étudiée. La présence d'un saut de variance dans la série tend à augmenter le nombre de fausses détections pour la plupart des techniques basées sur l'hypothèse que la variance est constante. Malgré que cette augmentation soit significative, elle demeure néanmoins faible.

3.3.4.2. Saut de paramètre de forme

La méthode BGHT permet de détecter un saut dans le paramètre d'échelle d'une série, en supposant le paramètre de forme connu et constant. Un saut dans le paramètre de forme pourrait éventuellement affecter la détection des inhomogénéités. Cet aspect devrait être étudié plus en profondeur.

4. Conclusions et perspectives de recherche

4.1. Conclusions

Le développement de techniques d'homogénéisation des données est crucial pour l'étude du climat et plusieurs méthodes existent déjà dans la littérature pour s'attaquer au problème. Toutefois, il était nécessaire de déterminer lesquelles seraient les plus appropriées pour les données de précipitation totales du Québec. En effet, on ne dispose pas encore de base de données de précipitations homogénéisées. Seule une pré-homogénéisation pour certains types de changements a été réalisée (Mékis et Hogg, 1999). Plusieurs techniques d'homogénéisation ont été appliquées à des séries synthétiques de précipitations totales annuelles afin d'identifier les techniques les plus prometteuses. Quelques études comparatives du même genre ont été réalisées auparavant pour les séries de températures, mais pas pour les précipitations. Les techniques qui ont bien performé auparavant à détecter des inhomogénéités dans les séries de températures, n'ont pas toutes bien performées dans la présente étude, ce qui illustre l'intérêt de vérifier que les techniques couramment utilisées et nouvellement développées sont adéquates et performantes pour le problème. Les précipitations sont plus problématiques parce qu'elles ont une plus grande variabilité spatio-temporelle. L'étude réalisée permettra de guider les praticiens (hydrologues, climatologues,...) dans le choix de la technique d'homogénéisation à appliquer sur des séries de précipitations totales annuelles.

Des techniques présentées dans la littérature de climatologie, d'hydrologie et de statistique ont été comparées. Parmi les méthodes classiques couramment utilisées pour faire de l'homogénéisation, celles basées sur le test de rapport des vraisemblances semblent les plus prometteuses (SNHT, BIVT, JARU). Une méthode non paramétrique basée sur une application successive du test de Wilcoxon (WILS), légèrement modifiée dans le but d'améliorer sa performance, s'est montrée également prometteuse.

Peu de techniques bayésiennes de détection de ruptures ont été appliquées au problème d'homogénéisation jusqu'à ce jour. Cependant, l'approche bayésienne est avantageuse puisqu'elle permet d'intégrer la connaissance du phénomène (e.g. le jugement de l'expert ou l'information contenue dans les métadonnées) dans l'analyse. Cette connaissance est disponible. En effet, plusieurs scientifiques ont homogénéisé des séries climatiques auparavant, et ont acquis une expérience qui permet d'évaluer a priori la probabilité que la série soit affectée par un type de changement en particulier. Les métadonnées permettent de déterminer les positions pour lesquelles il y a une forte probabilité d'occurrence d'une inhomogénéité. De ce fait, plusieurs méthodes bayésiennes ont été proposées dans cette thèse pour détecter des inhomogénéités dans les séries de précipitations totales. Elles ont été appliquées en utilisant des distributions a priori non informatives pour être en mesure de les comparer avec les techniques classiques. Parmi ces dernières, la première permet de détecter un changement dans un modèle de régression simple (BAYE). La deuxième permet de détecter un changement dans les paramètres d'un modèle de régression multiple (BARE), la troisième permet de détecter plusieurs changements simultanés dans un modèle de régression multiple (BAMS) et la quatrième est un test d'homogénéité bayésien (BNHT). Ces techniques ont été identifiées dans la littérature d'hydrologie et de statistique. Parmi les techniques bayésiennes proposées, la méthode BAYE ne devrait pas être utilisée puisque le risque de fausse détection encouru lors de son application est beaucoup trop élevé. La méthode BARE a un faible taux de détection. Toutefois, les méthodes BAMS et BNHT se sont montrées très prometteuses pour détecter des inhomogénéités dans des séries de précipitations totales annuelles, même plus prometteuses que les techniques classiques les plus performantes. La méthode BAMS a donné le meilleur taux de détection en présence de sauts multiples. BNHT a donné la meilleure performance sur les séries avec un saut et une performance équivalente à celle de BAMS en présence de sauts multiples. De plus, il est l'équivalent en simplicité du très populaire test d'homogénéité basé sur le rapport des vraisemblances (SNHT) développé par Alexandersson (1986). Ces résultats ont été obtenus en utilisant des

distributions a priori non informatives. En utilisant des distributions a priori informatives, on peut s'attendre à ce que des sauts de faible amplitude deviennent plus facilement détectables.

La nouvelle technique bayésienne qui permet la détection simultanée de plusieurs sauts constitue une autre contribution importante de cette thèse. En effet, des techniques qui permettent la détection simultanée de plusieurs sauts ont été développées dans le cadre classique, mais ces dernières sont performantes lorsque plusieurs séries voisines sont disponibles. Ces techniques ont été développées en Europe, où les réseaux d'observations sont très denses et donc, très facile d'identifier plusieurs séries voisines. Les réseaux d'observations au Canada sont beaucoup moins denses et l'identification d'un grand nombre de séries voisines peut s'avérer impossible. La nouvelle technique permettant la détection simultanée de plusieurs sauts est donc plus adapté aux réseaux d'observations situés au Canada.

Ces approches bayésiennes sont toutes basées sur l'hypothèse de normalité des données. Elles ont été appliquées à des séries de précipitations totales annuelles, mais elles pourraient s'avérer prometteuses pour homogénéiser d'autres variables climatiques qui respectent l'hypothèse de normalité comme par exemple, les températures.

Une nouvelle méthode a été proposée pour détecter des inhomogénéités dans les séries de précipitations totales avec une distribution asymétrique. Cette dernière est plus flexible que la plupart des techniques d'homogénéisation développées auparavant, y compris celles proposées dans cette thèse, qui sont basées sur l'hypothèse de normalité. La distribution gamma peut approximer la distribution normale. De plus, elle peut représenter différents cas d'asymétrie qui sortent du cadre imposé par la distribution normale. Ces cas d'asymétrie se rencontrent surtout aux pas de temps saisonniers et mensuels. Un cas d'étude a permis de montrer que cette technique semble appropriée pour détecter des inhomogénéités dans les séries de précipitations totales aux pas de temps annuel, saisonnier et mensuel.

En résumé, les principales contributions de cette thèse sont l'identification des techniques qui semblent les plus prometteuses pour l'homogénéisation des précipitations totales, le développement de nouvelles méthodes bayésiennes qui permettent la détection d'inhomogénéités, l'adaptation d'une technique qui permet la détection simultanée de plusieurs sauts et le développement d'une technique qui permet de détecter des inhomogénéités dans les séries de précipitations totales ayant une distribution asymétrique.

Les méthodes bayésiennes adaptées et développées dans cette thèse seront appliquées aux séries synthétiques et réelles utilisées par le groupe d'action COST – HOME. Ces séries ont été rendues disponibles en novembre 2008 et sont décrites dans Venema *et al.* (2008). Les méthodes développées pourront ainsi être comparées à toutes les techniques d'homogénéisation présentées dans la littérature pour évaluer leur performance à détecter des inhomogénéités dans des séries de températures et de précipitations.

4.2. Perspectives de recherche

4.2.1. Applications avec des distributions a priori informatives

Dans cette thèse, de nouvelles techniques d'homogénéisation basées sur l'approche bayésienne ont été développées. Ces techniques ont été appliquées sur des séries synthétiques en utilisant des distributions a priori non informatives pour pouvoir les comparer avec les techniques classiques. Un cas d'application a été réalisé dans lequel l'information provenant des métadonnées est incluse dans l'analyse. D'autres sources d'informations peuvent être disponibles. Dans le but d'exploiter l'approche bayésienne à son plein potentiel, il serait intéressant de réaliser des applications dans lesquelles différentes sources d'informations sont intégrées dans l'analyse. Par exemple, ces distributions a priori informatives pourraient être construites à partir des métadonnées, d'information

régionale, de jugement d'un expert, de segments de séries non utilisées dans l'analyse ou encore de séries voisines non utilisées dans l'analyse.

4.2.2. Vérification de la performance de la technique BGHT

Dans cette thèse, la technique BGHT a été appliquée sur des données réelles, alors que la performance de toutes les autres techniques a été vérifiée par simulation. Il serait important de vérifier sa performance par simulation sur des séries homogènes et inhomogènes. De plus, comme cette technique s'adresse à des séries de précipitations totales à différents pas de temps, il serait intéressant de vérifier l'effet de l'autocorrélation sur la performance de cette technique. Enfin, comme cette technique permet à la fois de traiter les cas d'asymétrie, mais aussi les cas où les données sont normales, il serait intéressant de l'appliquer sur des séries synthétiques qui représentent différents niveaux d'asymétrie.

4.2.3. Sauts multiples

Les méthodes BNHT et BGHT pourraient être modifiées pour permettre la détection de sauts multiples d'une manière plus sophistiquée. En effet, l'application d'une technique d'homogénéisation de manière itérative, sur des segments de plus en plus petits, peut s'avérer problématique puisque la puissance de détection diminue avec la longueur de la série testée. Une hiérarchie de sous-hypothèses telle que présentée dans Chen et Gupta (2001) pourrait être formulée pour permettre la détection d'un nombre inconnu de sauts et leurs positions. L'idée est de partager la vraisemblance selon les différentes hypothèses, en utilisant toujours la série de données entière. Par ailleurs, la méthode BAMS pourrait être adaptée à la distribution gamma. L'algorithme original présenté par Fearnhead (2006) ne présuppose aucune distribution. Seidou et Ouarda (2007) l'ont adapté à la régression linéaire, mais l'algorithme pourrait aussi être adapté à la distribution gamma.

Hors du cadre bayésien, l'algorithme de programmation dynamique présenté dans Hawkins (2001) permet de détecter des sauts multiples dans une distribution appartenant à la famille exponentielle. Le cas gamma pourrait être développé pour détecter des inhomogénéités dans les totaux de précipitations à différents pas de temps.

4.2.4. Autres types d'inhomogénéités

Dans cette thèse, l'emphase a été mise sur la détection d'inhomogénéités de type rupture abrupte de la moyenne ou rupture du paramètre d'échelle (BGHT). Parmi les techniques permettant une inhomogénéité plus graduelle (tendance), soit les modèles de régression, il serait intéressant de faire une comparaison afin d'identifier les plus appropriées pour ce type d'inhomogénéité. D'une manière similaire, il est possible d'observer une inhomogénéité de type saut de variance. Toutefois, la plupart des techniques d'homogénéisation ont été développées pour détecter une rupture dans la moyenne. Dans la littérature statistique, on retrouve des techniques développées pour détecter des ruptures de variance (Hsu, 1977; Inclan, 1993; Chen et Gupta, 1997). Parmi les techniques adaptées dans cette thèse, BAMS permet la détection d'un changement dans la variance. Il serait intéressant d'identifier dans la littérature statistique des techniques permettant la détection d'une rupture dans la variance et de vérifier leur performance. Des méthodes basées sur une décomposition en ondelettes permettent la détection de plusieurs types de changement et seraient également prometteuses pour différents types d'inhomogénéités.

4.2.5. Homogénéisation des séries de vitesse du vent

La littérature sur l'homogénéisation des données est très concentrée sur les séries de températures, de précipitations et de pression. D'autres variables climatiques, comme les séries de vent, peuvent contenir des inhomogénéités (Thomas *et al.*, 2005). La vitesse du vent peut être

représentée par une distribution gamma (Wilks, 2006). Il serait envisageable d'utiliser l'approche présentée dans Beaulieu *et al.* (2009c) pour détecter des inhomogénéités dans des séries de vitesse du vent.

4.2.6. Détection de changement dans différentes variables climatiques

Plusieurs variables climatiques ont une distribution appartenant à la famille exponentielle. Les distributions a posteriori pour d'autres distributions appartenant à la famille exponentielle pourraient être dérivées, suivant la méthodologie utilisée par Lee (1998). L'humidité relative et la nébulosité peuvent être représentées par la distribution beta (Wilks, 2006). La distribution de Poisson représente souvent l'occurrence d'évènements rares comme les tornades par exemple (Wilks, 2006). La distribution Weibull est communément utilisée pour représenter les extrêmes climatiques (Jandhyala *et al.*, 1999).

4.2.7. Précipitations totales journalières

La détection d'inhomogénéités dans des séries de précipitations totales journalières n'est pas possible avec les techniques présentées dans cette thèse. Un modèle Poisson-gamma pour caractériser les occurrences et l'intensité des précipitations journalières serait plus approprié. Suivant la méthodologie de Lee (1998), les distributions a posteriori pour la loi de Poisson pourraient être dérivées. Le test d'homogénéité pourrait être appliqué pour détecter des changements ou des inhomogénéités des les séries de précipitations totales journalières en utilisant un modèle Poisson-gamma.

5. Références

Action COST-HOME: Advances in homogenisation methods of climate series: an integrated approach (HOME). <http://www.homogenisation.org/>, site visité le 15 décembre 2008.

Aguilar, E., Auer, I., Brunet, M., Peterson, T.C., Wieringa, J. (2003) Guidelines on climate metadata and homogenization. WMO-TD No. 1186, World Meteorological Organization, Geneva, Switzerland.

Aguilar, E., Peterson, T.C., Obando, P.R., Frutos, R., Retana, J.A., Solera, M., Soley, J., Garcia, I.G., Araujo, R.M., Santos, A.R., Valle, V.E., Brunet, M., Aguilar, L., Alvarez, L., Bautista, M., Castanon, C., Herrera, L., Ruano, E., Sinay, J.J., Sanchez, E., Oviedo, G.I.H., Obed, F., Salgado, J.E., Vazquez, J.L., Baca, M., Gutierrez, M., Centella, C., Espinosa, J., Martinez, D., Olmedo, B., Espinoza, C.E.O., Nunez, R., Haylock, M., Benavides, H., Mayorga, R. (2005) Changes in precipitation and temperature extremes in Central America and northern South America, 1961-2003. *Journal of Geophysical Research Atmospheres* **110** (D23), D23107.

Akaike, H. (1973) Information theory and an extension of the maximum likelihood principle. Second international symposium on information theory.

Alexandersson, H. (1986) A homogeneity test applied to precipitation data. *Journal of Climatology* **6** (1), 661-675.

Alexandersson, H., Moberg, A. (1997) Homogenization of Swedish temperature data. Part 1: Homogeneity test for linear trends. *International Journal of Climatology* **17** (1), 25-34.

Allen, R.J., DeGaetano, A.T. (2000) A method to adjust long-term temperature extreme series for nonclimatic inhomogeneities. *Journal of Climate* **13** (20), 3680-3695.

Asselin, J., Ouarda, T.B.M.J., Fortin, V., Bobée, B. (1999) Une procédure bayésienne bivariée pour détecter un décalage de la moyenne. Rapport de recherche R-528, Chaire en hydrologie statistique, INRS-Eau, Québec, QC, Canada.

Auer, I., Bohm, R., Scheifinger, H., Ungersbock, M., Orlik, A., Jurkovic, A. (2004) Metadata and their role in homogenising. Proc. Fourth seminar for homogenization and quality control in climatological databases, 6-10 October 2003, Budapest, Hungary, WCDMP-56, WMO-TD No. 1236, World Climate Data and Monitoring Programme, World Meteorological Organization, Geneva, Switzerland.

Auer, I., Böhm, R., Jurkovicacute, A., Orlik, A., Potzmann, R., Schöner, W., Ungersböck, M., Brunetti, M., Nanni, T., Maugeri, M., Briffa, K., Jones, P., Efthymiadis, D., Mestre, O., Moisselin, J.M., Begert, M., Brazdil, R., Bochnicek, O., Cegnar, T., Gajicacute-Ccaronapka, M., Zaninovicacute, K., Majstorovicacute, Z., Szalai, S., Szentimrey, T., Mercall, L. (2005) A new instrumental precipitation dataset for the greater alpine region for the period 1800-2002. *International Journal of Climatology* **25** (2), 139-166.

Beaulieu, C., Ouarda, T.B.M.J., Seidou, O. (2007) Synthèse des techniques d'homogénéisation des séries climatiques et analyse d'applicabilité aux séries de précipitations. *Hydrological Sciences-Journal-des Sciences Hydrologiques* **52** (1), 18-37.

Beaulieu, C., Seidou, O., Ouarda, T.B.M.J., Zhang, X., Boulet, G., Yagouti, A. (2008) Intercomparison of homogenization techniques for precipitation data. *Water Resources Research* **44** (2), W02425.

Beaulieu, C., O. Seidou, T.B.M.J. Ouarda, Zhang, X. (2009a) Intercomparison of homogenization techniques for precipitation data continued: the comparison of two recent Bayesian changepoint models. Accepté dans *Water Resources Research*.

Beaulieu, C., Ouarda, T.B.M.J., Seidou, O. (2009b) A Bayesian Normal Homogeneity Test for the detection of artificial discontinuities in climatic series. En révision à *International Journal of Climatology*.

Beaulieu, C., Ouarda, T.B.M.J., Seidou, O. (2009c) A Bayesian Gamma Homogeneity Test for the detection of artificial discontinuities in precipitation series. À soumettre.

Bobée, B., Ashkar, F. (1991) *The Gamma family and derived distributions in hydrology*. Water Resources Publications, Littleton, United States.

Begert, M., Schlegel, T., Kirchhofer, W. (2005) Homogeneous temperature and precipitation series of Switzerland from 1864 to 2000. *International Journal of Climatology* **25** (1), 65-80.

Bois, P. (1986) Contrôle de séries chronologiques corrélées par étude du cumul des résidus de la corrélation. 2e Journées hydrologiques de l'ORSTOM, Montpellier, France.

Boroneant, C., Tomozeiu, R. (1999) Experience with homogeneity testing of temperature data at Bucuresti Filaret station. Proc. Second seminar for homogenization of surface climatological data, 9-13 November 1998, Budapest, Hungary, WMO-TD No. 962., World Meteorological Organization, Geneva, Switzerland.

Brandsma, T., Konnen, G.P. (2006) Application of nearest-neighbor resampling for homogenizing temperature records on a daily to sub-daily level. *International Journal of Climatology* **26** (1), 75-89.

Brunet-Moret, Y. (1979) Homogénéisation des précipitations. *Cahiers de l'ORSTOM, série Hydrologie* **16** (3-4), 147-170.

Brunetti, M., Maugeri, M., Monti, F., Nannia, T. (2006) Temperature and precipitation variability in Italy in the last two centuries from homogenised instrumental time series. *International Journal of Climatology* **26** (3), 345-381.

Buffoni, L., Maugeri, M., Nanni, T. (1999) Precipitation in Italy from 1833 to 1996. *Theoretical and Applied Climatology* **63** (1-2), 33-40.

Buishand, T.A. (1982) Some methods for testing the homogeneity of rainfall records. *Journal of Hydrology* **58** (1), 11-27.

Buishand, T.A. (1984) Tests for detecting a shift in the mean of hydrological time series. *Journal of Hydrology* **73** (2), 51-69.

Caussinus, H., Lyazrhi, F. (1997) Choosing a linear model with a random number of change-points and outliers. *Annals of the Institute of Statistical Mathematics* **49** (4), 761-775.

- Caussinus, H., Mestre, O. (2004) Detection and correction of artificial shifts in climate series. *Applied Statistics* **53** (3), 405-425.
- Chen, J., Gupta, A.K. (1997) Testing and locating variance change points with application to stock prices. *Journal of the American Statistical Association* **92**, 739-747.
- Chen, J., Gupta, A.K. (2000) *Parametric statistical change-point analysis*. Birkhauser, Boston, United States.
- Chen, J., Gupta, A.K. (2001) On change-point detection and estimation. *Communications in Statistics-Simulation and Computation* **30** (3), 665-697.
- Conrad, V. (1948) A new criterion of relative homogeneity of climatological series. *Theoretical and Applied Climatology* **1** (1), 1-8.
- Conway, D., Mould, C., Bewket, W. (2004) Over one century of rainfall and temperature observations in Addis Ababa, Ethiopia. *International Journal of Climatology* **24** (1), 77-91.
- Craddock, J.M. (1979) Methods of comparing annual rainfall records for climatic purposes. *Weather* **34** 332-346.
- Daly, C., Gibson, W.P., Taylor, G.H., Doggett, M.K., Smith, J.I. (2007) Observer bias in daily precipitation measurements at United States cooperative network stations. *Bulletin of the American Meteorological Society* **88** (6), 899-912.
- Datsenko, N.M., Moberg, A., Sonechkin, D.M. (2002) Objective time-scale-dependent homogenization of early instrumental temperature series. *Theoretical and applied climatology* **72** (1-2), 103-126.
- DeGaetano, A.T. (2006) Attributes of several methods for detecting discontinuities in mean temperature series. *Journal of Climate* **19** (5), 838-853.
- Della-Marta, P.M., Wanner, H. (2006) A method of homogenizing the extremes and mean of daily temperature measurements. *Journal of Climate* **19** (17), 4179-4197.
- Devine, K.A., Mékis, E. (2008) Field accuracy of Canadian rain measurements. *Atmosphere-Ocean* **46** (2), 213-227.
- Ducré-Robitaille, J.F., Boulet, G., Vincent, L.A. (2003) Comparison of techniques for detection of discontinuities in temperature series. *International Journal of Climatology* **23** (9), 1087-2003.
- Easterling, D.R., Peterson, T.C. (1992) Techniques for detecting and adjusting for artificial discontinuities in climatological time series: a review. Proc. Fifth International Meeting on Statistical Climatology, 22-26 June 1996, Toronto, Ontario, Canada.
- Easterling, D.R., Peterson, T.C. (1995) A new method for detecting undocumented discontinuities in climatological time series. *International Journal of Climatology* **15** (4), 369-377.
- Easterling, D.R., Peterson, T.C., Karl, T.R. (1996) On the development and use of homogenized climate datasets. *Journal of Climate* **9** (6), 1429-1434.

- Fearnhead, P. (2006) Exact and efficient Bayesian inference for multiple changepoint problems. *Statistics and Computing* **16** (2), 203-213.
- Göktürk, O.M., Bozkurt, D., Sen, Ö.L., Karaca, M. (2008) Quality control and homogeneity of Turkish precipitation data. *Hydrological Processes* **22** (16), 3210-3218.
- Gonzalez Hidalgo, J.C., de Luis, M., Vicente, S.M., Stepanek, P., Raventos, J., Saz, M.A., Cuadrat, J. M., Ferraz, J., Creus, J. (2004) Monthly precipitation data base in mediterranean climate area of Spain. Reconstruction processes and quality control. Proc. Fourth seminar for homogenization and quality control in climatological databases, 6-10 October 2003, Budapest, Hungary, WCDMP-56, WMO-TD No. 1236, World Climate Data and Monitoring Programme, World Meteorological Organization, Geneva, Switzerland
- Gonzalez-Rouco, J.F., Jiménez, J.L., Quesada, V., Valero, F. (2001) Quality control and homogeneity of precipitation data in the southwest of Europe. *Journal of Climate* **14** (5), 964-978.
- Groisman, P.Y., Easterling, D.R. (1994) Variability and trends of total precipitation and snowfall over the United States and Canada. *Journal of Climate* **7** (1), 184-205.
- Groisman, P.Y., Legates, D.R. (1994) The accuracy of United States precipitation data. *Bulletin of the American Meteorological Society* **75** (3), 215-227.
- Gullett, D.W., Vincent, L.A., Sajecki, P.J.F. (1990) Testing for homogeneity in temperature time series at Canadian climate stations. Report 90-4, Canadian Climate Center, Downsview, ON, Canada.
- Gullett, D.W., Vincent, L.A., Malone, L.H. (1991) Homogeneity testing of monthly temperature series. Application of multiphase regression models with mathematical changepoints. Report 91-10, Canadian Climate Center, Downsview, ON, Canada.
- Haimberger, L. (2007) Homogenization of radiosonde temperature time series using innovation statistics. *Journal of Climate* **20** (7), 1377-1403.
- Hanssen-Bauer, I., Forland, E.J. (1994) Homogenizing long Norwegian precipitation series. *Journal of Climate* **7** (2), 1001-1013.
- Hawkins, D.M. (1977) Testing a sequence of observations for a shift in location. *Journal of the American Statistical Association* **72** 180-186.
- Hawkins, D.M. (2001) Fitting multiple change-point models to data. *Computational Statistics and Data Analysis* **37** (3), 323-341.
- Heino, R. (1997) Metadata and their role in homogenization. Proc. First Seminar for Homogenization of Surface Climate Data. Hungarian Meteorological Service, Budapest, Hungary.
- Hinkley, D.V. (1969) Inference about the intersection in two-phase regression. *Biometrika* **56** (3), 496-504.
- Hsu, D.A. (1977) Tests for variance shifts at an unknown time point. *Applied Statistics* **26** (3), 179-184.

- Hubert, P., Carbonnel, J.P., Chaouche, A. (1989) Segmentation des séries hydrométéorologiques : application à des séries de précipitations et de débits de l'Afrique de l'Ouest. *Journal of Hydrology* **110** (3-4), 349-367.
- Inclan, C. (1993) Detection of multiple changes of variance using posterior odds. *Journal of Business and Economics Statistics* **11**, 189-300.
- Jandhyala, V.K., Fotopoulos, S.B., Evaggelopoulos, N. (1999) Change-point methods for weibull models with applications to detection of trends in extreme temperatures. *Environmetrics* **10** (5), 547-564.
- Jaruskova, D. (1996) Change-point detection in meteorological measurement. *Monthly Weather Review* **124** (7), 1535-1543.
- Jones, P.D., Raper, R.S., Diaz, H.F., Kelly, P.M., Wigley, T.M.L. (1986) Northern hemisphere surface air temperature variations: 1851-1984. *Journal of Climate and Applied Meteorology* **25** (2), 161-179.
- Karabork, M.C., Kahya, E., Komuscu, A.U. (2007) Analysis of Turkish precipitation data: homogeneity and the Southern Oscillation forcings on frequency distributions. *Hydrological Processes* **21** (23), 3203-3210.
- Karl, T.R., Quayle, R.J., Groisman, P.Y. (1993) Detecting climate variations and change : new challenges for observing and data management systems. *Journal of Climate* **6** (8), 1481-1494.
- Karl, T.R., Williams, C.N.J. (1987) An approach to adjusting climatological time series for discontinuous inhomogeneities. *Journal of Climate and Applied Meteorology* **26** (12), 1744-1763.
- Kehagias, A., Nidelkou, E., Petridis, V. (2006) A dynamic programming segmentation procedure for hydrological and environmental time series. *Stochastic Environmental Research and Risk Assessment* **20** (1-2), 77-94.
- Kendall, M.G. (1975) *Rank correlation methods*. Griffin, London, United Kingdom.
- Khaliq, M.N., Ouarda, T.B.M.J. (2007) A Note on the Critical Values of Standard Normal Homogeneity Test (SNHT). *International Journal of Climatology* **27** (5), 681-687.
- Kohler, M.A. (1949) On the use of double-mass analysis for testing the consistency of meteorological records and for making required adjustments. *Bulletin of the American Meteorological Society* **30** 188-189.
- Kruskal, W.H., Wallis, W.A. (1954) Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association* **47** (260), 583-621.
- Lamarque, P., Jourdain, S. (1994) Élaboration de longues séries climatologiques homogènes pour l'étude de l'évolution climatique. *La Météorologie* **8** (7), 61-68.
- Lanzante, J.R. (1996) Resistant, robust and non-parametric techniques for the analysis of climate data: theory and examples, including applications to historical radiosonde station data. *International Journal of Climatology* **16** (11), 1197-1226.

- Lanzante, J.R., Klein, S.A. (2003) Temporal homogenization of monthly radiosonde temperature data. Part I: Methodology. *Journal of Climate* **16** (2), 224-240.
- Lee, A.F.S., Heghinian, S.M. (1977) A shift of the mean level in a sequence of independent normal random variables: A Bayesian approach. *Technometrics* **19** (4), 503-506.
- Lee, C.B. (1998) Bayesian analysis of a change-point in exponential families with applications. *Computational Statistics and Data Analysis* **27** (2), 195-208.
- Li, Q.X., Liu, X.N., Zhang, H.Z., Peterson, T.C., Easterling, D.R. (2004) Detecting and adjusting temporal inhomogeneity in Chinese mean surface air temperature data. *Advances in Atmospheric Sciences* **21** (2), 260-268.
- Lubes-Niel, H., Masson, J.M., Paturel, J.E., Servat, E. (1998) Variabilité climatique et statistiques. Étude par simulation de la puissance et de la robustesse de quelques tests utilisés pour vérifier l'homogénéité de chroniques. *Revue des Sciences de l'Eau* **11** (3), 383-408.
- Lund, R., Reeves, J. (2002) Detection of undocumented changepoints: A revision of the two-phase regression model. *Journal of Climate* **15** (17), 2547-2554.
- Lund, R., Wang, X.L., Lu, Q., Reeves, J., Gallagher, C., Feng, Y. (2007) Changepoint detection in periodic and autocorrelated time series. *Journal of Climate* **20** (20), 5178-5190.
- Mann, H.B. (1945) Nonparametric tests against trend. *Econometrica* **13** (3), 245-259.
- Maronna, R., Yohai, V.J. (1978) A bivariate test for the detection of a systematic change in mean. *Journal of the American Statistical Association* **73** (363), 640-645.
- Mékis, E., Vincent, L. A. New developments in the homogenization of temperature and precipitation series in Canada. Proc. Fourth seminar for homogenization and quality control in climatological databases, 6-10 October 2003, Budapest, Hungary, WCDMP-56, WMO-TD No. 1236, World Climate Data and Monitoring Programme, World Meteorological Organization, Geneva, Switzerland.
- Mékis, E., Hogg, W.D. (1999) Rehabilitation and analysis of Canadian daily precipitation series. *Atmosphere-Ocean* **37** (1), 53-85.
- Menne, M.J., Duchon, C.E. (2001) A Method for Monthly Detection of Inhomogeneities and Errors in Daily Maximum and Minimum Temperatures. *Journal of Atmospheric and Oceanic Technology* **18** (7), 1136-1149.
- Menne, M.J., Williams, C.N.J. (2005) Detection of undocumented changepoints using multiple test statistics and composite reference series. *Journal of Climate* **18** (20), 4271-4286.
- Mestre, O. (2000) Méthodes statistiques pour l'homogénéisation des données climatiques. Thèse, Université Paul Sabatier, Toulouse, France.
- Metcalf, J.R., Routledge, B., Devine, K. (1997) Rainfall measurement in Canada: changing observational methods and archive adjustment procedures. *Journal of Climate* **10** (1), 92-101.

Moberg, A., Alexandersson, H. (1997) Homogenization of Swedish temperature data. Part II: homogenized gridded air temperature compared with a subset of global gridded air temperature since 1861. *International journal of climatology* **17** 35-54.

Moisselin, J.M., Schneider, M. (2002) Homogénéisation des séries françaises de précipitations couvrant le 20e siècle. *Houille Blanche-Revue internationale de l'eau* **6-7** 126-130.

Moisselin, J.M., Schneider, M., Canellas, C., Mestre, O. (2002) Les changements climatiques en France au XXe siècle. Étude des longues séries homogénéisées de données de température et de précipitations. *La Météorologie* **38** 45-56.

Muller-Westermeier, G. (2004) Statistical analysis of results of homogeneity testing and homogenisation of long climatological time series in Germany. Proc. Fourth seminar for homogenization and quality control in climatological databases, 6-10 October 2003, Budapest, Hungary, WCDMP-56, WMO-TD No. 1236, World Climate Data and Monitoring Programme, World Meteorological Organization, Geneva, Switzerland.

Ouarda, T.B.M.J., Rasmussen, P.F., Cantin, J.-F., Bobée, B., Laurence, R., Hoang, V.D., Barabé, G. (1999) Identification d'un réseau hydrométrique pour le suivi des modifications climatiques dans la province de Québec. *Revue des Sciences de l'Eau* **12** (2), 425-448.

Perreault, L., Haché, M., Slivitzky, M., Bobée, B. (1999) Detection of changes in precipitation and runoff over eastern Canada and U.S. using a Bayesian approach. *Stochastic Environmental Research and Risk Assessment* **13** (3), 201-216.

Peterson, T.C., Easterling, D.R. (1994) Creation of homogeneous composite climatological reference series. *International Journal of Climatology* **14** (6), 671-679.

Peterson, T.C., Easterling, D.R., Karl, T.R., Groisman, P., Nicholls, N., Plummer, N., Torok, S., Auer, I., Boehm, R., Gullett, D., Vincent, L., Heino, R., Tuomenvirta, H., Mestre, O., Szentimrey, T., Salinger, J., Forland, E.J., Hanssen-Bauer, I., Alexandersson, H., Jones, P., Parker, D. (1998) Homogeneity adjustments of in situ atmospheric climate data: a review. *International Journal of Climatology* **18** (13), 1493-1517.

Pettitt, A.N. (1979) A non-parametric approach to the change-point problem. *Applied Statistics* **28** (2), 126-135.

Potter, K.W. (1981) Illustration of a new test for detecting a shift in mean in precipitation series. *Monthly Weather Review* **109** (9), 2040-2045.

Rasmussen, P. (2001) Bayesian estimation of change points using the general linear model. *Water Resources Research* **37** (11), 2723-2731.

Reeves, J., Chen, J., Wang, X.L., Lund, R., Lu, Q.Q. (2007) A review and comparison of changepoint detection techniques for climate data. *Journal of Applied Meteorology and Climatology* **46** (6), 900-915.

Rhoades, D.A., Salinger, M.J. (1993) Adjustment of temperature and rainfall records for site change. *International Journal of Climatology* **13** (8), 899-913.

Schwarz, G. (1978) Estimating the dimension of a model. *The Annals of Statistics* **6** (2), 461-464.

- Seidou, O., Asselin, J.J., Ouarda, T.B.M.J. (2007) Bayesian multivariate linear regression with application to change point models in hydrometeorological variables. *Water Resources Research* **43** (8), W08401.
- Seidou, O., Ouarda, T.B.M.J. (2007) Recursion-based multiple changepoint detection in multiple linear regression and application to river streamflows. *Water Resources Research* **43** (7), W07404.
- Sevruk, B. (1996) Adjustment of tipping-bucket precipitation gauge measurement. *Atmospheric Research* **42** (1-4), 237-246.
- Slonosky, V.C., Jones, P.D., Davies, T.D. (1999) Homogenization techniques for European monthly mean surface pressure series. *Journal of Climate* **12** (8), 2658-2672.
- Solow, A.R. (1988) A bayesian approach to statistical inference about climate change. *Journal of Climate* **1** (5), 512-521.
- Solow, A.R. (1987) Testing for climate change: an application of the two-phase regression model. *Journal of Climate and Applied Meteorology* **26** (10), 1401-1405.
- Staudt, M., EstebanParra, M.J., CastroDiez, Y. (2007) Homogenization of long-term monthly Spanish temperature data. *International Journal of Climatology* **27** (13), 1809-1823.
- Szentimrey, T. (1999) Multiple analysis of series for homogenization (MASH). Proc. Second seminar for homogenization of surface climatological data, 9-13 November 1998, Budapest, Hungary, WMO-TD No. 962., World Meteorological Organization, Geneva, Switzerland.
- Szentimrey, T. (1997) Statistical procedure for joint homogenization of climatic time series. Proc. First seminar for homogenization of surface climatology data, 6-12 October 1996, Budapest, Hungary.
- Tayanc, M., Dalfes, H.N., Karaca, M., Yenigun, O. (1998) A comparative assesment of different methods for detecting inhomogeneities in Turkish temperature data set. *International Journal of Climatology* **18** (5), 561-578.
- Thomas, B.R., Kent, E.C., Swail, V.R. (2005) Methods to homogenize wind speeds from ships and buoys. *International Journal of Climatology* **25** (7), 979-995.
- Thompson, C.S. (1984) Homogeneity analysis of rainfall series: an application of the use of a realistic rainfall model. *International Journal of Climatology* **4** (6), 609-619.
- Tomozeiu, R., Busuioc, A., Marletto, V., Zinoni, F., Cacciamani, C. (2000) Detection of changes in the summer precipitation time series of the region Emilia-Romagna, Italy. *Theoretical and Applied Climatology* **67** (3-4), 193-200.
- Tuomenvirta, H. (2001) Homogeneity adjustments of temperature and precipitation series-finnish and nordic data. *International Journal of Climatology* **21** (4), 495-506.
- Venema, V., Mestre, O., Aguilar, E. (2008) Description of the COST-HOME monthly benchmark dataset with temperature and precipitation data for testing homogenisation algorithms. http://www.meteo.uni-bonn.de/mitarbeiter/venema/themes/homogenisation/description_monthly_benchmark_dataset.pdf

- Vincent, L.A. (1998) A technique for the identification of inhomogeneities in Canadian temperature series. *Journal of Climate* **11** (4), 1094-1105.
- Vincent, L.A., Zhang, X., Bonsal, B.R., Hogg, W.D. (2001) Homogenization of daily temperatures over Canada. *Journal of Climate* **15** (11), 1322-1334.
- Von Neumann, J. (1941) Distribution of the ratio of the mean square successive difference to the variance. *Annals of Mathematical Statistics* **12** (4), 367-395.
- Von Storch, H., Zwiers, F. (1999) *Statistical analysis in climate research*. Cambridge University Press, Cambridge, United Kingdom.
- Wald, A., Wolfowitz, J. (1943) An Exact Test for Randomness in the Non-Parametric Case Based on Serial Correlation. *Annals of Mathematical Statistics* **14** (4), 378-388.
- Wang, X.L. (2003) Comments on 'Detection of Undocumented Change-points: A revision of the Two-Phase regression model'. *Journal of Climate* **16** (20), 3383-3385.
- Wang, X.L., Wen, Q.H., Wu, Y.H. (2007) Penalized maximal t test for detecting undocumented mean change in climate data series. *Journal of Applied Meteorology and Climatology* **46** (6), 916-931.
- Wijngaard, J.B., Klein Tank, A.M.G., Können, G.P. (2003) Homogeneity of 20th century European daily temperature and precipitation series. *International Journal of Climatology* **23** (6), 679-692.
- Wilks, D.S. (2006) *Statistical methods in the atmospheric sciences*. Elsevier, United States.
- Yagouti, A., Boulet, G., Vescovi, L. (2006) Homogénéisation des séries de températures du Québec méridional et analyse de l'évolution du climat à l'aide d'indicateurs. Rapport de recherche, Consortium OURANOS. http://www.ouranos.ca/media/publication/8_Homogenisation.pdf
- Yang, D., Goodison, B.E., Metcalfe, J.R., Louie, P., Leavesley, G., Emerson, D., Hanson, C.L., Golubev, V.S., Elomaa, E., Gunther, T., Pangburn, T., Kang, E., Milkovic, J. (1999) Quantification of precipitation measurement discontinuity induced by wind shields on national gauges. *Water Resources Research* **35** (2), 491-508.
- Young, K.C. (1993) Detecting and removing inhomogeneities from long-term monthly sea level pressure time series. *Journal of Climate* **6** (6), 1205-1220.
- Yue, S., Pilon, P., Phinney, B., Cavadas, G. (2002) The influence of autocorrelation on the ability to detect trend in hydrological series. *Hydrological Processes* **16** (9), 1807-1829.
- Zurbenko, I., Porter, P.S., Rao, S.T., Ku, J.Y., Gui, R., Eskridge, R.E. (1996) Detecting discontinuities in time series of upper-air data: development and demonstration of an adaptive filter technique. *Journal of Climate* **9** (12), 3548-3560.
- Zwiers, F.W., Von Storch, H. (2004) On the role of statistics in climate research. *International Journal of Climatology* **24** (6), 665-680.

PARTIE 2 : ARTICLES

**6. Article 1: Synthèse des techniques d'homogénéisation des
séries climatiques et analyse d'applicabilité aux séries de
précipitations**

**Synthèse des techniques d'homogénéisation des séries climatiques et analyse d'applicabilité
aux séries de précipitations**

**A review of homogenization techniques for climate data and their applicability to
precipitation series**

Claudie Beaulieu*, Taha B.M.J. Ouarda* et Ousmane Seidou*

**Institut National de la Recherche Scientifique-Eau, Terre et Environnement, Université du Québec, Québec, QC,
Canada*

Soumis le 15 juillet 2005, accepté le 16 octobre 2006, publié le 1^{er} février 2007 dans le Journal
des Sciences Hydrologiques

Copyright © 2007 IAHS Press

Correspondance: Claudie Beaulieu, Chaire en hydrologie statistique, Chaire de recherche du Canada en estimation
des variables hydrologiques, INRS-ÉTÉ, Université du Québec, 490, de la Couronne, Québec (Québec) G1K 9A9,
CANADA. e-mail: claudie_beaulieu@ete.inrs.ca tel. (418) 654-2530 #4462

Résumé Au cours des dernières décennies, plusieurs méthodes d'homogénéisation ont été développées pour corriger les biais d'origines artificiels dans les séries climatiques. Elles ont été développées dans divers pays, pour différentes situations et avec des domaines d'application différents. On présente dans cet article une revue exhaustive des méthodes d'homogénéisation des séries climatiques publiées dans la littérature. Une analyse critique de ces méthodes ainsi qu'une réflexion sur leur applicabilité aux séries de précipitations totales annuelles sont également présentées.

Mots clés homogénéisation; série climatique; saut; tendance; test statistique; méthodes Bayésiennes

Abstract During the last decades, a considerable effort was spent on the development of homogenization techniques that can identify and correct anthropogenic bias in climatic series. These methods were developed in various countries, for different contexts and for different domains of applicability. The present paper is an exhaustive review of published homogenization techniques for climatic series. A critical analysis of the described methods and a discussion on their applicability to total annual precipitations are also presented.

Key words homogenization; climatic series; jump; trend; statistical test; Bayesian methods

NOTATIONS

β_{ij}	coefficient de la série voisine j dans le modèle de régression l
δ	amplitude du saut
λ_l	coefficient du temps dans le modèle de régression l
μ	moyenne
σ^2	variance
τ_l	ordonnée à l'origine dans le modèle de régression l
e_i	résidus de la régression au temps i
k	nombre de séries voisines
n	longueur de la série de base
p	position du saut
q	série de différences ou de ratios entre la série de base et ses voisines
x_{ij}	observation i de la série voisine j
$x_{1:i,j}$	segment des observations 1 jusqu'à i de la série voisine j
\bar{x}	moyenne de la série voisine (lorsqu'il n'y en a qu'une seule)
y_i	série de base au temps i
$y_{1:i}$	segment de la série de base des observations 1 jusqu'à i
\bar{y}	moyenne de la série de base

INTRODUCTION

Les caractéristiques statistiques des enregistrements à une station de mesure peuvent subir toutes sortes de perturbations artificielles qui ne reflètent pas les variations réelles du climat: déplacement de stations, remplacement d'instruments de mesure, changement d'heures d'observations ou encore modification de l'environnement immédiat de l'instrument de mesure. De ce fait, il arrive régulièrement que des décisions soient prises en se basant sur des données entachées d'erreurs.

Au cours des dernières décennies, plusieurs méthodes ont été développées pour corriger les biais d'origines artificiels des séries climatiques. En effet, le besoin de longues séries de données climatiques fiables se fait de plus en plus sentir dans divers domaines. Par exemple, les études des changements climatiques nécessitent la création de bases de données complètes avec lesquelles on pourra analyser de manière adéquate le signal climatique, suivre son évolution dans le temps et prévoir les changements futurs avec une incertitude minimale. Il est alors très important de développer des techniques robustes de détection de ces biais artificiels pour que les données utilisées se rapprochent le plus possible des observations qui auraient été faites sans perturbation des conditions de mesure. Le processus de détection et de correction des ruptures d'origine non climatique est appelé homogénéisation.

Ce travail présente une revue bibliographique des méthodes d'homogénéisation des données climatiques décrites dans la littérature. On en fera ensuite une analyse critique, suivie d'une réflexion sur leur applicabilité aux séries de précipitations enregistrées au Québec. Cette analyse facilitera le choix de la méthode à utiliser étant donné l'information disponible de la station à homogénéiser et de ses stations voisines. Cette revue de littérature constitue un travail préliminaire à une comparaison des méthodes d'homogénéisation des précipitations au Québec méridional.

Le reste de l'article est organisé comme suit : (a) explication de la problématique de l'homogénéisation des données; (b) présentation des méthodes d'homogénéisation utilisées pour les données climatiques; (c) présentation des études comparatives réalisées; (d) discussion des méthodes d'homogénéisation; (e) une analyse d'applicabilité pour l'homogénéisation des séries de précipitations totales annuelles et finalement (f) conclusion.

L'HOMOGENÉISATION DES DONNÉES

Le terme 'inhomogénéités' représente les variations non naturelles qui sont causées par des modifications dans les réseaux d'observations. L'homogénéisation consiste à détecter puis à corriger ces variations. Les sauts de moyenne et les tendances sont les deux types d'inhomogénéités les plus communes. Par exemple, les mesures de précipitations sont perturbées par le vent et l'effet Venturi produit au dessus du cône du pluviomètre. Donc, toute modification ayant un effet sur le vent induira des sauts dans les données (e.g. un changement d'emplacement, de la forme du capteur ou de sa hauteur au dessus du sol). Une modification dans l'environnement immédiat d'une station, induite par exemple par l'urbanisation, la reforestation, ou l'industrialisation, suscite des changements plus graduels (e.g. Peterson *et al.* 1998) et ce, en particulier sur les températures.

De toutes les méthodes d'homogénéisation de données climatiques développées dans la littérature, celles ayant la capacité de bien détecter des tendances sont très rares. En effet, une tendance est plus délicate à quantifier car il faut identifier correctement le début et la fin de cette tendance ainsi que son amplitude.

En pratique, il est assez difficile de déterminer si une rupture dans une série représente un changement dans le climat régional ou une variation due à un changement à la station de mesure. Deux types d'informations permettent de pencher pour l'une ou l'autre des explications : les enregistrements aux stations voisines (s'il y en a) et les métadonnées (s'il y en a également).

Utilisation de l'information aux stations voisines

En homogénéisation, on compare souvent les données d'une certaine station avec celles des stations voisines pour éviter qu'un changement climatique soit interprété comme une inhomogénéité. La série dont on veut vérifier l'homogénéité est appelée *série de base*. Les *séries voisines* sont supposées être climatiquement similaires à la série de base (souvent des observations de stations géographiquement voisines). Les séries voisines doivent être homogènes, sinon des inhomogénéités dans une de ces dernières pourraient être attribuées à la série de base. Or, il est difficile de garantir que les séries voisines sont exemptes d'inhomogénéités. Il existe des techniques qui permettent de minimiser les inhomogénéités potentielles et de former une *série de référence* homogène. Une *série de référence* est une fonction d'une ou de plusieurs séries voisines. On peut trouver dans Peterson *et al.* (1998) une revue des techniques utilisées pour minimiser les inhomogénéités potentielles de ces séries.

Pour les précipitations, il est souvent difficile de trouver des stations voisines qui appartiennent à la même région climatique que la série de base étant donné la forte variabilité spatio-temporelle de cette variable. Lorsque les stations voisines utilisées sont trop éloignées ou d'altitudes très différentes, les inhomogénéités présentes dans la série de base risquent d'être masquées par une trop grande divergence entre ces caractéristiques. La qualité de l'homogénéisation des précipitations dépend alors de la densité du réseau d'observations et de la topographie de la zone où ont été prises les données. De plus, le degré d'importance de ces caractéristiques dépend fortement du pas de temps des données que l'on veut homogénéiser.

Les métadonnées

Les métadonnées sont des informations historiques sur les conditions dans lesquelles les données à une station ont été enregistrées. Elles sont constituées des enregistrements de la station, des annuaires météorologiques, des fiches d'inspection, des photographies de la station et de son environnement, d'une entrevue avec la personne responsable d'une station, etc. (Aguilar *et al.*,

2003). En homogénéisation, les métadonnées sont consultées pour expliquer la cause d'une inhomogénéité. Cependant, les informations sont souvent incomplètes, inexistantes, erronées ou remplies d'informations non pertinentes. Il se peut alors qu'une inhomogénéité détectée soit impossible à valider par les métadonnées. A l'inverse, un changement recensé dans les métadonnées peut ne pas avoir influencé sensiblement la mesure.

MÉTHODES UTILISÉES POUR HOMOGENÉISER DES DONNÉES CLIMATIQUES

Diverses techniques ont été développées pour accommoder différents types de facteurs tels que la variable à homogénéiser, la variabilité spatiale et temporelle des données selon l'endroit où les stations sont situées, la longueur des séries et le nombre de données manquantes, les métadonnées disponibles et la densité du réseau d'observations (Aguilar *et al.*, 2003). Les techniques de détection des inhomogénéités varient également selon l'objectif pour lequel on les applique et la philosophie de chaque équipe de travail.

Malgré la grande diversité des méthodes d'homogénéisation, elles peuvent néanmoins être classifiées en deux catégories principales qui les rejoignent toutes: méthodes subjectives ou objectives. Lorsque l'emplacement d'une discontinuité est détecté à l'œil nu sur un graphique, la méthode appartient à la classe subjective même si des tests statistiques sont appliqués par la suite. Par contre, les méthodes objectives ne dépendent pas du jugement de l'utilisateur pour localiser les inhomogénéités. Les méthodes Bayésiennes, qui seront présentées plus loin, pourraient être considérées comme subjectives ou objectives selon qu'elles incorporent ou non l'opinion du modélisateur par le biais de l'information a priori informative. Comme aucune des méthodes présentées dans cet article n'utilise de distributions a priori informatives, les méthodes Bayésiennes sont présentées parmi les approches objectives.

Cette section présente les approches subjectives et objectives pour l'homogénéisation des données climatiques rencontrées dans la littérature. Par la suite, les études comparatives réalisées

sur ces méthodes sont discutées dans le but de mieux les classer. Enfin, les approches utilisées pour homogénéiser les précipitations sont également présentées par pays.

Approches subjectives

La plupart des approches subjectives sont basées sur l'examen visuel d'observations cumulées, de différences ou de ratios ou encore de déviations cumulées entre la série de base et des séries voisines pour que les inhomogénéités soient plus facilement détectables visuellement. Dans cette section, elles sont regroupées selon la variable que l'on inspecte sur le graphique.

Examen graphique des doubles accumulations. Plusieurs techniques se basent sur une analyse visuelle de valeurs cumulées. Par exemple, l'analyse des doubles accumulations (Kohler, 1949) consiste à faire une régression linéaire simple sur les valeurs cumulées de la série de base (y) en fonction des valeurs accumulées d'une série de référence (x). Ensuite, on produit un graphique des couples de points (x, y) sur lesquels on superpose la droite de régression. Si les séries sont homogènes, les points sont disposés aléatoirement autour de la droite de régression. Par contre, un changement à l'une ou l'autre des deux stations se remarque par une cassure de la pente. La technique des ellipses de Bois (Bois, 1986) constitue une extension de l'analyse des doubles accumulations. Par contre, on examine plutôt les résidus cumulés de la régression en fonction du temps. Une courbe de contrôle est calculée de façon à ce que si une valeur cumulée sort de cette limite de contrôle, l'hypothèse d'homogénéité est rejetée. Enfin, la méthode du vecteur des indices annuels (Brunet-Moret, 1979), aussi appelée méthode du vecteur régional, consiste à appliquer l'analyse des doubles accumulations en utilisant une série de données fictives comme série de référence. On forme donc une série de référence homogène qui comprend les effets de persistance, les tendances et les cycles saisonniers de la région climatique étudiée à partir de plusieurs stations appartenant à cette zone climatique. Par la suite, on applique l'analyse des doubles accumulations pour comparer la série de base à la série de référence créée.

Examen graphique des déviations cumulées. L'analyse graphique des déviations cumulées a été développée pour homogénéiser des séries de précipitations totales annuelles par Craddock (1979). On calcule les déviations de la série de base par rapport au ratio \bar{y}/\bar{x} , où y représente la série de base et x une série de référence homogène. Un graphique des déviations cumulées en fonction du temps est produit. Un saut de moyenne dans la série de base se remarque facilement par un changement de pente sur le graphique des déviations cumulées en fonction du temps. Craddock (1979) répète le processus plusieurs fois avec différentes séries de référence et compare les graphiques obtenus. Dans le même ordre d'idées, la méthode des sommes cumulatives parallèles consiste à calculer les différences entre les observations mensuelles d'une variable climatique à la station de base et celles à plusieurs stations voisines, puis de cumuler chaque série différenciée. Les différences cumulées sont placées parallèlement sur un même graphique. Lorsqu'un changement de pente apparaît sur une seule ligne du graphique, alors il y a probablement eu un changement à la station voisine en cause. Par contre, si plusieurs lignes du graphique témoignent du même changement de pente au même endroit, alors cette modification s'est probablement produite à la station de base. Une technique permettant d'estimer l'amplitude du saut ainsi que son intervalle de confiance est présentée dans Rhoades et Salinger (1993). De plus, on y présente diverses façons de déterminer le seuil significatif du saut dans le cas de stations isolées.

Examen graphique des séries de différences ou de ratios. Jones *et al.* (1986) propose une méthode qui consiste à examiner graphiquement les séries de différences entre la série de base et ses voisines. Les points de discontinuités sont identifiés par un examen graphique des différences mensuelles entre la série de base et chaque série de référence. Lorsque le même saut se répète sur plusieurs graphiques, alors il provient de la série de base. Ensuite, les métadonnées sont consultées et si les positions d'inhomogénéités détectées sont validées, la série est ajustée. Par ailleurs, Tayanc (1998) propose d'appuyer cette méthode de deux tests non paramétriques.

Lorsqu'un saut est apparent par un examen visuel, les tests de Kruskal-Wallis et de Wald-Wolfowitz décrits par Tayanc *et al.* (1998) sont appliqués sur les séries différenciées. Les résultats des deux tests donnent le niveau de signification des inhomogénéités détectées visuellement. Pour les précipitations, on peut analyser graphiquement les ratios de précipitations totales annuelles entre la série de base et une série de référence (Craddock, 1979). Pour une série homogène, les ratios devraient être distribués aléatoirement autour d'une certaine constante. Par contre, une discontinuité entraîne une déviation des valeurs par rapport à cette constante. Le but de cette méthode n'est pas de fournir un outil de correction, mais plutôt de montrer qu'il y a une discontinuité à expliquer dans la série.

Examen graphique des données filtrées. Le filtre de Kolmogorov-Zurbenko est applicable à des données contenant des tendances et des cycles saisonniers (Zurbenko *et al.*, 1996). Ce filtre a été testé directement sur des données prises en altitude (température et humidité). Le filtre est appliqué sur les données brutes. Lorsque la fenêtre mobile s'approche d'une région dans laquelle il y a une discontinuité, elle peut être identifiée par une croissance de la variabilité ou de l'amplitude de la pente de la série filtrée. Donc, les discontinuités de la série deviennent évidentes en représentant la série filtrée sur un graphique. Leur amplitude est estimée en calculant la différence des moyennes des valeurs avant et après le saut des données filtrées. Le seuil significatif des sauts est obtenu par un test sur la variance de la série filtrée (Zurbenko *et al.*, 1996).

Approches objectives

Les approches objectives peuvent se classer selon le modèle qui représente la rupture dans la série de base. On distingue (a) les modèles de régression linéaire avec des changements dans un ou plusieurs coefficients de régression, (b) les modèles avec une forme paramétrique différente de la régression linéaire et (c) les modèles sans forme paramétrique explicitement spécifiée. À l'intérieur de chaque classe, on définit des sous classes en fonction du nombre de ruptures (une

ou plusieurs) et de la nature de la rupture (changement de moyenne ou de tendance). Dans les sections suivantes, on regroupe les méthodes inventoriées selon la classification qu'on vient de décrire.

Modèles de régression linéaire avec des changements dans un ou plusieurs paramètres.

(a) un changement de moyenne

Plusieurs méthodes sont basées sur la recherche d'un saut dans une série de différences ou de ratios formée de la série de base avec une série de référence. Cela permet d'inclure l'information d'une série de référence et d'avoir un modèle très simple :

$$q_i = \begin{cases} \tau + e_i, & i = 1, \dots, p \\ \tau + \delta + e_i, & i = p + 1, \dots, n \end{cases} \quad (1)$$

où q_i est la valeur de la série de différences/ratios au temps i , τ est la moyenne, p est la position d'un saut potentiel et δ son amplitude. Les résidus du modèle sont représentés par e_i . Pour détecter la position du saut et/ou estimer son amplitude, plusieurs techniques peuvent être employées.

Dans un premier temps, lorsque la position p est connue, par l'information obtenue des métadonnées, on peut appliquer un test de différences de moyennes local pour vérifier la significativité de l'amplitude (Karl and Williams, 1987). On forme des séries de différences (température) ou de différences entre les logarithmes des rapports (précipitation) de la série de base avec ses voisines. Ensuite, un intervalle de confiance basé sur le test de Student est calculé pour la température et sur le test de Wilcoxon pour les précipitations. Par la suite, une procédure est appliquée pour corriger la série de base des sauts significatifs.

Lorsque la position du saut est inconnue, plusieurs méthodes peuvent résoudre ce problème. La méthode Bayésienne univariée pour un saut proposée par Lee et Heghinian (1977)

et expliquée par Ouarda *et al.* (1999) peut être employée. La distribution *a priori* est déterminée par la série de base. La procédure calcule, pour chaque année, la probabilité que le saut se produise à ce moment étant donné les observations recueillies. De la même manière, l'amplitude du saut possède sa fonction de densité *a posteriori*. Le seuil significatif de l'amplitude du saut est déterminé en calculant un intervalle de crédibilité Bayésien. Une extension de cette méthode pour détecter un saut dans une série de précipitations ou de débits a été développée (Perreault *et al.*, 1999). Selon cette méthode, on ne doit pas faire l'hypothèse qu'il y a eu un changement de moyenne dans la série. Dans cette version révisée de la méthode, les distributions *a priori* sont plus flexibles.

Des tests statistiques classiques peuvent également résoudre ce problème. Le test le plus utilisé est celui développé par Alexandersson (1986) pour homogénéiser des séries de précipitations totales annuelles ou de températures moyennes annuelles (Alexandersson and Moberg, 1997) en se basant sur un test du rapport des vraisemblances. On calcule une statistique pour toutes les positions de sauts possibles. La statistique maximale correspond à l'emplacement le plus probable pour un saut et son seuil significatif se détermine à partir d'un tableau de valeurs critiques (Alexandersson, 1986; Khaliq & Ouarda, 2006). En principe, les mathématiques du test sont applicables à une discontinuité par série. En pratique, on applique successivement le test jusqu'à ce que tous les segments de la série soient considérés homogènes. La procédure consistant à chercher le maximum de la statistique en fonction de la position s'applique à d'autres statistiques. La méthode de Jaruskova (1996) et les tests sur les déviations cumulées (Buishand, 1984; Buishand, 1982) se basent sur le même principe. Enfin, le test de Student séquentiel consiste à appliquer un test de Student pour comparer 5 observations avant et après chaque position de saut potentielle (Gullett *et al.*, 1990). Les endroits où la statistique calculée dépasse la valeur critique de la distribution de Student avec 8 degrés de liberté correspondent aux positions de sauts significatifs.

Une autre façon de procéder est de détecter un saut dans une série de base qui ne se produit pas dans la série de référence. Un modèle représentant une telle série peut s'exprimer ainsi :

$$y_i = \begin{cases} \tau + \beta x_i + e_i, & i = 1, \dots, p \\ \tau + \delta + \beta x_i + e_i, & i = p + 1, \dots, n \end{cases} \quad (2)$$

où y_i est la valeur de la série de base au temps i , x_i est la valeur de la série de référence au temps i , τ est la moyenne, β est le coefficient de la série de référence, p est la position d'un saut potentiel et δ son amplitude. Les résidus du modèle sont représentés par e_i . Pour détecter la position du saut et/ou estimer son amplitude, on peut employer l'approche bivariée et la méthode Bayésienne bivariée pour un saut.

L'approche bivariée est une technique statistique classique basée sur le rapport des vraisemblances (Maronna & Yohai, 1978; Potter, 1981; Young, 1993). La série de base et la série de référence appartiennent à une distribution normale bivariée et on teste l'hypothèse que les moyennes sont différentes. Pour détecter la position du changement, on utilise la procédure consistant à chercher le maximum de la statistique en fonction de la position.

La méthode Bayésienne bivariée pour un saut constitue une extension de la méthode Bayésienne univariée pour un saut. Asselin *et al.* (1999) ont développé ce modèle qui diffère du précédent en incluant une série de référence qui peut être corrélée avec la série de base. Les observations de la série de base et de la série de référence sont distribuées selon la loi normale bivariée. Les distributions *a priori* des paramètres de ce modèle sont les mêmes que celles du modèle univarié présenté dans la section précédente. Les détails concernant les densités marginales et la densité conjointe *a posteriori* de la date du changement et de son amplitude et sont présentés dans Asselin *et al.* (1999). Une autre approche Bayésienne pour un saut basée sur la distribution normale multivariée a été proposée pour détecter un changement de moyenne

commun dans plusieurs séries hydrologiques corrélées (Perreault *et al.*, 2000). Dans ce modèle, on teste l'hypothèse qu'il n'y a pas de changement dans les séries contre l'hypothèse qu'il y a un saut simultané à chacun des sites. Par la suite, Ouarda *et al.* (2005) ont développé un modèle de régression Bayésienne. Un changement de moyenne dans un modèle de régression multivariée est considéré. Cette méthode est basée sur des hypothèses *a priori* non informatives et les distributions *a posteriori* sont obtenues par l'échantillonnage de Gibbs. Encore une fois, la méthode vérifie la présence d'une rupture concomitante à plusieurs séries.

(b) un changement de moyenne et/ou de tendance

Il existe des méthodes d'homogénéisation basées sur des modèles de régression linéaire qui recherchent à la fois des sauts et des tendances. Ce sont des modèles dans lesquels on suppose les propriétés de normalité, d'indépendance et d'hétéroscédasticité des données et qui sont résolus par des techniques moindres carrés standard. Ceci s'applique à toutes les techniques basées sur des modèles de régression présentées dans cette section. Par exemple, la régression multiple se base sur l'application de plusieurs modèles de régression pour homogénéiser des séries de températures (Vincent, 1998). Le premier représente une série homogène. On considère que lorsque les résidus sont indépendants, le modèle appliqué ajuste bien les données. Dans le cas contraire, le modèle n'ajuste pas bien les données et il pourrait y avoir une inhomogénéité dans la série de base. Le test de Durbin-Watson ou un intervalle de confiance sur l'autocorrélation des résidus permettent de vérifier cette hypothèse. Lorsque la série de base est considérée inhomogène, on ajuste un autre modèle pour déterminer le type de discontinuité dans la série de base. Pour identifier un saut, le modèle suivant est appliqué :

$$y_i = \begin{cases} \tau + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + e_i & i = 1, \dots, p \\ \tau + \delta + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + e_i & i = p+1, \dots, n \end{cases} \quad (3)$$

où y_i est la valeur de la série de base au temps i , x_{ik} est la valeur de la série de référence k au temps i , τ est la moyenne, β_k est le coefficient de la série de référence k , p est la position du saut potentiel et δ son amplitude. Il y a n observations et k séries de référence. Les résidus du modèle sont représentés par e_i . L'emplacement du saut, p , est déterminé en ajustant le modèle pour toutes les positions possibles et en sélectionnant celui avec la plus petite somme des carrés résiduelle. Il s'agit du modèle qui ajuste le mieux les données au sens des moindres carrés et se valide avec le test de Fisher. L'estimation de l'amplitude du saut est donnée par δ et son seuil significatif se calcule avec la statistique de Student. Des modèles représentant une série avec une tendance et une série qui contient une tendance et un saut sont ajustés par la suite (Vincent, 1998).

Un modèle de régression à deux phases permet également de détecter un changement de moyenne et de tendance dans une série. La régression à deux phases a été exploitée par plusieurs chercheurs en climatologie (Wang, 2003; Lund and Reeves, 2002; Easterling and Peterson, 1995; Solow, 1987). La première technique proposée vise à détecter un changement de pente et/ou de moyenne dans une série climatique (Solow, 1987). D'abord, un modèle représentant une série homogène est ajusté. Le deuxième modèle ajusté représente une série dans laquelle il y a une discontinuité à un certain point p :

$$y_i = \begin{cases} \tau_1 + \lambda_1 i + e_i, & i = 1, \dots, p \\ \tau_2 + \lambda_2 i + e_i, & i = p + 1, \dots, n \end{cases} \quad (4)$$

où y_i est la valeur de la série de base au temps i , τ_1 et τ_2 sont les moyennes avant et après le changement, λ_1 et λ_2 sont les tendances avant et après le changement et p est la position du changement. Les résidus du modèle sont représentés par e_i . L'emplacement du saut est déterminé selon les moindres carrés. Plusieurs modifications ont été apportées :

- (i) Easterling et Peterson (1995) appliquent le modèle sur une série de différences créée avec des stations voisines, ils ne forcent pas les deux droites de régression à se joindre, ils appliquent la technique itérativement pour détecter plusieurs sauts et évaluent les seuils significatifs par une procédure de permutation multiple (Mielke *et al.*, 1981).
- (ii) Lund et Reeves (2002) proposent une statistique de Fisher révisée.
- (iii) Wang (2003) propose un modèle dans lequel les pentes sont égales avant et après la rupture.

Par ailleurs, la régression multiphase a été développée afin de détecter des sauts et des tendances dans des données de température mensuelle (Gullett *et al.*, 1991). Cette méthode consiste à ajuster sept modèles de régression différents dans lesquels la série de base constitue la variable dépendante et les variables explicatives sont le temps et des séries voisines. Le premier représente une série homogène tandis que les six autres évoquent une série de base comprenant de une à six discontinuités. Celui qui comprend une discontinuité se représente par :

$$y_i = \begin{cases} \tau_1 + \lambda_1 i + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + e_i, & i = 1, \dots, p \\ \tau_2 + \lambda_2 i + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + e_i, & i = p+1, \dots, n \end{cases} \quad (5)$$

où y_i est la valeur de la série de base au temps i , x_{ik} est la valeur de la série de référence k au temps i , τ_1 et τ_2 sont les moyennes avant et après le changement, λ_1 et λ_2 sont les tendances avant et après le changement, β_k est le coefficient de la série voisines k et p est la position du changement. Les résidus du modèle sont représentés par e_i . À chaque étape, on détermine les positions qui ajustent le mieux les données au sens des moindres carrés. Le nombre de discontinuités dans la série est déterminé en choisissant le modèle le plus simple qui améliore significativement le dernier modèle ajusté. Par la suite, le seuil significatif de chaque inhomogénéité potentielle est déterminé par la statistique de Student.

(c) plusieurs sauts

Dans le cas où on recherche plusieurs inhomogénéités dans la série de base, on ne fait plus face à un problème de test d'hypothèse, mais plutôt au choix de modèle ainsi qu'à la combinatoire. Dans cette section, on présente les techniques de détection de sauts multiples en séparant les problèmes de choix de modèle et de combinatoire.

(i) choix de modèle

Le choix des emplacements de sauts et l'estimation de leurs amplitudes peut se faire en utilisant une procédure de vraisemblance pénalisée par un critère comme le critère de Schwarz (1978), d'Akaike (1973) ou de Caussinus and Lyazrhi (1997). Les critères de Schwarz (BIC) et d'Akaike (AIC) sont des classiques en choix de modèle et sont utilisés pour faire de la sélection de prédicteurs en régression multiple. Le critère de Caussinus et Lyazrhi est spécifiquement conçu pour les modèles à rupture. D'ailleurs, il performe beaucoup mieux que AIC et BIC sur le modèle en marche d'escalier (Caussinus et Mestre, 2004). La méthode développée par Caussinus et Mestre (2004) se base sur ce critère pour le choix de modèle combinée à un algorithme de programmation dynamique. Les séries sont homogénéisées en étant tour à tour série de base alors que toutes les autres deviennent des séries voisines. Des séries de différences ou de ratios sont formées et lorsqu'un changement revient dans toutes les séries, on considère qu'il est attribuable à la série de base.

Une méthode Bayésienne ayant pour paramètres le nombre de changements et la position de chacun peut également être utilisée pour détecter des sauts multiples. L'approche Bayésienne multi-sauts est une adaptation de la méthode présentée dans Fearnhead (2004) qui cherche à détecter un changement de paramètres dans une séries de données distribuée selon une loi quelconque. Ce modèle a été adapté par Seidou et Ouarda (2006) au cas où les données sont normales et la moyenne de la série climatique considérée varie en fonction de plusieurs

covariables. L'observation i de la série de base est noté par $y_i (i = 1, \dots, n)$ et $x_{ij} (i = 1, \dots, n; j = 1, \dots, k)$ représente l'observation i de la série voisine j . La distribution de y_i est $N(\sum_{j=1}^k \beta_j x_{ij}, \sigma^2)$. On a considéré des distributions *a priori* uniforme ou normale pour les paramètres β_j et uniforme ou gamma pour σ . La probabilité de rupture *a priori* sur un tronçon est une binomiale. En sortie, le modèle fournit la densité de probabilité *a posteriori* du nombre de changements, et la position des sauts étant donné le nombre de changements le plus probable.

Pour détecter des sauts multiples, on a également la possibilité d'appliquer des tests localement comme le test de Student ou le test Scheffé. Il faut alors posséder des métadonnées complètes qui nous indiquent les positions de changements potentielles. Par ailleurs, on peut également appliquer une méthode développée pour détecter plusieurs sauts à la fois comme la méthode de Szentimrey (1996;1999). Cette dernière se base sur le principe du rapport de vraisemblance, mais dans le cas de sauts multiples. La particularité de cette méthode est le traitement des séries voisines qui ne sont pas considérées homogènes. La méthode consiste à attribuer un poids aux séries de différences formées à partir des séries voisines. Le poids optimal à allouer est obtenu en minimisant la variance des résidus des séries de différences. La procédure de détection s'applique et retourne les emplacements des changements potentiels. Il est également possible d'obtenir des intervalles de confiance pour l'emplacement du saut et son amplitude.

(ii) combinatoire

Ensuite, on fait face au problème de combinatoire. Puisqu'il serait impraticable de faire une recherche exhaustive de toutes les combinaisons de ruptures possibles, la plupart des auteurs utilisent une technique pas-à-pas simple. Cela consiste à détecter une rupture, la conserver en mémoire, en détecter une seconde, et ainsi de suite. La plupart des techniques développées pour détecter un seul saut s'appliquent de cette façon en pratique. Par exemple, la procédure de

segmentation de série hydrométéorologique (Hubert *et al.*, 1989) utilise cet algorithme. La technique cherche la segmentation optimale au sens des moindres carrés, celle pour laquelle l'écart entre la série et la segmentation considérée est minimale. La procédure consiste à calculer l'écart quadratique entre la série et la segmentation considérée. Une fois que la segmentation optimale est identifiée, on teste l'égalité des moyennes des segments consécutifs au moyen du test des contrastes de Scheffé. La segmentation obtenue sera acceptée seulement si le test de Scheffé est significatif. Une généralisation de cette méthode a été formulée par Kehagias *et al.* (2006). Dans cette dernière, on choisit le nombre de sauts en se basant sur le critère de Schwarz. Les techniques pas-à-pas deviennent problématique lorsque les ruptures consécutives sont trop proches.

Lorsque les données sont gaussiennes, la technique de programmation dynamique fournit un algorithme équivalent à la recherche exhaustive de toutes les combinaisons de ruptures (Hawkins, 2001). Cette technique est plus chère que du pas-à-pas simple, mais optimale.

Modèles avec une forme paramétrique différente de la régression linéaire. Des méthodes différentes de celles qui ont été présentées jusqu'à maintenant peuvent être utilisées pour l'homogénéisation des données climatiques. Thompson (1984) a développé une méthode pour homogénéiser une série de précipitations totales mensuelles basée sur un processus aléatoire de Poisson. La fréquence des précipitations est caractérisée par un processus de Poisson et son amplitude suit indépendamment une loi exponentielle. Cette méthode s'applique à une station qui a été déplacée à des dates connues. La série est divisée en plusieurs segments selon ces dates de changement. Le segment avec le plus d'années d'observations devient la série de référence et les autres séries seront comparées à celle-ci. En se basant sur l'information historique, les estimateurs du maximum de vraisemblance sont dérivés pour chaque segment de la série. Le test d'homogénéité se fait à partir du test du rapport de vraisemblance donné par Thompson (1984). On vérifie si les estimateurs du maximum de vraisemblance varient entre les segments. Un

segment inhomogène est corrigé par le ratio de précipitations moyennes mensuelles avant et après le changement de site. Pour vérifier que la série corrigée est effectivement homogène, la statistique de Mann-Whitney est calculée afin de tester l'hypothèse que les deux séries font partie de la même population. Ce processus est répété pour tous les segments.

Une méthode basée sur l'analyse des ondelettes a été proposée pour ajuster les observations de températures prises avant 1850 parce que les observations de cette époque sont souvent mal documentées (Datsenko *et al.*, 2002). Elle se définit par trois idées principales. En premier lieu, les séries climatiques sont décomposées en différentes échelles de temps et l'ajustement se fait sur ces différentes échelles. En deuxième lieu, on suppose que les inhomogénéités dans les séries sont nombreuses et qu'elles peuvent varier continuellement. En troisième lieu, on fait l'hypothèse que les données prises avant que l'on adopte des méthodes de calcul modernes sont de mauvaise qualité. La méthode proposée ajuste les séries même si l'amplitude des inhomogénéités détectées est inférieure à un certain seuil significatif en se basant sur l'idée que les ajustements apportés sur des données de bonne qualité seront pratiquement négligeables. L'analyse se réalise simultanément sur la plus longue période d'observations commune de la série de base ainsi que plusieurs séries prises à des stations voisines à l'aide de fonctions d'ondelettes. On divise les séries en deux parties : la première contient des données qui renferment des biais et des erreurs tandis que les données de la partie récente sont supposées homogènes. L'information historique peut servir à déterminer la date de séparation des séries. La méthode consiste à estimer les statistiques de la partie récente des séries, et à déterminer les ajustements optimaux à faire sur la première partie des séries par une stratégie Bayésienne.

Modèles sans forme paramétrique explicitement spécifiée. Les modèles basés sur l'application successive de tests non paramétriques entrent dans cette catégorie. Par exemple, la procédure du test de Mann-Whitney séquentiel corrige les inhomogénéités pas-à-pas (Lanzante, 1996). On inclue également dans cette section des méthodes hybrides qui sont constituées de

l'application simultanée de plusieurs tests paramétriques et/ou non paramétriques successifs qui ne vérifient pas les mêmes hypothèses. Ceci vise à éviter de réduire le nombre de fausses alarmes et de ne pas se méprendre sur le type de discontinuité détecté. Par exemple, Boroneant et Tomozeiu (1999) proposent d'appliquer successivement les tests de Pettitt, de Mann-Kendall et d'Alexandersson pour l'homogénéisation de séries de températures mensuelles. Le test de Pettitt est une approche non paramétrique pour identifier un changement de moyenne ou de variance dans une série (Pettitt, 1979). Le test de Mann-Kendall présenté dans Boroneant et Tomozeiu (1999) permet également de tester la stabilité de la moyenne et de la variance d'une série d'observations. Il s'agit d'une procédure non paramétrique dans laquelle on compare chaque observation de la série de base avec ses observations précédentes. Le test d'Alexandersson, tel que présenté précédemment, est à son tour appliqué. La méthode proposée par Wijngaard (2003) consiste à appliquer 4 tests d'homogénéité différents (test d'Alexandersson, test sur l'étendue des déviations cumulées standardisées, test de Pettit et ratio de Von Neumann) sur des données annuelles et classer les séries selon le nombre de tests significatifs. Les tests sont présentés en détails dans Wijngaard et al (2003). La station est classifiée selon le nombre de tests significatifs au seuil de 1%. Elle appartient à la première catégorie lorsque 1 ou aucun test n'est significatif, à la deuxième catégorie lorsque deux tests rejettent l'hypothèse nulle et à la troisième catégorie lorsque 3 ou 4 des tests sont significatifs. Les séries classées dans la catégorie 1 peuvent être utilisées ultérieurement car elles semblent homogènes. Les séries appartenant aux classes 2 ou 3 doivent être vérifiées en consultant les métadonnées.

ÉTUDES COMPARATIVES RÉALISÉES

Quelques études comparatives ont été effectuées sur des séries climatiques simulées (Ducre-Robitaille *et al.*, 2003; Slonosky *et al.*, 1999; Lubes-Niel *et al.*, 1998; Zurbenko *et al.*, 1996; Easterling and Peterson, 1992). Étant donné que les résultats de ces études dépendent fortement

de l'effectif et des caractéristiques des séries synthétiques générées, les conclusions doivent être considérées avec prudence.

La plus récente (Ducré-Robitaille *et al.*, 2003) compare huit techniques différentes : la méthode d'Alexandersson pour les sauts et pour les tendances, la régression multiple, la régression à deux phases (version proposée par Easterling et Peterson; 1995), le test de Student séquentiel, le test de Mann-Whitney séquentiel et la méthode Bayésienne univariée pour un saut. Le test de Mann-Whitney séquentiel utilisé ici est similaire au test de Student séquentiel en remplaçant seulement le test de Student par le test de Mann-Whitney. Ces méthodes ont été comparées sur des séries simulées de température avec un nombre aléatoire de sauts. Les résultats se résument ainsi :

- (a) Pour identifier les séries homogènes, la méthode d'Alexandersson pour les sauts, la régression multiple et l'approche Bayésienne ont été performantes.
- (b) Les sauts de grandes amplitudes sont toujours bien détectés. Par contre, les petits sauts sont plus facilement détectables par les méthodes d'Alexandersson et Bayésienne en utilisant une série de référence.
- (c) Les méthodes possédant la capacité de déceler une tendance (Alexandersson pour tendances et régression multiple) sont portées à mal classifier les petits sauts.
- (d) La méthode d'Alexandersson pour les sauts, la régression multiple et la régression à deux phases ont offert la meilleure performance pour identifier des sauts multiples.
- (e) Le modèle de régression à deux phases présenté par Easterling et Peterson (1995) se montre très sensible, mais la version révisée de ce modèle (Lund et Reeves, 2002) pourrait améliorer la performance du modèle.
- (f) Le test de Student séquentiel est désavantagé parce que la fenêtre mobile n'était pas assez large.

(g) Les méthodes incluant une ou des séries de référence performant mieux du fait qu'elles expliquent une part de la variabilité de la série de base.

Une étude comparative de trois méthodes d'homogénéisation a été réalisée sur des séries de pression moyenne en surface en Europe (Slonosky *et al.*, 1999). Les inhomogénéités ont été détectées au pas de temps annuel et l'ajustement des séries est fait mensuellement. Les techniques utilisées sont l'analyse graphique des séries de différences, la méthode de Caussinus et Mestre et la méthode d'Alexandersson. La documentation a été consultée pour valider les inhomogénéités identifiées. Les grandes conclusions de cette étude sont :

(a) La méthode de Caussinus et Mestre requiert d'un grand nombre de stations voisines pour fournir des facteurs de correction robustes. La méthode est problématique lorsque la station de base se trouve à un endroit isolé ou avec seulement quelques voisines (moins de six).

(b) La méthode d'Alexandersson est performante avec de bonnes séries de référence. L'efficacité de cette technique dépend alors de la disponibilité de séries de référence.

(c) L'examen graphique des séries de différences a donné les meilleurs résultats.

Lubes-Niel *et al.* (1998) ont comparé le test de corrélation sur les rangs de Kendall, le test de Pettitt, le test de Buishand, la procédure Bayésienne de Lee et Heghinian et la procédure de segmentation des séries hydrométéorologiques de Hubert et Carbonnel (1989) sur des séries simulées de précipitations annuelles. Ces méthodes ne fournissent pas les mêmes éléments de réponse quant à l'homogénéité d'une série. En effet, le test de Kendall sert à vérifier la présence d'une tendance dans les données, les tests de Pettitt et de Buishand permettent de positionner une rupture de moyenne et d'évaluer son seuil significatif, la méthode Bayésienne de Lee et Heghinian permet d'estimer la distribution de la position de la rupture et la technique de segmentation de Hubert et Carbonnel est conçue pour la recherche de changements multiples de moyenne dans une série. L'intérêt de leur travail n'était pas de choisir la meilleure procédure, mais plutôt d'étudier la puissance et la robustesse de ces cinq méthodes dans des conditions

contrôlées. La puissance des méthodes a été évaluée en introduisant des ruptures de moyenne (saut et tendance) dans des séries normales indépendantes. La robustesse a été évaluée en générant des séries normales indépendantes, lognormales indépendantes, gamma indépendantes, normales autocorrélées et des séries avec des ruptures d'écart-type. Voici les conclusions générales de cette étude :

(a) Les tests de Kendall, de Pettitt et de Buishand donnent des résultats conformes à l'erreur de type 1 choisi sur les séries stationnaires.

(b) La méthode Bayésienne avec une loi *a priori* non informative détecte un fort pourcentage (40%) de ruptures sur les séries stationnaires. Elle devrait être utilisée seulement lorsque les séries étudiées ont une forte probabilité de rupture *a priori*.

(c) Avec la procédure de segmentation, le pourcentage de séries rejetées alors qu'elles respectent l'homogénéité est acceptable.

(d) La puissance de toutes les méthodes est supérieure à 50% pour détecter des sauts de moyenne dépassant 75% de la valeur de l'écart-type.

(e) Les méthodes reposant sur l'indépendance des observations (Pettitt, Buishand, Lee et Heghinian) ne sont pas robustes lorsque les données contiennent une tendance. Ces méthodes montrent également un manque de robustesse lorsque les données sont autocorrélées.

(f) Les méthodes basées sur la normalité des données (Buishand, Lee et Heghinian, Hubert et Carbonnel) offrent une bonne performance lorsque cette hypothèse n'est pas respectée.

(g) Les méthodes exigeant l'invariance de l'écart-type (procédure Bayésienne, Pettitt, Buishand) sont relativement robustes lorsque les séries contiennent des ruptures d'écart-type et que la moyenne est stationnaire.

(Easterling and Peterson, 1992) ont choisi un ensemble de méthodes indépendantes de l'information historique disponible, facilement automatisables et qui considèrent une série de référence formée de stations voisines et les ont appliquées sur des séries réelles et simulées de

températures. Ils ont étudié entre autres la méthode d'Alexandersson, le test bivarié, le test de Student séquentiel, l'analyse des doubles accumulations, les sommes cumulatives parallèles et la régression à deux phases. Ils ont cherché à déterminer la performance de chacune des techniques à identifier correctement l'année d'une discontinuité et d'estimer l'amplitude du facteur de correction à appliquer. Les méthodes qui ont bien identifié une discontinuité ont ensuite été testées pour détecter des inhomogénéités multiples. Par la suite, elles ont été appliquées aux données réelles pour comparer les résultats avec l'information historique disponible. Les grandes conclusions de ce travail sont :

- (a) Toutes ces méthodes se sont montrées efficaces pour détecter un saut d'amplitude 2σ .
- (b) La régression à deux phases semble plus appropriée pour détecter des tendances que des sauts.
- (c) Pour identifier un saut de faible amplitude, la méthode d'Alexandersson, le test bivarié et le test de Student ont offert la meilleure performance.
- (d) Le test de Student est moins approprié pour détecter des sauts multiples. En effet, les valeurs critiques augmentent considérablement vers la fin de la série et ainsi le test est inapproprié pour examiner les segments plus courts de la série de base.
- (e) La technique d'Alexandersson et le test bivarié ont été retenus pour détecter des sauts multiples. Lorsque le nombre de discontinuités augmente, la méthode d'Alexandersson fournit une meilleure performance.
- (f) Par ailleurs, l'efficacité de la méthode d'Alexandersson diminue lorsque les séries de référence ont des écarts-type de 0.5 à 1.5 fois celui de la série de base.
- (g) Pour les changements plus graduels, le test bivarié et la méthode d'Alexandersson tendent à identifier un saut dans le milieu de cette tendance.
- (h) Les deux méthodes précédentes ont été appliquées à des stations possédant de l'information historique et elles ont identifié des discontinuités non documentées.

Enfin, Zurbenko *et al.* (1996) ont comparé leur filtre (Kolmogorov-Zurbenko) avec le critère de Schwarz pour détecter des discontinuités dans des séries de données en altitude. Ils ont appliqué les techniques sur des séries simulées contenant des patrons saisonniers. Les deux méthodes ont la propriété d'estimer directement la dimension du modèle sans passer par des tests d'hypothèse. Le filtre a offert une meilleure performance que le critère de Schwarz sur les séries avec des tendances cycliques. Par contre, le critère de Schwarz offre une performance supérieure lorsque les données sont distribuées normalement. Les deux méthodes ont eu du mal à détecter de petites discontinuités. Zurbenko *et al.* (1996) suggèrent d'utiliser le critère de Schwarz pour faire une analyse exploratoire en support à d'autres méthodes et à l'information historique.

ANALYSE CRITIQUE DES MÉTHODES

Les forces et les faiblesses des méthodes d'homogénéisation dépendent de plusieurs facteurs et plusieurs regroupements (qui se recoupent) sont possibles. Nous allons en discuter en les classifiant de trois façons : (a) leur caractère objectif ou subjectif, (b) l'utilisation de l'information des stations voisines et des métadonnées et (c) le nombre et le type des inhomogénéités pour lesquels leur théorie est conçue.

Classification selon leur caractère objectif ou subjectif

Méthodes objectives. Les méthodes dites objectives se basent sur des considérations mathématiques pour repérer les inhomogénéités et peuvent ne pas les détecter si certaines de leurs hypothèses sont violées. Certaines méthodes (Approche bivariée, Méthode d'Alexandersson, Test de Student séquentiel, les méthodes basées sur la régression) requièrent que la variance de la série soit constante. Les méthodes non paramétriques comme le test de Wilcoxon séquentiel ont l'avantage d'être basées sur des concepts de statistique non paramétrique robustes et indépendants de toute distribution. De plus, les observations extrêmes ont moins d'effet sur le résultat du test. Cependant, il peut arriver qu'on perde en puissance ce

qu'on gagne en robustesse. Une étude de robustesse quant aux postulats exigés pourrait être effectuée pour vérifier jusqu'à quel point leur non-respect affecte les résultats. Ensuite, l'application séquentielle des tests de Wilcoxon et de Student peut modifier l'erreur de type 1, mais ce problème peut se régler en diminuant le seuil critique de chaque méthode. Enfin, la méthode d'Alexandersson et le test de Student séquentiel sont sensibles au coefficient de variation de la série (écart-type/moyenne). Le test d'Alexandersson se montre plus ou moins sensible selon que ce coefficient est faible ou élevé (Alexandersson and Moberg, 1997). Une variance forte fera en sorte que la statistique de Student soit faible et ainsi la série est déclarée homogène sans l'être nécessairement.

Autant les approches basées sur la régression que celles basées sur les méthodes statistiques classiques (sauf les tests non-paramétriques) reposent sur la normalité des données. De plus, la segmentation des séries sous-entend la normalité des données. Cependant, même si les données ne sont pas normales, les séries de différences ou de rapports peuvent tout de même respecter la normalité. Par ailleurs, la méthode Bayésienne multi-sauts possède l'avantage de pouvoir être adaptée sans difficulté à une distribution non normale. De plus, les codes mettant en œuvre toutes les méthodes statistiques classiques s'exécutent rapidement, tandis que le temps de calcul des méthodes Bayésiennes est considérablement plus élevé. Il demeure néanmoins raisonnable pour une application sur un petit ensemble de séries. Enfin, plusieurs techniques tendent à détecter des inhomogénéités au début et à la fin des séries. Pour contrer ce problème, les inhomogénéités inexplicables identifiées en début ou en fin de série peuvent être ignorées.

Méthodes subjectives. Le fait qu'elles soient basées sur le jugement de l'utilisateur constitue à la fois un point fort et un point faible. La personne qui utilise une technique subjective peut évaluer le poids à donner à certains éléments selon leur importance (la qualité des métadonnées et des séries voisines ou l'amplitude d'une rupture par rapport à la variance de la série). Par ailleurs, une personne non experte risque de commettre des erreurs de jugement.

Néanmoins, ces méthodes constituent une bonne façon de s'approprier les données à titre exploratoire.

Classification selon l'utilisation de l'information aux stations voisines et des métadonnées

Méthodes utilisant l'information à une ou plusieurs séries voisines. L'efficacité des méthodes qui utilisent une ou plusieurs séries voisines dépend de la quantité de stations voisines et de la qualité de leurs données. Une inhomogénéité éventuelle dans l'une de celles-ci aura moins d'impact sur l'ajustement du modèle lorsque les séries voisines sont nombreuses. Par exemple, lorsqu'il y a une seule série voisine (e.g. analyse des doubles accumulations ou analyse graphique des ratios), il est impossible de savoir laquelle de la série de base ou de la série de référence contient une inhomogénéité. De ce fait, elles se montrent efficaces lorsque les dates de changements sont connues (Kohler, 1949). De plus, si la série de référence est mal choisie, des variations naturelles causant des changements de pente peuvent être interprétées comme des inhomogénéités. La méthode de Tayanc est également adéquate lorsque les métadonnées sont exhaustives. En effet, lorsque la station visée n'est pas documentée, cette méthode permet seulement de classer la station comme douteuse. Les méthodes graphiques incluant plusieurs séries voisines seront beaucoup plus appropriées si les métadonnées sont incomplètes (sommations cumulatives parallèles, analyse graphique des séries de différence).

Quoiqu'il soit quasi impossible de s'assurer de l'homogénéité d'une station voisine, il est important d'en tester l'hypothèse pour éviter d'interpréter un changement climatique comme une inhomogénéité. Par contre, lorsque toutes les stations du réseau étudié ont des changements d'instruments simultanés, les tests d'homogénéité absolus sont préférables aux tests relatifs qui se basent sur des stations voisines. Lorsque le nombre de stations à homogénéiser est élevé, on peut considérer des méthodes comme celles de Szentimrey ou celle de Caussinus et Mestre qui, tout en homogénéisant simultanément un ensemble de stations, ne se basent pas sur l'hypothèse que la ou les séries voisines utilisées sont homogènes.

Méthodes utilisant les métadonnées. Ces méthodes comprennent les approches subjectives, mais aussi les approches objectives dont les résultats doivent être corroborés par les métadonnées avant qu'une correction ne soit effectuée. Les points forts et les points faibles de ces méthodes sont les mêmes que ceux des approches subjectives.

Classification selon le nombre et le type des inhomogénéités pour lesquels leur théorie est conçue

Méthodes conçues à l'origine pour détecter une seule inhomogénéité à la fois. Les mathématiques de la plupart de ces méthodes visent à détecter un seul saut de moyenne dans la série. Il est possible que ces méthodes soient moins performantes à détecter des sauts multiples plutôt qu'un seul saut de moyenne. En présence de sauts multiples, la correction des données peut être biaisée. L'estimation d'un saut peut être faussé par la présence d'autres sauts dans la série. Lorsqu'une méthode développée pour détecter un changement de moyenne est appliquée séquentiellement pour détecter des sauts multiples, une attention particulière doit être accordée à la façon d'estimer l'amplitude des sauts pour ne pas laisser d'autres sauts éventuels interférer dans les calculs. En effet, lorsque les facteurs de correction sont calculés pendant la segmentation de la série, ils ne sont pas nécessairement déterminés seulement avec des segments homogènes.

Méthodes conçues à l'origine pour détecter plusieurs inhomogénéités simultanément. Les méthodes qui peuvent détecter des sauts multiples en une application (méthode Bayésienne multivariée multi-sauts, filtre de Kolmogorov-Zurbenko, critère de Schwarz, segmentation des séries, méthode de Szentimrey, méthode de Caussin et Mestre) ont l'atout de fournir des estimations non biaisées par d'autres discontinuités. En effet, elles permettent d'éviter l'interférence d'éventuels sauts non encore détectés.

Méthodes conçues pour détecter plusieurs types d'inhomogénéités. Certaines méthodes ont la capacité de détecter à la fois des sauts et des tendances (méthode d'Alexandersson, régression multiple, régression à deux phases). Cependant, le test

d'Alexandersson pour les sauts semble détecter un saut près du milieu d'un segment en réalité tendancieux. De plus, le test d'Alexandersson destiné aux tendances conduit à une classification de plusieurs sauts consécutifs comme une tendance. La régression à deux phases se montre plus adéquate pour détecter des tendances que des sauts (Easterling and Peterson, 1992). Par ailleurs, les méthodes possédant la capacité de déceler une tendance sont portées à mal classifier les petits sauts (Ducré-Robitaille *et al.*, 2003). La détection de sauts de faibles amplitudes est problématique dans la plupart des cas. Par contre, en travaillant avec les déviations cumulées, un changement de moyenne est plus prononcé que dans une série d'observations brutes. Par conséquent, les tests sur les déviations cumulées permettent de détecter plus facilement des inhomogénéités de faible amplitude.

ANALYSE D'APPLICABILITÉ AUX SÉRIES DE PRÉCIPITATIONS TOTALES ANNUELLES

Cette section vise à identifier les méthodes qui peuvent homogénéiser des séries de précipitations totales annuelles. Premièrement, certaines d'entre elles comme l'analyse des doubles accumulations, la méthode du vecteur régional, l'analyse graphique des ratios, l'analyse graphique des déviations cumulées, la méthode d'Alexandersson, l'approche bivariée et les tests sur les déviations cumulées ont été développées spécialement pour l'homogénéisation des précipitations totales annuelles. De plus, les méthodes de Szentimrey, de Jaruskova et de Caussinus-Mestre permettent d'homogénéiser plusieurs types de variables incluant les précipitations. Plusieurs applications ont été faites sur des données de précipitations. Par exemple, la méthode d'Alexandersson a servi à homogénéiser des séries de précipitations en Norvège (Hanssen-Bauer and Forland, 1994), en France (Lamarque et Jourdain, 1994), en Suède et en Finlande (Peterson *et al.*, 1998). La procédure de Boroneant et Tomozeiu (1998) a permis d'homogénéiser des séries de précipitations en Italie (Tomozeiu *et al.*, 2000). Wijngaard *et al.*

(2003) ont appliqué leur méthode sur des séries de précipitations localisées en Europe. Puisque les précipitations sont très variables, Wijngaard *et al.* (2003) ont extrait le nombre de jours humides par année (précipitations de plus de 1 mm) et appliqué leur méthode sur cette variable. Cette variable a pour effet de réduire la variabilité et ainsi, seuls les sauts assez grands peuvent être détectés. De plus, cette méthode a été choisie pour tester l'homogénéité de séries de pluie en Éthiopie (Conway *et al.*, 2004). Chez Météo-France, les séries de précipitations sont homogénéisées selon la méthode de Caussinus et Mestre. En Autriche, l'homogénéité du réseau de mesure de précipitations a été vérifiée au moyen du test de Craddock sur les déviations cumulées combiné à une analyse des métadonnées (Peterson *et al.*, 1998). En Suisse, les séries de précipitations ont été homogénéisées en combinant 12 méthodes et la consultation des métadonnées en utilisant le logiciel THOMAS (Begert *et al.*, 2005). Au Canada, les précipitations ont été homogénéisées pour des changements d'instruments systématiques à tout le réseau et pour des modifications dans la façon de prendre les observations (Mekis and Hogg, 1999). Aux États-Unis, les métadonnées ainsi que la méthode de Karl et Williams sont combinées. La méthode de Rhoades et Salinger (1993) est appliquée en Nouvelle-Zélande et dans les îles du Pacifique. Enfin, la régression à deux-phases selon le modèle de Wang (2003) a été utilisée par un regroupement de scientifiques pour homogénéiser des extrêmes de température et de précipitations en Amérique Centrale et au nord de l'Amérique du Sud (Aguilar *et al.*, 2005).

Certaines techniques développées afin d'homogénéiser des séries de températures moyennes annuelles pourraient performer également pour des séries de précipitations totales annuelles. Cela va dépendre de la distribution des données (normale ou non-normale). Par exemple, l'applicabilité de la régression multiple, de la méthode Bayésienne multivariée multi-sauts, du filtre Kolmogorov-Zurbenko, du critère de Schwarz, du test de Student et de Wilcoxon séquentiel n'a pas encore été testé. Les méthodes citées dans cet article ont été développées dans le but précis de tester ou de corriger l'homogénéité de séries climatiques. Par ailleurs, d'autres

techniques statistiques qui servent à analyser les ruptures dans les séries chronologiques pourraient également accomplir cette tâche (Raimondo and Tajvdi, 2004; Gombay, 2003; Chen and Gupta, 2000, 2001; Lavielle and Moulines, 1997; Ghorbanzadeh, 1995).

CONCLUSIONS

La correction des inhomogénéités dans les séries climatiques est cruciale pour tous types d'analyses sur des variables climatiques. Plus particulièrement, il est important de posséder des séries de précipitations de qualité pour prendre des décisions basées sur des données fiables dans les recherches sur les changements climatiques ou même d'autres domaines comme l'irrigation de l'eau ou la production d'énergie. Avec la prise de conscience que des changements artificiels dans une série de données peuvent biaiser les analyses, le besoin de jeux de données homogènes se fait sentir.

Un grand nombre de méthodes d'homogénéisation ont été présentées dans cette revue de littérature. Les méthodes objectives ont été approfondies parce qu'elles sont plus complètes. Néanmoins, il ne faut pas négliger la possibilité d'utiliser une approche subjective pour bien s'approprier les données et éviter de commettre des erreurs. La tendance dans les dernières méthodes développées consiste à combiner plusieurs tests très différents et de déclarer une série inhomogène seulement lorsque plusieurs tests sont significatifs. Cela évite de réduire le nombre de fausses alarmes. De plus, la consultation des métadonnées est essentielle malgré l'utilisation d'une méthode sophistiquée. L'ajustement des séries climatiques est une tâche délicate et il faut connaître les raisons des corrections apportées.

Le choix de la méthode optimale à utiliser ou même la combinaison des méthodes à utiliser dépend fortement des caractéristiques statistiques de la station. Pour aider les utilisateurs dans le choix d'une méthode, des études comparatives de la robustesse et de la puissance des méthodes sur des données de précipitations totales synthétiques seraient grandement appréciées.

Par exemple, on pourrait évaluer la robustesse des méthodes lorsque l'hypothèse de normalité des données (ou des séries de différences/rapports, selon la méthode) n'est pas vérifiée. Cependant, il est fort possible que ce point ne soit pas crucial. Les données synthétiques pourraient également contenir des changements de variance et/ou des inhomogénéités dans les séries voisines. On pourrait ainsi avoir une idée de l'impact du non-respect des postulats propres à chaque méthode. L'aptitude de différentes méthodes d'homogénéisation des précipitations totales annuelles à identifier des séries homogènes et inhomogènes n'a pas encore été analysée.

REMERCIEMENTS

Les auteurs désirent remercier le Consortium OURANOS, le Conseil de Recherche en Sciences Naturelles et Génie du Canada (CRSNG) et le Centre d'Études Nordiques (CEN) pour avoir financé cette recherche. Les auteurs aimeraient remercier MM. Abderrahmane Yagouti et Gilles Boulet du Ministère du Développement durable, de l'Environnement et des Parcs du Québec pour leur collaboration. Les auteurs désirent également remercier M. Olivier Mestre et un autre réviseur anonyme pour leurs commentaires et suggestions qui ont contribué à améliorer la qualité de l'article.

RÉFÉRENCES

- Aguilar E., Auer I., Brunet M., Peterson T.C., & Wieringa J. (2003) Guidelines on climate metadata and homogenization. *World Meteorological Organization, Geneva, Switzerland, Report WMO-TD 1186*.
- Aguilar E., Peterson T.C., Obando P.R., Frutos R., Retana J.A., Solera M., Soley J., Garcia I.G., Araujo R.M., Santos A.R., Valle V.E., Brunet M., Aguilar L., Alvarez L., Bautista M., Castanon C., Herrera L., Ruano E., Sinay J.J., Sanchez E., Oviedo G.I.H., Obed F., Salgado J.E., Vazquez J.L., Baca M., Gutierrez M., Centella C., Espinosa J., Martinez D., Olmedo B., Espinoza C.E.O., Nunez R., Haylock M., Benavides H., & Mayorga, R. (2005) Changes in precipitation and temperature extremes in Central America and northern South America, 1961-2003. *J. Geophys. Res. Atmospheres*. **110**, (D23):23107.
- Akaike H. (1973) Information theory and an extension of the maximum likelihood principle. *Proc. of the Second International Symposium on Information Theory*, Budapest, Hungary, pp. 267-281.
- Alexandersson H. (1986) A homogeneity test applied to precipitation data. *J. Climatol.* **6**, 661-675.
- Alexandersson H. & Moberg A. (1997) Homogenization of swedish temperature data. Part 1: Homogeneity test for linear trends. *Int. J. Climatol.* **17**, 25-34.
- Asselin J., Ouarda T.B.M.J., Fortin V., & Bobée B. (1999) Une procédure Bayésienne bivariée pour détecter un décalage de la moyenne. *INRS-Eau, Chaire en hydrologie statistique, Québec, Québec, Canada, Rapport de recherche R-528*.
- Begert M., Schlegel T., & Kirchhofer W. (2005) Homogeneous temperature and precipitation series of Switzerland from 1864 to 2000. *Int. J. Climatol.* **25**:65-80.
- Bois P. (1986) Contrôle de séries chronologiques corrélées par étude du cumul des résidus de la corrélation. *Deuxièmes Journées Hydrologiques de l'ORSTOM, Montpellier, France*.
- Boroneant C. & Tomozeiu R. (1998) Experience with homogeneity testing of temperature data at Bucuresti Filaret station. *Proc. of the Second Seminar for Homogenization of Surface Climatological Data, Budapest, Hungary*.
- Brunet-Moret Y. (1978) Homogénéisation des précipitations. *Cahiers De L'ORSTOM, Série Hydrologie*. **16**, 147-170.
- Buishand T.A. (1982) Some methods for testing the homogeneity of rainfall records. *J. Hydrol.* **58**, 11-27.
- Buishand T.A. (1984) Tests for detecting a shift in the mean of hydrological time series. *J. Hydrol.* **73**, 51-69.
- Caussinus H. & Mestre O. (2004) Detection and correction of artificial shifts in climate series. *Journal of the Royal Statistical Society Series C: Applied Statistics* **53**, 405-425.

- Caussinus H. & Lyazrhi F. (1997) Choosing a linear model with a random number of change-points and outliers. *Ann. Inst. Statist. Math.* **49**, 761-775.
- Chen J. & Gupta A.K. (2001) On change-point detection and estimation. *Communications in Statistics-Simulation and Computation* **30**, 665-697.
- Chen J. & Gupta A.K. (2000) *Parametric statistical change-point analysis*. Birkhauser, Boston, USA
- Conway D., Mould C., & Bewket W. (2004) Over one century of rainfall and temperature observations in Addis Ababa, Ethiopia. *Int. J. Climatol.* **24**, 77-91.
- Craddock J.M. (1979) Methods of comparing annual rainfall records for climatic purposes. *Weather* **34**, 332-346.
- Datsenko N.M., Moberg A., & Sonechkin D.M. (2002) Objective time-scale-dependent homogenization or early instrumental temperature series. *Theo. Appl. Climatol.* **72**, 103-126.
- Ducré-Robitaille J.F., Boulet G., & Vincent L.A. (2003) Comparison of techniques for detection of discontinuities in temperature series. *Int. J. Climatol.* **23**, 1087-2003.
- Easterling D.R. & Peterson T.C. (1992) Techniques for detecting and adjusting for artificial discontinuities in climatological time series: a review. *Proc. of the Fifth International Meeting on Statistical Climatology, 22-26 june 1996, Toronto, Ontario, Canada*.
- Easterling D.R. & Peterson T.C. (1995) A new method for detecting undocumented discontinuities in climatological time series. *Int. J. Climatol.* **15**, 369-377.
- Fearnhead P. (2004) Exact and Efficient Bayesian Inference for Multiple Change-point. *Unpublished work*.
- Ghorbanzadeh D. (1995) Un test de détection de rupture de la moyenne dans un modèle gaussien. *Statistique appliquée* **2**, 67-76.
- Gombay E. (2003) Sequential Change-Point Detection and Estimation. *Sequential Analysis* **22**, 203-222.
- Gullett D.W., Vincent L., & Malone L.H. (1991) Homogeneity testing of monthly temperature series. Application of multiphase regression models with mathematical change-points. *Canadian Climate Center, Downsview, Ontario, Canada, Report 90-4*.
- Gullett D.W., Vincent L.A., & Sajecki P.J.F. (1990) Testing for homogeneity in temperature time series at canadian climate stations. *Canadian Climate Center, Downsview, Ontario, Canada, Report 91-10*.
- Hanssen-Bauer I. & Forland E.J. (1994) Homogenizing Long Norwegian Precipitation series. *J. Climate.* **7**, 1001-1013.
- Hawkins D.M. (2001) Fitting multiple change-point models to data. *Computational Statistics & Data Analysis.* **37**, 323-341.

- Hubert P., Carbonnel J.P., & Chaouche A. (1989) Segmentation des séries hydrométéorologiques-Application a des séries de précipitations et de débits de l'Afrique de l'Ouest. *J. Hydrol.* **110**, 349-367.
- Jaruskova D. (1996) Change-point detection in meteorological measurement. *Monthly Weather Rev.* **124**, 1535-1543.
- Jones P.D., Raper R.S., Diaz H.F., Kelly P.M., & Wigley T.M.L. (1986) Northern hemisphere surface air temperature variations:1851-1984 *J. Climate Appl. Met.* **25**, 161-179.
- Karl T.R. & Williams C.N.Jr. (1987) An approach to adjusting climatological time series for discontinuous inhomogeneities. *J. Climate Appl. Met.* **26**, 1744-1763.
- Kehagias A., Nidelkou E., & Petridis V. (2006) A dynamic programming segmentation procedure for hydrological and environmental time series. *Stochastic Environmental Research and Risk Assessment.* **20**, 77-94.
- Khaliq M.N. & Ouarda T.B.M.J. (2006) A note on the critical values of the Standard Normal Homogeneity Test (SNHT). *Submitted to the International Journal of Climatology.*
- Kohler M.A. (1949) On the use of double-mass analysis for testing the consistency of meteorological records and for making required adjustments. *Bul. Am. Met. Soc.* **30**, 188-189.
- Lamarque P. & Jourdain S. (1994) Élaboration de longues séries climatologiques homogènes pour l'étude de l'évolution climatique. *La Météorologie.* **8**, 61-68.
- Lanzante J.R. (1996) Resistant, robust and non-parametric techniques for the analysis of climate data: theory and examples, including applications to historical radiosonde station data. *Int. J. Climatol.* **16**, 1197-1226.
- Lavielle M. & Moulines É. (1997) Détection de ruptures multiples dans la moyenne d'un processus aléatoire. *Comptes rendus de l'Académie des Sciences de Paris* **324**, 239-243.
- Lee A.F.S. & Heghinian S.M. (1977) A shift of the mean level in a sequence of independent normal random variables: A Bayesian approach. *Technometrics* **19**, 503-506.
- Lubes-Niel H., Masson J.M., Paturel J.E., and Servat E. (1998) Variabilité climatique et statistiques. Étude par simulation de la puissance et de la robustesse de quelques tests utilisés pour vérifier l'homogénéité de chroniques. *Rev. Sci. Eau* **11**, 383-408.
- Lund R. & Reeves J. (2002) Detection of undocumented changepoints: A revision of the two-phase regression model. *J. Climate* **15**, 2547-2554.
- Maronna R. & Yohai V.J. (1978) A bivariate test for the detection of a systematic change in mean. *J. American Statistical Association* **73**, 640-645.
- Mekis E & Hogg WD. (1999) Rehabilitation and analysis of canadian daily precipitation series. *Atmosphere-Ocean* **37**, 53-85.

- Mielke P.W., Berry K.J., & Brier G.W. (1981) Application of multi-response permutation procedures for examining seasonal changes in monthly mean sea-level pressure patterns. *Monthly Weather Rev.* **109**, 120-126.
- Ouarda T.B.M.J., Rasmussen P.F., Cantin J-F., Bobée B., Laurence R., Hoang V.D., & Barabé G. (1999) Identification d'un réseau hydrométrique pour le suivi des modifications climatiques dans la province de Québec. *Rev. Sci. Eau* **12**, 425-448.
- Ouarda T.B.M.J., Asselin J. & Seidou O. (2005) Bayesian multivariate linear regression with application to changepoint models in hydrometeorological variables. Model Development. *INRS-ETE, Chair in Statistical Hydrology/Canada Research Chair on the Estimation of Hydrological Variables, Québec, Québec, Canada, Research Report R-838*.
- Perreault L., Haché M., Slivitzsky M., & Bobée B. (1999) Detection of changes in precipitation and runoff over eastern Canada and U.S. using a Bayesian approach. *Stochastic environmental research and risk assesment* **13**, 201-216.
- Perreault L., Parent É., Bernier J., Bobée B., & Slivitzky M. (2000) Retrospective multivariate Bayesian change-point analysis: A simultaneous single change in the mean of several hydrological sequences. *Stochastic Environmental Research and Risk Assessment* **14**, 243-261.
- Peterson T.C., Easterling D.R., Karl T.R., Groisman P., Nicholls N., Plummer N., Torok S., Auer I., Boehm R., Gullett D., Vincent L., Heino R., Tuomenvirta H., Mestre O., Szentimrey T., Salinger J., Forland E.J., Hanssen-Bauer I., Alexandersson H., Jones P., & Parker D. (1998) Homogeneity adjustments of in situ atmospheric climate data: a review. *Int. J. Climatol.* **18**, 1493-1517.
- Pettitt A.N. (1979) A non-parametric aproach to the change-point problem. *Appl. Statist.* **28**, 126-135.
- Potter K.W. (1981) Illustration of a new test for detecting a shift in mean in precipitation series. *Monthly Weather Rev.* **109**, 2040-2045.
- Raimondo M. & Tajvdi N. (2004) A peaks over threshold model for change-point detection by wavelets. *Statistica Sinica* **14**, 395-412.
- Rhoades D.A. & Salinger M.J. (1993) Adjustment of temperature and rainfall records for site change. *Int. J. Climatol.* **13**, 899-913.
- Schwarz G. (1978) Estimating the dimension of a model. *The annals of statistics* **6**, 461-464.
- Seidou O. & Ouarda T.B.M.J. (2006) Recursion-based Multiple Changepoint Detection in Multivariate Linear Regression and Application to River Streamflow. *INRS-ETE, Chair in Statistical Hydrology/Canada Research Chair on the Estimation of Hydrological Variables, Québec, Québec, Canada, Research Report R-843*.
- Slonosky V.C., Jones P.D., & Davies T.D. (1999) Homogenization techniques for European monthly mean surface pressure series. *J. Climate* **12**, 2658-2672.

- Solow A.R. (1987) Testing for climate change: an application of the two-phase regression model. *J. Climate Appl. Met.* **26**, 1401-1405.
- Szentimrey T. (1996) Statistical procedure for joint homogenization of climatic time series. *Proc. of the First Seminar for Homogenization of Surface Climatology Data, Budapest, Hungary.*
- Szentimrey T. (1999) Multiple analysis of series for homogenization (MASH). *Proc. of the Second Seminar for Homogenization of Surface Climatological Data, Budapest, Hungary.*
- Tayanc M., Dalfes H.N., Karaca M., & Yenigun O. (1998) A comparative assesment of different methods for detecting inhomogeneities in Turkish temperature data set *Int. J. Climatol.* **18**, 561-578.
- Thompson C.S. (1984) Homogeneity analysis of rainfall series: an application of the use of a realistic rainfall model. *J. Climatol.* **4**, 609-619.
- Tomozeiu R., Busuioc A., Marletto V., Zinoni F., & Cacciamani C. (2000) Detection of changes in the summer precipitation time series of the region Emilia-Romagna, Italy. *Theo. Appl. Climatol.* **67**, 193-200.
- Vincent L.A. (1998) A technique for the identification of inhomogeneities in Canadian temperature series. *J. Climate* **11**, 1094-1105.
- Von Neumann J. (1941) Distribution of the ratio of the mean square successive difference to the variance. *Ann. Math. Statist.* **13**, 367-395.
- Wang X.L. (2003) Comments on 'Detection of Undocumented Changeoints: A revision of the Two-Phase regression model'. *J. Climate* **16**, 3383-3385.
- Wijngaard J.B., Klein Tank A.M.G., & Können G.P. (2003) Homogeneity of 20th century european daily temperature and precipitation series. *Int. J. Climatol.* **23**, 679-692.
- Young K.C. (1993) Detecting and removing inhomogeneities from long-term monthly sea level pressure time series. *J. Climate* **6**, 1205-1220.
- Zurbenko I., Porter P.S, Rao S.T., Ku J.Y., Gui R., & Eskridge R.E. (1996) Detecting discontinuities in time series of upper-air data: development and demonstration of an adaptive filter technique. *J. Climate* **9**, 3548-3560.

7. Article 2: Intercomparison of homogenization techniques for precipitation data

Une comparaison de huit méthodes d'homogénéisation des séries climatiques est présentée dans cet article. L'objectif est de déterminer lesquelles semblent les plus prometteuses pour détecter des inhomogénéités dans les séries de précipitations totales annuelles observées au centre et au sud de la province de Québec et ses alentours, au Canada. La performance des différentes techniques a été évaluée par simulations de Monte Carlo sur des séries synthétiques homogènes et inhomogènes. Ces dernières ont été générées de manière à reproduire les caractéristiques statistiques des précipitations totales annuelles typiques du centre et du sud de la province de Québec et ses alentours. Certaines techniques comparées ont donné une meilleure performance générale que les autres techniques : le test bivarié, la méthode d'Alexandersson et la méthode de Jaruskova. Le test de Student séquentiel et la régression à deux-phases ont donné les moins bons résultats. Une procédure qui prend avantage des meilleures techniques est proposée.

Intercomparison of homogenization techniques for precipitation data

Claudie Beaulieu¹, Ousmane Seidou¹, Taha B.M.J. Ouarda¹, Xuebin Zhang², Gilles Boulet³ and
Abderrahmane Yagouti³

¹*Institut National de la Recherche Scientifique-Eau, Terre et Environnement, Université du Québec, Québec, QC, Canada*

²*Climate Research Division, Science and Technology Branch, Environment Canada, Toronto, ON, Canada*

³*Direction du suivi de l'état de l'environnement, Ministère de l'Environnement du Québec, Québec, QC, Canada*

Submitted the October 10th 2006, accepted the 20th September 2007, published the 19th February
2008 in Water Resources Research

Copyright © 2008 American Geophysical Union

Corresponding author: Claudie Beaulieu, Chair in statistical Hydrology, Canada Research Chair on the Estimation of Hydrological Variables, INRS-ETE, University of Quebec, 490, de la Couronne, Québec (Québec) G1K 9A9, CANADA. e-mail: claudie.beaulieu@ete.inrs.ca tel. (418) 654-2530 #4462

Abstract

This paper presents an inter-comparison of eight statistical tests to detect inhomogeneities in climatic data. The objective was to select those which are more suitable for precipitation data in the southern and central regions of the province of Quebec, Canada. The performances of these methods were evaluated by simulation on several thousands of homogeneous and inhomogeneous synthetic series. These series were generated to reproduce the statistical characteristics of typical precipitations observed in the southern and central parts of the province of Quebec and nearby areas, Canada. It was found that none of these methods was efficient for all types of inhomogeneities, but some of them performed substantially better than others: the bivariate test, the Jaruskova's method and the standard normal homogeneity test. Techniques such as the Student sequential test and the two-phase regression led to the worst performances. The analysis of the performances of each method in several situations allowed the design of an optimal procedure which takes advantage of the strengths of the best performing techniques.

Keywords: homogenization, precipitation series, simulation, discontinuities.

1. Introduction

Hydroclimatic data records often undergo artificial disturbances that do not reflect the real climate variations. These disturbances can be related for instance to station relocation, instrument replacement, change in observation procedures or modification in the immediate environment of the site. Homogenization is the technique of detecting and correcting these artificial disturbances. A climate series is considered homogeneous when the measurement conditions of the station had not varied with time. Different types of homogenization techniques are presented in the literature: statistical techniques or techniques based on expert judgment, techniques based principally on metadata (archive of a station) or based on the concept of relative homogeneity (it consists of a comparison of the series to homogenize with a reference series to separate the artificial change from the regional climate signal). The different issues related to the homogenization process are discussed by *Peterson et al.* [1998] and by *Aguilar et al.* [2003]. In this paper, we focus on statistical techniques which use a reference series.

Different types of changes will introduce different types of inhomogeneities in the series. For a variable such as precipitation, a change of exposition or a relocation of the gauge are the type of changes that are the most likely to introduce an inhomogeneity in the series. For an exhaustive review of the degree of influence of each type of change on different variables, the reader is referred to *Heino* [1997].

In various fields, the need for long and reliable climatic data series is high. During the last decades, several efforts were made to develop techniques to correct anthropogenic changes in climate series. For example, climate change studies require the creation of complete databases in order to adequately analyze the climatic signal, and estimate the future change with a minimal uncertainty. The need to develop robust homogenization techniques and to identify the most suitable method for each type of variable (e.g temperature or precipitation) is thus obvious. A

study of appropriate techniques for temperature series was conducted by *Ducré-Robitaille et al.* [2003].

A comparative study of several homogenization methods for precipitation is carried out in this paper. The techniques are systematically applied to several thousands of synthetic data sets having the same statistical characteristics as the recorded series of total annual precipitation. Stations located in the southern and central regions of the province of Quebec (Canada) and nearby areas were selected to estimate these statistical characteristics. First, the homogenization techniques were selected using some practical considerations such as speed and technical characteristics. Then, several sets of homogeneous and inhomogeneous synthetic series were generated using a technique derived by *Easterling and Peterson* [1992]. Once the series were generated, each selected homogenization method was applied to each synthetic series, and the resulting performance was used as the basis of the inter-comparison. Finally, recommendations are formulated concerning the best combination of homogenization methods for the application to total annual precipitation.

2. Data

The data sets are synthetic series of precipitation with the same statistical characteristics (average, standard deviation, order 1 autocorrelation and spatial cross-correlation) as typical series of annual total precipitation observed in southern and central Quebec and nearby areas (Canada). The parameters of the generation scheme were chosen to fit the statistical characteristics of precipitation data series recorded at some selected stations in the province or immediate surroundings. These stations have high quality records (long observation series with little missing data). All the observation series of the selected stations passed the Shapiro-Wilk normality test, supporting our choice of a normal distribution in the generation scheme.

Homogeneous synthetic series (mean and variance are constant) and inhomogenous (one or multiple shifts, trend, shift in standard deviation) were first generated. Correlated neighbour series were also generated using a technique that will be described further in the text [*Easterling and Peterson, 1992; Vincent, 1998; Ducre-Robitaille et al., 2003*].

2.1. Base series

2.1.1. Homogeneous

Homogeneous base series were generated to study the sensitivity of the techniques on homogeneous series. Lag one autoregressive variables, z_i , were first generated using the following model:

$$z_i = \phi_1 z_{i-1} + e_i \quad (1)$$

where ϕ_1 is the autocorrelation coefficient and e_i is a normally distributed residual with zero mean and variance $1 - \phi_1^2$. The mean and variance of the real total annual precipitation were introduced in the z_i series.

The statistical characteristics to be reproduced are a) a mean total annual precipitation of 1089 mm, b) a standard deviation of 142 mm and c) a lag one autocorrelation of 0.02. These values are the average characteristics of the selected stations. Even though the autocorrelation was not significant, the series were generated using an autoregressive model instead of a normal independent model to represent the real data series as much as possible. A total number of 10 000 homogeneous series (5000 60-year long and 5000 100-year long) were generated this way.

2.1.2. Series with a single shift

Series with a single shift in the mean were generated to study the ability of the methods to detect the position and to estimate the magnitude of a single shift. The procedure for the selection of the magnitude and position of a shift is described in the next paragraph. A series with a single shift can be represented by:

$$y_i = \begin{cases} y_i^* - \delta_{p_1} \sigma, & i = 1, \dots, p_1 - 1 \\ y_i^*, & i = p_1, \dots, n \end{cases} \quad (2)$$

δ_{p_1} and p_1 were randomly generated using the following distributions:

$$\delta_{p_1} = \text{sign}(u - 1/2) \cdot 3 \cdot b, \quad u \sim U(0,1), \quad b \sim \text{BETA}(2,2) \quad (3)$$

$$p_1 = 10 + ud, \quad ud \sim \text{DUNIF}(n - 20) \quad (4)$$

where y_i represents the i^{th} observation of an inhomogenous series of size n , y_i^* the i^{th} observation of a homogeneous series, σ is the standard deviation of the last segment of the series, δ_{p_1} the magnitude of the shift, u is a uniform variable, b is a beta variable, p_1 is the position of the shift and ud is a discrete uniform variable. The magnitude varies randomly between -3 and 3 standard deviations (Eq. (3)). The distribution of the magnitudes was chosen to have few cases for which the magnitude is close to zero, to avoid a repetition of assessing the methods when the series was close to being homogeneous, and to save computing time. Furthermore, since the distribution of the magnitudes of real inhomogeneities is never known, the distribution was chosen to represent all potential magnitudes. Figure 1 presents the distribution of the magnitudes of the shifts for the 100-year long series with a single shift. The position was generated from a discrete uniform distribution that was truncated to avoid the presence of shifts during the first ten or the last ten years of the series (Eq. (4)). A total number of 50 000 series (25 000 60-year long and 25 000 100-year long series) were generated this way. Since the techniques were developed to detect a single shift, a higher number of synthetic series were generated for this case. The number of series for each case was chosen to have enough repetitions of the different scenarios and according to the interest of each type of series. Figure 2a presents an example of a series with a single shift.

2.1.3. Series with multiple shifts

Series with multiple shifts (2 and 3) were also generated. A series with two shifts can be represented by:

$$y_i = \begin{cases} y_i^* - \delta_{p_1} \sigma, & i = 1, \dots, p_1 - 1 \\ y_i^* - \delta_{p_2} \sigma, & i = p_1, \dots, p_2 - 1 \\ y_i^*, & i = p_2, \dots, n \end{cases} \quad (5)$$

The positions and magnitudes of the shifts were generated using the following distributions:

$$\delta_{p_1}, \delta_{p_2} = \text{sign}(u - 1/2) \cdot 3 \cdot b, \quad u \sim U(0,1), \quad b \sim \text{BETA}(2,2) \quad (6)$$

$$p_1 = 10 + ud, \quad ud \sim \text{DUNIF}(n - 31) \quad (7)$$

$$p_2 = 10 + p_1 + ud, \quad ud \sim \text{DUNIF}(n - 20 - p_1) \quad (8)$$

where y_i represents the i^{th} observation of an inhomogeneous series of size n , y_i^* the i^{th} observation of a homogeneous series, and σ the standard deviation of the last segment of the series. $\delta_{p_1}, \delta_{p_2}, p_1$ and p_2 respectively denote the magnitudes and positions of the shifts. u, b and ud respectively denote a uniform variable, a beta variable and a discrete uniform variable. The magnitudes were generated in the same way as for a single shift. Discrete uniform distributions were used to generate the positions, but their parameters were adapted to the number of shifts to generate. The minimum interval between two consecutive shifts was set to ten years. Series with three shifts were generated using a similar procedure. A total number of 15 000 series with two shifts were generated and 15 000 series with three shifts in the mean.

2.1.4. Series with a trend

When there is a trend in the data, the homogenization methods may interpret it as one or several consecutive shifts. Series with trends were generated to study the behaviour of the homogenization methods on such series. In spite of the fact that most of the homogenization

methods are not developed to detect trends, we were interested in their performance to identify a change inside the trend. A series with a trend can be represented by:

$$y_i = \begin{cases} y_i^* - \delta_{p_1:p_2} \sigma, & i = 1, \dots, p_1 - 1 \\ y_i^* - \delta_{p_1:p_2} \sigma - mi, & i = p_1, \dots, p_2 - 1 \\ y_i^*, & i = p_2, \dots, n \end{cases} \quad (9)$$

The following distributions were used to generate the trended series:

$$\delta_{p_1:p_2} = \text{sign}(u - 1/2) \cdot 3 \cdot b, \quad u \sim U(0,1), \quad b \sim \text{BETA}(4,2) \quad (10)$$

$$p_1 = 10 + ud, \quad ud \sim \text{DUNIF}(n - 31) \quad (11)$$

$$p_2 = 10 + p_1 + ud, \quad ud \sim \text{DUNIF}(n - 20 - p_1) \quad (12)$$

where y_i represents the i^{th} observation of an inhomogeneous series of length n , y_i^* the i^{th} observation of a homogeneous series, σ the standard deviation of the last part of the series, $\delta_{p_1:p_2}$ the magnitude of the trend, and m the slope. u , b and ud respectively denote a uniform variable, a beta variable and a discrete uniform variable. p_1 and p_2 represent the beginning and the end of the trend. The magnitude was chosen to lie between -3 and 3 standard deviations. The positions of the beginning and the end of the trend were generated with the same technique as that for series with two shifts. A total number of 10 000 synthetic series with a random trend were generated. Figure 2b presents an example of series with this type of discontinuity.

2.1.5. Series with a shift of variance

Series with a shift of variance were generated to determine which methods are sensitive to this type of discontinuity. The studied methods were not initially designed to detect changes in the variance and most of them are based on the hypothesis that the variance is constant. It is thus interesting to check their robustness to violations of the latter postulate. The position of the variance shift was randomly selected from a discrete uniform distribution (Eq. 4). The magnitudes are generated from a BETA(8,2) distribution multiplied by a random sign and

divided by 2. The magnitudes of the generated variance shifts lie between 0 and 50 % of the standard deviation. A total number of 10 000 series with a shift of variance were generated. Figure 2c represents an example of a synthetic series with a shift of variance.

2.2. Neighbour series

For every base series three correlated neighbour series were generated in two steps. First, three homogeneous series (independent of the base series) were generated:

$$w_i = \phi_1 w_{i-1} + e_i \quad (13)$$

where ϕ_1 is the lag one autocorrelation coefficient and e_i is a normally distributed residual with zero mean and variance $1 - \phi_1^2$. Then, a correlation structure was introduced between the base series and the neighbour series:

$$w_i = \psi z_i + w_i \quad (14)$$

where z_i represents the standardized total precipitation at the base station for year i , w_i the standardized total precipitation at a neighbour station for year i , and ψ is a correlation coefficient between the neighbour series and the base series. w_i was then standardized to ensure it has zero mean and standard deviation 1. The neighbour series possess the same statistical characteristics as the base series. The correlation coefficient (0.7) was determined by simulation to reproduce a spatial cross-correlation of 0.55. This value was chosen because the mean spatial cross-correlation in the set of selected stations that are located at a distance less than 300 km is 0.55. This distance was chosen to test the techniques in the worst conditions. Correlation between stations varies enormously and is likely to affect the performance of the homogenization methods, which are expected to perform better when the base and neighbour series are highly correlated. However, the station network density in Quebec is relatively low given the large size of the province (1 542 056 km²). Consequently, it can be difficult in some regions to find several

neighbours for the same base series. To represent this reality as closely as possible, the number of neighbour stations was set to 3.

3. Methods

Homogenization techniques compared in this work were selected according to the following set of criteria. First, the methods should be objective. Since the techniques were applied to thousands of series, subjective methods could not be used. Second, they should be able to detect multiple shifts (as these can be observed in practice) and estimate them. Some methods developed for one shift were nevertheless selected, but they were adapted for multiple shifts using a segmentation approach. Third, the techniques should allow the use of one or several neighbour series. Finally, the algorithms must be available in the literature and have a reasonable running time.

An extensive literature review was performed [*Beaulieu et al.*, 2007] and eight methods were selected:

- standard normal homogeneity test [*Alexandersson*, 1986; *Khaliq and Ouarda*, 2007]
- multiple regression [*Vincent*, 1998]
- two-phase regression [*Easterling and Peterson*, 1995; *Lund and Reeves*, 2002]
- bivariate test [*Maronna and Yohai*, 1978; *Potter*, 1981]
- sequential Wilcoxon test [*Karl and Williams*, 1987; *Lanzante*, 1996; *Ducré-Robitaille et al.*, 2003]
- sequential t-test [*Gullett et al.*, 1990]
- Jaruskova's method [*Jaruskova*, 1996]
- Bayesian approach [*Rasmussen*, 2001]

The hypothesis of normality has to be respected to apply these tests except for the sequential Wilcoxon test which is nonparametric. All methods were coded to ignore the shifts detected amongst the first ten or last ten observations. The authors consider that it is not reasonable to estimate the magnitude of an inhomogeneity with less than ten observations on each side of the inhomogeneity.

3.1. Standard normal homogeneity test (SNHT)

A series of ratios between the base and the neighbour series is created:

$$q_i = y_i / \left[\left(\sum_{j=1}^k \rho_j^2 x_{ij} \bar{y}_{1:n} / \bar{x}_{1:n,j} \right) / \sum_{j=1}^k \rho_j^2 \right] \quad \begin{array}{l} i = 1, \dots, n \\ j = 1, \dots, k \end{array} \quad (15)$$

where the value of the year i of the base series is represented by y_i , x_{ij} is the i^{th} observation of the neighbour series j . There are k nearby sites with n observations each. The correlation coefficient between the base series and the neighbour series j is noted ρ_j . The hypothesis that the standardized ratios follow a normal distribution with zero mean and a variance of 1 is tested against the hypothesis that there is a shift in the mean of the series. To find the position of the change, a weighted average series is created:

$$Q_i = i\bar{v}_{1:i}^2 + (n-i)\bar{v}_{i+1:n}^2, \quad i = 1, \dots, n-1 \quad (16)$$

where $\bar{v}_{1:i}$ and $\bar{v}_{i+1:n}$ are the average of the standardized ratios for segments 1:i and i+1: n . The test statistic, $Q_{p_1} = \max_{i=1, \dots, n-1} \{Q_i\}$, is significant if it exceeds the associated critical value [Khaliq and Ouarda, 2007] and the shift is located at position p_1 . For this study, this test was successively applied to all synthetic series with a critical level of 5 %.

3.2. Multiple linear regression (MREG)

This approach is based on the application of four regression models representing different types of inhomogeneities [Vincent, 1998]. In this study only two models are used. The first one represents a homogeneous base series. When residuals are independent, the model provides a

good fit to the data and the series is considered homogeneous. The homogeneity of the series is verified with an independence test on the residuals (e.g. a confidence interval on the lag one autocorrelation of the residuals). On the other hand, if residuals are autocorrelated, it indicates that the model does not fit the data well and that the base series could be inhomogeneous. In this case, the model describing a shift in the base series is applied:

$$y_i = \begin{cases} \tau + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + e_i & i = 1, \dots, p_1 - 1 \\ \tau + \delta_{p_1} + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + e_i & i = p_1, \dots, n \end{cases} \quad (17)$$

where y_i represents the base series, x_{ij} represents the neighbour series j , and τ and $\beta_j (j = 1, \dots, k)$ are the least squares parameters. There are k neighbour series with n observations each. The residuals, e_i , follow a normal distribution with zero mean and constant variance. The position of the shift, p_1 , is determined by fitting the model for all possible positions and by selecting the results with the smallest residual sum of squares. Again, if residuals are independent, then there is probably a shift at position p_1 . A Fisher test comparing the fit of the homogeneous model with the model representing a shift is applied. If the model with a shift is better, the significance level of the shift is evaluated with a Student statistic. The same process is continued on each side of the shift until all segments are considered homogeneous. This method was applied to all synthetic series with a 5 % critical level.

3.3. Two-Phase regression (REG2)

Several versions of the two-phase regression can be found in the literature [*Solow*, 1987; *Easterling and Peterson*, 1995; *Lund and Reeves*, 2002; *Wang*, 2003]. For the purpose of this work, the model proposed by *Lund and Reeves* [2002] was used. Two regression models are fitted with time as the explanatory variable. The first model represents a homogeneous series. The second model represents a discontinuous series displaying a shift at time p_1 :

$$y_i = \begin{cases} \tau_1 + \lambda_1 i + e_i, & i = 1, \dots, p_1 - 1 \\ \tau_2 + \lambda_2 i + e_i, & i = p_1, \dots, n \end{cases} \quad (18)$$

where y_i represents the i^{th} observation of the base series, τ_1 , τ_2 , λ_1 and λ_2 are respectively the intercepts and slopes before and after the change. The position of the shift is chosen by fitting the model for all possible values of p_1 , by computing the Fisher statistics, comparing all fitted models to that of the homogeneous one, and then by choosing the one which gives the maximum Fisher statistic (*Fmax*). The Fisher test gives information regarding the contribution of a new variable in a regression model. In this case, it means that the significance of the introduction of a step at position p_1 is verified. The critical values of the Fmax statistic were obtained by simulation and provided by *Lund and Reeves* [2002]. The same process is repeated until all segments are found homogeneous or have less than ten observations. The method described in the above section was applied to the difference series between the base series and the neighbour series, with a critical level of 5 %.

3.4. Bivariate test (BIVT)

This method was developed by *Maronna and Yohai* [1978] and applied to homogenization problems by *Potter* [1981]. The technique is based on the postulate that the base series (y_i , $i = 1, \dots, n$) and a reference series (x_i , $i = 1, \dots, n$) belong to the same bivariate normal distribution. It is hypothesized that there is a shift in the base series that does not occur in the reference series. The method is based on the following series of ratios:

$$q_i = \frac{i(n-i)\delta_i^2 F_i}{S_x S_y - S_{xy}^2}, \quad i = 1, \dots, n-1 \quad (19)$$

where

$$\delta_i = \frac{\left[S_x (\bar{y}_{1:n} - \sum_{j=1}^i y_j / i) - S_{xy} (\bar{x}_{1:n,1} - \sum_{j=1}^i x_{j1} / i) \right] n}{(n-i)F_i}, \quad i = 1, \dots, n-1 \quad (20)$$

$$F_i = S_x - (x_i - \bar{x})^2 ni / (n-i) \quad (21)$$

$$S_{xy} = \sum_{i=1}^n (x_{i1} - \bar{x}_{1:n,1})(y_i - \bar{y}_{1:n}) \quad (22)$$

$$S_x = \sum_{i=1}^n (x_{i1} - \bar{x}_{1:n,1})^2 \quad (23)$$

$$S_y = \sum_{i=1}^n (y_i - \bar{y}_{1:n})^2 \quad (24)$$

The statistic of the test is given by $Q_{p_1} = \max_{i=1, \dots, n-1} \{ |q_i| \}$. The critical values of Q_{p_1} are obtained by simulation [Maronna and Yohai, 1978]. When the test is positive, it is assumed that a shift occurred at year p_1 . The approach was iteratively applied on all sets of synthetic series, assuming a critical level of 5 %. Since the technique allows the use of a single reference series, the average of the three synthetic neighbour series was used as reference series.

3.5. Sequential Student test (STUS)

The sequential Student test consists of using a moving window and testing successively the equality of the means of the first half and the second half of the observations in the window [Gullett et al., 1990]. Following the recommendations given by *Ducré-Robitaille et al.* [2003], the size of the moving window was increased to 20 years to obtain a better performance. Then, we test the equality of the means by using 10 years before and after every potential position:

$$Q_i = \frac{\bar{q}_{i-10:i-1} - \bar{q}_{i:i+9}}{\sqrt{s_{i-10:i-1}^2 / 10 + s_{i:i+9}^2 / 10}}, \quad i=11, \dots, n-9 \quad (25)$$

where $\bar{q}_{i-10:i-1}$, $\bar{q}_{i:i+9}$, $s_{i-10:i-1}^2$ and $s_{i:i+9}^2$ are the means and variances of the ten observations located before and after the position i in the ratios series. The maximum value of the Student statistic corresponds to the position of the shift. The Student statistic is significant if it exceeds the critical

value of the Student distribution with 18 degrees of freedom. When the statistic is significant, the series is split into two segments and the process is repeated until all shifts are detected. This procedure was substituted for the original one because, by extracting simultaneously all the significant statistics, the same shift is identified several times since successive Student statistics are highly correlated. Furthermore, since the test is applied to the same series several times, the probability to meet a type 1 error (to reject the null hypothesis while it is true) is increased. That is why the critical levels used to get a global critical level of 5 % were computed by simulation. The critical levels used were 0.225% and 0.0875% for series of length 60 and 100 years respectively. Finally, the method was applied to the series of ratios between the base series and the neighbour series, since the ratios are usually used with precipitation.

3.6. Sequential Wilcoxon test (WILS)

The Wilcoxon test has been extensively used for the homogenization of climate data [Karl and Williams, 1987; Lanzante, 1996; Ducre-Robitaille et al., 2003]. The most recent version of the method was used in this work. It consists of computing successively the Wilcoxon statistic and estimating its significance level using a normal approximation:

$$Q_i = \frac{R_i - i(n+1)/2}{\sqrt{i(n-i)(n+1)/12}}, \quad i = 11, \dots, n-9 \quad (26)$$

where n represents the length of the tested series, $R_i = \sum_{j=1}^i r_j$ and r_i are the ranks of the first part of the series. The maximum of the series and its position are then extracted ($Q_{p_i} = \max_{i=11, \dots, n-9} \{|Q_i|\}$). If the statistic is significant, then there is a change of mean at this position. The series is then split into two segments, and the same procedure is applied on each of the new series. The same operations are performed on each obtained segment until all segments are found homogeneous or have a length smaller than 10 observations. As in the case of the sequential Student test, the critical levels were modified to have a critical level of 5 %. The

critical levels used were 0.44% and 0.289% for series of length 60 and 100 years respectively. In the work of *Ducré-Robitaille et al.* [2003], this method was applied to a series of differences between the base series and a reference series because the variable of interest was the temperature. For the purpose of this work, we used a series of ratios between the base series and the neighbour series.

3.7. Jaruskova's method (JARU)

This method was proposed by *Jaruskova* [1996] to detect a shift in a meteorological series. Several alternative approaches were presented in the work of *Jaruskova* [1996], but the model for which the date of change is unknown was selected in this work. The method consists of building a difference series between the base series and a reference series, and then testing the hypothesis that there is a change in the mean of the difference series. The following statistic is computed for all possible positions for a shift in the series.

$$Q_i = \sqrt{\frac{(n-i)i}{n}} \frac{(\bar{q}_{1:i} - \bar{q}_{i+1:n})}{s_i} \quad i=1, \dots, n-1 \quad (27)$$

where

$$s_i^2 = \frac{1}{n-2} \left[\sum_{j=1}^i (q_j - \bar{q}_{1:i})^2 + \sum_{j=i+1}^n (q_j - \bar{q}_{i+1:n})^2 \right] \quad (28)$$

and q_i represents the difference between the base station and a reference station for the year i and n is the length of the series. We assume that the difference series is normally distributed. The maximum of the series, $Q_{p_1} = \max_{i=1, \dots, n-1} \{|Q_i|\}$, is extracted. The shift is significant if the statistic exceeds the critical value of the distribution [*Jaruskova*, 1996]. This method was applied to all sets of synthetic series at the critical level of 5 %, ignoring the inhomogeneities in the first ten or last ten years. The reference series is the mean of the three synthetic neighbour series.

3.8. Bayesian method (BAYE)

Many Bayesian models have been proposed in the literature [Asselin *et al.*, 1999; Perreault *et al.*, 1999; Rasmussen, 2001; Ouarda *et al.*, 2005]. In this study we used the technique presented by Rasmussen [2001]. It consists of inferring the parameters of a linear regression model using an analytical Bayesian approach. Several alternative models were considered in Rasmussen (2001), but the one that was used in this study is:

$$y_i = \begin{cases} \tau_1 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + e_i & i = 1, \dots, p_1 - 1 \\ \tau_2 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + e_i & i = p_1, \dots, n \end{cases} \quad (29)$$

where y_i represents the base series, x_{ij} represents the i^{th} observation of the j^{th} neighbour series, τ_1 , τ_2 and β_j ($j=1, 3$) respectively represent the intercept point before the change, the intercept point after the change and the neighbour series coefficients. The model assumes that the data are independent and normally distributed. The prior probability densities on the regression parameters and on the position of change are non-informative (uniformly distributed with bounds: $(-\infty, \infty)$). The position of the change is chosen to be the mode of the posterior distribution. For the purpose of this work, the Bayesian inference is performed on the position of the change (p_1) as well as on the parameter vector $\boldsymbol{\theta} = [\tau_1, \tau_2, \beta_1, \beta_2, \beta_3]^T$. Let us denote the base series by the vector Y in this description. The posterior probability density of the position of the change is given by:

$$pr(p_1 | Y) = \left\{ \left| G_{p_1-1}^T G_{p_1-1} \right|^{-1/2} [Y^T Y - Y^T G_{p_1-1} (G_{p_1-1}^T G_{p_1-1})^{-1} G_{p_1-1}^T Y] \right\} \\ * \left\{ \sum_{i=1}^{n-1} \left| G_i^T G_i \right|^{-1/2} [Y^T Y - Y^T G_i (G_i^T G_i)^{-1} G_i^T Y]^{-(n-5)/2} \right\}^{-1} \quad (30)$$

where

$$G_{p_1-1}^T = \begin{pmatrix} 1 & 0 & x_{1,1} & x_{1,2} & x_{1,3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & x_{p_1-1,1} & x_{p_1-1,2} & x_{p_1-1,3} \\ 0 & 1 & x_{p_1,1} & x_{p_1,2} & x_{p_1,3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & x_{n,1} & x_{n,2} & x_{n,3} \end{pmatrix} \quad (31)$$

The first two columns of the G_i matrix contain the indicative variables. The other columns contain the observations of the neighbour series. n is the length of the vector and x_{ij} the observations of the neighbour series j . The posterior density of the parameters of $\boldsymbol{\theta}$ is given by:

$$pr(\boldsymbol{\theta} | Y, p_1) = \frac{\left\{ \frac{\left[\Gamma(\nu + 5/2) |G_{p_1-1}^T G_{p_1-1}|^{1/2} \right]}{\left[\left(\Gamma(1/2) \right)^5 \Gamma(\nu/2) (c\sqrt{\nu})^5 \right]} \right\}}{\left\{ 1 + \frac{(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T G_{p_1-1}^T G_{p_1-1} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})}{\nu c^2} \right\}^{(\nu+5)/2}} \quad (32)$$

with

$$\hat{\boldsymbol{\theta}} = (G^T G)^{-1} G^T Y \quad (33)$$

$$c^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 / (n-2) \quad (34)$$

and ν is the number of degrees of freedom ($n-5$), Γ represents the Gamma function and c^2 is the unbiased estimate of the noise variance. As there is no analytical expression for the posterior distribution of the magnitude of change, it was computed with Monte-Carlo simulations. A Bayesian credibility interval was used to verify the significance of the shift. This method was applied to all sets of synthetic series.

4. Performance evaluation

The selected homogenization methods were developed to detect a single shift in a series. In the presence of multiple shifts, the estimation of the magnitude may be biased. To avoid this, the magnitudes can be obtained by computing the difference of the means of the segments before and after the shift. For all sets of synthetic series, the homogenization methods were applied to find the position of the shifts. The magnitudes were then estimated by difference of means. Furthermore, the performance of each technique was evaluated differently according to the type of synthetic series (homogeneous, with a single step, with multiple steps, with a change of standard deviation and with a trend). In this section, the criteria used to evaluate the performance of the homogenization techniques for each set of synthetic series are presented. The reader must note that since the performance of the different techniques is compared inside each set of series, with the same number of repetitions, there is no bias in the statistics that are presented in the result section.

4.1. Homogeneous series

The homogenization methods were applied to two sets of homogeneous series (60-year long and 100-year long) and their performance was evaluated by the percentage of type-1 error (the number of cases for which the homogeneity hypothesis is rejected while it is true).

4.2. Series with a single shift

The selected methods were applied to two sets of synthetic series containing a shift with random position and magnitude. The numbers of **correctly identified**, **well identified** and **well positioned** shifts were computed. We consider that a shift is correctly identified when its position is exact and the relative difference between the estimated magnitude and the real magnitude is less than 20 % of the real magnitude. A shift is well identified when the estimated position is less than 2 years from the real position and the absolute error on the estimation of the magnitude is

lower or equal to 50% of the real magnitude. A well positioned shift is located between 0 and 2 years from the exact location of the shift and there is no measure of the accuracy of the magnitude of the shift. Furthermore, the differences between the real position and magnitude and the estimated position and magnitude were computed independently. Four types of error series were produced: the position error, the magnitude error, the absolute position error and the absolute magnitude error. When a technique did not detect a shift, the position error and absolute position error are fixed to the length of the series (either 60 or 100) while the magnitude error and absolute magnitude error are fixed to 3 (the highest possible magnitude).

4.3. Series with multiple shifts

Most of the methods compared in this work are not designed to detect multiple shifts. This issue is addressed in practice using a segmentation approach. The performance of the methods to identify multiple shifts on synthetic series with two or three shifts was thus tested. A performance criterion was designed to estimate the capacity of the techniques to correctly position all the shifts, without omission or false detection. This criterion measures a distance between the positions of the real shifts and the detected shifts. It can be expressed as follows:

$$C = \begin{cases} \frac{1}{nd} \sum_{i=1}^{nd} (p_i^d - p_i)^2, & nr = nd \\ \frac{1}{nr} \left[\sum_{i=1}^{nd} (p_i^d - p_i)^2 + |nr - nd|(n-1)^2 \right], & nr > nd \\ \frac{1}{nd} \left[\sum_{j=1}^{nr} (p_j^d - p_j)^2 + |nr - nd|(n-1)^2 \right], & nr < nd \end{cases} \quad (35)$$

where $p_i^d, i=1, \dots, nd$ and $p_j, j=1, \dots, nr$ represent respectively the positions of the detected and real shifts and n is the length of the series. The pairs (p_i^d, p_j) are chosen to minimize the criterion. When the number of detected and real shifts is the same, the criterion is the sum of squares of the differences between the pairs which minimize the criterion. When the number of detected shifts is different from the number of real shifts, the value $(n-1)^2$ is added for every

wrongly detected shift. This value corresponds to the square of the maximum possible distance between two shifts. C is equal to zero when all shifts are correctly positioned. When C is close to zero, the detected shifts are located near the real positions. A high value of C indicates that some shifts in the series are not detected or are wrongly detected. The performance criterion C was computed for the sets of synthetic series with two and three shifts.

4.4. Series with a shift of variance

The techniques presented in this paper make the assumption that the variance is constant throughout the series. Therefore, a change of variance could affect the results of a homogenization procedure. To investigate the robustness of the methods regarding this postulate, synthetic series (10 000) with a shift in standard deviation were generated. As in the case of homogeneous series, the percentage of falsely detected shifts was computed.

4.5. Series with a trend

An inhomogeneity can also take the form of a trend. However, it is impossible to compare the methods selected for this work to identify trends because only MREG and REG2 are developed to detect this type of inhomogeneity. The number of shifts that are positioned inside the trend (two positions before the beginning and two after the end of the trend) were rather computed. This aims to show that gradual inhomogeneities can be interpreted as one or several consecutive shifts by most techniques.

5. Results

5.1. Homogeneous series

The percentages of falsely detected shifts were approximately between 1 % (MREG, JARU) and 5 % (STUS) for the classical techniques while BAYE gave a higher percentage (more than 20 %) of false detections (Tables 1 and 2). The high number of wrongly detected shifts can be explained by the fact that the Bayesian model of *Rasmussen* [2001] makes the implicit hypothesis that there is necessarily a shift in the series. Indeed, the prior probability of no change is automatically fixed to $1/n$, which is very negligible. Therefore, this method practically assumes that there is always a change in the series, and forces to position it somewhere since the sum of probabilities of all possible positions is constrained to be 1. In the case of homogeneous series, the probabilities for a change are concentrated towards the extremities of the series. These probabilities would be near 100% if the extremities were not ignored. However, in spite of the decision to ignore shifts close to the two extremities, the percentage of false detections remained very high. Most of the percentages of falsely detected shifts presented in tables 1 and 2 are significantly different from the type I error of 5% that was used to apply the tests. If the false detection at the extremities were not removed, then the percentage of falsely detected shifts would be around 5%. For STUS, the extremities have less impact on the detection rate since a moving window is used.

Figure 3 presents the magnitudes and positions of wrongly detected shifts on 100-year long homogeneous series. These magnitudes are presented according to the position of the real shift. Figures corresponding to the 60-year long series are not presented as the results were very similar to those corresponding to the 100-year series. It can be noticed that for all methods, the magnitudes of falsely detected shifts rarely exceeded one standard deviation.

In a similar study dealing with the homogenization of temperature series [Ducre-Robitaille et al., 2003], REG2 and WILS displayed high false detection rates in opposition to the results presented in this paper. For REG2, this can be explained by the fact that the Fisher revised statistic [Lund and Reeves, 2002] was used in the present work. This revised statistic has higher critical values than the original one [Solow, 1987; Easterling and Peterson, 1995] which was used in the study of Ducre-Robitaille et al. [2003], and hence gives more conservative results and decreases the percentage of false detection. For WILS, this result can be explained by the fact that we reduced the critical level of the Wilcoxon test to have a global critical level around 5 % in this study. The critical levels used for WILS and STUS were obtained by simulation and depend on the length of the series.

5.2. Series with a single shift

Tables 3 and 4 present the total percentages of correctly identified, well identified and well positioned shifts for all methods. It can be seen in these tables that most techniques were able to position the shift. Indeed, the percentage of well positioned shifts was of the order of 75% for most methods except for REG2 and STUS. Thus, the homogenization methods were efficient to approximately identify the position of a shift. However, the correct estimation of magnitudes appeared to be more problematic. Indeed, the difference of percentage between the correctly identified and well identified shifts varies between 13% and 39%. To investigate the magnitudes and/or problematic positions, the percentages of correctly and well identified shifts were analyzed for various classes of magnitude and position. Figures 4-5 present the results of this analysis for the 100-year long series. Results for the 60-year long series are very similar to those of the 100-year long series and are consequently not presented.

All these techniques identified well the shifts with magnitudes greater than two standard deviations. For shifts with magnitudes less than a standard deviation, the percentage of well identified shifts decreased very quickly. Furthermore, the performance of STUS and REG2

dropped down for shifts with a magnitude of 1.5 standard deviations and less. This was probably due to the narrow moving window, which degraded the performance of STUS. Similar results were reported by *Ducré-Robitaille et al.* [2003] on temperature series. On the other hand, there was no position that seemed to affect the performance of the techniques. In the work of *Ducré-Robitaille et al.* [2003], shifts located at position 5 were less easily identified than shifts located in the middle of the series. By introducing shifts starting at position 10 instead of 5, the effect of the position on the percentage of identified shifts was attenuated.

The errors and absolute errors in the position and magnitude were also analyzed separately. Tables 5 and 6 present the descriptive statistics of the absolute errors in position and magnitude for each technique. The Kruskal-Wallis test [*Lehman, 1998*] with a 5% critical level was used to verify the significance of the differences in absolute errors for the various techniques. Since the differences were significant, the Conover-Inman [*Conover, 1999*] procedure was used to make multiple pairwise comparisons between the absolute errors obtained from the various techniques (tables 7 and 8).

The absolute errors in position and magnitude are significantly different from one technique to another. Furthermore, for 60-year long series, the absolute errors in the position and magnitude obtained with BAYE are significantly the smallest (tables 5 and 7). BAYE is followed by the methods BIVT, JARU and SNHT which are not significantly different. The techniques WILS, REGM, REG2 and STUS follow and lead to absolute errors in the position and magnitude that are significantly different from all other techniques.

For 100-year long series, the absolute errors in the position and magnitude obtained with BAYE and BIVT are the smallest and are not significantly different from each other (tables 6 and 8). They are followed by JARU, SNHT, WILS, REGM, REG2 and STUS. Figure 6 presents the magnitude and position errors on 100-year long series with one shift obtained from each technique. The errors rather than absolute errors are presented as they are easier to visualize. The

shifts that were not detected are not presented in this figure. For all techniques, the errors are concentrated around the origin. For most techniques, the magnitude errors lie between -1 and 1 standard deviation with few cases outside this interval. The position errors interval ranges as far as -80 and 80 for most techniques. The technique REG2 has the errors that are the most scattered.

5.3. Series with multiple shifts

Figures 7a)-b) present a histogram of the performance criterion (Eq. 35) obtained on both sets of synthetic series. For every class of the criterion, eight bands representing the various techniques are presented. The criterion has some preferential values since the classes of the criterion represent different cases and the peaks mean that some cases occur more often. The most successful methods are those with the lowest values of the performance criterion. Tables 9 and 10 present the descriptive statistics of C (Eq. 35). Table 9 indicates that the criterion median for five methods is very low. This means that in half of the synthetic series with two shifts, these methods well positioned all the shifts without detecting nonexistent shifts or omitting real ones. For the series with three shifts, the median criterion was higher. Indeed, when the number of real shifts increases, it becomes more difficult to identify all of them. The maximum criterion was 9801 and corresponds to the case that all the real shifts were not detected. An analysis of variance of Kruskal-Wallis was realized with a critical level of 5 % to compare the criteria obtained with the various methods. Since the criteria obtained with the different techniques are significantly different, the Conover-Inman procedure was applied to make multiple pairwise comparisons. For the series with two shifts, the smaller criteria were obtained with BIVT (table 9) and for the series with three shifts, BAYE gave the smaller criteria (table 10).

The performance of the various methods according to the distance separating both shifts was then investigated. For every method, the mean criterion was computed according to various classes of distance between two shifts (Figure 8). It was noticed that the distance between the shifts seems to have an influence on the performance criterion. It seems that close or far shifts

were less easily identified than two shifts of a mean distance. When the two shifts are too close, the segment between the two shifts is short. Then, it can be more difficult to detect a change with few observations. When the two shifts are far away, it means that they are located in the extremities of the series and it is more problematic to detect a step in this case.

Notably, for STUS, the distance did not have any effect because of the moving window. The technique REG2 seemed to perform better when positioning two far shifts than two close ones. We were also interested to verify if the signs of the shifts have an effect on the performance of the techniques: Is it easier to identify shifts of opposite signs (a positive shift followed by a negative one) or shifts with the same sign? The performance of the methods regarding the signs of the shifts was analysed. For every method, the mean criterion was calculated according to all possible combinations of signs (Figure 9). It was effectively noticed that the combination of the signs seems to influence the performance criterion. Indeed, in both cases of three shifts of the same sign, the average criterion was higher than in the other cases. When a shift was followed by another one of an opposite sign, the performance seems better except for BAYE. Finally, the cases where two shifts of the same sign were followed or preceded by a shift of an opposite sign led to a better performance than three shifts of the same sign.

5.4. Series with a shift of variance

Table 13 presents the wrongly detected shifts on the synthetic series with a change in variance. It is considered that a method that is robust to a change of variance should give approximately the same rate of false detection in the presence of a shift of standard deviation. The rates of false detection were compared with those obtained from the homogeneous series by a confidence interval on the difference between the two proportions:

$$\begin{aligned} \hat{p}_H - \hat{p}_I - Z_{\alpha/2} \sqrt{\frac{\hat{p}_H(1-\hat{p}_H)}{n_H} + \frac{\hat{p}_I(1-\hat{p}_I)}{n_I}} &\leq p_H - p_I \\ &\leq \hat{p}_H - \hat{p}_I + Z_{\alpha/2} \sqrt{\frac{\hat{p}_H(1-\hat{p}_H)}{n_H} + \frac{\hat{p}_I(1-\hat{p}_I)}{n_I}} \end{aligned} \quad (36)$$

where \hat{p}_H and \hat{p}_I are the rates of false detection in the homogeneous series and in the series with a change in variance, n_H and n_I are the number of homogeneous series and inhomogeneous series respectively and $Z_{\alpha/2}$ is the standard normal value corresponding to a 5% critical level. When the interval contains the value zero, there is no significant difference between the two proportions. The results are presented in table 14. Most of the differences of proportions were significant at a 5 % critical level except for STUS, WILS and BAYE. This was expected for WILS because it is a nonparametric technique. On the other hand, STUS is based on the assumption of the equality of variances, and hence the results for this method were unexpected. Once again, this can be a result of the moving window given that the change of standard deviation has less impact on a small part of the series. Finally, it seems that a change in variance in the base series increases the probability of the type 1 error for most of the methods, but this increase is relatively minor. In summary, the shifts of variance seem to raise slightly the percentage of false detection.

5.5. Series with a trend

Tables 15 and 16 present respectively the number of cases for which one shift and two shifts or more were detected inside the trend. It was noticed that in most cases, the trend was interpreted as an abrupt shift of mean (table 15). This occurs less often when the trend has a weak magnitude (0-0.5 standard deviation). It also happened that the trend was interpreted as several consecutive shifts when the magnitude is high (table 16). To avoid this kind of mistake, a graphical method combined with an objective method could be used. Graphically, it is easier to identify the type of change (abrupt or gradual).

6. Discussion

When the metadata is incomplete, there is little information about the presence, number, type, position and magnitude of potential inhomogeneities in the base series. Therefore, this work aimed to identify methods able to detect homogeneous and inhomogeneous series. Several types of synthetic series (homogeneous, one shift, multiple shifts, a trend, a shift of variance) representing the typical annual total precipitation of the southern and central regions of the province of Quebec (Canada) and nearby areas were generated.

The performance of the studied methods can be summarized according to some criteria. First, it is very important to have a small percentage of false detection (5 % and less), to avoid introducing inhomogeneities in a series that is homogeneous in reality. Also, a homogenization method should be able to identify a shift in a series. The technique should also be able to position a shift in at least 75 % of the cases. Shifts of high magnitudes (2 standard deviations and more), are expected to be identified in nearly 100 % of the cases. It is also expected to observe a reasonable percentage of well identified shifts of 1 standard deviation and more. For multiple shifts, the method should also be able to position several shifts without omission or false detection. Finally, the methods that can be applied in real conditions without additional modifications are advantageous because they allow a gain of time.

The application on series with a shift of variance showed that most of the methods being compared in this work are not very sensitive to this type of inhomogeneity. The results of the application on synthetic series containing a trend showed that a gradual change is often interpreted as one or several consecutive shifts inside the trend. This is not a weakness because the compared methods were not designed for this purpose.

Since our results indicate that none of the methods performed well in all cases, the design of an optimal procedure using the strengths of some of the techniques was considered. This

procedure consists of the sequential application of some selected techniques. If we judge that it is better to omit inhomogeneities in a series than to introduce new artificial ones, JARU should be applied first because of its good capacity to identify homogeneous series (tables 1 and 2). Although MREG gave an equivalent performance, JARU is preferred because on series with a single shift, JARU was less conservative than MREG (the absolute errors in position and magnitude obtained with JARU are significantly smaller than those obtained with MREG, section 5.2). If the series is found inhomogeneous with JARU, BIVT can be applied. In the case of series with a single shift, BIVT had the second best performance for the absolute errors in position and magnitude (tables 5 and 6). On series with two shifts, BIVT had the best performance (table 9) and for series with three shifts, BAYE was better (table 10). Given that BAYE also detected a high number of non-existent shifts in the homogeneous series, BIVT was preferred. Indeed, it has a weak percentage of false detection (tables 1 and 2). With such a procedure, the probability of identifying all existing shifts is increased, and the risk of false detection is reduced. It is important to mention that large shifts (greater than one standard deviation), have a very low false detection rate with all methods. It is nevertheless recommended to confirm the results with the metadata. It should be stressed that this proposed procedure is rather conservative, and is based on the idea that it is better to omit existing shifts than to correct a series for shifts that are not real.

This comparison included only objective techniques because the application of subjective methods (involving the judgment of an expert) on thousands of synthetic series is an impracticable task. Nevertheless, the use of subjective approaches should not be automatically rejected as these approaches may sometimes be appropriate to analyse the data and interpret the results. We also insist on the use of metadata, when available, to validate and identify the cause of the detected inhomogeneities.

A limitation of the presented techniques is that they require the presence of homogeneous neighbouring stations, while it may not be always the case in practice. The authors believe that it is important to use neighbour series to avoid a misinterpretation of a regional climate change. Nevertheless, in the cases where neighbour series are not available or inhomogeneous, techniques developed to homogenize isolated stations and to create homogeneous reference series from neighbour series can be used. For a review of these techniques, see *Peterson et al.* [1998].

The compared techniques require data to be normally distributed (except WILS). The synthetic series were generated from a normal distribution. Nevertheless, the introduction of inhomogeneities in the series could have affected the distribution and hence, the performance of the techniques.

The techniques selected for this work are based on the analysis of either differences or ratios between the base series and neighbour series. Since some tests use ratios and others use differences, the performance could be affected because of the choice of variable. The objective of this work was to compare the techniques as they are presented in the literature. The sensitivity of the techniques to the use of ratios or differences was not studied in this work. Future work can focus on the study of the sensitivity of the various techniques to the choice of variables.

Finally, the various homogenization methods were applied under the specific conditions of the province of Quebec, Canada, and the results of this study may only be valid under these conditions. The same techniques could lead to different performances on series with a different distribution, a different autocorrelation structure and/or different correlation with neighbour series.

7. Conclusions

Homogeneous precipitation series are essential, particularly when data are used in climate models or to assess climate change and associated environmental and socio-economics impacts. The performance of eight homogenization techniques on synthetic precipitation series with similar characteristics to typical series observed in southern and central Quebec and surrounding areas in Canada were compared in this work. The results of this study will be of use for future activities dealing with the homogenization of precipitation series in Canada. It was found that techniques which gave a good performance on temperature series like the multiple regression [Ducré-Robitaille *et al.*, 2003], were not necessarily appropriate for precipitation data.

Three methods had similar performances with all sets of synthetic series (BIVT, JARU and SNHT). Some techniques cannot be applied efficiently to all types of series. For instance, MREG performed well for the identification of a homogeneous series and was good to identify a single shift. However, in the presence of multiple shifts, the performance of this method was poor. The technique BAYE performed well for the identification of one or multiple shifts, but detected too many nonexistent shifts. Finally, STUS and REG2 were able to detect homogeneous series, but did not perform well on series with single and multiple shifts. An optimal procedure using the strengths of the various methods was proposed.

The mathematics of the studied methods were first developed to detect a single shift. A sequential application of these techniques can disadvantage them because they are not designed to detect multiple shifts. However, it is common to have several inhomogeneities in hydroclimatic series. The development of methods that are able to identify one or multiple changes of several types (shift or trend) is desirable.

In this work, it was shown that the selected homogenization techniques are robust in presence of a change of variance. Nevertheless, the homogenization methods are based on other

assumptions such as the normality and the homogeneity of neighbour series. The violation of these assumptions may also alter the performance of the techniques. There is no existing work which addresses these aspects. Future work in these directions is desirable.

Acknowledgments. The authors wish to thank the National Sciences and Engineering Research Council of Canada (NSERC), the OURANOS Consortium and the Canada Research Chair Program for funding this research. The authors would also like to thank Mrs Lucie Vincent of the Meteorological Service of Canada and Dr Paul Whitfield of Environment Canada for their useful comments. The authors would also like to thank the Associate Editor and four anonymous reviewers for their valuable comments and suggestions.

Notation

β_j	coefficient of the j^{th} neighbour series in the regression model l
Γ	Gamma function
δ_{p_i}	magnitude of the shift at position p_i
θ	vector of parameters
$\hat{\theta}$	vector of estimated parameters
λ_l	time coefficient in the regression model l
ρ_j	correlation coefficient between the base series and neighbour series j
σ	standard deviation of the last segment of the base series
τ_l	intercept of the regression model l
ν	number of degrees of freedom
ϕ_1	lag one autocorrelation coefficient
ψ	correlation constant between the base series and the neighbour series
b	Beta variable
<i>BETA</i>	Beta distribution
C	position criterion
<i>DUNIF</i>	Discrete Uniform distribution
e_i	residual at time i
k	number of neighbour series
m	slope of the trend
n	length of the base series
nd	number of shifts detected
nr	number of real shifts

p_i^d	position of the i^{th} detected shift
p_i	position of the i^{th} real shift
q_i	difference/ratios series between the base series and the neighbours at time i
Q_{p_i}	test statistic associated with the i^{th} shift
r_i	rank of the i^{th} observation
R_i	sum of the ranks of observations 1 to i
s^2	variance in a series
S	sum of squares of the differences
$sign$	sign (positive or negative)
u	Uniform variable
U	Uniform distribution
ud	Discrete Uniform variable
w_i	standardized neighbour series at time i
$x_{1:i,j}$	observations from 1 to i of the neighbour series j
$y_{1:i}$	observations from 1 to i of the base series
$\bar{y}_{1:i}$	mean of the segment from 1 to i of variable y
y_i^*	homogeneous base series at time i
z_i	standardized base series at time i

References

- Aguilar, E., I. Auer, M. Brunet, T. C. Peterson, and J. Wieringa (2003), Guidelines on climate metadata and homogenization. *Report WMO-TD 1186*, 50 p., World meteorological Organization, Geneva, Switzerland.
- Alexandersson, H. (1986), A homogeneity test applied to precipitation data, *J. Climatol.*, 6, 661-675.
- Asselin, J., T. B. M. J. Ouarda, V. Fortin, and B. Bobée (1999), Une procédure Bayésienne bivariée pour détecter un décalage de la moyenne, *Research Report R-528*, 33 p., INRS-ETE, Quebec, Qc.
- Beaulieu, C., T. B. M. J. Ouarda, and O. Seidou (2007), Synthèse des techniques d'homogénéisation des séries climatiques et analyse d'applicabilité aux séries de précipitations, *Hydrol. Sci. J.*, 52(1), 18-37.
- Conover, W. J., M. O'Sullivan (Eds), and B. Wiley (Eds) (1999), *Practical Nonparametric Statistics*, 3rd ed., 584 p., Wiley, New York.
- Ducré-Robitaille, J. F., G. Boulet, and L. A. Vincent (2003), Comparison of techniques for detection of discontinuities in temperature series, *Int. J. Climatol.*, 23, 1087-2003.
- Easterling, D. R., and T. C. Peterson (1992), Techniques for detecting and adjusting for artificial discontinuities in climatological time series: a review, Proc. Fifth International Meeting on Statistical Climatology, Amer. Meteor. Soc., Toronto, Ont., 22-26 June.
- Easterling, D. R. and T. C. Peterson (1995), A new method for detecting undocumented discontinuities in climatological time series, *Int. J. Climatol.*, 15, 369-377.
- Gullett, D. W., L. A. Vincent, and P. J. F. Sajecki (1990), Testing for homogeneity in temperature time series at Canadian climate stations, *Canadian Climate Centre Rep. 90-4*, 43 p., Atmospheric Environment Service, Downsview, Ont.

- Heino, R. (1997), Metadata and their role in homogenization. Proc. First Seminar for Homogenization of Surface Climate Data, Hungarian meteorological service, Budapest, Hungary.
- Jaruskova, D. (1996), Change-point detection in meteorological measurement, *Mon. Wea. Rev.*, *124*, 1535-1543.
- Karl, T. R. and C. N. Jr. Williams (1987), An approach to adjusting climatological time series for discontinuous inhomogeneities, *J. Climate Appl. Meteor.*, *26*, 1744-1763.
- Khaliq, M. N., and T. B. M. J. Ouarda (2007), A note on the critical values of the Standard Normal Homogeneity Test (SNHT), *Int. J. Climatol.*, *27*(5), 681-687.
- Lanzante, J. R. (1996), Resistant, robust and non-parametric techniques for the analysis of climate data: theory and examples, including applications to historical radiosonde station data, *Int. J. Climatol.*, *16*, 1197-1226.
- Lehman, E.L., and H. J. M. D'Abrera (1998), *Nonparametrics statistical methods based on ranks*, Revised first edition, 463 p., Prentice Hall, Upper Saddle River, New Jersey.
- Lund, R., and J. Reeves (2002), Detection of undocumented changepoints: A revision of the two-phase regression model, *J. Climate*, *15*, 2547-2554.
- Maronna, R., and V. J. Yohai (1978), A bivariate test for the detection of a systematic change in mean, *J. Amer. Statis. Assoc.*, *73*, 640-645.
- Ouarda, T. B. M. J., J. Asselin, and O. Seidou (2005), Bayesian multivariate linear regression with application to changepoint models in hydrometeorological variables. Model Development, *Research Report R-838*, 39 p., INRS-ETE, Quebec, Qc.
- Perreault, L., M. Haché, M. Slivitzsky, and B. Bobée (1999), Detection of changes in precipitation and runoff over eastern Canada and U.S. using a Bayesian approach, *Stoch. Environ. Res. Risk Assess.*, *13*, 201-216.

- Peterson, T.C., D. R. Easterling, T. R. Karl, P. Groisman, N. Nicholls, N. Plummer, S. Torok, I. Auer, R. Boehm, D. Gullett, L. Vincent, R. Heino, H. Tuomenvirta, O. Mestre, T. Szentimrey, J. Salinger, E. J. Forland, I. Hanssen-Bauer, H. Alexandersson, P. Jones, and D. Parker (1998), Homogeneity adjustments of in situ atmospheric climate data: a review, *Int. J. Climatol.*, *18*, 1493-1517.
- Potter, K.W. (1981), Illustration of a new test for detecting a shift in mean in precipitation series, *Mon. Wea. Rev.*, *109*, 2040-2045.
- Rasmussen, P. (2001), Bayesian estimation of change points using the general linear model, *Water Resour. Res.*, *37*, 2723-2731.
- Solow, A. R. (1987), Testing for climate change: an application of the two-phase regression model, *J. Climate Appl. Met.*, *26*, 1401-1405.
- Vincent, L. A. (1998), A technique for the identification of inhomogeneities in Canadian temperature series, *J. Climate*, *11*, 1094-1105.
- Wang, X. L. (2003), Comments on 'Detection of Undocumented Changepoints: A revision of the Two-Phase regression model', *J. Climate*, *16*, 3383-3385.

Figure captions

Figure 1. Histogram of the magnitudes of the shifts introduced in 100-year long series with a single shift.

Figure 2. Example of generated synthetic candidate series a) with a shift at position 56 b) with a trend starting at position 30 and ending at position 83 c) with a change in standard deviation at position 33.

Figure 3. Position and magnitude of falsely detected shifts by each technique when applied to 100-year long homogeneous series a) SNHT b) MREG c) REG2 d) BIVT e) STUS f) WILS g) JARU h) BAYE.

Figure 4. Percentage of correctly identified shifts, as defined in section 4.2, according to their position and magnitude obtained by each technique when applied to 100-year long series with a single shift a) SNHT b) MREG c) REG2 d) BIVT e) STUS f) WILS g) JARU h) BAYE.

Figure 5. Percentage of well identified shifts, as defined in section 4.2, according to their position and magnitude obtained by each technique when applied to 100-year long series with a single shift a) SNHT b) MREG c) REG2 d) BIVT e) STUS f) WILS g) JARU h) BAYE.

Figure 6. Magnitude and position errors, calculated as described in section 4.2, obtained by each technique when applied to 100-year long series with a single shift a) SNHT b) MREG c) REG2 d) BIVT e) STUS f) WILS g) JARU h) BAYE. Shifts that were not detected by the techniques are not represented in this figure.

Figure 7. Histogram of the positioning criterion C (Eq. 35, section 4.3) obtained by all techniques when applied to series with multiple shifts a) two shifts b) three shifts.

Figure 8. Mean positioning criterion C (Eq. 35, section 4.3) according to the distance between two shifts obtained by all techniques.

Figure 9. Mean positioning criterion C (Eq. 35, section 4.3) according to the order of signs of the three shifts obtained by all techniques.

Tables

Table 1. Falsely detected shifts (%) by each technique when applied to 60-year long homogeneous series.

Magnitude (standard deviation)	SNHT	MREG	REG2	BIVT	STUS	WILS	JARU	BAYE
0-0.25	0.1	0.1	1.5	0.0	1.7	0.3	0.0	0.7
0.25-0.5	0.4	0.4	1.2	0.4	1.9	1.1	0.1	11.8
0.5-1	2.0	0.4	1.2	1.7	1.5	2.1	0.9	8.5
1-2	0.0	0.0	0.1	0.1	0.1	0.3	0.0	0.2
>2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Total	2.5*	0.9*	4.0*	2.2*	5.2	3.8*	1.0*	21.2*

* Significantly different from the expected percentage of type I error (5% critical level)

Table 2. Falsely detected shifts (%) by each technique when applied to 100-year long homogeneous series.

Magnitude (standard deviation)	SNHT	MREG	REG2	BIVT	STUS	WILS	JARU	BAYE
0-0.25	0.3	0.2	2.1	0.2	2.3	0.8	0.1	1.7
0.25-0.5	2.0	0.7	1.5	2.0	1.8	1.7	0.5	18.3
0.5-1	1.7	0.3	0.8	1.7	0.8	2.2	0.6	7.4
1-2	0.0	0.0	0.0	0.0	0.1	0.1	0.0	0.1
>2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Total	4.0*	1.2*	4.4	3.9*	5.0	4.8	1.2*	27.5*

* Significantly different from the expected percentage of type I error (5% critical level)

Table 3. Correctly identified, well identified and well positioned shifts (%) (as defined in section 4.2) by each technique when applied to 60-year long series with a single shift.

	SNHT	MREG	REG2	BIVT	STUS	WILS	JARU	BAYE
Correctly identified	60.0	56.0	51.9	61.6	32.1	57.3	61.7	58.6
Well identified	81.1	71.8	64.7	81.7	70.8	80.5	81.0	81.9
Well positioned	81.6	72.0	64.8	82.2	71.2	80.9	81.4	83.3

Table 4. Correctly identified, well identified and well positioned shifts (%) (as defined in section 4.2) by each technique when applied to 100-year long series with a single shift.

	SNHT	MREG	REG2	BIVT	STUS	WILS	JARU	BAYE
Correctly identified	61.5	59.3	57.4	62.7	30.3	57.9	63.0	57.8
Well identified	83.5	76.8	73.5	83.9	67.4	82.4	83.4	83.2
Well positioned	83.9	76.9	73.6	84.3	67.5	82.7	83.6	85.2

Table 5. Descriptive statistics of the absolute errors in position and magnitude (defined in section 4.2) for each technique when applied to 60-year long series with a single shift.

Absolute error	Statistic	SNHT	REGM	REG2	BIVT	STUS	WILS	JARU	BAYE
Position	Mean	6.9	14.8	17.7	6.8	14.8	6.4	7.7	5.0
	Median	0	0	0	0	1	0	0	0
	Standard Deviation	18.0	25.4	26.5	18.0	25.0	17.2	19.2	15.0
	Mean Rank*	93840	103110	109340	92380	120780	95960	93180	91140
Magnitude	Mean	0.3	0.7	0.9	0.3	0.8	0.3	0.4	0.2
	Median	0	0	0	0	0.1	0	0	0
	Standard Deviation	0.9	1.3	1.3	0.9	1.2	0.9	1.0	0.7
	Mean Rank*	93120	102970	109280	91710	124890	95190	92600	90250

* The absolute errors in position and magnitude differ significantly according to the techniques used (Kruskal-Wallis test, 5% critical level).

Table 6. Descriptive statistics of the absolute errors in position and magnitude (defined in section 4.2) for each technique when applied to 100-year long series with a single shift.

Absolute error	Statistic	SNHT	REGM	REG2	BIVT	STUS	WILS	JARU	BAYE
Position	Mean	7.7	18.3	18.7	7.5	28.9	7.5	9.2	5.2
	Median	0	0	0	0	1	0	0	0
	Standard Deviation	25.0	38.0	37.4	24.7	44.5	24.4	27.6	19.6
	Mean Rank*	93750	100370	102900	92720	127630	97110	93340	92180
Magnitude	Mean	0.2	0.6	0.6	0.2	0.9	0.2	0.3	0.1
	Median	0	0	0	0	0.1	0	0	0
	Standard Deviation	0.7	1.1	1.1	0.7	1.3	0.7	0.8	0.6
	Mean Rank*	92890	100170	102860	91840	132130	96430	92720	90980

* The absolute errors in position and magnitude differ significantly according to the techniques used (Kruskal-Wallis test, 5% critical level).

Table 7. Pairwise comparison of the absolute errors in position and magnitude (defined in section 4.2) for each technique when applied to 60-year long series with a single shift.

Absolute error	Technique	SNHT	REGM	REG2	BIVT	STUS	WILS	JARU	BAYE
Position	SNHT	0	1	1	1	1	1	0	1
	REGM		0	1	1	1	1	1	1
	REG2			0	1	1	1	1	1
	BIVT				0	1	1	0	1
	STUS					0	1	1	1
	WILS						0	1	1
	JARU							0	1
	BAYE								0
Magnitude	SNHT	0	1	1	1	1	1	0	1
	REGM		0	1	1	1	1	1	1
	REG2			0	1	1	1	1	1
	BIVT				0	1	1	0	1
	STUS					0	1	1	1
	WILS						0	1	1
	JARU							0	1
	BAYE								0

1 = significantly different, 0 = not significantly different (Conover-Inman test, 5% critical level)

Table 8. Pairwise comparison of the absolute errors in position and magnitude (defined in section 4.2) for each technique when applied to 100-year long series with a single shift.

Absolute error	Technique	SNHT	REGM	REG2	BIVT	STUS	WILS	JARU	BAYE
Position	SNHT	0	1	1	1	1	1	0	1
	REGM		0	1	1	1	1	1	1
	REG2			0	1	1	1	1	1
	BIVT				0	1	1	0	0
	STUS					0	1	1	1
	WILS						0	1	1
	JARU							0	1
	BAYE								0
Magnitude	SNHT	0	1	1	1	1	1	0	1
	REGM		0	1	1	1	1	1	1
	REG2			0	1	1	1	1	1
	BIVT				0	1	1	0	0
	STUS					0	1	1	1
	WILS						0	1	1
	JARU							0	1
	BAYE								0

1 = significantly different, 0 = not significantly different (Conover-Inman test, 5% critical level)

Table 9. Descriptive statistics of the positioning criterion C (Eq. 35, section 4.3) for each technique when applied to series with two shifts.

Statistic	SNHT	MREG	REG2	BIVT	STUS	WILS	JARU	BAYE
Mean	1768	2757	3044	1730	3268	1812	1928	1813
Median	2	4901	4901	2	4901	3	2	18
Standard Deviation	2483	2864	2831	2489	3219	2503	2643	2264
Minimum	0	0	0	0	0	0	0	0
Maximum	9801	9801	9801	9801	9801	9801	9801	9801
Mean rank*	54607	62960	66917	53537	74179	56446	54808	56550

* The mean positioning criterion C differs significantly according to the techniques used (Kruskal-Wallis test, 5% critical level)

Table 10. Descriptive statistics of the positioning criterion C (Eq. 35, section 4.3) for each technique when applied to series with three shifts.

Statistic	SNHT	MREG	REG2	BIVT	STUS	WILS	JARU	BAYE
Mean	2439	3291	3654	2378	3476	2565	2683	2216
Median	3267	3267	3267	3267	3267	3267	3267	2451
Standard Deviation	2443	2500	2451	2451	2723	2479	2633	2488
Minimum	0	0	0	0	0	0	0	0
Maximum	9801	9801	9801	9801	9801	9801	9801	9801
Mean rank*	54941	64350	69104	53625	71305	57850	56548	52280

* The mean positioning criterion C differs significantly according to the techniques used (Kruskal-Wallis test, 5% critical level)

Table 11. Pairwise comparison of the the positioning criterion C (Eq. 35, section 4.3) for each technique when applied to series with two shifts.

Technique	SNHT	REGM	REG2	BIVT	STUS	WILS	JARU	BAYE
SNHT	0	1	1	1	1	1	0	1
REGM		0	1	1	1	1	1	1
REG2			0	1	1	1	1	1
BIVT				0	1	1	1	1
STUS					0	1	1	1
WILS						0	1	0
JARU							0	1
BAYE								0

1 = significantly different, 0 = not significantly different (Conover-Inman test, 5% critical level)

Table 12. Pairwise comparison of the the positioning criterion C (Eq. 35, section 4.3) for each technique when applied to series with three shifts.

Technique	SNHT	REGM	REG2	BIVT	STUS	WILS	JARU	BAYE
SNHT	0	1	1	1	1	1	1	1
REGM		0	1	1	1	1	1	1
REG2			0	1	1	1	1	1
BIVT				0	1	1	1	1
STUS					0	1	1	1
WILS						0	1	1
JARU							0	1
BAYE								0

1 = significantly different, 0 = not significantly different (Conover-Inman test, 5% critical level)

Table 13. Falsely detected shifts (%) by each technique when applied to series with a change in variance.

Technique	SNHT	MREG	REG2	BIVT	STUS	WILS	JARU	BAYE
% Rejected	5.4	2.9*	7.7*	5.4	4.7	5.1	2.4*	27.7*

* Significantly different from the expected percentage of type I error (5% critical level)

Table 14. Difference (%) of proportions between falsely detected shifts in 100-year long homogeneous series and series with a change in variance with the associated 95% confidence interval (Eq. 36, section 5.4).

Technique	Lower Bound	Difference	Upper Bound
SNHT*	0.7	1.4	2.1
MREG*	1.2	1.7	2.1
REG2*	2.5	3.3	4.1
BIVT*	0.8	1.5	2.2
STUS	-1	-0.3	0.4
WILS	-0.5	0.3	0.9
JARU*	0.8	1.2	1.6
BAYE	-1.3	0.2	1.7

* Significant (5% critical level)

Table 15. Number of cases (%) for which one shift is detected inside the trend.

Magnitude (standard deviation)	SNHT	MREG	REG2	BIVT	STUS	WILS	JARU	BAYE
0-0.5	39.1	4.3	13.0	30.4	8.7	30.4	30.4	52.2
0.5-1	84.2	54.1	28.5	84.9	13.2	83.5	83.0	79.4
1-2	77.3	89.8	65.7	76.0	47.0	76.8	81.5	63.2
2-3	56.1	80.9	77.0	55.3	64.5	52.6	58.2	52.2

Table 16. Number of cases (%) for which at least two shifts are detected inside the trend.

Magnitude (standard deviation)	SNHT	MREG	REG2	BIVT	STUS	WILS	JARU	BAYE
0-0.5	0.0	0.0	0.0	4.3	0.0	0.0	0.0	4.3
0.5-1	2.6	1.2	0.0	2.6	0.0	4.3	1.4	13.6
1-2	21.9	6.1	1.8	23.4	2.7	22.3	17.7	36.5
2-3	43.9	19.0	6.4	44.7	12.3	47.3	41.7	47.8

Figures

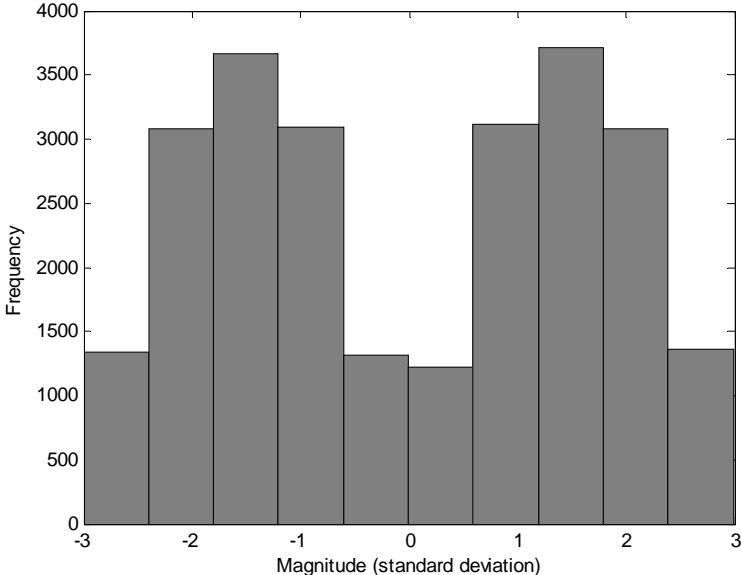


FIG. 1.

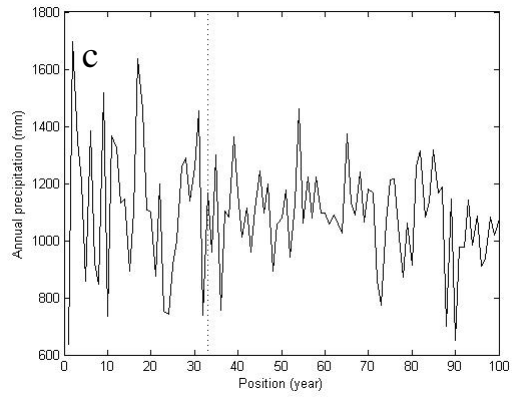
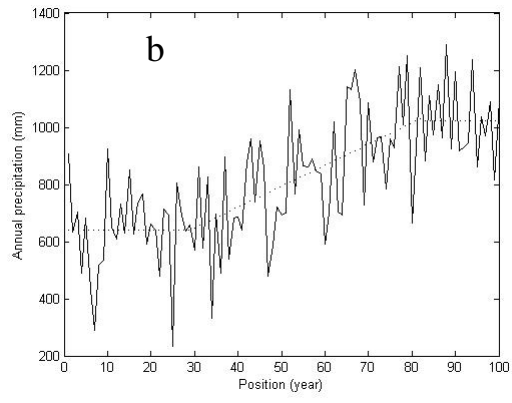
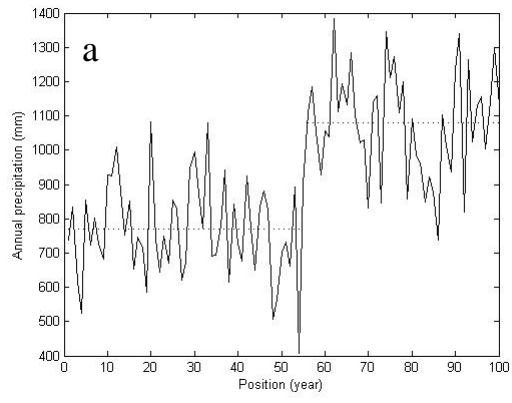


FIG. 2.

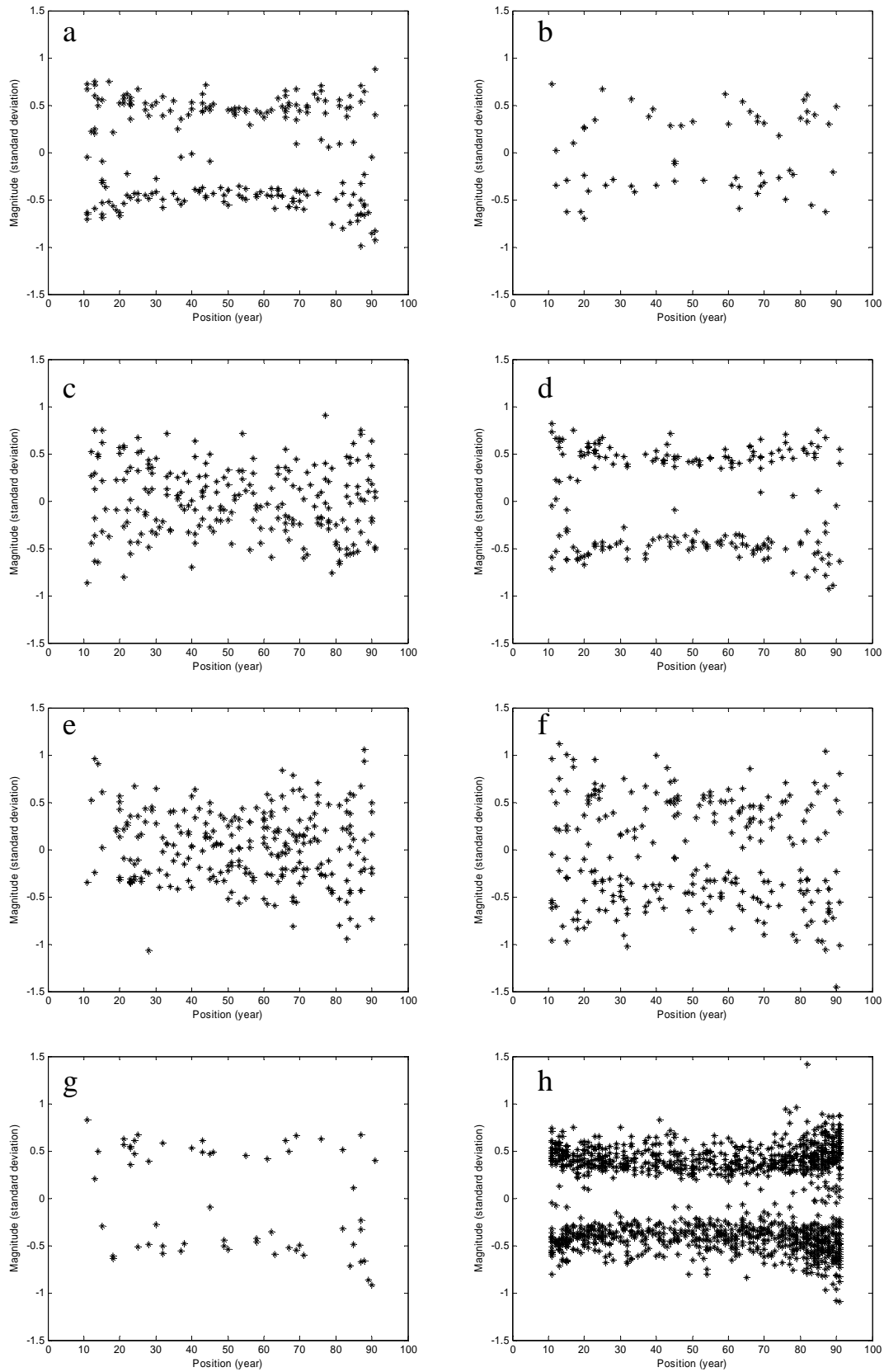


FIG. 3.

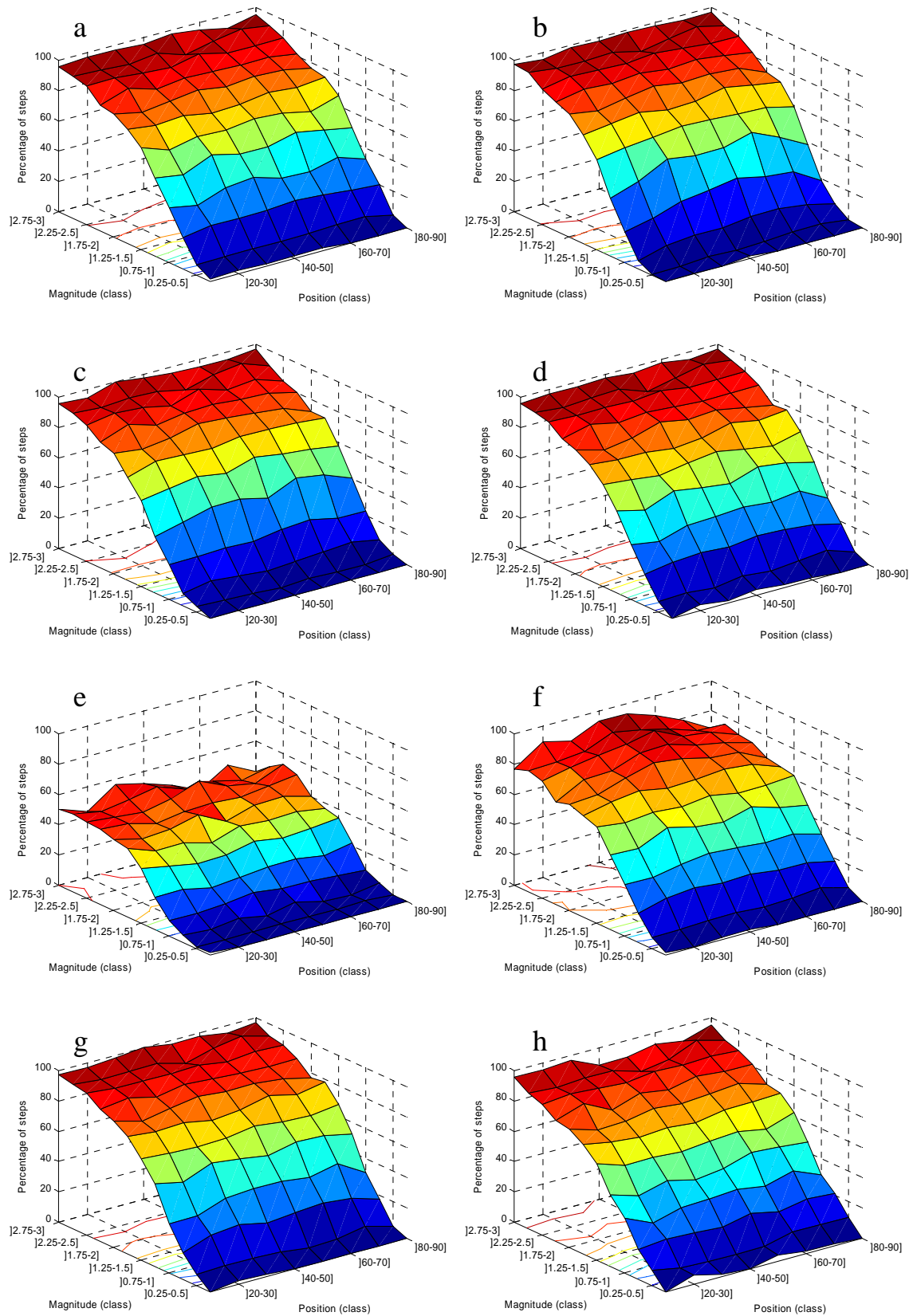


FIG. 4.

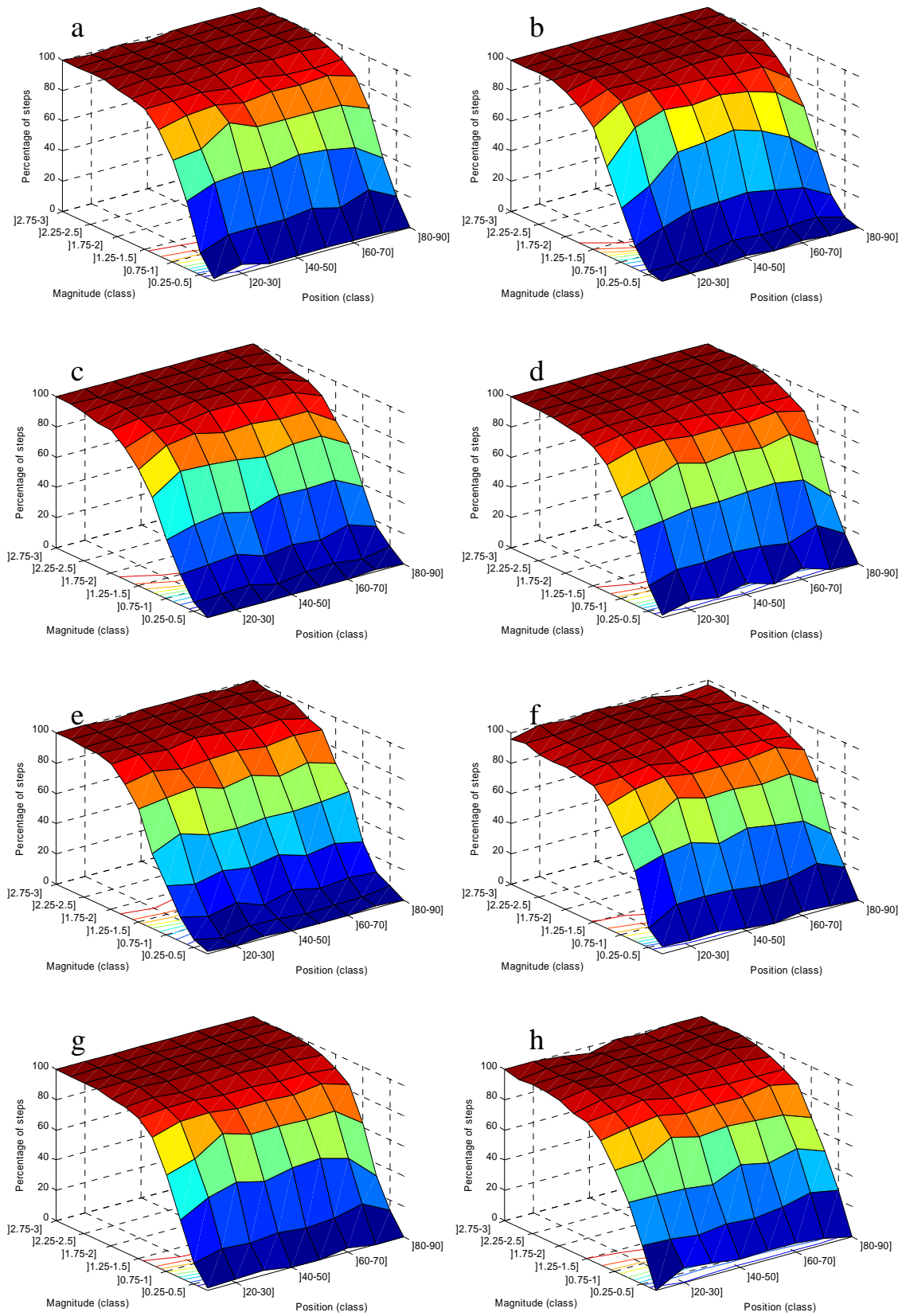


FIG. 5.

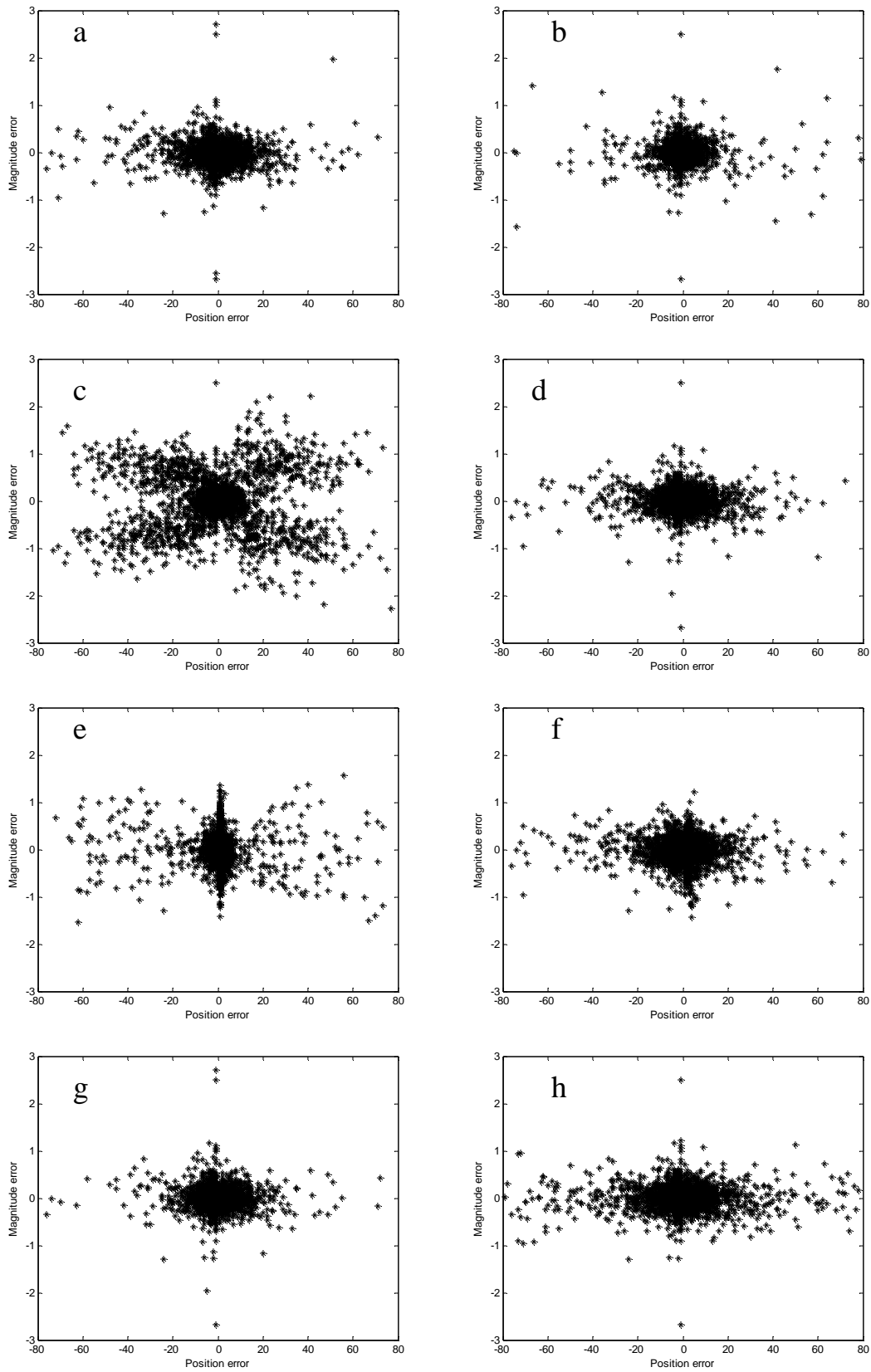


FIG. 6.

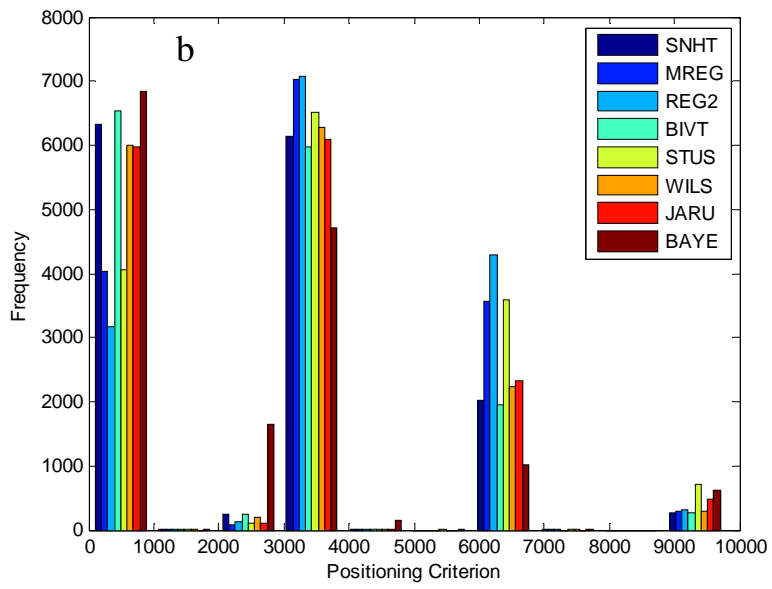
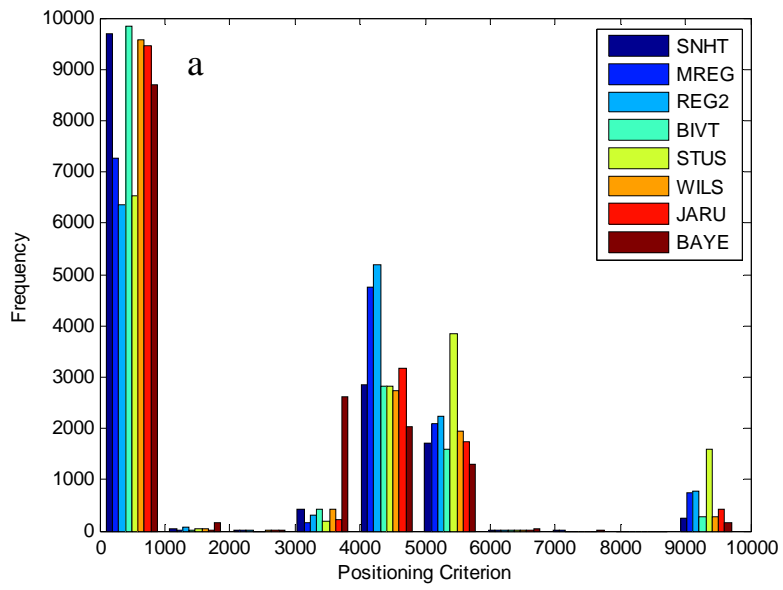


FIG. 7.

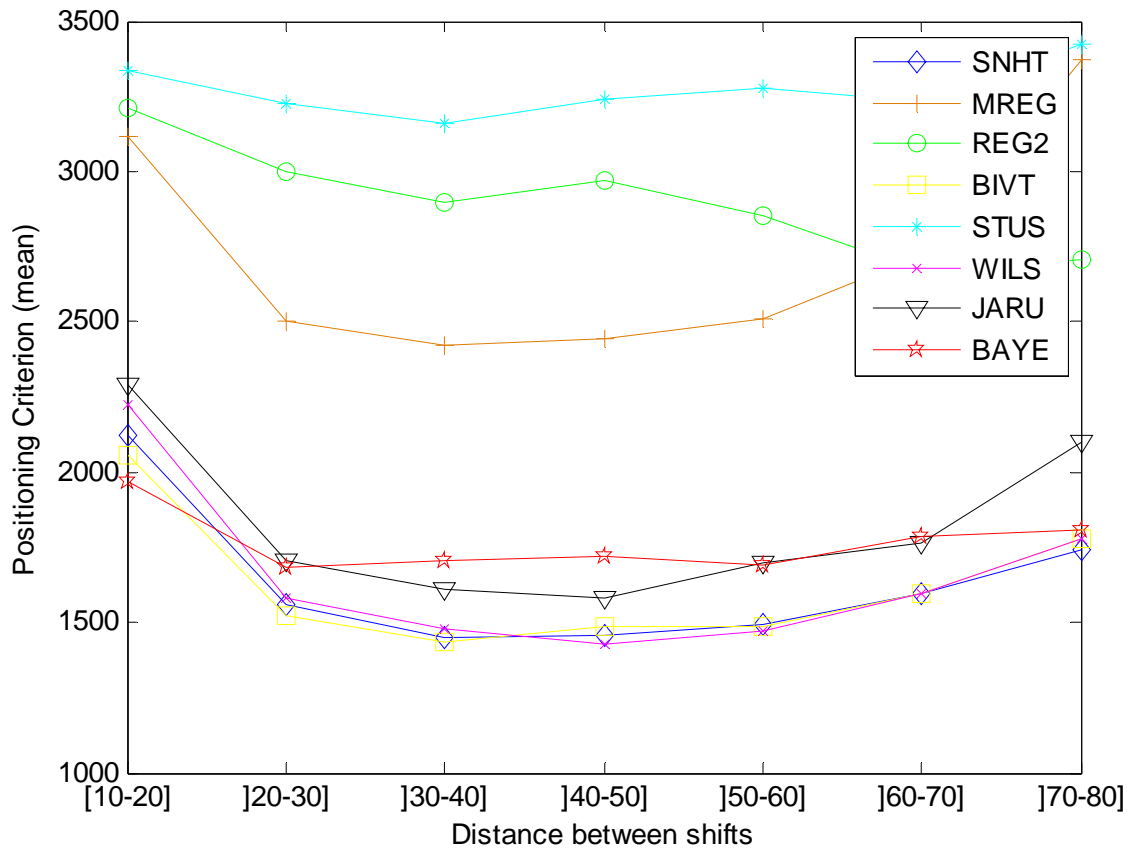


FIG. 8.

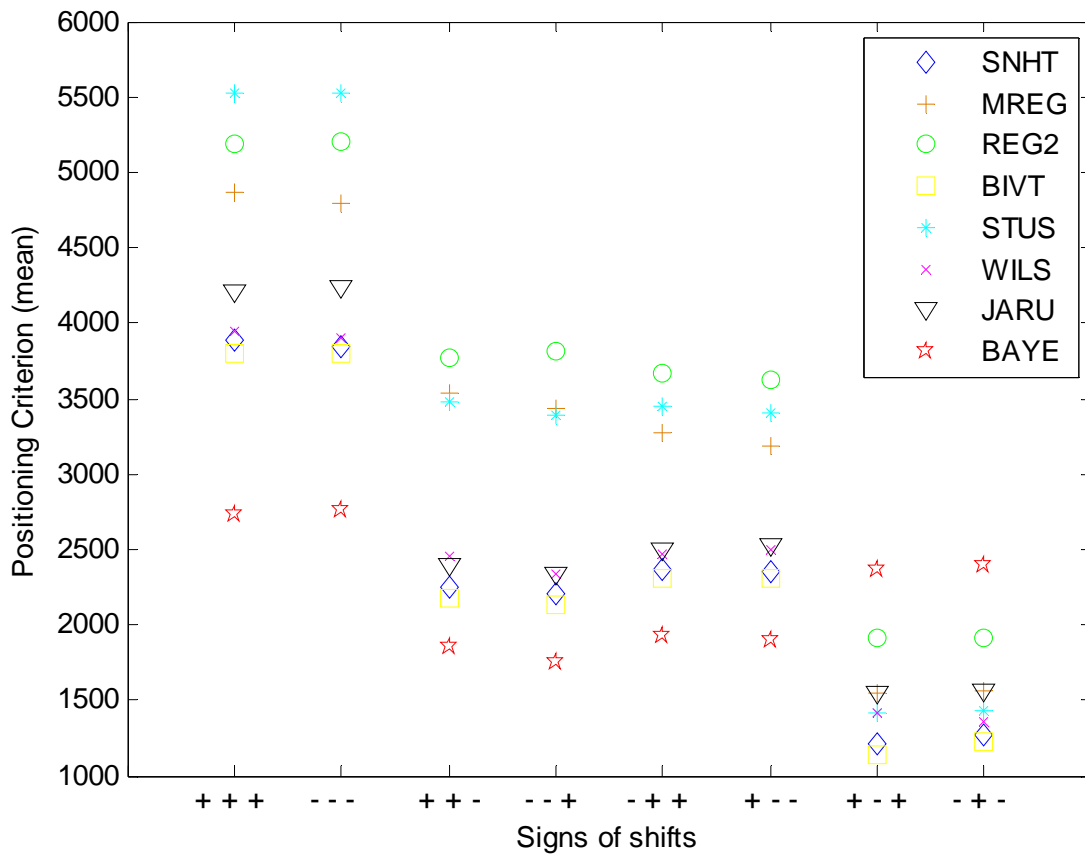


FIG. 9.

8. Article 3: Intercomparison of homogenization techniques for precipitation data continued: the comparison of two recent Bayesian changepoint models

Deux nouvelles méthodes bayésiennes pour détecter des inhomogénéités dans des séries climatiques sont proposées dans cet article. Ces deux dernières sont comparées à huit techniques d'homogénéisation présentées dans un travail précédent. La première technique permet de détecter des sauts multiples dans une relation de régression linéaire multiple. La seconde technique permet de détecter un changement dans un modèle de régression multiple. Leur capacité à identifier des séries homogènes et inhomogènes a été évaluée dans une étude de simulation de Monte Carlo. Différents ensembles de séries synthétiques (homogènes, avec un saut, avec plusieurs sauts) qui représentent les caractéristiques statistiques des précipitations totales annuelles typiques observées dans le sud et le centre de la province de Québec et ses alentours, au Canada, ont été générées. Les deux techniques ont donné des faibles risques de fausses détections lorsqu'elles ont été appliquées sur les séries homogènes. De plus, les deux techniques se sont montrées efficaces pour détecter un saut. Sur les séries avec des sauts multiples, la méthode bayésienne développée spécifiquement pour détecter des sauts multiples a mieux performé que toutes les autres méthodes. La seconde technique développée pour un seul changement a été appliquée de manière itérative pour la détection de plusieurs sauts. Cependant, cette technique semble inappropriée pour détecter plusieurs sauts. Un cas d'application réel à une série de précipitations totales annuelles est également présenté et l'inhomogénéité détectée est validée avec les métadonnées disponibles.

Intercomparison of homogenization techniques for precipitation data continued: the comparison of two recent Bayesian changepoint models

Claudie Beaulieu¹, Ousmane Seidou², Taha B.M.J. Ouarda¹ and Xuebin Zhang³

¹*Institut National de la Recherche Scientifique-Eau, Terre et Environnement, Université du Québec, Québec, QC, Canada*

²*Department of Civil Engineering, University of Ottawa, Ottawa, ON, Canada*

³*Climate Research Division, Science and Technology Branch, Environment Canada, Toronto, ON, Canada*

Submitted to Water Resources Research the October 6th, 2008

Corresponding author: Claudie Beaulieu, Chair in statistical Hydrology, Canada Research Chair on the Estimation of Hydrological Variables, INRS-ETE, University of Quebec, 490, de la Couronne, Québec (Québec) G1K 9A9, CANADA. E-mail: claudie.beaulieu@ete.inrs.ca; tel. (418) 654-2530 #4462

Abstract

In this paper, two new Bayesian changepoint techniques are described and compared to eight homogenization techniques presented in previous work to detect inhomogeneities in climatic series. An inhomogeneity can be defined as a changepoint (a time point in a series such that the observations have a different distribution before and after this time) in the data series induced from changes in measurement conditions at a given station. It is important to be able to detect and correct an inhomogeneity, as it can interfere with the real climate change signal. The first technique is a Bayesian method of multiple changepoint detection in a multiple linear regression. The second one allows the detection of a single changepoint in a multiple linear regression. The ability of the two techniques to identify homogeneous and inhomogeneous series was evaluated using simulated data series. Various sets of synthetic series (homogeneous, with a single shift, with multiple shifts) representing the typical total annual precipitation observed in the southern and central parts of the province of Quebec and nearby areas, Canada, were generated for the purpose of this study. The two techniques gave small false detection rates on the homogeneous series. Furthermore, the two techniques proved to be efficient for the detection of a single shift in a series. For the series with multiple shifts, the Bayesian method of multiple changepoint detection performed better than all the other techniques. The second Bayesian technique for the detection of a single changepoint seemed inappropriate for the detection of multiple shifts. An application to a real data set is also provided and validated with the available metadata.

Keywords: homogenization, precipitation series, Bayesian technique, changepoint, multiple linear regression, multiple shifts.

1. Introduction

The increasing interest in climate modeling has triggered the development and testing of more performing homogenization techniques. These techniques are intended to identify and sometimes correct inhomogeneities in data series that do not reflect the real climate variations. An inhomogeneity can be defined as a changepoint (a time point in a series such that the observations have a different distribution before and after this time) in the data series induced from changes in measurement conditions at a given station. These inhomogeneities (also called artificial discontinuities) may be induced from changes in measurement conditions at a given station such as relocation, instrument replacement, changes in observation procedures or modification in the immediate environment of the site. There is a difference between an inhomogeneous series and a series that is non-stationary, as a series can be non-stationary and homogeneous at the same time (the changepoints are only caused by real climatic variations). The user of climatic data series is often unaware of the presence or absence of inhomogeneities in the series, which is why the homogenization techniques are needed. The artificial discontinuities can interfere with the real climate change signal and lead to poor climatic or hydrological model calibration or biased studies of climate trends and variability. Therefore, the detection and correction of these inhomogeneities is important before undertaking any kind of climate analysis.

Several techniques have been developed for the detection of inhomogeneities in climate series. For a complete review, the reader is referred to *Peterson et al.* [1998], *WMO* [2003], *Beaulieu et al.* [2007] and *Reeves et al.* [2007]. Most of the homogenization techniques are addressed in classical or Bayesian statistics, supported by parametric or nonparametric models. The changepoint model often represents a shift in the mean of the base series (the series that is tested for a potential inhomogeneity) and that does not occur in the neighbor series. Some techniques allow to identify the position of the shift and to determine its significance using a

procedure based on the likelihood ratio test [*Maronna and Yohai, 1978; Potter, 1981; Alexandersson, 1986*]. The changepoint analysis can also be performed with a regression model which represents a change in the mean and/or in the trend of the series [*Easterling and Peterson, 1995; Vincent, 1998; Lund and Reeves, 2002; Wang, 2003*]. The neighbor series are used to identify shifts in the base series that are not reproduced in the immediate neighborhood. These shifts are assumed to be artificially caused and not representative of the regional climatic signal. Hence, the neighbor series are climatically similar to the base series and must be homogeneous. Otherwise, an inhomogeneity in a neighbor series could be allocated to the base series. However, there are a few techniques developed for isolated stations that do not rely on neighbor series. For more details on these techniques, the reader is referred to *Peterson et al. [1998]*.

Most homogenization techniques were developed for data series which contain at most one changepoint. In general, for the detection of multiple changes, a segmentation procedure is applied. The drawback with the segmentation approach is that the same test applied several times on the same observations can increase the risk of false detection. Furthermore, the size of the segments is reduced during the analysis, each time diminishing the detection power of the method. To avoid these inconvenients, a few techniques were developed to detect multiple shifts in several series [*Szentimrey 1996; 1999; Caussinus and Mestre, 2004*]. The series are mutually compared to each other by switching the role (neighbor or base) of the series at each step of the procedure. The technique proposed by *Szentimrey [1996; 1999]* allows the inference on the number of shifts by minimizing the variance of the difference series. The technique of *Caussinus and Mestre [2004]* allows the determination of the number of changepoints with a criterion developed by *Caussinus and Lyazrhi [1997]*. These techniques are suitable for cases with several neighbor series, but they are less efficient with few series (less than six) [*Slonosky et al., 1999*].

Few changepoint Bayesian techniques have been specifically applied to homogenization problems. The technique proposed by *Lee and Heghinian [1977]* and extended in *Perreault et al.*

[1999; 2000] was applied in comparison studies to detect inhomogeneities in synthetic temperature series. This technique performed well to detect a single shift, but tends to combine the sequential shifts in the presence of multiple inhomogeneities [*Ducré-Robitaille et al.*, 2003; *DeGaetano*, 2006]. *Beaulieu et al.* [2008] applied the Bayesian technique proposed by *Rasmussen* [2001] to identify inhomogeneities in synthetic precipitation series. This technique performed well to detect a single or multiple shifts, but lead to a very high percentage of false detections when the series are homogeneous. Then, the Bayesian techniques that were compared in simulation studies to solve homogenization problems such as *Ducré-Robitaille et al.* [2003], *DeGaetano* [2006] and *Beaulieu et al.* [2008] did not give better results than popular techniques such as the standard normal homogeneity test [*Alexandersson*, 1986] or the bivariate test [*Potter*, 1981].

There are many changepoint techniques presented in the literature. For a complete review of the different approaches, the reader can refer to *Chen and Gupta* [2001]. Bayesian linear regression models were addressed by *Bacon and Watts* [1971], *Chin Choy and Broemeling* [1980], *Smith and Cook* [1980], *Hsu* [1982], *Holbert* [1982], *Solow* [1988], *Kim* [1991] and *Rasmussen* [2001]. *Seidou et al.* [2007] presented a technique which has a general formulation and which allows to test for the hypothesis of no-change. In the case of multiple shifts, Bayesian techniques to infer the positions of the change when the number of shifts is known have been proposed by *Stephens* [1994]. *Barry and Hartigan* [1992; 1993] developed a technique to detect multiple changepoints with an algorithm based on product partition models. *Green* [1995] proposed the use of reversible jump Markov Chain Monte Carlo (MCMC) methods for multiple changepoint analysis. *Chib* [1998] and *Lavielle and Lebarbier* [2001] used a new formulation for the changepoint problem and propose to solve it with MCMC. *Seidou and Ouarda* [2007] applied the algorithm based on recursions proposed by *Fearnhead* [2006], to detect an unknown number of shifts in a multiple linear regression model. Changepoint linear regression models can be used

to detect inhomogeneities in climatic series by setting the base series as the dependent variable and the neighbor series as the independent variables.

The objectives of this work are to introduce two new Bayesian techniques [*Seidou and Ouarda, 2007; Seidou et al., 2007*] for the homogenization of climate series and to verify their capacity to discriminate homogeneous and inhomogeneous series having the statistical characteristics of precipitation data series observed in the southern and central parts of the province of Quebec and nearby areas, Canada. These two techniques were applied to various types of synthetic precipitation series: a set of homogeneous series, a set of series with a single shift and a set of series with multiple shifts, to evaluate their performances. For each type of synthetic series (homogeneous, one changepoint, two changepoints or three changepoints), the performances of these two new Bayesian techniques were compared to those of the best of eight techniques reviewed in *Beaulieu et al. [2008]*. These methods were not included in this comprehensive intercomparison exercise because they were under development at the time the comparison work was carried out. However, they were shown to be applicable to several practical problems that previously available methods could not handle. An application to real world data series is also presented.

The paper is organized as follows: section 2 gives a detailed description of the two techniques. The synthetic data sets are described in section 3. Section 4 presents the evaluation of the detection skills and the results of the application to the synthetic data sets are summarized in section 5. In section 6, an application to series of total annual precipitation is given. Finally, the discussion and conclusions are presented in section 7.

2. Theoretical background

In this paper, the Bayesian approach to changepoint detection is considered. Bayesian methods have a different approach from classical techniques. Through a prior distribution, they allow the integration of some knowledge (informative prior distribution) or lack of information (non-informative prior distribution) about the phenomenon being studied. Then, the information provided by the prior distribution and the information provided by the observations are combined into a posterior distribution, which is used to make inference about the parameters. Some advantages of using Bayesian analysis over classical analysis is that it allows a formal use of non-experimental sources of information and it provides the full posterior probability distribution for the parameter of interest [Berger, 1985]. For example, Bayesian changepoint methods provide the full posterior probability distribution of the position of the shifts, which can be multimodal or skewed. After specifying a loss function, an estimate of the shift's position can be obtained. This provides more information than an estimate of the shift's position and its significance level, which is obtained with the classical changepoint techniques.

In this section, the two techniques are presented. The first technique proposed in the present work is a Bayesian linear regression model designed to detect multiple changepoints [Seidou and Ouarda, 2007]. Its main feature is that it allows the detection of an unknown number of shifts. The second technique proposed in the present work is a Bayesian linear regression model which allows for the detection of a single change [Seidou et al., 2007]. It improves the model presented in Rasmussen [2001], which assumes the presence of a changepoint and infer its position, by extending it to the case for which a changepoint does not occur with certainty. The two techniques are described in the two following sub-sections.

2.1. Bayesian multiple changepoint detection in multiple linear regression (BAMS)

This technique is an adaptation of *Fearnhead* [2006] which presents a general procedure to detect the number and positions of the changepoints in a data series with a nonparametric approach. Let $\mathbf{y} = (y_1, y_2, \dots, y_n)^T$ be the sample, n the sample size, m the number of changepoints, $\tau_0 = 0, \tau_1, \dots, \tau_m, \tau_{m+1} = n$ the positions of the changepoints and $\mathbf{y}_{i:j}$ the observations from time i to time j . The superscript T indicates the transpose. We also denote $g(\cdot)$ the probability distribution of the time interval between consecutive changepoints and $g_0(\cdot)$ the probability distribution of the first changepoint. The j^{th} segment is then $\mathbf{y}_{(\tau_{j-1}+1):\tau_j}$ with parameter Φ_j . The posterior probabilities of the changepoints positions were derived by *Fearnhead* [2006]:

$$\begin{cases} \Pr(\tau_1 | \mathbf{y}_{1:n}) = P(1, \tau_1)Q(\tau_1 + 1)g_0(\tau_1) / Q(1) \\ \Pr(\tau_j | \tau_{j-1}, \mathbf{y}_{1:n}) = P(\tau_{j-1} + 1, \tau_j)Q(\tau_j + 1)g(\tau_j - \tau_{j-1}) / Q(\tau_{j-1} + 1) \end{cases} \quad (1)$$

where $P(t, s)$, $s \geq t$ is the probability that t and s (representing the lower and upper limits of a given segment) are in the same segment and :

$$\begin{aligned} P(t, s) &= \Pr(\mathbf{y}_{t:s}; t, s \text{ in the same segment}) \\ &= \int \prod_{i=t}^s f(y_i | \Phi) \pi(\Phi) d\Phi \end{aligned} \quad (2)$$

and $Q(t)$ is the likelihood of the segment $\mathbf{y}_{t:n}$ given a changepoint at $t-1$. $Q(t)$ $t = 1, \dots, n$ and

$P(t, s)$, $s \geq t$ are linked by these recursive equations:

$$\begin{cases} Q(1) = \sum_{s=1}^{n-1} P(1, s)Q(s+1)g_0(s) + P(1, n)(1 - G_0(n-1)) \\ Q(t) = \sum_{s=1}^{n-1} P(t, s)Q(s+1)g_0(s+1-t) + P(t, n)(1 - G(n-t)) \end{cases} \quad (3)$$

where $G(t) = \sum_{i=1}^t g(i)$ and $G_0(t) = \sum_{i=1}^t g_0(i)$. It is assumed that the observations are independent

conditional on the changepoints and parameter values. The position of the changepoint is inferred

based on the coherence $P(t,s)$ of all segments $[t,s]$, where t varies from 1 to n and s varies from t to n .

The procedure was adapted by *Seidou and Ouarda* [2007] to multiple linear regression, for normally distributed data having a mean which varies according to several covariates. The response variable is noted by $y_j (j=1, \dots, n)$, or $\mathbf{y}_{n \times 1}$ in vectorial form, while $x_{ij} (i=1, \dots, d^*; j=1, \dots, n)$ represents the j^{th} observation of the i^{th} explanatory variable ($\mathbf{X}_{d^* \times n}$ in matricial form). There are n observations and d^* explanatory variables. For homogenization problems, the response variable is the base series, while the neighbor series are the explanatory variables. The relationship between the base series and the neighbor series can be represented by

$$y_j = \sum_{i=1}^{d^*} \theta_i x_{ij} + \varepsilon_j, \quad j = 1, \dots, n \quad (4)$$

The response variable is normally distributed $N(\sum_{i=1}^{d^*} \theta_i x_{ij}, \sigma^2)$. Given the parameter vector

$\Phi = [\theta_1, \theta_2, \dots, \theta_{d^*}, \sigma]$, the density of the response variable is:

$$f(y_i | \Phi) = (2\pi\sigma^2)^{-1/2} \exp \left(-\frac{\left(y_i - \sum_{j=1}^{d^*} \theta_j x_{ij} \right)^2}{2\sigma^2} \right) \quad (5)$$

Following *Rasmussen* [2001], we have:

$$\Pr(\mathbf{y}_{t:s} | \Phi) = \prod_{i=1}^s f(y_i | \Phi) = (2\pi\sigma^2)^{-(s-t+1)/2} \exp \left[-\frac{(\mathbf{y}_{t:s} - \mathbf{X}_{t:s} \boldsymbol{\theta})^T (\mathbf{y}_{t:s} - \mathbf{X}_{t:s} \boldsymbol{\theta})}{2\sigma^2} \right] \quad (6)$$

From Eq. (3) and (6):

$$\begin{aligned} P(t,s) &= \Pr(\mathbf{y}_{t:s}; t, s \text{ in the same segment}) = \int \prod_{i=t}^s f(y_i | \Phi) \pi(\Phi) d\Phi \\ &= \int (2\pi\sigma^2)^{-(s-t+1)/2} \exp \left[-\frac{(\mathbf{y}_{t:s} - \mathbf{X}_{t:s} \boldsymbol{\theta})^T (\mathbf{y}_{t:s} - \mathbf{X}_{t:s} \boldsymbol{\theta})}{2\sigma^2} \right] \pi(\Phi) d\Phi \end{aligned} \quad (7)$$

where $P(t, s)$ is the posterior probability that t and s ($s \geq t$) are in the same segment. The prior probability distribution for the position of the change is uniform, given the preceding changepoint:

$$\pi(\tau_{i+1} | \tau_i) = \frac{1-p}{n-\tau_i} \quad (8)$$

where p represents the prior probability for no-change. Uniform priors are assumed for $\boldsymbol{\theta}$.

Therefore, the prior distribution of $\boldsymbol{\Phi}$ depends only on σ ($\pi(\boldsymbol{\Phi}) = \pi(\theta_1, \dots, \theta_{d^*}, \sigma) \propto \pi(\sigma)$). It is

assumed that the prior distribution of $\boldsymbol{\Phi}$ has this particular form:

$$\pi(\boldsymbol{\Phi}) = \pi(\sigma) = p(\sigma | a, c) = \frac{\sigma^{-a} \exp(-\frac{c}{2\sigma^2})}{2^{\frac{a-3}{2}} c^{\frac{a-1}{2}} \Gamma(\frac{a-1}{2})}, a > 1, c > 0 \quad (9)$$

where a and c are the hyperparameters and Γ represents the Gamma function. This prior was used instead of the inverted gamma distribution, which is the standard prior for Bayesian linear models, because it allows to obtain an analytic form for $P(t, s)$. The analytic expression of $P(t, s)$ is obtained after substituting Eq. (8) in Eq. (7) and integrating out σ and $\boldsymbol{\theta}$ in Eq. (7):

$$P(t, s) = 2^{\frac{d^*}{2}} \pi^{\frac{d^*-t+s-1}{2}} \frac{(\boldsymbol{\epsilon}_{s:t}^T \boldsymbol{\epsilon}_{s:t} + c)^{\frac{-(t-s+a)}{2}} \Gamma\left(\frac{t-s+a}{2}\right)}{c^{\frac{(a-1)}{2}} |\mathbf{X}_{s:t}^T \mathbf{X}_{s:t}|^{1/2} \Gamma\left(\frac{a-1}{2}\right)} \quad (10)$$

where $\boldsymbol{\epsilon}_{t:s}$ is a part of the vector of random errors between t and s and $\mathbf{X}_{t:s}$ represents the rows between t and s of the explanatory variables.

The hyperparameter a was chosen after a sensitivity analysis. The level of information in the prior can be controlled by changing the value of a . It was noticed, after a few trials, that with the non informative prior $a = 1 + \varepsilon$, this technique performed well on homogeneous series with a weak percentage of false detection. When an informative prior ($a \geq 3$) was used, the technique had a high percentage of false detection on the homogeneous series and a high percentage of

detection on series with real shifts. For the present study, a different approach was used. It consists in applying the technique with a non informative prior ($a = 1.1$) to determine if the series is homogeneous or not. If the series is considered inhomogeneous, the technique is reapplied with informative priors ($a = 5$) to detect the position and magnitude of the shifts. The parameter c , representing the dimension of the variance, was set to the variance obtained by least square estimates of the linear regression equations,

$$(c = \boldsymbol{\varepsilon}_{1:n}^T \boldsymbol{\varepsilon}_{1:n} = \mathbf{y}_{1:n}^T \mathbf{y}_{1:n} - \mathbf{X}_{1:n} \mathbf{X}_{1:n}^T (\mathbf{X}_{1:n}^T \mathbf{X}_{1:n})^{-1} \mathbf{y}_{1:n}^T \mathbf{y}_{1:n}).$$

To make inference on the positions of changepoints, a set of M possible scatter schemes ($E = \{S_k, k = 1 : M\}$) of the changepoints on the segment using the posterior probability mass of the first changepoint, and the conditional probability mass of subsequent changepoints was simulated. The set of scatter schemes is represented by E . Indeed, M should be large enough to obtain a reliable distribution for the positions of the changepoints. The k^{th} element of E , S_k , is a set of \tilde{m}_k changepoints $S_k = \{\tilde{t}_1^k, \tilde{t}_2^k, \dots, \tilde{t}_{\tilde{m}_k}^k\}$. An efficient simulation algorithm for E is given by *Fearnhead* [2006]:

1. For a sample of size M , initiate M samples with a changepoint at $t = 0$.
2. For $t = 0, \dots, n - 2$, repeat the following steps:
 - a) Compute the number n_t of samples for which the last changepoint was at time t ;
 - b) If $n_t > 0$, compute $\Pr(\tau | \tau_{j-1} = t, \mathbf{y}_{1:n})$;
 - c) Sample n_t times from $\Pr(\tau | \tau_{j-1} = t, \mathbf{y}_{1:n})$ and use the values to update the n_t samples of changepoints which have a changepoint at time t ;

This algorithm is very efficient since $\Pr(\tau | \tau_{j-1} = t, \mathbf{y}_{1:n})$ has to be computed only one time regardless of the number of samples required from it. Inference on the number and positions of

the changepoints is readily carried out using the M samples. For instance, the probability of having i changepoints is approximated by:

$$\Pr(m = i) \approx \text{card}(\{k \mid \text{card}(S_k) = i\}) / M \quad (11)$$

The posterior probability of having the k^{th} changepoint at position t given m changepoints can be approximated by:

$$\Pr(\tau_i = t \mid m) \approx \frac{\text{card}(\{k \mid (\text{card}(S_k) = m) \& (\tilde{t}_i^k = t)\})}{\text{card}(\{k \mid \text{card}(S) = m\})} \quad (12)$$

where $\text{card}(S)$ stands for the number of elements of the set S . The estimators of the number and positions of changepoints are the modes of their posterior distributions.

2.2. Bayesian changepoint in multiple linear regression (BARE)

This technique is presented in *Seidou et al.* [2007]. The model was designed to infer the position of a single changepoint in the parameters of a multiple regression equation. The model can be represented in matrix form by:

$$\mathbf{y}_t = \mathbf{X}_t \boldsymbol{\theta}_t^{(p)} + \mathbf{v}_t \quad (13)$$

where \mathbf{y}_t represents the response vector ($r \times 1$), \mathbf{X}_t is the matrix of explanatory variables matrix ($r \times d^*$) of the neighbor series, \mathbf{v}_t represents the vector of residuals that are independent and identically normally distributed ($N[0, \boldsymbol{\Sigma}_v]$) and $\boldsymbol{\theta}_t^{(p)}$ is the parameter vector where:

$$\boldsymbol{\theta}_t^{(p)} = \begin{cases} \boldsymbol{\beta}_1^*, & 1 \leq t < p, \\ \boldsymbol{\beta}_2^*, & p \leq t \leq n, \end{cases} \quad (14)$$

under the following constraints:

$$\boldsymbol{\beta}_1^* = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_0)^T \text{ and } \boldsymbol{\beta}_2^* = (\boldsymbol{\beta}_2, \boldsymbol{\beta}_0)^T. \quad (15)$$

The dimensions of the vectors $\boldsymbol{\theta}_t^{(p)}$, $\boldsymbol{\beta}_1^*$, $\boldsymbol{\beta}_2^*$, $\boldsymbol{\beta}_0$, $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$ are respectively $(d^* \times 1)$, $(d^* \times 1)$, $(d^* \times 1)$, $(d_0^* \times 1)$, $(d_1^* \times 1)$ and $(d_1^* \times 1)$. The constraints imply that $d^* = d_0^* + d_1^*$. The model infers

a change in the vector $\theta_t^{(p)}$ at the position p . The vector of parameters β_0 is assumed constant before and after the changepoint. The vectors β_1 and β_2 contain the values of the remaining parameters before and after the changepoint. Finally, the vectors β_1^* and β_2^* represent the regression parameters before and after the change. Hence, the model represents a changepoint in the vector $\theta_t^{(p)}$, from the subvector β_1 to the subvector β_2 . The subvector β_0 remains part of $\theta_t^{(p)}$. When this model is applied to detect inhomogeneities in a climate series, the base series is represented by y_t , and X_t is the matrix of the neighbor series.

In *Seidou et al.* [2007], non informative prior distributions for the regression parameters and the variance were considered. The prior for the changepoint position is a uniform distribution. The posterior distributions were obtained by Gibbs sampling [*Gelfand and Smith*, 1990]. The method was developed to detect at most one changepoint, but it can be applied for multiple changepoints using a segmentation approach. As the Gibbs sampling is used, the model can take into account missing observations.

3. Simulation study

3.1. Synthetic series

The two techniques were applied to synthetic series of precipitation that were generated for previous comparative studies of several homogenization techniques. The statistical properties reproduced in these series are a) a mean total annual precipitation of 1089 mm, b) a standard deviation of 142 mm and c) a lag one autocorrelation of 0.02. These values are the average characteristics of a set of selected stations located in central Quebec and surroundings having long time series with few missing data. For each base series, three correlated neighbor series are generated to reproduce a spatial cross-correlation of 0.55. This value is the mean spatial cross-

correlation in the set of selected stations that are located at a distance less than 300 km. Different data sets are used here: homogeneous series, series with a single or multiple shifts. The magnitude and positions of the shifts were generated randomly to represent as many configurations as possible. The detailed generation scheme of the synthetic series is presented in Beaulieu *et al.* [2008]. An example of series with two shifts is presented in figure 1.

3.2. Application of the techniques

The two techniques were applied to the synthetic series. The mode of the posterior distributions was used as the Bayes parametric estimator for the presence and positions of the shifts. In practice, it is common to apply sequentially a technique developed for a single shift, to be able to detect multiple shifts. This procedure was used to detect multiple shifts in the comparative studies of Ducré-Robitaille *et al.* (2003) and Beaulieu *et al.* (2008). To detect multiple shifts with BARE, this procedure was used. When a shift is detected, the series is divided into two segments, the technique is reapplied to each segment, and this procedure is repeated until all segments of the series are considered homogeneous or too short to be tested. For both techniques, the magnitudes were estimated by computing the difference of the means of the segments preceding and following a shift. The estimated magnitudes are then compared to the real magnitudes introduced in the synthetic series. A prior probability of change of 0.5 was used with both techniques, to be able to compare the results with classical techniques. It can be considered as non informative as it represents the case for which the user has no information about the presence or absence of a shift in the series. The minimal length between two consecutive shifts was set to 10 observations and the shifts identified in the first ten or last ten observations in the series were ignored to make the results comparable to those obtained in Beaulieu *et al.* [2008].

3.3. Evaluation of the detection skills

The detection skills of the techniques were assessed on all the synthetic series. For all types of synthetic series, the number of shifts detected is compared to the real number of shifts without taking into account the position or the magnitude of the shifts detected.

For the homogeneous series, the performance is verified by computing the false detection (type I error) rate. It is the percentage of cases for which the null hypothesis is rejected (homogeneity) when it is true.

For the series with a single shift, the errors in position and magnitude were computed (real position/magnitude – estimated position/magnitude). The absolute errors were also calculated. When a shift is not detected, the position error and the absolute position error are fixed to the length of the series (100) and the magnitude error and the absolute magnitude error are fixed to 3 standard deviations (highest generated magnitude). The numbers of **correctly identified, well identified** and **well positioned** shifts were also computed. A shift is correctly identified when the estimated position is exact and the relative difference between the estimated magnitude and the real magnitude is less than 20 % of the real magnitude. A shift is considered well identified when it is located within ± 2 years of the true location and the absolute error on the estimation of the magnitude is lower or equal to 50% of the real magnitude. A shift is well positioned when it is located within ± 2 years of the true position without any measure of accuracy for the magnitude.

For the series with multiple shifts, the ability of the techniques to correctly position all the shifts, without omission or false detection is evaluated with a performance criterion proposed by *Beaulieu et al.* [2008]. This criterion measures a distance between the locations of the true shifts and the detected shifts. It can be expressed as follows:

$$C = \begin{cases} \frac{1}{nd} \sum_{i=1}^{nd} (p_i^d - p_i)^2, & nr = nd \\ \frac{1}{nr} \left[\sum_{i=1}^{nd} (p_i^d - p_i)^2 + |nr - nd|(n-1)^2 \right], & nr > nd \\ \frac{1}{nd} \left[\sum_{j=1}^{nr} (p_j^d - p_j)^2 + |nr - nd|(n-1)^2 \right], & nr < nd \end{cases} \quad (16)$$

where $p_i^d, i=1, \dots, nd$ and $p_j, j=1, \dots, nr$ represent respectively the positions of the detected and real shifts, nd is the number of detected shifts, nr is the number of real shifts, and n is the length of the series. The pairs (p_i^d, p_j) are chosen to minimize the criterion. When the exact number of shifts is detected ($nr = nd$), the criterion is the sum of squares of the differences between the pairs which minimize the criterion. When $nr < nd$ or $nr > nd$, $(n-1)^2$ is added each time that a true shift is neglected or that a false shift is detected. This value corresponds to the square of the maximum possible distance between two shifts. C is equal to zero when all shifts are correctly positioned ($p_i^d = p_i, i=1, \dots, nd = nr$). When C is close to zero, the detected shifts are located near the true positions ($p_i^d \approx p_i, i=1, \dots, nd = nr$). A high value of C indicates that some shifts in the series are not detected or are wrongly detected ($nd \neq nr$). The performance criterion C was computed for all series with two and three shifts and the overall performance is the mean of the criterion over each set of synthetic series.

The performances of the two techniques are compared with those of the techniques that were found the best to identify each type of synthetic series in *Beaulieu et al.* [2008]. These techniques are: the Jaruskova's technique [*Jaruskova, 1996*], the bivariate approach [*Maronna and Yohai, 1978; Potter, 1981*] and a Bayesian technique [*Rasmussen, 2001*]. In the remainder of the text, these techniques are respectively named JARU, BIVT and BAYE.

These criteria used to assess the detection skills were chosen to be able to compare the results with those of *Beaulieu et al.* [2008]. However, from a Bayesian point of view, the use of

these criteria could be contested, as the information given by the posterior probability distribution is not used in its full potential. However, to be able to compare Bayesian methods with classical methods, some criteria have to be set. In the comparative studies of homogenization techniques of *Ducré-Robitaille et al.* [2003] and *DeGaetano* [2006], which involved Bayesian and classical techniques, similar criteria were used to assess the performance of the different techniques.

4. Results

4.1. Ability to identify the exact number of shifts

For all synthetic series, the number of shifts detected was compared with the true number of shifts that were introduced in the series. Figure 2 presents the percentage of series for which 0, 1, ..., 5 shifts were detected for the homogeneous series (0 shift) and the series with 1, 2 and 3 shifts. The position and magnitude of the detected shifts are not the focus here. For the homogeneous series and the series with a single shift, both techniques detected, in a large majority, the exact number of shifts (0 and 1). For the series with two and three shifts, BAMS detected, almost systematically, the exact number of shifts, but BARE detected one shift in most cases.

4.2. False detection rates in the homogeneous series

Table 1 presents the false detection rates obtained with the two techniques. The best results observed in the work of *Beaulieu et al.* [2008] are also presented in table 1 for comparison purposes. Both techniques gave a percentage of falsely detected shifts significantly smaller than 5%, the same as for JARU. Figure 3 presents the magnitude and position of the falsely detected shifts on the homogeneous series. It can be seen in this figure that the number of falsely detected shifts tends to increase in the extremities as well as their magnitudes. It is also interesting to note that the magnitudes of the falsely detected shifts rarely exceed 1 standard deviation.

4.3. Ability to identify a series with a single shift

Table 2 presents the descriptive statistics of the absolute errors in position and magnitude for the two techniques. The mean absolute errors in position/magnitude are respectively 12.7/0.4 and 19.2/0.6 for BAMS and BARE. The results were compared with those obtained by the three techniques (BIVT, JARU and BAYE) that led to the best performances in Beaulieu *et al.* (2008). The results are also presented in table 2. The mean absolute errors in position/magnitude obtained with the various techniques were significantly different (Kruskal-Wallis test, 5% critical level). The Conover-Inman procedure was used (5% critical level) to make multiple pairwise comparisons between the mean absolute errors in position/magnitude obtained with the five techniques. The ranking of the absolute errors in position can be expressed as follows:

$$\underline{\text{BAYE BIVT JARU}} < \text{BAMS} < \text{BARE} \quad (17)$$

The techniques underlined are not significantly different. The same ranking was obtained for the absolute errors in magnitude. The techniques BIVT, JARU and BAYE remain those with the smallest errors. The absolute errors in position and magnitude obtained with BAMS and BARE are significantly different, BAMS having a better performance than BARE. Consequently, the two Bayesian techniques have a good capacity to detect series with a single shift, but some classical techniques perform slightly better. The type II error rates (failure to reject the homogeneity hypothesis when it is false) are respectively 11.5% and 17.1% for BAMS and BARE.

Table 3 presents the rates of correctly identified, well identified and well positioned shifts for the two techniques as well as BAYE, BIVT and JARU. BAMS and BARE do not give the best performances. Nevertheless, the two techniques are able to well position a shift (75% and more). The percentage of well identified shifts is similar to the percentage of well positioned shifts. Hence, when the detected shift is well positioned (± 2 years of the true location), the

magnitude is well estimated most of the time. However, there is a large gap between the rates of correctly identified shifts and the well identified or well positioned shifts (between 23% and 28%). Figure 4 presents the percentages of well identified shifts for various classes of magnitude and position obtained on series with a single shift. The two techniques have a high percentage of well identified shifts when the magnitude is larger than two standard deviations (almost 100%). When the magnitude is less than a standard deviation, the percentage of well identified shifts decreased quickly (50% and less). The percentage of well identified shifts does not seem to be affected by the position of the shift.

4.4. Ability to identify a series with multiple shifts

Tables 4 and 5 present the descriptive statistics of the positioning criterion C (Eq. 15) obtained on the series with two and three shifts with BAMS and BARE. The results of the techniques that had the best performances in the study of Beaulieu *et al.* (2008) are also presented. These techniques are: BIVT for the series with two shifts (table 4) and BAYE for the series with three shifts (table 5). The most successful techniques are those with the lowest values of the performance criterion. The criteria obtained with the different techniques are significantly different (Kruskal-Wallis test, 5% critical level). Hence, multiple pairwise comparisons were used to rank the results obtained with the different techniques (Conover-Inman procedure, 5% critical level). For the series with two shifts, the multiple pairwise comparison procedure gives:

$$\underline{\text{BAMS BIVT}} < \text{BARE} \quad (18)$$

The techniques underlined are not significantly different. BAMS led to the smallest criterion and its value is not significantly different from the value obtained with BIVT. For the series with three shifts, the ordering is:

$$\text{BAMS} < \text{BAYE} < \text{BARE} \quad (19)$$

BAMS led to the smallest criterion. The positioning criterion obtained with BAMS is significantly smaller than those obtained with all others techniques. It can be concluded that, BAMS is the most performing technique for the series with two and three shifts. In table 5, the median criterion is very low. It means that in half of the series with two shifts, BAMS well positioned all the shifts without detecting nonexistent shifts or omitting real ones. In table 6, the median criterion is higher. Indeed, the difficulty to identify all the shifts increases with the number of real shifts. The maximum criterion is 9801 and corresponds to the case for which all the real shifts were not detected.

5. Example of application to real data

In this section, an application of BAMS and BARE to detect a shift in the mean of a precipitation series is presented. A comparison is also made with some classical techniques which already proved their efficiency. The purpose of this application was to assess the appropriateness of the methods for practical use.

The total annual precipitation series of Quebec City (station 7016294), Quebec, Canada, was tested for homogeneity. This station was chosen among 35 stations located in southern and central Quebec and nearby areas (Canada) with high-quality data (few missing values, long observation series) that were extracted from Environment Canada's database. The Quebec station was chosen as the base series because of the availability of neighbor series and of metadata (station history) for this station. The metadata are used to identify the cause of an inhomogeneity.

The base station is located at latitude 46.8, longitude -71.38 and at an altitude of 70 meters. Two neighbor stations, Shawinigan Falls (7018000) and La Pocatière (7054095), were identified in the set of 35 stations by considering the distance from the base station, the elevation difference, the observation period, the correlations and the correlations computed from the first

difference series (table 6). The two neighbor series are used as explanatory variables in the two regression models (BAMS and BARE), while the dependent variable is the base series. The neighbor series allow to searching for a change which occurs only in the base series. When the neighbor series are too far from the base series or have a high elevation difference, the inhomogeneities in the base series may be hidden by the discordance between the series, due to the large spatial variability of the precipitation series. The correlations are computed for the period 1944-1982, the common observation period for all series. The correlations computed from the first difference series are high (0.89 and 0.88), which indicates a strong relationship between the neighbor series and the base series without inhomogeneities. The correlations are lower (0.72 and 0.53); hence the base or the neighbor series may be inhomogeneous. The disagreement between the correlations and the correlations computed from the first difference series is an indicator of the presence of potential inhomogeneities in the base or in the neighbor series [Vincent, 1998]. Figure 5 presents the locations of the base station and the two neighbor stations. Figure 6 presents the series of the base and the two neighbor stations.

BAMS and BARE were applied to the base series with the neighbor series as the explanatory variables. Both techniques assume IID Gaussian time series. All the series seemed normally distributed (Shapiro-Wilk test, 5% critical level). Since annual series were used, the independence can be assumed. As the two techniques are applied along with classical techniques, non-informative priors are necessary to allow a reliable comparison. The prior probability of change was set to 0.5 and the minimal length between two shifts to 10 years. The posterior probability distributions for the number of shifts and for the positions of the shifts given the number of shifts are presented in figure 7. The probabilities of the presence of a single changepoint are the highest for both techniques (0.9 with BAMS and 1.0 with BARE). With BAMS, a small probability is attributed to the presence of two shifts (0.1). Given the presence of

a single changepoint, the posterior probability that the change occurred in 1959 is 0.4 with BAMS. With BARE, the most probable date of change is 1956 with a probability of 0.6.

The posterior distributions of the date of change obtained by the two techniques are bimodal; they both have two peaks in 1956 and 1959. The two detected shifts are in the same neighborhood. The two techniques could have detected the same shift at 3 years interval. The presence of two shifts in the series is also possible, but only one is detected because of their proximity. Hence, when two changepoints are too close in a time series, a technique may detect only one shift, which is in fact a blend of the two changepoints [*Menne and Williams, 2005*]. If they have different magnitudes, the larger shift could obliterate the smaller one. This is emphasized if the magnitudes of the shifts have the same signs. Consecutive shifts having the same signs are more difficult to detect than successive shifts with alternate signs [*Beaulieu et al., 2008*].

For comparison purposes, classical techniques that were found efficient in previous work [*Beaulieu et al., 2008*] were also applied at a 5% critical level: BIVT, JARU and SNHT; the standard normal homogeneity test [*Alexandersson, 1986*]. All these techniques also detected a single shift in 1959.

The metadata was investigated to find the cause of a shift for these years. There is no information about a potential cause of change for the years 1955-1956, but it is indicated that the Quebec station was relocated around 1958, with a change in the elevation of the station. For precipitation series, a small change in the instrument height can induce a very important shift in a series [*Heino, 1997*]. Hence, this relocation could be the cause of the detected shifts. The metadata could have been used to build the prior distribution. As changes are documented in the metadata, a higher prior probability of change could have been used. In this case, it would not affect the results and the shift would remain detected. However, shifts having a smaller magnitude could be more easily detected using an informative prior distribution. The metadata

could also be used to set the minimum length between two segments as the minimum length between two changes documented.

6. Discussion, conclusions and future works

This work aimed to verify the capacity of the two Bayesian techniques to detect homogeneous and inhomogeneous series. Their capacity to identify homogeneous and inhomogeneous series was verified by applying the two techniques to several sets of synthetic series representing the typical total annual precipitation of the southern and central regions of the province of Quebec (Canada) and nearby areas. The synthetic series were also used in a previous comparative study aiming to identify the most appropriate techniques to homogenize precipitation series. The use of the same synthetic series allowed the comparison of the two Bayesian techniques with other techniques that were found to be the most suitable in the previous work.

The performances of the two techniques can be summarized according to some criteria. First, a weak percentage of false detection is very important. The two techniques presented in this work gave false detection rates smaller than 5%, which is the usual critical level used with classical techniques. Furthermore, the magnitudes of the falsely detected shifts rarely exceed 1 standard deviation (table 1). Hence, the probability to introduce large artificial shifts is very low. Second, the techniques should be able to identify a single shift in a series. The percentages of positioned shifts were around 80% (BAMS) and 74% (BARE) for the series with a single shift (table 3). The large shifts (magnitudes of 2 standard deviations and more) were well identified most of the time (figure 4). The probability to identify a shift with a small magnitude is lower. The absolute errors in position and magnitude obtained from each technique were also studied (table 2). In the comparative study of *Beaulieu et al.* [2008], some techniques had mean absolute

errors in position and magnitude smaller than the two Bayesian techniques. The two techniques have shown to be efficient for the detection of series with a single shift, but some techniques are more appropriated for this case (BAYE, BIVT and JARU). Finally, the techniques should be able to detect multiple shifts in a series, without omission or false detection. BARE seems disadvantaged in the case of multiple shifts and BAMS gave the best performance. In all cases of series with multiple shifts, BAMS gave the smaller criteria (tables 4 and 5). Even though the mean criterion obtained with BAMS was the smallest for the series with two shifts, it was not significantly different from BIVT, the technique that was identified as the best in this case in *Beaulieu et al.* [2008]. For the series with three shifts (table 5), BAMS was better than the best technique (BAYE) presented in *Beaulieu et al.* [2008] for the same case. It is also interesting to note that for the homogeneous series and the series with a single shift, BAMS and BARE detected in strong majority the exact number of shifts. For the series with two and three shifts, the probability to detect the exact number of shifts with BAMS is the highest, while BARE detected a single shift in the majority of the series (figure 2).

An example of application to a real data set was included to illustrate the homogenization procedures and to compare the two approaches to other performing techniques. The two Bayesian techniques were applied using non-informative priors for the real example as well as for the synthetic series to have results comparable to those obtained with classical methods. In homogenization, the metadata, some data not used in the analysis (the extremities) and regional information can be incorporated in the prior distribution to give information about the parameters and to reduce the uncertainty. The metadata can indicate the position of a potential shift as well as its associated incertitude. Usually, the metadata is used to validate and identify the cause of an inhomogeneity, as it was done in the application presented in the present work (preceding section). The use of metadata in the choice of the prior distribution is possible with Bayesian techniques and could improve their performances. For example, if some modifications are

documented, the prior probability for a change could be set higher and if there is no shift documented, the prior probability for a change could be set lower. The minimum distance between two consecutive shifts could be set as the minimum length between two changes documented in the metadata. It can be expected that the shifts having a small magnitude will be more easily detected with informative prior distributions.

The main advantage of BAMS is the simultaneous detection of several shifts. Most homogenization techniques were developed to detect at most one changepoint and are applied iteratively to detect multiple shifts. It was shown that it is advantageous to use a technique which allows the simultaneous detection of several shifts, such as BAMS, as it was the best performing technique in this case. Furthermore, other techniques developed to detect multiple shifts are performing when several neighbor series are available. This technique does not necessitate a high number of neighbor series to perform such as the technique developed Caussinus and Mestre (2004) for example. Another advantage of this technique is that it allows the detection of different types of inhomogeneities. It was tested for the detection of one or multiple shifts in the mean, but trends and shifts in the variance can also be detected with this technique. This should be the subject of a future work.

These two techniques were applied with neighbor series to be able to compare the results with those obtained with classical methods. A benefit of the two techniques is that it is also possible to apply these techniques without neighbor stations. Most homogenization techniques require the presence of homogeneous neighboring series to represent the regional climate. The use of a neighbor series that is inhomogeneous, could alter the detection and correction of a shift. This issue is discussed in *Menne and Williams* [2005] with simulated series and in *Reeves et al.* [2007] with an example of application to real data series.

A limitation of the two techniques, as well as several of the other techniques, is that they require data to be normally distributed. The normality assumption can be contested for several

types of climatic series. Nevertheless, a Box-Cox transformation [Box and Cox, 1964] can be used to normalize the data. For the total annual precipitation series, the central limit theorem is often evoked, as the data represents the sum of several events [Bendjoudi and Hubert, 1998]. For monthly or daily precipitation series, the normality hypothesis is even less realistic. The robustness of the two techniques to other distributions was not verified in the present work and no existing work has been done in this direction. The independence hypothesis is also made with these two techniques. Once again, with a smaller time scale, this assumption may be less realistic. The changepoint detection techniques which were developed for independent error series tend to detect more false shifts in the presence of autocorrelation [Lund *et al.*, 2007]. The effect of autocorrelation on these two techniques is not known yet. Application of the two techniques to series with different autocorrelation values would be necessary to quantify the effect of autocorrelation on the false detection rates. The extension of the two techniques to less restrictive models is highly desirable.

In Wang *et al.* [2007], it was shown that the false detection rates and the probabilities of detection according to the shift positions generally have a U-shape. For the false detection rates, it can be seen in figure 3 that the two techniques tend to detect more false shifts at the extremities of the series. The magnitudes of the falsely detected shifts tend also to increase at the extremities of the series. However, the power of detection seems mainly affected by the magnitude of the shift, and slightly affected by the position of the shift (figure 4). Wang *et al.* [2007] proposed a penalized statistic to correct the U-shape problem for a classic homogenization technique. With this penalized statistic, the false detection probabilities are the same for all positions in a homogeneous series and the confidence levels are the same for all positions of the detected changepoints. This issue was not addressed in the present work, but should be the subject of a future study.

These two techniques were applied under the specific conditions of the province of Quebec, Canada, and nearby areas. The results of this work may only be valid under these conditions. The same techniques could lead to different performances on series with a different length, more than three shifts, other shift magnitudes, a different distribution, a different autocorrelation structure and/or different correlations with neighbor series. However, these techniques will be useful to detect a single (BARE) or multiple shifts (BAMS) in precipitation series similar to the synthetic series generated. They can also be used with other climatic variables (e.g. temperature), with hydrological variables (e.g. streamflow) or more generally, to solve different changepoint problems which can be expressed in a multiple linear regression form as long as the hypotheses of the linear regression are respected.

Acknowledgments. The authors wish to thank the Fonds Québécois de Recherche sur la Nature et les Technologies, the National Sciences and Engineering Research Council of Canada (NSERC), the OURANOS Consortium and the Canada Research Chair Program for funding this research. The authors would also like to thank Mrs L. Vincent of the Meteorological Service of Canada for providing the metadata and Mr G. Boulet of Environment Quebec and Mr A. Yagouti of Transport Canada for their helpful comments.

Notations

β_1^*	regression parameters before the changepoint (BARE)
β_2^*	regression parameters after the changepoint (BARE)
β_0	component of the vector of regression parameters that is constant before and after the changepoint (BARE)
β_1	component of the vector of regression parameters that changes to β_2 after the changepoint (BARE)
β_2	component of the vector of regression parameters that replaces β_1 after the changepoint (BARE)
Γ	Gamma function
$\boldsymbol{\varepsilon}$	vector of random errors in the linear regression equation (BAMS)
$\boldsymbol{\varepsilon}_{t:s}$	part of the vector of random errors between s and t (BAMS)
$\boldsymbol{\theta}$	vector of regression parameters (BAMS)
$\boldsymbol{\theta}_t^{(p)}$	vector of regression parameters at time t given p (BARE)
\mathbf{v}_t	vector of residuals of the linear regression (BARE)
$\pi(\boldsymbol{\Phi})$	prior of $\boldsymbol{\Phi}$ (BAMS)
σ^2	variance (BAMS)
$\boldsymbol{\Sigma}_y$	variance-covariance matrix (BARE)
τ_k	position of the k^{th} changepoint (BAMS)
$\boldsymbol{\Phi}$	vector of parameters of the linear regression equation (BAMS)
a	parameter of the prior distribution of $\boldsymbol{\Phi}$ (BAMS)
c	parameter of the prior distribution of $\boldsymbol{\Phi}$ (BAMS)

C	positioning criterion
d^*	number of explanatory variables (including the intercept if any)
d_0^*	number of explanatory variables for which the regression coefficients are constant before and after the change (BARE)
d_1^*	number of explanatory variables for which the regression coefficients change (BARE)
E	set of generated scatter schemes (BAMS)
$f(y_i \Phi)$	conditional distribution of y given Φ (BAMS)
$g(t)$	probability distribution of the time interval between consecutive changepoints (BAMS)
$g_0(t)$	probability distribution of the first changepoint (BAMS)
k	number of generated scatter schemes (BAMS)
M	number of scatter schemes to generate with the posterior distributions of the positions of changepoints (BAMS)
m_u	number of changes in the u^{th} generated series (BAMS)
\hat{m}_u	estimate of the number of changes in the u^{th} generated series (BAMS)
\tilde{m}_k	number of changes in the k^{th} generated scatter scheme during the simulation of the changepoints
n	number of observations
N	Normal distribution
nd	number of shifts detected in the series
nr	number of true shifts in the series
p	position of the changepoint (BARE)

p_i	positions of the true shifts ($i = 1, \dots, nr$)
p_i^d	positions of the detected shifts ($i = 1, \dots, nd$)
$P(t, s)$	probability that t and s are in the same segment $s \geq t$ (BAMS)
$Q(t)$	likelihood of the segment $\mathbf{Y}_{t:n}$ given a changepoint at $t-1$ (BAMS)
r	number of response variables (BARE)
S_k	k^{th} scatter scheme generated with the posterior distributions of the positions of changepoints (BAMS)
t	time
\tilde{t}_i^k	estimate of the i^{th} change in the k^{th} generated scatter scheme (BAMS)
t_i^k	i^{th} change in the k^{th} generated scatter scheme (BAMS)
u	number of the generated series $\{\mathbf{Y}\}_u$ in the validation procedure (BAMS)
\mathbf{X}	matrix of explanatory variables (BAMS)
\mathbf{X}_t	t th row of the matrix of explanatory variables (BARE)
$\mathbf{X}_{t:s}$	rows t to s of the matrix of explanatory variables (BAMS)
\mathbf{Y}	vector of the response variable (BAMS)
\mathbf{Y}_t	t th row of the vector of the response variable (BARE)
$\mathbf{Y}_{t:s}$	rows t to s of the vector of the response variable (BAMS)
$\{\mathbf{Y}\}_u$	u^{th} generated series in the validation procedure (BAMS)

References

- Alexandersson, H. (1986), A homogeneity test applied to precipitation data, *J. Climatol.*, 6, 661-675.
- Bacon, D. W., and D. G. Watts (1971), Estimating the transition between two intersecting straight lines. *Biometrika Trust*, 58(3), 525-534.
- Barry, D., and J. A. Hartigan (1992), Product partition models for change point problems. *Ann. Stat.*, 20(1), 260-279.
- Barry, D., and J. A. Hartigan (1993), A Bayesian analysis for change point problems. *J. Amer. Statis. Assoc.*, 88(421), 309-319.
- Beaulieu, C., T. B. M. J. Ouarda, and O. Seidou (2007), A review of homogenization techniques for precipitation data and their applicability to precipitation series (in French), *Hydrol. Sci. J.*, 52(1), 18-37.
- Beaulieu, C., O. Seidou, T. B. M. J. Ouarda, X. Zhang, G. Boulet, and A. Yagouti (2008), Intercomparison of homogenization techniques for precipitation data. *Water Resour. Res.*, 44, W02425.
- Bendjoudi, H., and P. Hubert (1998), Statistical distribution of total annual rainfall. Is the Gaussian distribution inadequate? (in French). *Rev. Sci. Eau*, 4, 617-630.
- Berger, J. O. (1985), *Statistical decision theory and Bayesian analysis*, Springer series in statistics, Springer, second edition, 617 p.
- Box, G. E. P., and D. R. Cox (1964), An analysis of transformations. *J. Roy. Statist. Soc., Series B*, 26, 211-246.
- Caussinus, H., and F. Lyazrhi (1997), Choosing a linear model with a random number of change-points and outliers. *Ann. Inst. Statist. Math.*, 49, 761-775.

- Caussinus, H., and O. Mestre (2004), Detection and correction of artificial shifts in climate series. *J. Roy. Statist. Soc., Series C*, 53, 405-425.
- Chen, J., and A. K. Gupta (2001), On change-point detection and estimation. *Communications in Statistics-Simulation and Computation*, 30(3), 665-697.
- Chib, S. (1998), Estimation and comparison of multiple change-point models. *J. Econometrics*, 86, 221-241.
- Chin Choy, J. H., and L. D. Broemeling (1980), Some Bayesian inferences for a changing linear model. *Technometrics*, 22(1), 71-78.
- DeGaetano, A. T. (2006), Attributes of several methods for detecting discontinuities in mean temperature series. *J. Climate*, 19, 838-853.
- Ducré-Robitaille, J. F., G. Boulet, and L. A. Vincent (2003), Comparison of techniques for detection of discontinuities in temperature series. *Int. J. Climatol.*, 23, 1087-2003.
- Easterling, D. R. and T. C. Peterson (1995), A new method for detecting undocumented discontinuities in climatological time series, *Int. J. Climatol.*, 15, 369-377.
- Fearnhead, P. (2006), Exact and Efficient Bayesian Inference for Multiple Changepoint. *Stat. Comput.*, 16, 203-213.
- Gelfand, A. E. and A. F. M. Smith (1990), Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.*, 85(410), 398-409.
- Green, P. J. (1995), Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82(4), 711-732.

- Gullett, D. W., L. A. Vincent, and P. J. F. Sajecki (1990), Testing for homogeneity in temperature time series at Canadian climate stations, *Canadian Climate Centre Rep. 90-4*, 43 p., Atmospheric Environment Service, Downsview, Ont.
- Heino, R. (1997), Metadata and their role in homogenization. Proc. First Seminar for Homogenization of Surface Climate Data, Hungarian meteorological service, Budapest, Hungary.
- Holbert, D. (1982), A Bayesian analysis of a switching linear model. *Journal of Econometrics*, 19, 77-87.
- Hsu, D. A. (1982), A Bayesian robust detection shift in the risk structure of stock market returns. *J. Amer. Statis. Assoc.*, 77(377), 29-39.
- Kim, D. (1991), A Bayesian significance test of the stationnarity of regression parameters. *Biometrika Trust*, 78(3), 667-675.
- Jaruskova, D. (1996), Change-point detection in meteorological measurement, *Mon. Wea. Rev.*, 124, 1535-1543.
- Lanzante, J. R. (1996), Resistant, robust and non-parametric techniques for the analysis of climate data: theory and examples, including applications to historical radiosonde station data, *Int. J. Climatol.*, 16, 1197-1226.
- Lavielle, M., and E. Lebarbier (2001), An application of MCMC methods for the multiple change-points problem. *Signal Processing*, 81, 39-53.
- Lee, A. S. F. and S. M. Heghinian (1977), A shift of the mean level in a sequence of independent normal random variables. *Technometrics*, 19, 503-506.

- Lund, R., and J. Reeves (2002), Detection of undocumented changepoints: A revision of the two-phase regression model, *J. Climate*, 15, 2547-2554.
- Lund, R., X. L. Wang, Q. Lu, J. Reeves, C. Gallagher, and Y. Feng (2007), Change point detection in periodic and autocorrelated time series. *J. Climate*, 20, 5178-5190.
- Maronna, R., and V. J. Yohai (1978), A bivariate test for the detection of a systematic change in mean, *J. Amer. Statis. Assoc.*, 73, 640-645.
- Menne, M. J., and C. N. J. Williams (2005), Detection of undocumented changepoints using multiple test statistics and composite reference series. *J. Climate*, 18, 4271-4286.
- Perreault, L., M. Haché, M. Slivitzsky, and B. Bobée (1999), Detection of changes in precipitation and runoff over eastern Canada and U.S. using a Bayesian approach, *Stoch. Environ. Res. Risk Assess.*, 13, 201-216.
- Perreault, L., J. Bernier, B. Bobée, and E. Parent (2000), Bayesian change-point analysis in hydrometeorological time series. Part 1. The normal model revisited. *J. Hydrol.*, 235, 221-241.
- Peterson, T.C., D. R. Easterling, T. R. Karl, P. Groisman, N. Nicholls, N. Plummer, S. Torok, I. Auer, R. Boehm, D. Gullett, L. Vincent, R. Heino, H. Tuomenvirta, O. Mestre, T. Szentimrey, J. Salinger, E. J. Forland, I. Hanssen-Bauer, H. Alexandersson, P. Jones, and D. Parker (1998), Homogeneity adjustments of in situ atmospheric climate data: a review, *Int. J. Climatol.*, 18, 1493-1517.
- Potter, K.W. (1981), Illustration of a new test for detecting a shift in mean in precipitation series, *Mon. Wea. Rev.*, 109, 2040-2045.

- Rasmussen, P. (2001), Bayesian estimation of change points using the general linear model, *Water Resour. Res.*, 37, 2723-2731.
- Reeves, J., J. Chen, X. L. Wang, R. Lund, and Q. Lu (2007), A review and comparison of changepoint detection techniques for climate data. *J. App. Meteor. Climatol.*, 46, 900-915.
- Seidou, O., J. J. Asselin, and T. B. M. J. Ouarda (2007), Bayesian multivariate linear regression with application to changepoint models in hydrometeorological variables. *Water Resour. Res.*, 43, W08401, doi:10.1029/2005WR004835.
- Seidou, O., and T. B. M. J. Ouarda (2007), Recursion-based multiple changepoint detection in multiple linear regression and application to river streamflows. *Water Resour. Res.*, 43, W07404, doi:10.1029/2006WR005021.
- Slonosky, V. C., P. D. Jones, and T. D. Davies (1999), Homogenization techniques for European monthly mean surface pressure series. *J. Climate*, 12, 2658-2672.
- Smith, A. F. M. and D. G. Cook (1980), Straight lines with a change-point: A Bayesian analysis of some renal transplant data. *Appl. Stat.*, 29(2), 180-189.
- Solow, A. R. (1988), A Bayesian approach to statistical inference about climate change. *J. Climate*, 1, 512-521.
- Stephens, D. A. (1994), Bayesian retrospective multiple-changepoint identification. *Appl. Stat.*, 43(1), 159-178.
- Szentimrey, T. (1996), Statistical procedure for joint homogenization of climatic time series. Proceedings, *First Seminar for Homogenization of Surface Climatology Data*, Budapest, Hungary, Hungarian Meteorological Service.

- Szentimrey, T. (1999), Multiple analysis of series for homogenization (MASH). Proceedings, *Second Seminar for Homogenization of Surface Climatological Data*, Budapest, Hungary, Hungarian Meteorological Service.
- Vincent, L. A. (1998), A technique for the identification of inhomogeneities in Canadian temperature series, *J. Climate*, *11*, 1094-1105.
- Wang, X. L. (2003), Comments on 'Detection of Undocumented Changepoints: A revision of the Two-Phase regression model', *J. Climate*, *16*, 3383-3385.
- Wang, X. L., Q. H. Wen, and Y. Wu (2007), Penalized maximal t-test for detecting undocumented mean change in climate data series. *J. App. Meteor. Climatol.*, *46*, 916-931.
- WMO, (2003), Guidelines on climate metadata and homogenization. *Report WMO-TD 1186*, World Meteorological Organization, Geneva, Switzerland, 50 p.

Figure captions

FIG. 1. Example of synthetic series with two shifts.

FIG. 2. Percentage of series for which 0,1,...,5 shifts were detected by the two techniques when there are 0, 1, 2 or 3 true shifts.

FIG. 3. Falsely detected shifts by the two techniques when applied to the homogeneous series a) BAMS b) BARE.

FIG. 4. Percentage of well identified shifts according to their position and magnitude obtained by the two techniques when applied to series with a single shift a) BAMS b) BARE.

FIG. 5. Location map of the base station (Quebec) and the two neighbor stations (Shawinigan Falls and La Pocatiere).

FIG. 6. Total annual precipitation of the base station and the two neighbor stations: a) Quebec city (base), b) Shawinigan Falls (neighbor), c) La Pocatiere (neighbor).

FIG. 7. Posterior probability distributions for (1) the number of shifts and (2) the position of the shift obtained with the technique a) BAMS and b) BARE.

Tables

TABLE 1. Falsely detected shifts (%) in the homogeneous series according to the estimated magnitude of the shift.

Magnitude (standard deviation)	BAMS	BARE	JARU
0-0.25	1.04	0.18	0.08
0.25-0.5	0.96	0.16	0.46
0.5-1	0.44	0.16	0.64
1-2	0.06	0.00	0.00
>2	0.00	0.00	0.00
Total	2.50*	0.50*	1.18*

* Significantly smaller than the expected percentage of type I error obtained with a classical technique (5% critical level).

TABLE 2. Descriptive statistics of the absolute errors in position and magnitude obtained when the techniques are applied to series with a single shift.

Absolute error	Statistic	BAMS	BARE	BAYE	BIVT	JARU
	Mean	12.7	19.2	5.2	7.5	9.2
Position	Median	0	1	0	0	0
	Standard Deviation	32.0	38.4	19.6	24.7	27.6
	Mean	0.4	0.6	0.1	0.2	0.3
Magnitude	Median	0	0	0	0	0
	Standard Deviation	1.0	1.1	0.6	0.7	0.8

The absolute errors in position and magnitude obtained with all techniques differ significantly (Kruskal-Wallis test, 5% critical level).

TABLE 3. Correctly identified, well identified and well positioned shifts (%) obtained when the techniques are applied to series with a single shift.

Identification of the shift	BAMS	BARE	BAYE	BIVT	JARU
Correctly identified	56.5	46.2	57.8	62.7	63.0
Well identified	79.2	74.3	83.2	83.9	83.4
Well positioned	80.3	74.3	85.2	84.3	83.6

TABLE 4. Descriptive statistics of the positioning criterion C (Eq. 15) obtained when the techniques are applied to series with two shifts.

Statistic	BAMS	BARE	BIVT
Mean	1702	5625	1730
Median	3	4901	2
Standard deviation	2508	2928	2789
Minimum	0	0	0
Maximum	9801	9801	9801

The positioning criterion obtained with all techniques differ significantly (Kruskal-Wallis test, 5% critical level).

TABLE 5. Descriptive statistics of the positioning criterion C (Eq. 15) obtained when the techniques are applied to series with three shifts.

Statistic	BAMS	BARE	BAYE
Mean	2056	7198	2216
Median	2453	6534	2451
Standard deviation	2210	2156	2488
Minimum	0	0	0
Maximum	9801	9801	9801

The positioning criterion obtained with all techniques differ significantly (Kruskal-Wallis test, 5% critical level).

TABLE 6. Distance, elevation difference and correlation between the base series and the neighbor series.

Neighbor series	Distance (km)	Elevation difference (m)	Correlation	Correlation (first difference series)
Shawinigan Falls (7018000)	105	23	0.72	0.89
La Pocatiere (7054095)	119	40	0.53	0.88

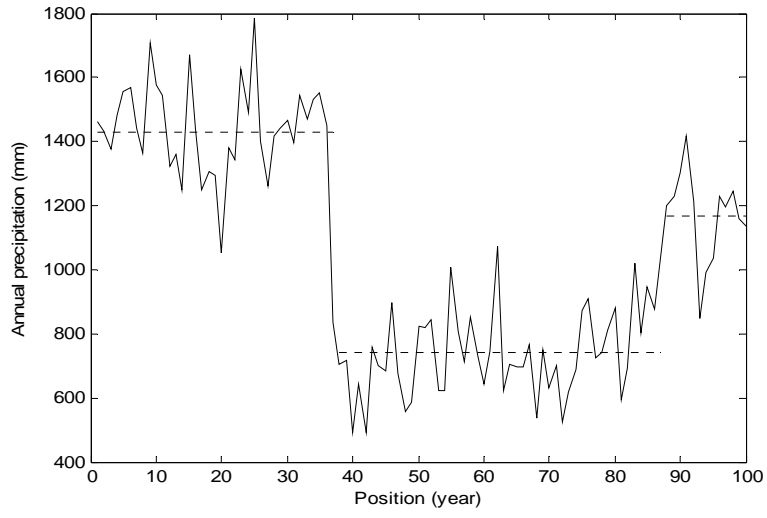


FIG. 1.

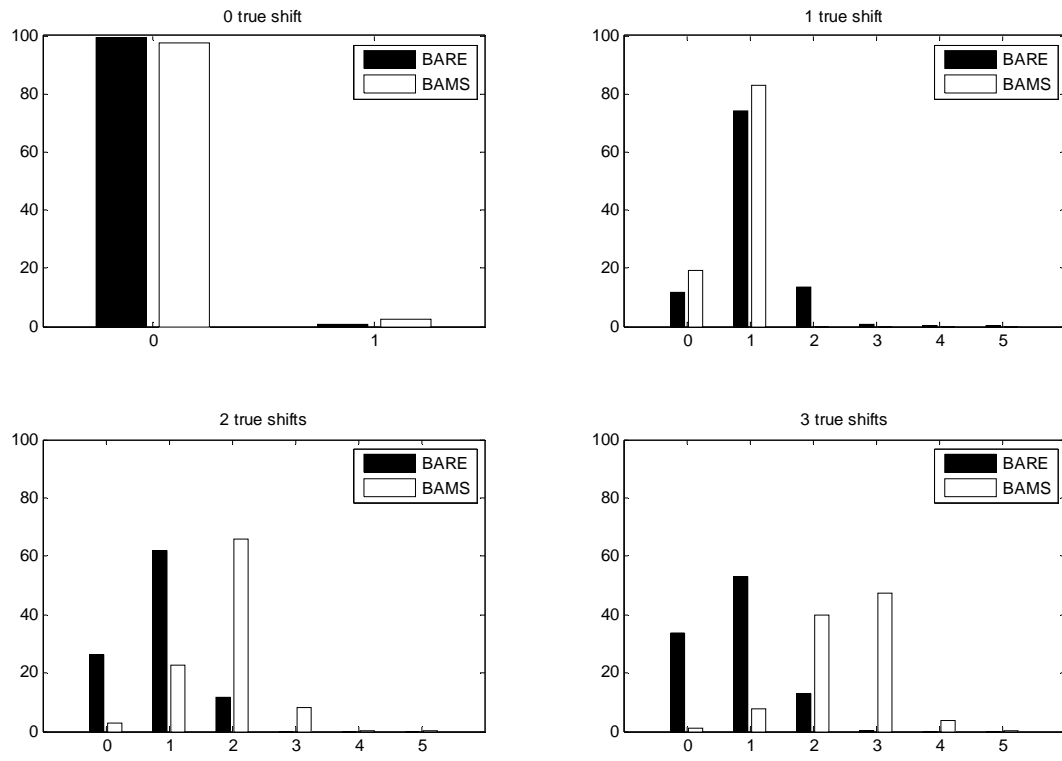


FIG. 2.

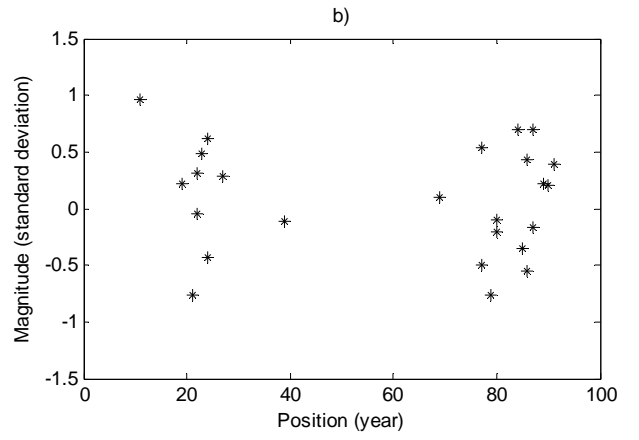
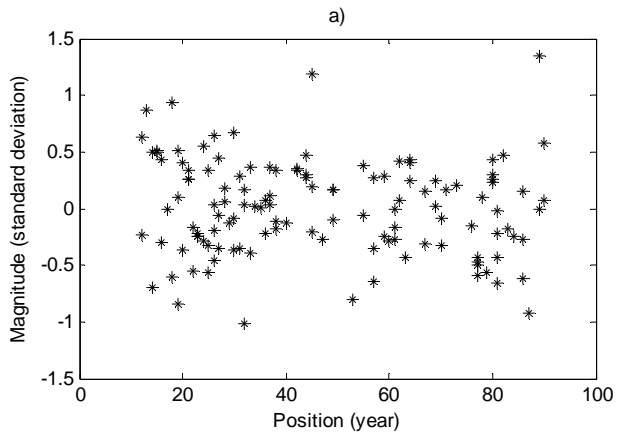


FIG. 3.

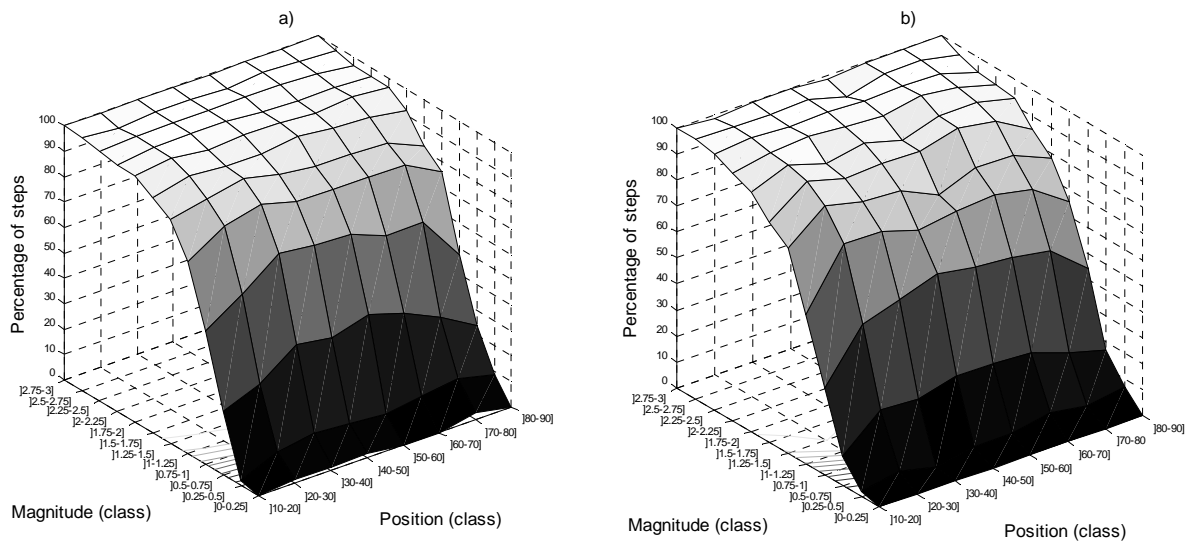


FIG. 4.

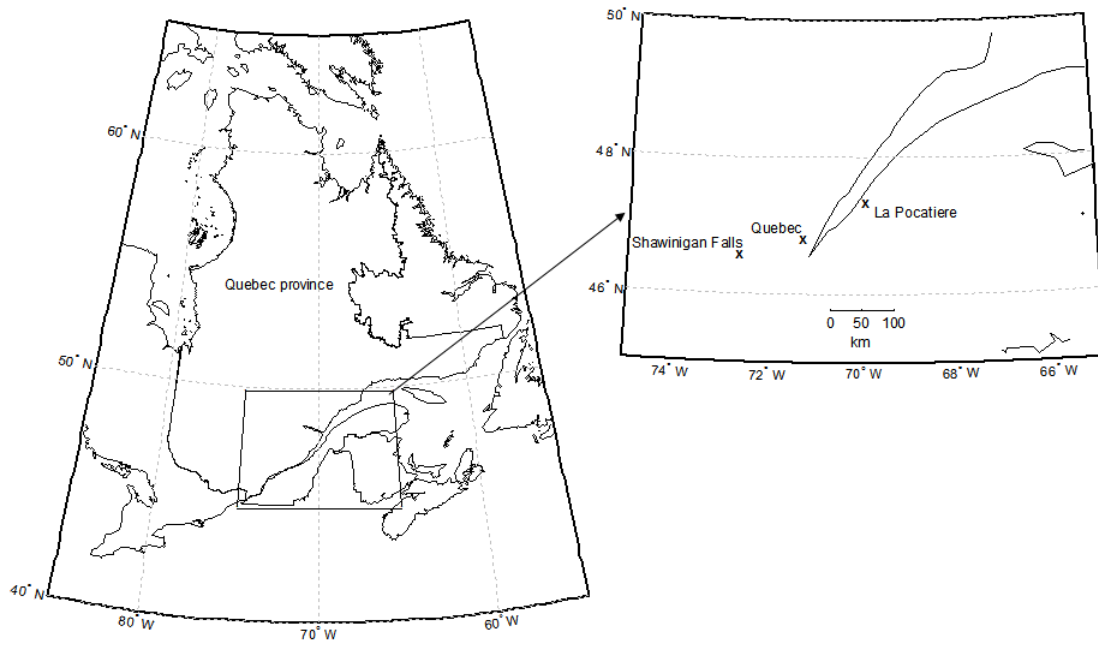


FIG. 5.

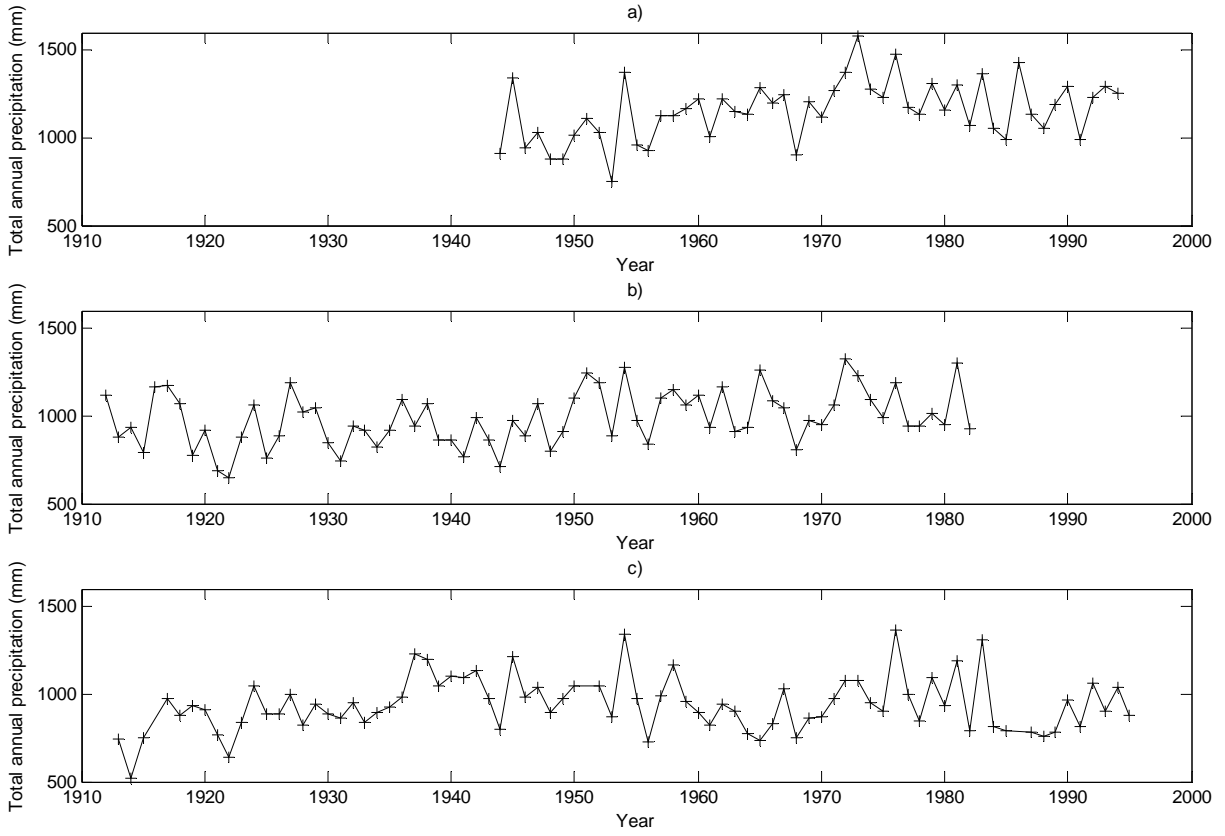


FIG. 6.

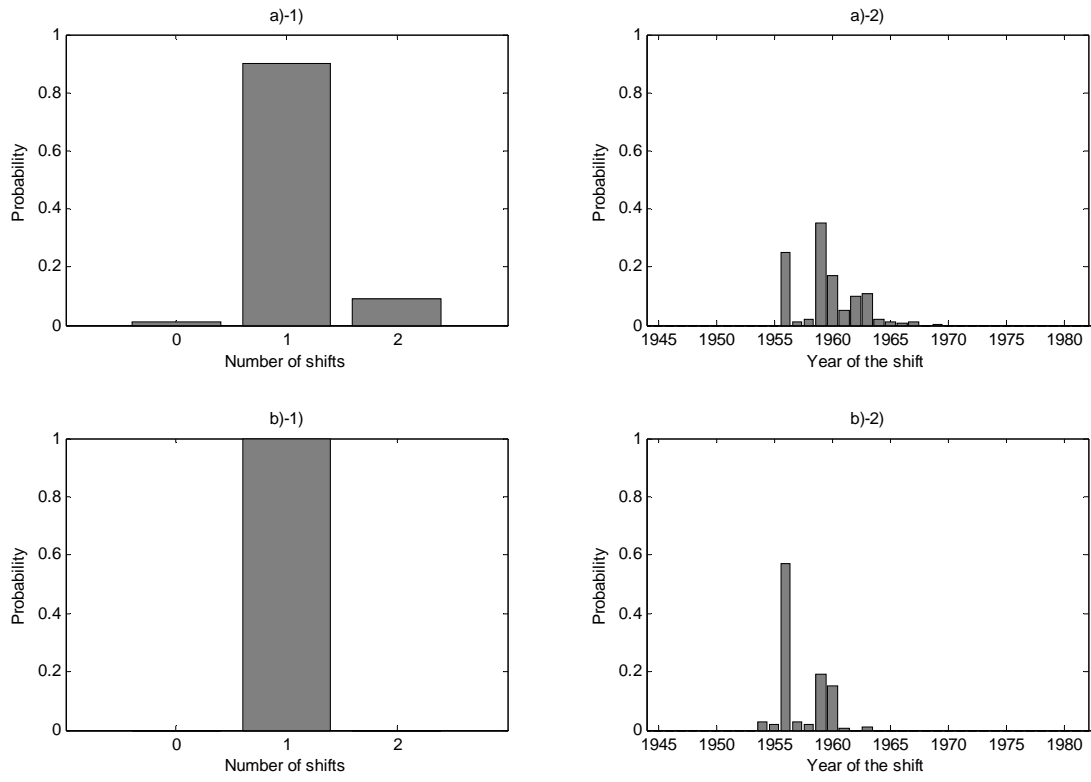


FIG. 7.

9. Article 4: A Bayesian Normal Homogeneity Test for the detection of artificial discontinuities in climatic series

Un test d'homogénéité bayésien pour détecter des inhomogénéités dans des séries climatiques est proposé dans cet article. La capacité du test à identifier des séries homogènes et inhomogènes a été évaluée dans une étude de simulation de Monte Carlo. Le test a été appliqué à des séries synthétiques qui représentent les précipitations totales annuelles typiques observées au centre et au sud de la province de Québec et ses alentours, au Canada. Ce test est très simple à utiliser et permet d'intégrer différentes sources d'information (métadonnées, jugement de l'expert, information régionale) dans l'analyse. La probabilité de non-changement a priori, l'information a priori sur la date du changement et la distribution a priori des données peuvent être spécifiés. La sensibilité de la technique aux paramètres de la distribution a priori de la position du changement et à la probabilité de non-changement ont été étudiés. Le test mène à de faibles pourcentages de fausses détections avec une probabilité de non-changement a priori élevée. Sur les séries avec un saut, le test a donné de très bons pourcentages de détection. Sur les séries avec des sauts multiples, le test a été appliqué avec une procédure de segmentation. Avec une probabilité de non-changement a priori élevée, le test s'est montré efficace pour détecter des sauts multiples. Un cas d'étude sur une série de précipitations totales annuelles est présenté.

**A Bayesian Normal Homogeneity Test for the detection of artificial discontinuities in
climatic series**

Claudie Beaulieu¹, Taha B.M.J. Ouarda¹, Ousmane Seidou²

¹INRS-ETE, University of Quebec, Quebec, QC, Canada

²Department of Civil Engineering, University of Ottawa, Ottawa, ON, Canada

Submitted to International Journal of Climatology the December 19th, 2008

Corresponding author: Claudie Beaulieu, Chair in statistical Hydrology, Canada Research Chair on the Estimation of Hydrological Variables, INRS-ETE, University of Quebec, 490, de la Couronne, Québec (Québec) G1K 9A9, CANADA. e-mail: claudie.beaulieu@ete.inrs.ca; (after March 23, 2009: claubeau@yahoo.com) tel. (418) 654-2530

#4462

Abstract

A Bayesian Normal Homogeneity Test for the detection of artificial discontinuities in climatic series is presented. The test is simple to use and allows the integration of prior knowledge on the date of change from various sources of information (metadata, expert belief, regional information) in the analysis. The performance of the new test was evaluated on synthetic series with similar statistical properties as observed total annual precipitation in the southern and central parts of the province of Quebec, Canada. Different priors were used to investigate the sensitivity of the test to the choice of priors. It was found that a) high prior probability of no-change yield low false detection rates on the homogeneous series; b) the test has a very high power of detection on series with a single shift; c) when applied to series with multiple shifts with a segmentation procedure and a high probability of no-change, the test proved to be very efficient in detecting multiple shifts. A case study is also presented and validated with available metadata.

Keywords: homogenization, climatic series, Bayesian technique, changepoint, exponential family, normal distribution.

1. Introduction

Raw climatic series are likely to exhibit artificial shifts due to modifications in the measurement procedures of the stations (e.g. relocation of a station, change of observer, modification in the immediate environment of the site, etc.). If left uncorrected, these artificial shifts may introduce a bias in climate studies such as trend or extreme value analysis. The problems induced by these artificial shifts in climatic data are the motivation for developing homogenization techniques able to detect and correct these artificial shifts. For a comprehensive review of all these techniques, the reader is referred to Peterson *et al.* (1998), Aguilar *et al.* (2003), Beaulieu *et al.* (2007) and Reeves *et al.* (2007).

Homogenization techniques can detect artificial changes by looking at metadata (history of the station), at relative variations with respect to neighbor series or at both of them. Other approaches such as side by side comparisons of instruments, statistical studies of instrument changes or use of a single station can also be used (Peterson *et al.*, 1998; Aguilar *et al.*, 2003), but are not discussed in this paper. Specific information contained in the metadata can provide the researcher with valuable knowledge of when a discontinuity is likely to have occurred and what may have caused it. Unfortunately, metadata are often incomplete, missing or even sometimes erroneous (Peterson *et al.*, 1998). Therefore, techniques that do not rely entirely on the metadata but can incorporate it in the analysis when available are preferable.

Neighbor series are often used in homogenization to represent the regional climate. They are used to isolate the real regional climate change from the inhomogeneities that are present in the base series (i.e. series to be tested for homogeneity). Homogenization techniques based on a comparison of the base series with neighbor series are presented for example in Potter (1981), Alexandersson (1986), Easterling and Peterson (1995), Vincent (1998), Szentimrey (1999) and Caussinus and Mestre (2004). When using such methods, metadata may be consulted independently of the analysis to identify the potential causes of inhomogeneities. However, an

ideal method would simultaneously use information from both metadata and neighbor series in the analysis to minimize the risk of false detection. Bayesian approaches provide a straightforward framework to combine all available sources of information (metadata, expert knowledge) or lack of information (metadata missing or incomplete) through a prior distribution about the phenomenon being studied. The information provided by the prior distribution is updated with the observations to give a posterior distribution, which is used to make inference about the parameters of the data model (in our case, the changepoint location).

Some Bayesian techniques were applied to solve homogenization problems. The technique presented in Lee and Heghinian (1977) and Perreault *et al.* (1999) was applied in comparison studies to detect discontinuities in synthetic temperature series. This technique performed well to detect a single shift, but tends to combine the sequential shifts in the presence of multiple discontinuities (Ducré-Robitaille *et al.*, 2003; DeGaetano, 2006). Beaulieu *et al.* (2008a) applied the Bayesian technique proposed by Rasmussen (2001) to identify inhomogeneities in synthetic precipitation series. This technique performed well to detect single or multiple shifts, but led to a very high percentage of false detections in homogeneous series. Beaulieu *et al.* (2008b) applied two Bayesian linear regression models to detect inhomogeneities in synthetic precipitation series. The technique proposed by Seidou and Ouarda (2007) was very good to detect multiple shifts, but less efficient in the presence of a single shift. The technique presented in Seidou *et al.* (2007) led to a poor power of detection. Previous Bayesian techniques that were applied to the same problem did not lead to a good overall performance. Furthermore, they were always applied with a non-informative prior distribution, which disadvantages the Bayesian approach, especially when additional information is available.

The general objective of this work is to present a new Bayesian Normal Homogeneity Test to detect homogeneities in climatic series. This technique is simpler than previous Bayesian techniques that were applied to the same problem and easy to use for non-Bayesians. The specific

objectives of this work are 1) to verify if the proposed Bayesian technique is efficient for the homogenization of precipitation series and 2) to study its sensitivity to the choice of the prior distribution. The proposed technique was originally developed by Lee (1998) to detect changes in the parameters of a distribution belonging to the exponential family.

The remainder of this paper is organized as follows: the Bayesian Normal Homogeneity Test is described in section 2, the simulation study is presented in section 3, the results are presented in section 4, a real data series is analyzed in section 5 and finally, a discussion and some conclusions are respectively presented in sections 6 and 7.

2. Description of the technique

2.1. Bayesian approach

Given a prior distribution π which summarizes the information about the parameter θ and a vector of observations \mathbf{x} having a probability density $f(\mathbf{x}|\theta)$, the Bayes theorem allows to actualize $\pi(\theta)$ with the observations:

$$\pi(\theta|\mathbf{x}) = \frac{f(\mathbf{x}|\theta)\pi(\theta)}{\int f(\mathbf{x}|\theta)\pi(\theta)d\theta} \quad (1)$$

The choice of the prior distribution can be either informative or non-informative. The knowledge about the phenomenon can be expressed through an informative prior. The lack of information and the associated uncertainty can be included in a non-informative prior distribution. The estimation of the parameter can be made with different Bayesian parametric estimators such as the posterior mean or the posterior mode. However, the posterior mode is preferred for hypothesis testing (Robert, 1992).

2.2. Bayesian changepoint detection in natural exponential family

The technique used in this work allows the detection of a change in the parameter of a distribution belonging to the natural exponential family (Lee, 1998). This problem has also been addressed in the Bayesian framework by Kander and Zacks (1966), Smith (1975) and Ghorbanzadeh and Lounes (2001). However, the approach of Lee (1998) is the most appropriate for the resolution of the present problem.

A probability distribution is said to belong to the natural exponential family if its probability density function can be written as:

$$f(x|\theta) = \exp\{\theta x + \varphi(\theta) + S(x)\} \quad (2)$$

where $\varphi(\theta)$ and $S(x)$ are two functions which depend on the form of the distribution. Consider the following changepoint model:

$$x_i \sim \begin{cases} F_{\theta_1}, & i = 1, \dots, k \\ F_{\theta_2}, & i = k + 1, \dots, n \end{cases} \quad (3)$$

where the x_i 's are independent variables, F_{θ_1} and F_{θ_2} represent the distributions before and after the shift with parameters θ_1 and θ_2 respectively, k is the unknown position of the changepoint and n is the length of the series. The joint distribution of $\mathbf{x} = (x_1, \dots, x_k, x_{k+1}, \dots, x_n)$ is then:

$$f(\mathbf{x} | \theta) = \exp \left\{ \theta_1 \sum_{i=1}^k x_i + \theta_2 \sum_{i=k+1}^n x_i + k\varphi(\theta_1) + (n-k)\varphi(\theta_2) + \sum_{i=1}^n S(x_i) \right\} \quad (4)$$

The hypothesis that there is no change in the parameter of the distribution ($\theta_1 = \theta_2$) is tested against the hypothesis that there is a change in the parameter of the distribution at time k :

$$\begin{aligned} H_0 &: k = n, \\ H_1 &: 1 \leq k \leq n-1 \end{aligned} \quad (5)$$

In Lee (1998), the prior probability distribution for k is given by:

$$g_0(k) = \begin{cases} p, & k = n \\ \frac{1-p}{n-1}, & k \neq n \end{cases} \quad (6)$$

where p represents the prior probability of no-change ($0 \leq p \leq 1$). Hence, $k = n$ represents the case where no-change has occurred in the series, while $k \neq n$ represents the case where a change has occurred at position k . The prior probability for the position of the change is uniformly distributed over $[1, \dots, n-1]$. A conjugate prior is used for the parameters θ_1 and θ_2 . For the exponential family, the conjugate prior has the form:

$$\pi(\theta | m, c) \propto \exp \{ mc\theta + m\varphi(\theta) \} \quad (7)$$

where m and c are hyperparameters such that $m \geq 0$ and $c \in X$. X represents the sample space of the x_i 's. Lee (1998) set the prior information about θ_1 and θ_2 to take into account the position of the changepoint:

$$\begin{aligned}\pi(\theta_1 | k) &= \pi(\theta_1 | m_{1k}, c_{1k}), \quad k = n \\ \pi(\theta_1, \theta_2 | k) &= \prod_{i=1}^2 \pi(\theta_i | m_{ik}, c_{ik}), \quad k \neq n\end{aligned}\tag{8}$$

where $\pi(\theta_i | m_{ik}, c_{ik}) \propto \exp\{m_{ik}c_{ik}\theta_i + m_{ik}\varphi(\theta_i)\}$, $i = 1, 2$, $m_{ik} \geq 0$, $c_{ik} \in X$. The joint prior distribution of (k, θ_1, θ_2) is given by:

$$\begin{aligned}\pi(k, \theta_1) &= p \frac{\exp\{m_{1k}c_{1k}\theta_1 + m_{1k}\varphi(\theta_1)\}}{\tilde{\pi}(\Theta | m_{1k}, c_{1k})}, \quad k = n \\ \pi(k, \theta_1, \theta_2) &= \frac{1-p}{n-1} \frac{\exp\left\{\sum_{i=1}^2 [m_{ik}c_{ik}\theta_i + m_{ik}\varphi(\theta_i)]\right\}}{\prod_{i=1}^2 \tilde{\pi}(\Theta | m_{ik}, c_{ik})}, \quad k \neq n\end{aligned}\tag{9}$$

where

$$\tilde{\pi}(\Theta | m_{ik}, c_{ik}) = \int_{\Theta} \exp\{m_{ik}c_{ik}\theta_i + m_{ik}\varphi(\theta_i)\} d\theta_i, \quad i = 1, 2\tag{10}$$

and Θ represents the parameter space of θ_1 and θ_2 . According to the Bayes theorem (Eq. 1), and from the joint distribution of \mathbf{x} (Eq. 4) and the joint prior distribution (Eq. 9), the posterior distribution of the position of the changepoint k is:

$$\pi(k | \mathbf{x}) \propto \begin{cases} p \frac{\tilde{\pi}\left(\Theta | m_{1k} + n, \left(m_{1k}c_{1k} + \sum_{i=1}^n x_i\right) / m_{1k} + n\right)}{\tilde{\pi}(\Theta | m_{1k}, c_{1k})}, & k = n \\ \frac{1-p}{n-1} \frac{\tilde{\pi}\left(\Theta | m_{1k} + k, \left(m_{1k}c_{1k} + \sum_{i=1}^k x_i\right) / m_{1k} + k\right)}{\tilde{\pi}(\Theta | m_{1k}, c_{1k})} \\ \times \frac{\tilde{\pi}\left(\Theta | m_{2k} + (n-k), \left(m_{2k}c_{2k} + \sum_{i=k+1}^n x_i\right) / m_{2k} + (n-k)\right)}{\tilde{\pi}(\Theta | m_{2k}, c_{2k})} \end{cases}, \quad k \neq n \quad (11)$$

The values of the hyperparameters m_{ik} can be thought of as the sample size of priors. Then they can be set such that $m_{1k} = k$ and $m_{2k} = n - k$, which corresponds to the size of the two samples (x_1, \dots, x_k) and (x_{k+1}, \dots, x_n) (Lee, 1998). To estimate the hyperparameters c_{ik} , Lee (1998) proposed to use the maximum likelihood type II (ML-II) approach for prior selection, such as presented in Berger (1985). It consists in performing a maximization of the marginal distribution of x , $m(\mathbf{x} | \pi)$, over the hyperparameters c_{ik} , $i = 1, 2$, $k = 1, \dots, n$. When $m_{1k} = k$ and $m_{2k} = n - k$, to maximize $m(\mathbf{x} | \pi)$ over π is equivalent to maximizing the two following functions on c_{1k} and c_{2k} (Lee, 1998):

$$\frac{\tilde{\pi}(\Theta | 2k, (c_{1k} + \bar{x}_k)/2)}{\tilde{\pi}(\Theta | k, c_{1k})} \text{ or } \frac{\tilde{\pi}(\Theta | 2(n-k), (c_{2k} + \bar{x}_k^*)/2)}{\tilde{\pi}(\Theta | n-k, c_{2k})} \quad (12)$$

where $\bar{x}_k = \sum_{i=1}^k x_i / k$ and $\bar{x}_k^* = \sum_{i=k+1}^n x_i / (n-k)$. The logarithm of these functions gives an expression

which is similar to the function:

$$h(s) = \log\left(\tilde{\pi}(\Theta | 2k, (c + s)/2)\right) - \log\left(\tilde{\pi}(\Theta | k, s)\right) \quad (13)$$

c is \bar{x}_k or \bar{x}_k^* and s represents the hyperparameters to maximize. By taking the derivative function, the maximum values can be obtained. For the normal case, $h(s)$ is concave downward and the ML-II hyperparameters are easily obtained.

2.3. Bayesian Normal Homogeneity Test (BNHT)

For the normal case, the test enables the detection of a change in the mean of a normal series. The changepoint model can be represented by:

$$x_i \sim \begin{cases} N(\mu_1, \sigma^2), & i = 1, \dots, k \\ N(\mu_2, \sigma^2), & i = k + 1, \dots, n \end{cases} \quad (14)$$

where μ_1 and μ_2 are the mean of the series before and after the shift and σ^2 is the variance of the series that is assumed to be known. The observations must be independent.

One may know the exact interval where a change could have occurred in the series. Furthermore, one may be aware of the higher risk of false detection in the extremities of the series, pointed out notably in Wang *et al.* (2006). Hence, the following prior probability distribution for k is proposed instead of the original one presented in Eq. (6):

$$g_1(k) = \begin{cases} p, & k = n \\ \frac{1-p}{n-a-b-1}, & k = a+1, \dots, n-b-1 \\ 0, & \text{otherwise} \end{cases} \quad (15)$$

where a and b represent the support of the distribution. When a and b are set to 0, Eq. (15) is equivalent to Eq. (6). Otherwise, it is impossible to detect a shift in the a first observations and in the b last observations. A similar prior distribution was used by Holbert (1982) for changepoint detection in a multiple linear regression model.

Conjugate priors are used for the means before and after the changepoint and have a $N(c/\sigma^2, 1/m\sigma^2)$ distribution. Lee (1998) found that the ML-II hyperparameters c and m are

$\hat{c}_{1k} = \bar{x}_n$ and $\hat{m}_{1k} = n$ for the case $k = n$. When $k \neq n$, the ML-II hyperparameters are respectively $\hat{c}_{1k} = \bar{x}_k$, $\hat{m}_{1k} = k$, $\hat{c}_{2k} = \bar{x}_k^*$ and $\hat{m}_{2k} = n - k$ for the means before and after the shift. With $g_1(k)$ as prior distribution for the position of the change, the posterior distribution of k given \mathbf{x} is:

$$\pi(k | \mathbf{x}) \propto \begin{cases} p \cdot \exp \left\{ - \sum_{i=1}^n (x_i - \bar{x}_n)^2 / 2\sigma^2 \right\}, & k = n \\ \frac{1-p}{n-a-b-1} \cdot \exp \left\{ - \left[\sum_{i=1}^k (x_i - \bar{x}_k)^2 + \sum_{i=k+1}^n (x_i - \bar{x}_k^*)^2 \right] / 2\sigma^2 \right\}, & k = a+1, \dots, n-b-1 \\ 0, & \text{otherwise} \end{cases} \quad (16)$$

The variance can be estimated by the pooled sample variances (Lee, 1998):

$$\hat{\sigma}_p^2 = \min_{1 \leq k < n} \left[\sum_{i=1}^k (x_i - \bar{x}_k) + \sum_{i=k+1}^n (x_i - \bar{x}_k^*) \right] / (n-2) \quad (17)$$

The use of the pooled sample variance has also been suggested by Reeves *et al.* (2007) to improve the Standard Normal Homogeneity Test (Alexandersson, 1986). The motivation of using the pooled variance is that in the presence of a changepoint, the overall sample variance $\hat{\sigma}^2$ is biased and is an inconsistent estimator of σ^2 , and in which case, the variance should be estimated by $\hat{\sigma}_p^2$.

The position of the shift can be estimated from the posterior distribution by taking the highest posterior probability (posterior mode). A credibility interval of the position can be computed as well. The BNHT can be applied to series of ratios/differences between the base series and the neighbor series to avoid the detection of a regional climatic change.

An alternative prior distribution that could be used for the position of the change is the triangular distribution:

$$g_2(k) = \begin{cases} p, & k = n \\ \frac{2(1-p)(k-a)}{(b-a)(d-a)}, & a \leq k \leq d \\ \frac{2(1-p)(b-k)}{(b-a)(b-d)}, & d \leq k \leq b \\ 0, & \text{otherwise} \end{cases} \quad (18)$$

where d represents the most probable position of the shift (the mode of the distribution). This distribution could be used instead of the uniform in the case where the exact date of change is known. This would be more informative. In this case, the posterior distribution of k given \mathbf{x} would be:

$$\pi(k | \mathbf{x}) \propto \begin{cases} p \cdot \exp\left\{-\sum_{i=1}^n (x_i - \bar{x}_n)^2 / 2\sigma^2\right\}, & k = n \\ \frac{2(1-p)(k-a)}{(b-a)(d-a)} \cdot \exp\left\{-\sum_{i=1}^k (x_i - \bar{x}_k)^2 + \sum_{i=k+1}^n (x_i - \bar{x}_k^*)^2 / 2\sigma^2\right\}, & a \leq k \leq d \\ \frac{2(1-p)(b-k)}{(b-a)(b-d)} \cdot \exp\left\{-\sum_{i=1}^k (x_i - \bar{x}_k)^2 + \sum_{i=k+1}^n (x_i - \bar{x}_k^*)^2 / 2\sigma^2\right\}, & d \leq k \leq b \\ 0, & \text{otherwise} \end{cases} \quad (19)$$

3. Simulation study

3.1. Synthetic series

For comparison purposes, the BNHT was applied to the same synthetic series of precipitation that were generated for previous comparative studies of several homogenization techniques (Beaulieu *et al.*, 2008a, 2008b). To estimate the statistical properties to reproduce with the model, a set of stations located in central Quebec and surroundings having long time series with few missing data were selected and their average characteristics were computed. The statistical properties reproduced in these series are a) a mean total annual precipitation of 1089 mm, b) a standard deviation of 142 mm and c) a lag one autocorrelation of 0.02. Four different data sets were generated to represent the different cases that could occur in the reality: homogeneous series, series with a single shift, and series with two and three shifts. The magnitude and positions of the random shifts were generated to represent as many configurations as possible. The positions are drawn from a truncated discrete uniform distribution $[11, n-10]$. The magnitudes are generated from a Beta distribution and lie between -3 and 3 standard deviations. For each base series, three correlated neighbor series were generated to reproduce a spatial cross-correlation of 0.55. This value is the mean spatial cross-correlation in the set of selected stations that are located at a distance less than 300 km. The detailed generation scheme of the synthetic series is presented in Beaulieu *et al.* (2008a).

3.2. Application of the technique

The BNHT was applied to the series of ratios between the base series and the mean of the neighbor series. As all neighbor series have the same correlation with the base series, it is normal that they all have the same weight in the series of ratios. The BNHT was applied to all data sets by varying the parameters of the prior distribution of the position to study the sensitivity of the technique to the choice of prior. As it is a simulation study, non-informative priors for the

position (Eq. 15) are used. For all data sets, the posterior distributions were computed with different prior probabilities of no-change p : 0.01, 0.05, 0.10, 0.25, 0.5, 0.75, 0.90, 0.95 and 0.99. The parameters a and b , used to set the interval of detection, were set to different values according to the type of data set. For the homogenous series, a and b were set to 0, 5, 10, 15 and 25, to allow the detection of a change in at least half of the series length. For the other series, a and b were set to 0, 5 and 10 only, as no shifts were introduced in the first ten or last ten observations in the series. In Beaulieu *et al.* (2008a, 2008b), the techniques were applied by ignoring the shifts detected in the first or last ten observations. For the series with multiple shifts, the techniques were applied with a minimum length between two consecutive shifts set to ten observations. To be able to compare the results obtained with the BNHT with those obtained in Beaulieu *et al.* (2008a, 2008b), the BNHT was applied using a and b set to 0.

3.3. Performance evaluation

For the homogeneous series, the performance is assessed by computing the false detection rate, which is the percentage of cases for which the null hypothesis (homogeneity) is rejected, while it is true. For the series with a single shift, the number of well positioned shifts (i.e. located within ± 2 years of the true position) is computed. For the series with multiple shifts, the ability of the technique to correctly position all the shifts, without omission or false detection is evaluated with a performance criterion proposed by Beaulieu *et al.* (2008a):

$$C = \begin{cases} \frac{1}{nd} \sum_{i=1}^{nd} (p_i^d - p_i)^2, & nr = nd \\ \frac{1}{nr} \left[\sum_{i=1}^{nd} (p_i^d - p_i)^2 + |nr - nd|(n-1)^2 \right], & nr > nd \\ \frac{1}{nd} \left[\sum_{j=1}^{nr} (p_j^d - p_j)^2 + |nr - nd|(n-1)^2 \right], & nr < nd \end{cases} \quad (20)$$

where $p_i^d, i = 1, \dots, nd$ and $p_j, j = 1, \dots, nr$ represent respectively the positions of the detected and real shifts, nd is the number of detected shifts, nr is the number of real shifts, and n is the length of the series. The pairs (p_i^d, p_j) are chosen to minimize the criterion. When the exact number of shifts is detected ($nr = nd$), the criterion is the sum of squares of the differences between the pairs which minimize the criterion. When $nr < nd$ or $nr > nd$, $(n-1)^2$ is added each time that a true shift is omitted or that a false shift is detected. This value is the square of the maximum possible distance between two shifts. When all shifts are correctly positioned ($p_i^d = p_i, i = 1, \dots, nd = nr$), C is equal to zero. When the detected shifts are located near the true positions ($p_i^d \approx p_i, i = 1, \dots, nd = nr$), C is close to zero. When some shifts in the series are not detected or are falsely detected ($nd \neq nr$), C is high. The performance criterion C was computed for all series with two and three shifts and the overall performance is the mean of the criterion over each set of synthetic series.

4. Results

4.1. False detection rates in the homogeneous series

Table 1 presents the false detection rates obtained with different prior parameters. It can be seen that the technique is very sensitive to the prior probability of no-change. Hence, with a very low prior probability of no-change (0.01), a false detection occurs in about 100% of the synthetic series. With a very high prior probability of no-change (0.99), a false detection rarely occurs (0.1-0.2% of false detection). With a less informative prior probability of no-change (0.5), the false detection rates lie between 6 and 8 %. The technique seems less sensitive to the interval of detection. When the shifts detected in the extremities are ignored, the percentage of false detections is under 5% with a prior probability of no-change between 0.5 and 0.99. Figure 1 presents the histograms of the positions of the falsely detected shifts when the prior probability of no-change is fixed to 0.25. This figure shows that false detections rates display a typical U-shape. When a and b are set to 0, false detections at the extremities increase. This effect is reduced when the interval of detection is decreased (when a and b are increased). Therefore, using a null prior distribution in areas close to the ends is useful in reducing the number of false detections at the extremities of the series, but does not completely remove the U-shape effect.

4.2. Series with a single shift

Table 2 presents the percentage of well positioned shifts in the series with a single shift. With a low prior probability of no-change, the percentage of well positioned shifts is around 93%. With a very high prior probability of no-change, the percentage of well positioned shifts is around 85%. When the shifts detected in the extremities are ignored, the percentage of well positioned shifts slightly diminishes. Figure 2 presents the percentage of well positioned shifts according to their position and magnitude when the parameters a and b are set to 0. As the results are very similar for other values of a and b , it was not judged necessary to present these

figures. The percentage of detection increases along with the magnitude of the shift. Shifts having a high magnitude (i.e more than one standard deviation) are almost always detected. The percentage of detection starts to decrease around a magnitude of about 1 standard deviation. Figure 2 also shows that a low prior probability of no-change (0.01) allows the detection of shifts having a small magnitude that would not be detected with a higher probability of no-change (0.99). The position of the shift does not seem to have a high impact on the power of detection.

4.3. Series with multiple shifts

Tables 3 and 4 presents the descriptive statistics of the positioning criterion (C) computed from Eq. 20 for the series with two and three shifts. Figure 3 presents the mean positioning criterion (C) for the series with two and three shifts. The positioning criterion seems to depend on the prior probability of no-change and on the way detections at the extremities are handled (Friedman's test, 5% critical level). With a minimum segment length of ten observations, the mean positioning criterion reaches its minimum. When the prior probability of no-change is small ($p < 0.5$), the mean positioning criterion is high. This is due to a large number of false detections. Hence, when p is small, the mean number of shifts detected is higher than the real number of shifts (tables 5 and 6). Furthermore, the risk of false detection increases as the test is reapplied several times to the same series. When the prior probability of no-change is around 0.5-0.75, the mean positioning criterion is the smallest (figure 3). With a high prior probability of no-change ($p > 0.9$), the mean positioning criterion seems to increase as p increases. This is due to undetected shifts. Hence, when p is high, the mean number of detected shifts is smaller than the real number of shifts (tables 5 and 6). The prior probability of no-change has less impact when the BNHT is applied with a minimum length of ten observations between two shifts. Finally, it

seems that it is better to set either a high prior probability of no-change (0.5-0.95) and a and b to ten, or to set a minimum length of ten observations between two shifts to detect multiple shifts.

4.4. Previous results

Previous comparative studies of several homogenization techniques on the same synthetic series have been presented in Beaulieu *et al.* (2008a, 2008b). In these two previous studies, the shifts detected in the extremities were ignored and a minimum of ten observations was imposed between two consecutive shifts. To be able to compare the BNHT with the other techniques, it was applied using the same procedure. Furthermore, only the results obtained with a prior probability of no-change of 0.5 are compared, as it corresponds to the non-informative case. The other Bayesian techniques compared previously were also applied with non-informative priors.

For the homogeneous series, the false detection rate is significantly smaller than 5%, as with the other good techniques presented in Beaulieu *et al.* (2008a, 2008b). For the series with a single shift, the percentage of shifts positioned is 89.2% with a prior probability of no-change of 0.5 and by ignoring the shifts detected in the extremities. The highest percentage of well positioned shifts in Beaulieu *et al.* (2008a, 2008b) was 85.2%. By testing the equality of these two proportions, it is found that the percentage of well positioned shifts is significantly higher with the BNHT (5% critical level). In Beaulieu *et al.* (2008a, 2008b), the best results in the series with multiple shifts were obtained with the Bayesian approach for multiple shifts (BAMS) of Seidou and Ouarda (2007). These results were a mean C of 1702 and a median C of 3 in the series with two shifts, and, in the series with three shifts, a mean C of 2056 and a median C of 2453. The median C obtained with the BNHT are not significantly different than those obtained with BAMS (Wilcoxon test, 5% critical level).

5. Case study

An application of the BNHT to a real case study is presented in this section. The total annual precipitation series of Quebec City (station 7016294), Quebec, Canada, is tested for homogeneity. This station was chosen because it has no missing values and a long series of observations, and also because neighbor series and metadata are available for this station. The observations and the metadata were provided by Environment Canada. The same case was studied previously in Beaulieu *et al.* (2008b) with other techniques.

The base station is located at latitude 46.8, longitude -71.38 and at an altitude of 70 meters. Two neighbor stations, Shawinigan Falls (7018000) and La Pocatiere (7054095), were identified in the set of 35 stations by considering the distance from the base station, the elevation difference, the observation period, the correlations and the correlations computed from the first difference series. In the metadata, there are no potential causes of inhomogeneities documented for the two neighbor stations. The analysis was performed for the common period of observations: 1944-1982. For this period, the correlations between Quebec-Shawinigan Falls and Quebec-La Pocatiere are respectively: 0.72 and 0.53. However, the correlations computed from the first difference series are higher: 0.89 and 0.88 for Quebec-Shawinigan Falls and Quebec-La Pocatiere respectively. This indicates that there may be a shift in the base series which affects the correlation (Vincent, 1998). Figure 4 presents the data series of the base and the two neighbor stations. For the Quebec station, there are many documented relocations; two of which represent changes in elevation (1958 and 1977). For precipitation series, a small change in the instrument height can induce a very important shift in a series (Heino, 1997). Hence, these two relocations could have introduced artificial shifts in the Quebec precipitation series.

The BNHT was applied to the series of ratios between the base series and the mean of the two neighbor series. The uniform prior distribution was used for the position (Eq. 15). Once again, different priors were used to verify the sensitivity of the technique to the choice of the

prior parameters. Figure 5 presents the posterior probability distributions obtained. A change is detected in 1958 in all cases. The 95% Bayesian credibility interval is given by the years {1958,1959,1960,1961,1962}. The change detected in 1958 corresponds to a relocation documented in the metadata. The equality of all posterior densities indicates that the test is robust to the choice of prior. Even with a very high prior probability of no-change (0.99), the shift is detected. The triangular prior distribution (Eq. 18) is also used by setting the mode at position 14 (1958) and by varying the other prior parameters. The posterior distributions remain almost exactly the same than with the uniform prior. Thus, it was not judged necessary to present these results as well in figure 5. In Beaulieu *et al.* (2008b), different homogenization techniques were applied to the same case study and also found a single significant shift in 1958. The observations of 1944-1958 could be corrected by multiplying them by the correction factor: \bar{q}_a/\bar{q}_b , where \bar{q}_a and \bar{q}_b are respectively the mean of the series of ratios ($q_i, i = 1, \dots, n$) after and before the change. The advantages of the proposed approach are its simplicity, the possibility to include different sources of information and the comprehensive interpretation of the prior and posterior distributions.

The Shapiro-Wilk and Wald-Wolfowitz tests were applied to the series of ratios to verify their normality and independence, as the technique relies on these hypotheses. Both tests are rejected at the 5% critical level. The normality test is probably rejected due to the presence of a shift in the mean of the series, which affects the symmetry of the distribution (figure 6). The Wald-Wolfowitz statistic checks the randomness hypothesis for a two-valued data sequence. In the presence of an important shift, the sequence is not random, as the smallest values are concentrated in one segment and the highest values are concentrated in the other segment. The usual statistical tests can only be performed as an indicator on both sides of the detected changepoint, after the changepoint test is applied (Perreault *et al.*, 1999). These two tests were re-

applied to the two segments of the series of ratios, separated after 1958, and the normality and independence hypothesis are respected (5% critical level). Furthermore, the equality of the variances before and after the shift was tested. The two variances do not seem different (Fisher test, 5% critical level).

The other documented change in 1977 could also have affected the homogeneity of the series. Furthermore, an undocumented change could have introduced an artificial shift in the series. Hence, the technique was re-applied to the two segments of the series with the uniform prior. As the two segments of the series are short, the parameters of the interval of detection were set to 0. In the second segment, the triangular prior distribution was also used by setting the mode at position 19 (1977). Figure 7 presents the posterior probabilities for the position of the change. The first column in figure 7 represents the results of the changepoint analysis performed on the first segment (before 1958). A change is detected in 1945. As it is the second observation in the series, it is reasonable to interpret it as a false detection in an extremity. The second column in figure 7 presents the posterior probability distributions for the last segment of the series. In this case, the metadata indicates that an inhomogeneity could have occurred in 1977. However, the posterior distribution of the position of the shift does not exhibit a peak in 1977, even if we force the detection with the triangular distribution. When the prior probability of no-change is equal to or higher than 0.25, the last segment of the series is considered homogeneous. Then, it seems that the change documented in 1977 did not introduce any artificial shift in the series.

6. Discussion

6.1. Use of metadata and neighbour series

Homogenization techniques presented in the literature rely on either the metadata or on a comparison with neighbor series to perform the analysis. When using relative homogenization techniques (based on neighbor series), the metadata is often consulted after a change is detected to determine its cause. Some techniques use metadata to identify which segment of the series will be tested (Peterson *et al.*, 1998). The BNHT allows a more formal use of the metadata, by incorporating it into the prior distribution. This is an advantage of the BNHT over the other homogenization techniques. It is also possible to apply the BNHT without neighbor stations. It should however be acknowledged that the use of a neighbor series that is inhomogeneous, could alter the detection and correction of a shift (Menne and Williams, 2005; Reeves *et al.*, 2007).

6.2. Priors and subjectivity

The prior distribution for the position of the shift used in the simulation study was uniform. The advantage is that it can accommodate both the case for which the position of the potential shift is known and the case for which it is not. However, it was shown in the case study that other prior distributions could be relevant. For example, a triangular distribution could be used to give the highest prior probability to the position for which a change is recorded in the metadata. Furthermore, the distributions for the means are conjugate. The hyperparameters are chosen according to the maximum likelihood type-II approach for prior selection. Once again, other approaches for prior selection could be used and could lead to different results. This is a common criticism of the Bayesian approach: it contains a certain amount of subjectivity. Hence, the specification of the prior distribution and of the likelihood function, entails a succession of subjective judgments (Goldstein, 2006). However, even a classical use of statistics can also end

up being subjective as the choice of the technique to use for a given problem, as well as the choice of the neighbor series, heavily depend on the modeler and can lead to different results.

6.3. Multiple shifts

The BNHT was developed for at most one changepoint, as it is the case for most homogenization techniques. Nevertheless, it was applied to multiple shifts using a segmentation procedure, as it is often done in practice. The test could be extended to detect multiple shifts in a more fashionable way. For example, a hierarchy of sub-hypotheses as presented in Chen and Gupta (2001) could be constructed. However, the detection of multiple inhomogeneities in precipitation series is usually more difficult than in temperature series, because of their larger spatial and temporal variability. For example, Müller-Westermeier *et al.* (2004) performed a homogeneity analysis of temperature and precipitation series in Germany for each month of the year. They rarely detected more than one inhomogeneity in precipitation series.

6.4. U-shape

In Wang *et al.* (2007), it was shown that the false detection rates and the probabilities of detection according to the shift positions generally have a U-shape. It was shown in figure 1 that BNHT tends to detect more shifts at the extremities of the series. The parameters of the interval of detection help to reduce this effect, but do not eliminate it entirely. This has to be taken into consideration when applying this technique. It was shown that false detections are reduced, when the shifts detected in the extremities are ignored. Nevertheless, the test could be modified by introducing a penalty in the likelihood function.

6.5. Normality and independence hypotheses

The BNHT was developed for independent series. The changepoint detection techniques which were developed for independent error series tend to detect more false shifts in the presence of autocorrelation (Lund *et al.*, 2007). The effect of autocorrelation on the BNHT is not known yet. Application of the test to series with different autocorrelation values would be necessary to quantify the effect of autocorrelation on the false detection rates. However, for annual series such as those used in this work, the independence assumption is realistic. Therefore, it was not judged necessary to study the effect of the autocorrelation on this technique. Furthermore, the normal case was used as this distribution is reasonable to represent total annual precipitation and also commonly used to represent temperature series.

7. Conclusions

7.1. Summary of results

The BNHT allows for the detection of a change in the mean of a normal series. The performance of this test and its sensitivity to the choice of the prior parameters were assessed in a simulation study. For the homogeneous series, the test was very sensitive to the choice of the prior probability of no-change. The percentages of false detections were small (<5%) when the prior probability of no-change was set to at least 0.75, or to at least 0.5 when the shifts detected in the extremities are ignored. For the series with a single shift, the technique was more robust to the choice of the prior parameters and had a high power of detection in all cases. It gave better results than those obtained in Beaulieu *et al.* (2008a, 2008b). Furthermore, this technique is able to detect a shift with a small magnitude, especially when the prior probability of no-change is small. For the series with multiple shifts, the technique gave positioning criteria equivalent to the best techniques previously compared. The BNHT was applied to synthetic series of precipitation, but it could be used to detect inhomogeneities or a change in the mean in other climatic series that are normally distributed (e.g. temperature).

A case study was also presented to illustrate the application of the BNHT to a real data series. Different priors were used to verify their effect on the posterior density of the shift's position. A change documented in the metadata in 1958 was detected with all different priors. The same shift was also detected by other homogenization techniques applied to the same case study in Beaulieu *et al.* (2008b).

7.2. Recommendations

According to the results of the simulation study and the case study, some recommendations can be made for further use of the BNHT. If there is no information about a potential change in the metadata (i.e. either that nothing could have affected the homogeneity at

this station or that a change is not documented), then, the BNHT should be applied with a high prior probability of no-change (≥ 0.75) to minimize the risk of false detection. The parameters of the interval of detection should then be set to avoid false detection at the extremities of the series. Hence, even with a high prior probability of no-change, this technique still has a high power of detection. If there is a documented change in the metadata that could affect the homogeneity of the series, the prior probability of no-change could be set low and the parameters of the interval of detection could concentrate the probabilities around the date of the potential shift. This would allow the detection of the shifts having a small magnitude that would not be detected with a high prior probability of no-change. In the case study, it was shown that a changepoint can affect the results of the normality and independence tests. Then, these hypotheses should be verified in each segment after the detection of a shift.

7.3. Future work

The general formulation of the technique as presented in Lee (1998) allows the detection of a change in the parameters of a distribution belonging to the exponential family. The Gamma distribution also belongs to the exponential family and is a more natural choice to represent variables with asymmetric distributions such as monthly or seasonal precipitation. Future work should focus on the application of the technique to detect inhomogeneities in precipitation series based on the Gamma distribution. Such test would be useful to detect changes in the parameter of the distribution (belonging to the exponential family) of a larger range of climatic series. For example, it could be used to detect changes in the intensity of tornado counts (Poisson distributed), in wind speed (Gamma), in cloudiness (Beta) or in climatic extremes (Weibull). The posterior densities for other distributions should constitute the subject of future work.

Acknowledgements

The authors wish to thank the Fonds Québécois de Recherche sur la Nature et les Technologies (FQRNT), the National Sciences and Engineering Research Council of Canada (NSERC) and the Canada Research Chair Program for funding this research. The authors would also like to thank Mrs L. Vincent of the Meteorological Service of Canada for providing the metadata.

Notations

Θ	parameter space for θ_1 and θ_2
θ	parameter of a distribution
θ_1	parameter of a distribution before the shift
θ_2	parameter of a distribution after the shift
μ_1	mean of the series before the shift
μ_2	mean of the series after the shift
$\pi(\theta)$	prior distribution of the parameter θ
$\pi(\theta_1 k)$	prior distribution of the parameter θ_1 given the position of the shift (under the no-change hypothesis)
$\pi(\theta_1, \theta_2 k)$	prior distribution of the parameters θ_1 and θ_2 given the position of the shift (under the change hypothesis)
$\pi(\theta m, c)$	prior distribution of the parameter θ with the hyperparameters m and c
$\pi(k \mathbf{x})$	posterior distribution of k given the observations \mathbf{x}
σ^2	variance
$\hat{\sigma}^2$	sample variance
$\hat{\sigma}_p^2$	pooled sample variance
$\varphi(\theta)$	function of θ in the exponential family form
a	prior parameter indicating the position of the beginning of the interval of detection
b	prior parameter indicating the position of the end of the interval of detection
C	positioning criterion
c	hyperparameter for the prior distribution of the mean
d	prior parameter indicating the mode of the prior triangular distribution
\hat{c}_{1k}	ML-II hyperparameter for the prior distribution of the mean μ_1
\hat{c}_{2k}	ML-II hyperparameter for the prior distribution of the mean μ_2
F_{θ_1}	distribution before the shift
F_{θ_2}	distribution after the shift
$f(\mathbf{x} \theta)$	probability density of \mathbf{x} given the parameter θ

$g_0(k)$	prior probability distribution of the position of the shift in Lee (1998)
$g_1(k)$	prior probability distribution of the position of the shift uniform over $[a, n - b]$
$g_2(k)$	prior probability distribution of the position of the shift triangular with a mode of d and over the support $[a, n - b]$
k	position of the shift
m	hyperparameter for the prior distribution of the mean
\hat{m}_{1k}	ML-II hyperparameter for the prior distribution of the mean μ_1
\hat{m}_{2j}	ML-II hyperparameter for the prior distribution of the mean μ_2
N	Normal distribution
n	number of observations
nd	number of shifts detected in the series
nr	number of true shifts in the series
p	prior probability of no-change
p_i	positions of the true shifts ($i = 1, \dots, nr$)
p_i^d	positions of the detected shifts ($i = 1, \dots, nd$)
\bar{q}_a	mean of the series of ratios after the change
\bar{q}_b	mean of the series of ratios before the change
q_i	i th ratios between the base and neighbor series
$S(x)$	function of x in the exponential family form
X	sample space for \mathbf{x}
\mathbf{x}	vector of observations
x_i	i th observation of the series
\bar{x}_k	mean of the first k observations
\bar{x}_k^*	mean of the last $n - k$ observations
\bar{x}_n	overall mean of the series

References

- Aguilar E, Auer I, Brunet M, Peterson TC, Wieringa J. 2003. Guidelines on climate metadata and homogenization. *Report WMO-TD 1186*, World Meteorological Organization, Geneva, Switzerland.
- Alexandersson H. 1986. A homogeneity test applied to precipitation data. *Journal of Climatology* **6**: 661-675.
- Beaulieu C, Ouarda TBMJ, Seidou O. 2007. A review of homogenization techniques for precipitation data and their applicability to precipitation series (in French). *Hydrological Sciences Journal* **52**: 18-37.
- Beaulieu C, Seidou O, Ouarda TBMJ, Zhang X, Boulet G, Yagouti A. .2008a. Intercomparison of homogenization techniques for precipitation data. *Water Resources Research* **44**: W02425.
- Beaulieu C, Seidou O, Ouarda TBMJ, Zhang X. 2008b. Bayesian linear regression changepoint models for the homogenization of climatic series. *Submitted to Water Resources Research*.
- Berger JO. 1985. *Statistical decision theory and Bayesian analysis*. 2nd edition, Springer: New York.
- Caussinus H, Mestre O. 2004. Detection and correction of artificial shifts in climate series. *Journal of the Royal Statistical Society, Series C* **53**: 405-425.
- Chen J, Gupta AK. 2001. On change point detection and estimation. *Communications in Statistics-Simulation and Computation* **30**: 665-697.
- DeGaetano AT. 2006. Attributes of several methods for detecting discontinuities in mean temperature series. *Journal of Climate* **19**: 838-853.

- Ducré-Robitaille JF, Boulet G, Vincent LA. 2003. Comparison of techniques for detection of discontinuities in temperature series. *International Journal of Climatology* **23**: 1087-2003.
- Easterling DR, Peterson TC. 1995. A new method for detecting undocumented discontinuities in climatological time series. *International Journal of Climatology* **15**: 369-377.
- Ghorbanzadeh G, Lounes R. 2001. Bayesian analysis for detecting a change in exponential family. *Applied Mathematics and Computation* **124**: 1-15.
- Goldstein JO. 2006 Subjective Bayesian analysis: principles and practice. *Bayesian Analysis* **1**: 403-420.
- Heino R. 1997. Metadata and their role in homogenization. *Proceedings of the First Seminar for Homogenization of Surface Climate Data*, Hungarian Meteorological Service, Budapest, Hungary.
- Holbert D. 1982. A Bayesian analysis of a switching linear model. *Journal of Econometrics* **19**: 77-87.
- Kander Z, Zacks S. 1966. Tests procedures for possible changes in parameters of statistical distributions occurring at unknown time points. *The Annals of Mathematical Statistics* **37**: 1196-1210.
- Lee ASF, Heghinian SM. 1977. A shift of the mean level in a sequence of independent normal random variables. *Technometrics* **19**: 503-506.
- Lee CB. 1998. Bayesian analysis of a change-point in exponential families with applications. *Computational Statistics and Data Analysis* **27**: 195-208.

- Lund R, Wang XL, Lu Q, Reeves J, Gallagher C, Feng Y. 2007. Changepoint detection in periodic and autocorrelated time series. *Journal of Climate* **20**: 5178-5190.
- Menne MJ, Williams CNJ 2005. Detection of undocumented changepoints using multiple test statistics and composite reference series. *Journal of Climate* **18**: 4271-4286.
- Müeller-Westermeier G. 2004. Statistical analysis of results of homogeneity testing and homogenization of long climatological time series in Germany. *Proceedings of the Fourth seminar for homogenization and quality control in climatological databases*, Budapest, Hungary, *Report WCDMP 56, WMO-TD 1236*, World Climate Data and Monitoring Programme, World Meteorological Organization, Geneva, Switzerland.
- Perreault L, Haché M, Slivitzsky M, Bobée B. 1999. Detection of changes in precipitation and runoff over eastern Canada and U.S. using a Bayesian approach. *Stochastic Environmental Research and Risk Assessment* **13**: 201-216.
- Peterson TC, Easterling DR, Karl TR, Groisman P, Nicholls N, Plummer N, Torok S, Auer I, Boehm R, Gullett D, Vincent L, Heino R, Tuomenvirta H, Mestre O, Szentimrey T, Salinger J, Forland EJ, Hanssen-Bauer I, Alexandersson H, Jones P, Parker D. 1998. Homogeneity adjustments of in situ atmospheric climate data: a review. *International Journal of Climatology* **18**: 1493-1517.
- Potter KW. 1981. Illustration of a new test for detecting a shift in mean in precipitation series. *Monthly Weather Review* **109**: 2040-2045.
- Rasmussen P. 2001. Bayesian estimation of change points using the general linear model. *Water Resources Research* **37**: 2723-2731.

- Reeves J, Chen J, Wang XL, Lund R, Lu Q. 2007. A review and comparison of changepoint detection techniques for climate data. *Journal of Applied Meteorology and Climatology* **46**: 900-915.
- Robert C. 1992. *L'analyse statistique Bayésienne*. Economica: Paris.
- Seidou O, Asselin JJ, Ouarda TBMJ. 2007. Bayesian multivariate linear regression with application to changepoint models in hydrometeorological variables. *Water Resources Research* **43**: W08401.
- Seidou O, Ouarda TBMJ. 2007. Recursion-based multiple changepoint detection in multiple linear regression and application to river streamflows. *Water Resources Research* **43**: W07404.
- Smith AFM. 1975. A Bayesian approach to inference about a change-point in a sequence of random variables. *Biometrika* **62**: 407-416.
- Szentimrey T. 1999. Multiple analysis of series for homogenization (MASH). *Proceedings of the Second Seminar for Homogenization of Surface Climatological Data*, Hungarian Meteorological Service, Budapest, Hungary.
- Vincent LA. 1998. A technique for the identification of inhomogeneities in Canadian temperature series. *Journal of Climate* **11**: 1094-1105.
- Wang XL, Wen QH, Wu Y. 2007. Penalized maximal t-test for detecting undocumented mean change in climate data series. *Journal of Applied Meteorology and Climatology* **46**: 916-931.

Figure captions

Figure 1. Histograms of the false detections in the homogeneous series when the prior probability of no-change is set to 0.25.

Figure 2. Percentage of well positioned shifts according to their position and magnitude when a and b are set to 0.

Figure 3. Mean positioning criteria obtained in the series with a) two shifts b) three shifts.

Figure 4. Total annual precipitation of the base station and the two neighbor stations: a) Quebec city (base), b) Shawinigan Falls (neighbor), c) La Pocatiere (neighbor).

Figure 5. Posterior probability distributions for the position of the shift in the series of ratios.

Figure 6. Histogram of the ratios between the base and the mean of the two neighbor series.

Figure 7. Posterior probability distributions for the position of the shift in the two segments of the series of ratios (first column: 1944-1958, second column: 1959-1982).

Tables

Table 1. Percentage (%) of false detections in the homogeneous series, with different prior probabilities of no-change (p), different prior parameters of the interval of detection (a, b) and with a minimum segment length (segmin).

p	a, b						segmin
	0	5	10	15	20	25	10
0.01	94.7	100	100	100	100	100	55.3
0.05	67.7	64.9	62.1	61.7	63.0	65.3	40.0
0.10	42.6	38.3	36.1	35.1	35.0	36.6	25.5
0.25	19.4	16.9	15.7	15.0	14.5	14.7	11.8
0.50	7.8	6.9	6.6	6.4	6.2	6.1	5.2*
0.75	3.2*	2.8*	2.8*	2.7*	2.7*	2.6*	2.1*
0.90	1.4*	1.2*	1.1*	1.0*	1.1*	1.0*	0.9*
0.95	0.7*	0.6*	0.5*	0.5*	0.5*	0.5*	0.4*
0.99	0.2*	0.1*	0.1*	0.1*	0.2*	0.1*	0.1*

* Significantly smaller than 5% (5% critical level).

Table 2. Percentage (%) of well positioned shifts in the series, with a single shift with different prior probabilities of no-change (p), different prior parameters of the interval of detection (a, b) and with a minimum segment length (segmin).

p	a, b			segmin
	0	5	10	10
0.01	92.5	92.8	93.1	91.5
0.05	92.1	92.3	92.6	91.3
0.10	91.5	91.6	91.9	90.8
0.25	90.5	90.6	90.8	90.0
0.50	89.6	89.7	89.9	89.2
0.75	88.5	88.6	88.8	88.2
0.90	87.4	87.5	87.7	87.1
0.95	86.6	86.8	86.9	86.4
0.99	84.9	85.0	85.1	84.7

Table 3. Descriptive statistics of the positioning criteria obtained in the series with two shifts, with different prior probabilities of no-change (p), different prior parameters of the interval of detection (a, b) and with a minimum segment length (segmin).

p	a, b									segmin		
	0			5			10			10		
	mean	med	std	mean	med	std	mean	med	std	mean	med	std
0.01	7666.1	8167.5	1425.9	8199.0	8167.9	108.5	6551.6	6534.2	332.6	2692.2	3267.0	2146.1
0.05	7213.2	7840.8	1784.5	7529.7	8167.5	1447.2	5847.4	6534.0	1795.3	2533.7	3267.0	2164.5
0.10	5969.8	6534.7	2283.0	6184.5	6535.5	1948.8	4357.4	4900.8	2321.7	2080.2	1120.5	2148.4
0.25	3523.6	3547.3	2505.2	3867.6	4900.5	2262.9	2671.8	3267.0	2312.5	1675.1	6.5	2145.0
0.50	2326.5	3267.0	2387.6	2481.1	3267.0	2322.1	1928.3	22.5	2297.1	1610.3	2.0	2214.5
0.75	2041.8	13.0	2468.4	1982.8	12.5	2438.3	1754.6	2.0	2445.0	1710.6	2.0	2303.4
0.90	2071.9	4.5	2606.0	2000.6	4.5	2591.0	1882.0	2.0	2601.7	1871.0	2.0	2376.1
0.95	2158.1	4.5	2696.9	2100.8	4.5	2685.8	2018.1	2.0	2693.6	1979.8	2.0	2410.3
0.99	2422.6	8.0	2868.6	2381.2	4.5	2863.2	2325.7	2.5	2866.3	2215.8	4.5	2463.6

* The prior probability of no-change and the way to handle the extremities both influence the positioning criteria (Friedman's test, 5% critical level).

Table 4. Descriptive statistics of the positioning criteria obtained in the series with three shifts, with different prior probabilities of no-change (p), different prior parameters of the interval of detection (a, b) and with a minimum segment length (segmin).

p	a, b									segmin		
	0			5			10			10		
	mean	med	std	mean	med	std	mean	med	std	mean	med	std
0.0	7041.	7539.	1390.	7395.	7351.		4973.	4902.		2012.	2450.	1766.
1	7	2	1	6	5	165.6	4	0	482.6	2	5	6
0.0	6719.	7151.	1639.	6893.	7351.	1308.	4632.	4901.	1254.	2002.	2450.	1799.
5	5	7	3	5	1	1	8	2	9	7	5	8
0.1	5720.	6128.	1949.	5655.	6125.	1794.	3469.	3922.	1931.	1888.	2450.	1885.
0	1	8	9	4	9	4	1	0	2	9	3	1
0.2	3459.	3379.	2135.	3603.	3920.	1946.	2233.	2450.	1944.	1894.	2450.	2056.
5	0	3	5	4	4	8	0	8	1	4	3	6
0.5	2510.	2466.	2142.	2576.	2457.	2031.	1989.	2450.	2138.	2103.	2451.	2221.
0	6	5	6	3	5	4	3	3	8	8	3	6
0.7	2458.	2470.	2359.	2374.	2452.	2335.	2147.	2450.	2409.	2365.	3267.	2342.
5	4	6	4	8	3	0	4	3	8	1	0	3
0.9	2694.	3267.	2581.	2610.	3267.	2587.	2473.	3267.	2615.	2652.	3267.	2425.
0	1	0	9	2	0	2	5	0	5	5	0	3
0.9	2894.	3267.	2698.	2818.	3267.	2703.	2719.	3267.	2714.	2839.	3267.	2464.
5	2	0	7	7	0	4	0	0	8	7	0	1
0.9	3328.	3267.	2842.	3269.	3267.	2849.	3206.	3267.	2863.	3213.	3267.	2490.
9	8	0	9	8	0	3	5	0	8	3	0	0

* The prior probability of no-change and the way to handle the extremities both influence the positioning criteria (Friedman's test, 5% critical level).

Table 5. Number of detected shifts in the series with two shifts, with different prior probabilities of no-change (p), different prior parameters of the interval of detection (a, b) and with a minimum segment length (segmin).

p	a, b						segmin	
	0		5		10		10	
	mean	std	mean	std	mean	std	mean	std
0.01	11.7	4.9	12.3	0.8	6.1	0.6	2.7	1.1
0.05	10.2	5.1	10.5	3.3	5.5	1.5	2.6	1.0
0.10	6.8	4.0	6.8	3.3	4.0	1.8	2.3	0.9
0.25	3.3	1.9	3.4	1.8	2.6	1.2	2.0	0.7
0.50	2.2	1.0	2.3	1.0	2.1	0.8	1.8	0.6
0.75	1.9	0.8	1.9	0.7	1.8	0.6	1.7	0.5
0.90	1.7	0.7	1.7	0.6	1.7	0.6	1.7	0.5
0.95	1.6	0.6	1.6	0.6	1.6	0.6	1.6	0.5
0.99	1.5	0.6	1.5	0.6	1.5	0.6	1.6	0.5

Table 6. Number of detected shifts in the series with three shifts, with different prior probabilities of no-change (p), different prior parameters of the interval of detection (a, b) and with a minimum segment length (segmin).

p	a, b						segmin	
	0		5		10		10	
	mean	std	mean	std	mean	std	mean	std
0.01	12.6	4.6	12.3	0.8	6.1	0.6	3.3	1.0
0.05	11.6	4.8	11.2	2.7	5.8	1.2	3.2	1.0
0.10	8.6	4.1	8.2	3.2	4.8	1.6	3.0	1.0
0.25	4.6	2.2	4.7	2.0	3.4	1.2	2.6	0.8
0.50	3.1	1.4	3.3	1.4	2.8	1.0	2.5	0.8
0.75	2.6	1.1	2.6	1.0	2.5	0.9	2.4	0.8
0.90	2.4	1.0	2.3	0.9	2.3	0.9	2.3	0.8
0.95	2.2	0.9	2.2	0.9	2.2	0.9	2.2	0.8
0.99	2.0	0.9	2.0	0.9	2.0	0.9	2.0	0.8

Figures

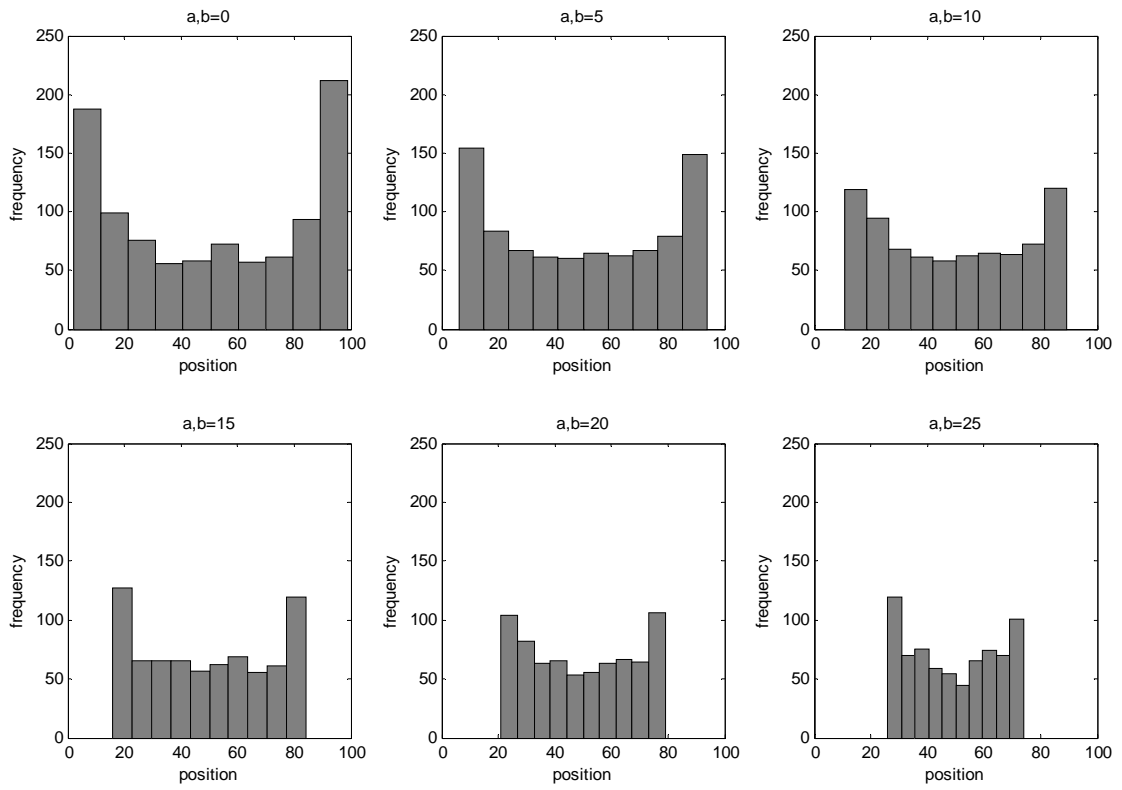


Figure 1.

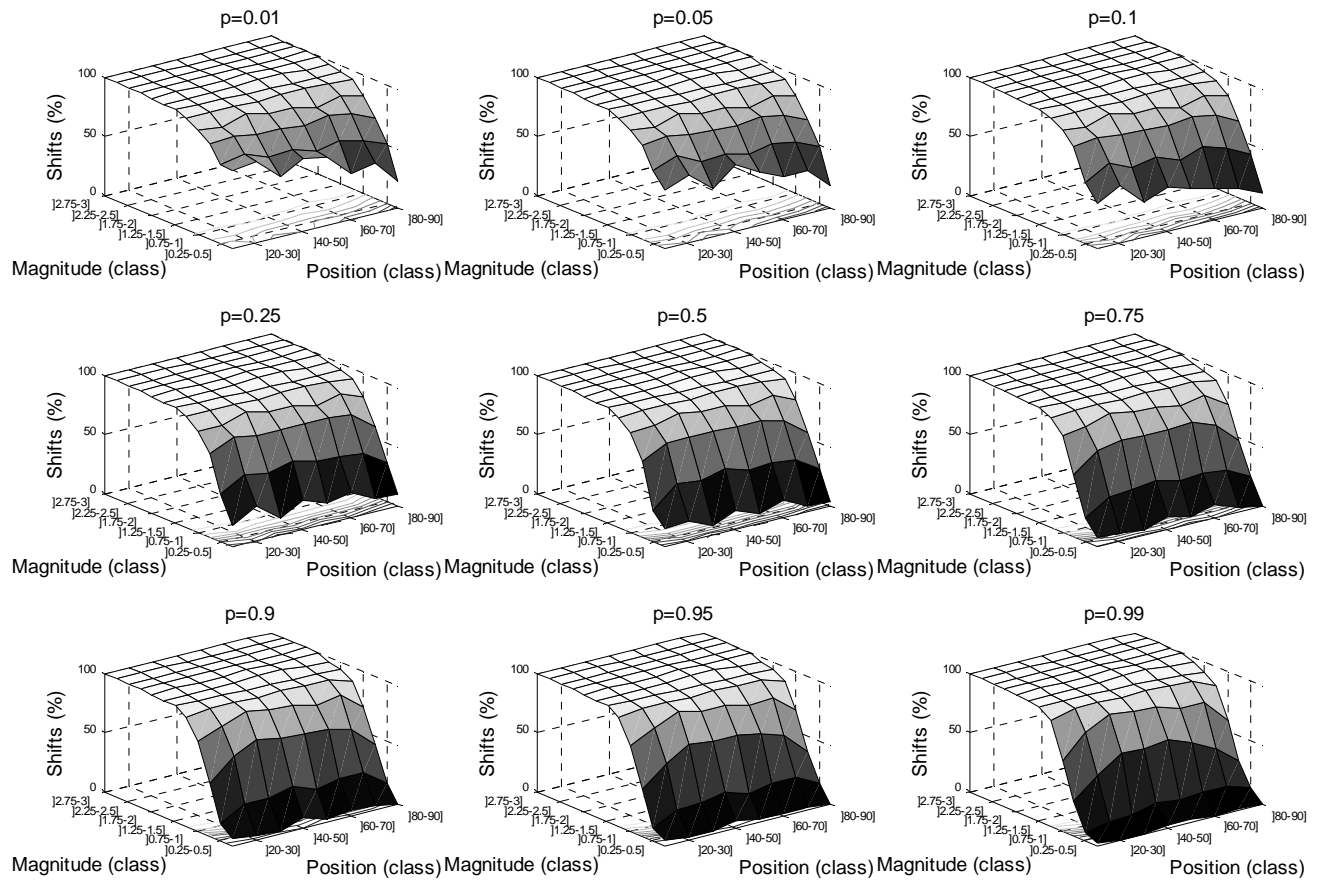


Figure 2.

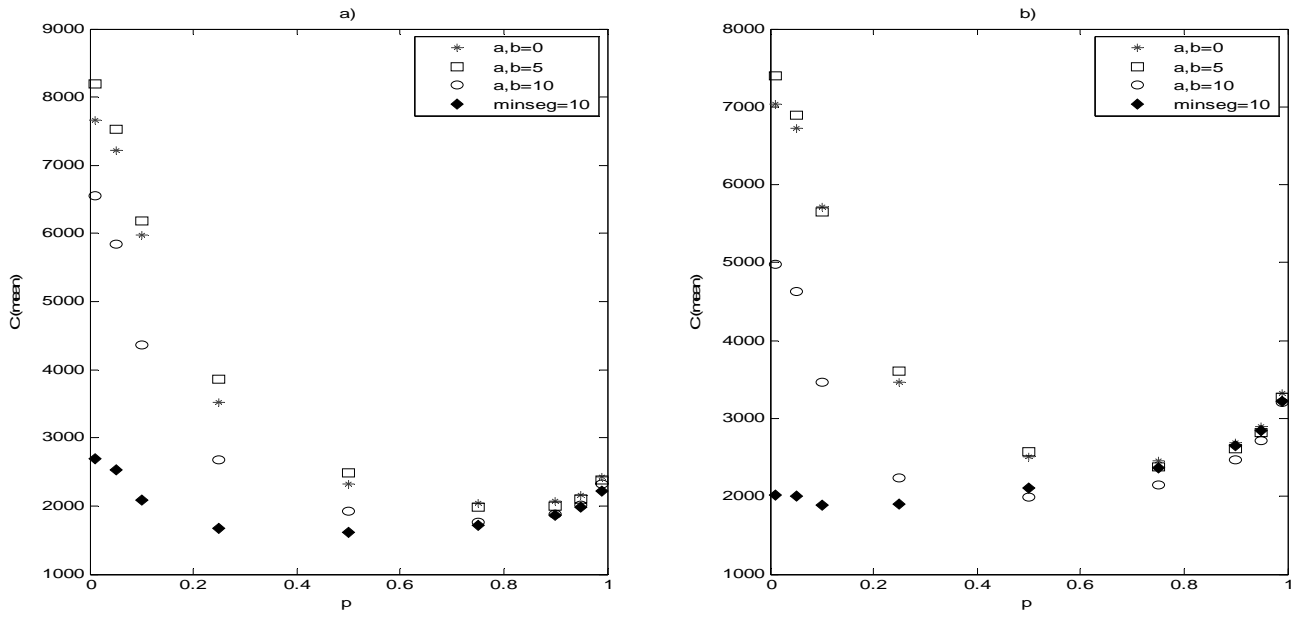


Figure 3.

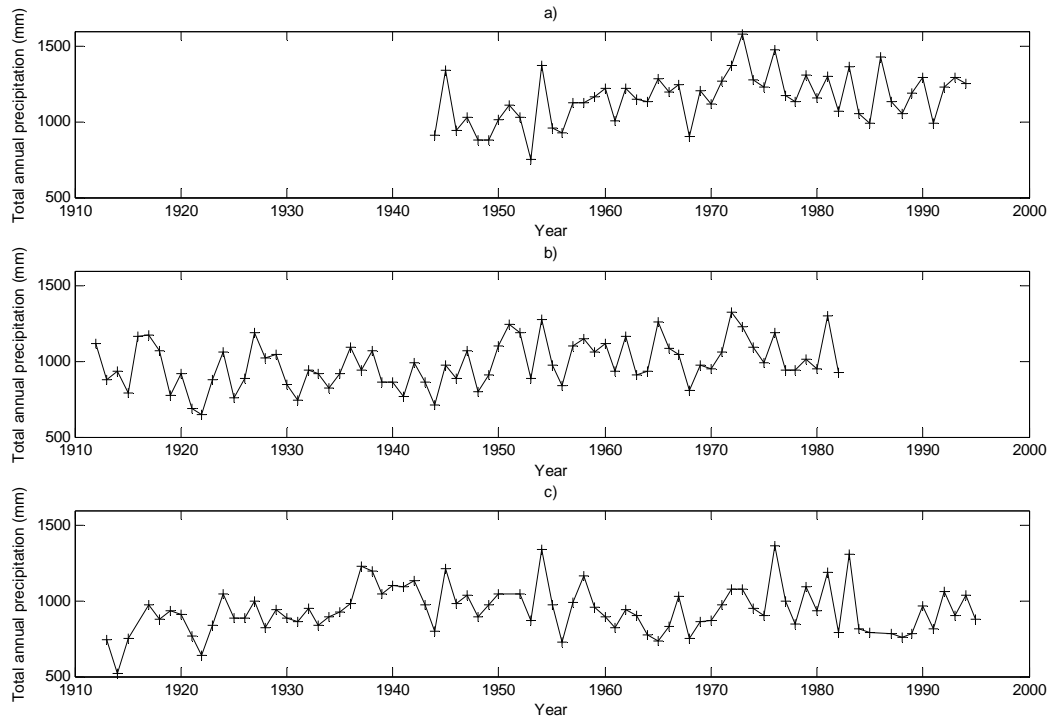


Figure 4.

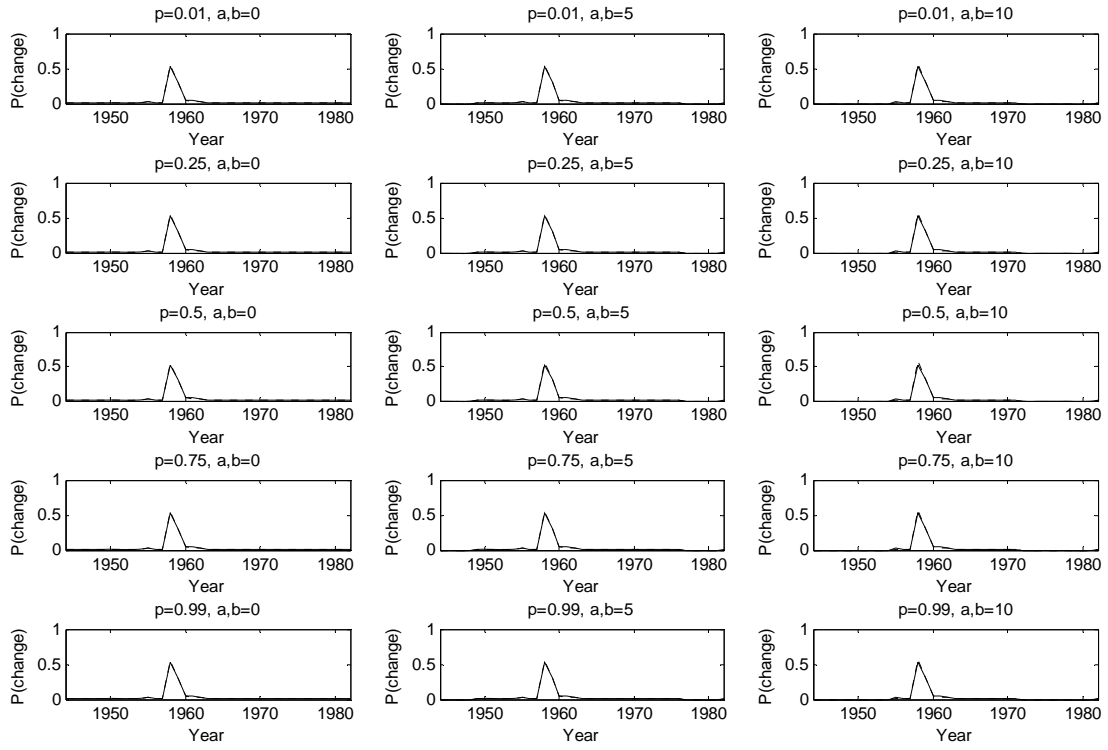


Figure 5.

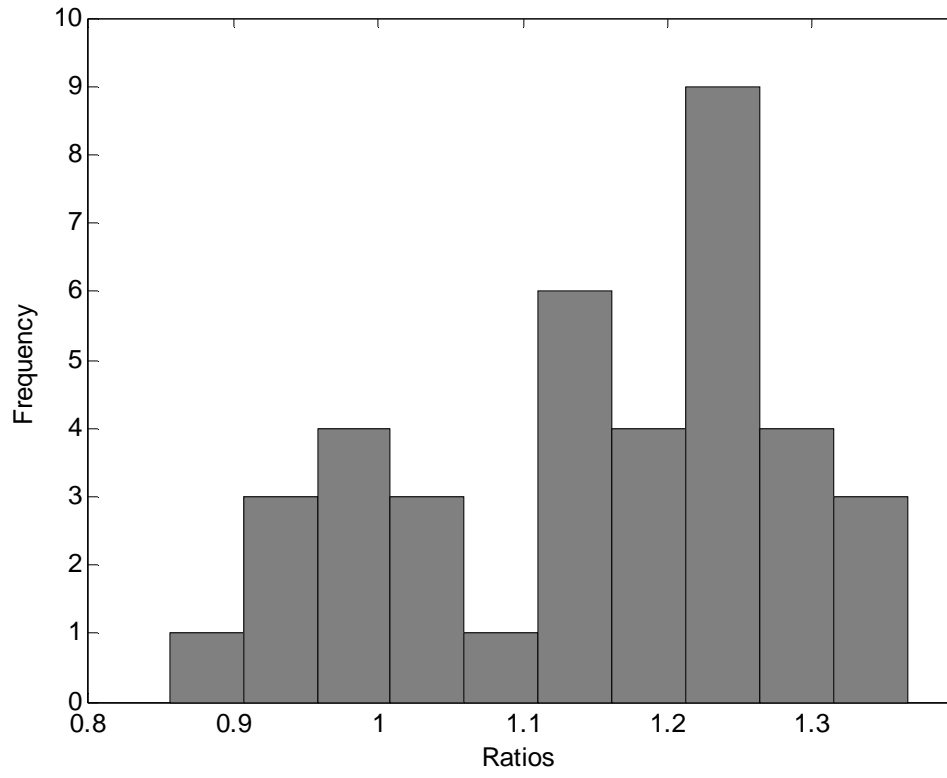


Figure 6.

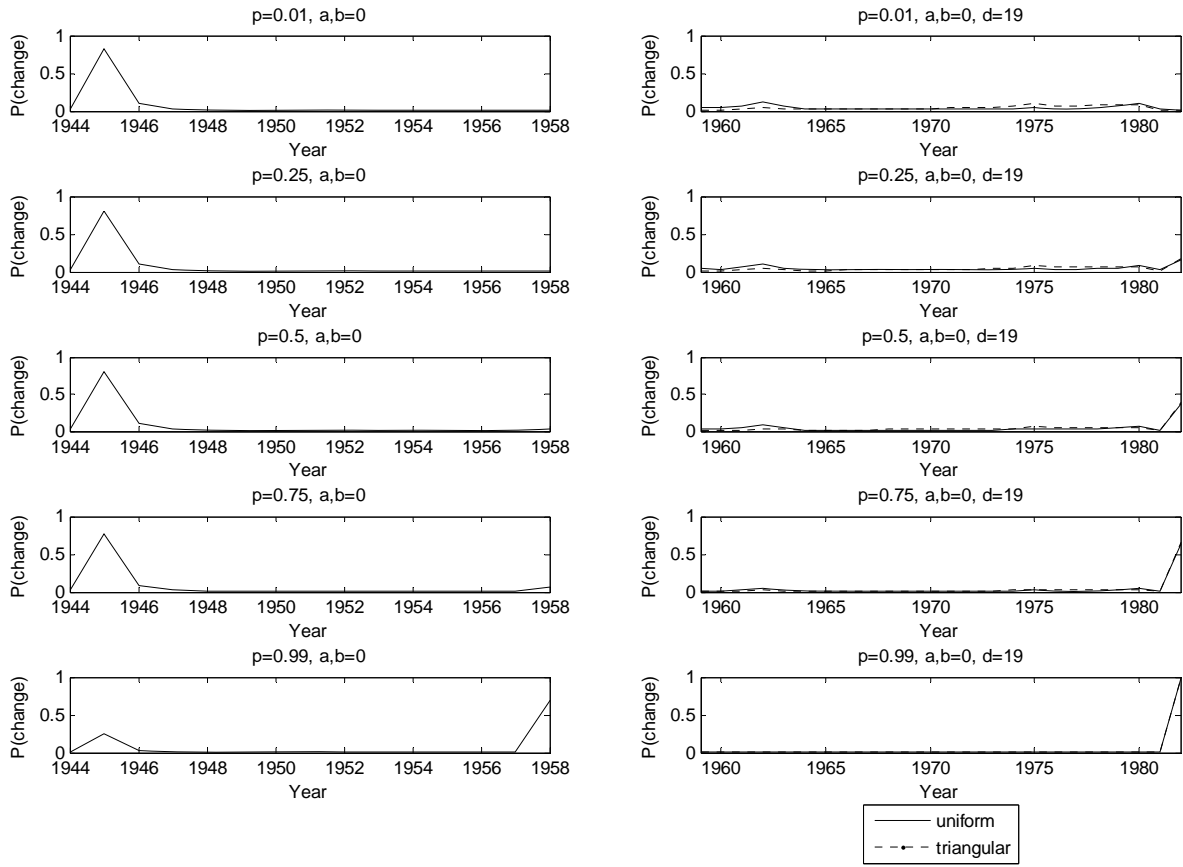


Figure 7.

10. Article 5: A Bayesian Gamma Homogeneity Test for the detection of artificial discontinuities in precipitation series

Un test d'homogénéité bayésien qui permet de détecter un changement de paramètre d'échelle dans une distribution gamma est présenté dans cet article. L'application du test pour détecter des inhomogénéités dans des séries de précipitations totales est proposée. Le test a été appliqué à des séries de totaux de précipitations à différents pas de temps (annuel, saisonnier et mensuel). L'avantage de ce test est qu'il ne présuppose pas que les rapports de précipitations totales sont normalement distribués, hypothèse qui n'est pas automatiquement vérifiée. Ce test est plus flexible que les autres méthodes d'homogénéisation développées dans le cadre normal. En effet, la distribution gamma permet d'approximer la distribution normale et peut aussi représenter des cas pour lesquels la distribution normale est inadéquate (asymétrie). Le test proposé est simple et facile à utiliser et permet d'intégrer différentes sources d'information dans l'analyse. Un cas d'étude est présenté dans lequel le test est appliqué pour analyser l'homogénéité des totaux de précipitations d'une série de la province de Québec, au Canada. L'homogénéité est vérifiée aux pas de temps annuel, saisonnier et mensuel. Le test proposé s'est montré efficace pour détecter des inhomogénéités aux différents pas de temps.

**A Bayesian Gamma Homogeneity Test for the detection of artificial discontinuities in
precipitation series**

Claudie Beaulieu¹, Taha B.M.J. Ouarda¹, Ousmane Seidou²

¹INRS-ETE, University of Quebec, Quebec, QC, Canada

²Department of Civil Engineering, University of Ottawa, Ottawa, ON, Canada

To be submitted

Corresponding author: Claudie Beaulieu, Chair in statistical Hydrology, Canada Research Chair on the Estimation of Hydrological Variables, INRS-ETE, University of Quebec, 490, de la Couronne, Québec (Québec) G1K 9A9, CANADA. e-mail: claudie.beaulieu@ete.inrs.ca; (after March 31st, 2009: claubeau@yahoo.com) tel. (418) 654-2530 #4462

Abstract

In this paper, a Bayesian Gamma Homogeneity Test for the detection of artificial discontinuities in precipitation series is presented. The test allows for the detection of a change in the scale parameter of a gamma series. It is applied to detect inhomogeneities in precipitation series at different time scales (annual, seasonal and monthly). The main advantage of the test is that it does not assume that the series of ratios of precipitation totals are normally distributed, a hypothesis that is not always verified. The test is more flexible than other homogenization techniques developed in the normal framework. The gamma distribution can approximate the normal distribution and can represent situations where the normal distribution is inadequate (asymmetric distribution). The proposed test is simple to use and also allows the integration of additional sources of information (metadata, expert belief, regional information) in the analysis through the Bayesian framework. The test was applied to analyze the homogeneity of precipitation totals in a case study in the province of Quebec, Canada, at different time scales. It was shown that the test is efficient to detect inhomogeneities in the annual, seasonal and monthly series.

Keywords: homogenization, precipitation series, Bayesian technique, changepoint detection, gamma distribution, exponential family, scale parameter.

1. Introduction

Reliable precipitation series are important in many applications in hydrology and climatology. However, some modifications in the measurement procedures at the climatic stations can induce artificial shifts in the observation series (e.g. relocation of a station, change of equipment and instruments, change of observer, modification in the immediate environment of the site, etc.). Thus, homogenization techniques can be used to detect and correct these artificial shifts. For a complete review of all these techniques, the reader is referred to Peterson *et al.* (1998), WMO (2003), Beaulieu *et al.* (2007) and Reeves *et al.* (2007).

Most homogenization techniques presented in the literature were developed under the normality hypothesis (Potter, 1981; Alexandersson, 1986; Easterling and Peterson, 1995; Vincent, 1998; Mestre, 2000; Wang *et al.*, 2007; Beaulieu *et al.*, 2008a). When the normality hypothesis is not respected, it is possible to use a non parametric test for changepoint detection such as the Wilcoxon test (Karl et Williams, 1987; Lanzante, 1996) or the test developed by Pettitt (1979). However, it was shown in previous comparative studies of several homogenization techniques that some parametric techniques are more powerful than non parametric tests (Ducre-Robitaille *et al.*, 2003; Beaulieu *et al.*, 2008b).

The parametric technique proposed by Thompson (1984) does not assume the normality of the time series. It was developed by to homogenize monthly precipitation totals, which were represented by a Poisson process for the occurrences, and by an independent exponential distribution for the event amount. Thompson (1984) proposed a test to determine if the parameters of the model change after a station relocation and applied it to a set of stations located in New Zealand. However, this homogenization technique does not allow for the detection of changes that are not documented. As the metadata is often missing or incomplete, it is advantageous to use a technique that is able to detect inhomogeneities independently of the metadata. Furthermore, the Poisson-Exponential model may not be appropriate in all climatic

regions. The gamma distribution is also commonly used to represent precipitation series at different time scales.

A new test to detect inhomogeneities in precipitation series is presented in this work. This test is based on the gamma distribution. This distribution was chosen because of its flexibility and because it is a natural choice to represent both precipitation totals and precipitation ratios. The test is based on a Bayesian approach and thus, it allows for the integration of different sources of information into the analysis. The objectives of this work are to present the Bayesian Gamma Homogeneity Test and to illustrate its usefulness to detect inhomogeneities in precipitation series at different time scales. The rest of this paper is organized as follows: section 2 presents the literature review, section 3 presents a description of the test, section 4 presents the case study and finally, section 5 and 6 present respectively a discussion of the results and the conclusions.

2. Literature review

2.1. Precipitation distribution

In general, total annual precipitation can be considered normal (Von Storch and Zwiers, 1999). The central limit theorem is invoked to validate this hypothesis. At smaller time scales, the distribution of the precipitation series is asymmetric, which makes the gamma distribution a more natural choice (Wilks, 2006). This distribution is next to the normal distribution in simplicity (only two-parameters), and, at the same time, is flexible as it covers a wide skewness range (Mooley, 1973). The gamma distribution has been used for a long time to represent seasonal or monthly precipitation totals (Ison *et al.*, 1971; Bridges and Han, 1972; Ropelewski *et al.*, 1985; Wilks, 1990; Bobée and Ashkar, 1991; Wilks and Eggleston, 1992; Groisman and Easterling, 1994; Briggs and Wilks, 1996). In Yue and Hashino (2007), several distributions were compared to fit the annual, seasonal and monthly precipitation series of a set of stations in Japan. Among the two-parameter distributions, the gamma was almost always the most appropriate one to fit

precipitation series at the various time scales. Groisman and Easterling (1994) have represented the annual and monthly precipitation series of several stations in Canada and in the United States with the gamma distribution. They have shown that the shape parameter is constant in the same region. More generally, this is a common assumption in hydrology: the shape parameter is often assumed constant in the same region (Buishand, 1989; Wilby et Wigley, 2002; Katz *et al.* 2002).

The gamma distribution is disadvantageous when the series of observations contains some zero-values (Juras, 1994). This problem occurs especially in arid and semi-arid regions with entire months without any precipitation. The estimation of the scale and shape parameters with the maximum likelihood method is then impossible. The method of moments is also inappropriate (Thom, 1958). Wilks (1990) presents a technique to compute the maximum likelihood parameters by treating the zeros as censored data. Some authors model separately the occurrence (the discrete component) and the quantity (the continuous component) of daily and/or monthly precipitation series with a Poisson-gamma or Poisson-exponential model (Ozturk, 1981; Alexandersson, 1985; Revfeim, 1985; Dunn, 2004). In wet regions, such as Canada, where months without precipitations are rare, this type of model would be more appropriate for daily series.

2.2. Ratios distribution

In homogenization, the analysis is often performed on a series of ratios or differences between the base series (series that is tested for homogeneity) and the neighbor series. Ratios are more commonly used for precipitation series (Alexandersson, 1986). Even if the base and neighbor series are normally distributed, it does not imply that the ratios/differences will be normally distributed. In theory, the difference between two normal variables follows a normal distribution, but the distribution of the ratios of two normal variables with different means and variances cannot be obtained analytically. However, the hypothesis of normality of the ratios is

made notably in the techniques developed by Alexandersson (1986) and Mestre (2000) and it is verified most of the time. The distribution of precipitation ratios at smaller time scales (seasonal, monthly) gets more asymmetric and can hardly be represented by the normal distribution. The distribution of the ratios of two gamma variables having a common shape parameter is not known. Nevertheless, the gamma distribution is more appropriate than the normal distribution as it has an asymmetric shape. Furthermore, the range of the normal distribution is infinite, while the gamma distribution is left-bounded by zero, which is more appropriate to represent precipitation ratios.

2.3. Changepoint detection in gamma series

The literature of changepoint analysis is quite large. Most changepoint techniques were developed for normal series. There were few changepoint techniques developed for gamma series. Hsu (1979) presented a test to detect a change in the scale parameter of a gamma series with a known and constant shape parameter. The Bayesian approach for this test was developed by Diaz (1982). More generally, some techniques were developed to detect a change in the parameter of a distribution belonging to the exponential family. This family of distributions includes the normal, gamma, exponential, Poisson, weibull, beta, binomial, etc. These techniques are especially interesting as they can be adapted to several distributions. Kander and Zacks (1966) developed a Bayesian test for a change in the parameter of the binomial and exponential distributions. Smith (1975) developed a Bayesian test to detect a change in the parameter of the binomial and normal distributions. Lee (1998) presented the posterior probabilities to infer concerning a change in the mean of a normal distribution and in the scale parameter of a gamma distribution. Ghorbanzadeh and Lounes (2001) explored the Poisson and exponential cases with a Bayesian approach. Worsley (1986) used a different approach based on the likelihood ratio test.

The technique developed by Lee (1998) is particularly interesting for precipitation series as the normal and gamma distributions are treated and can be used to represent precipitation ratios.

3. Presentation of the technique

3.1. Bayesian approach

The Bayes theorem can be expressed by:

$$\pi(\theta | \mathbf{x}) = \frac{f(\mathbf{x} | \theta) \pi(\theta)}{\int f(\mathbf{x} | \theta) \pi(\theta) d\theta} \quad (1)$$

where $\pi(\theta)$ is the prior distribution which summarizes the information about the parameter θ , \mathbf{x} is a vector of observations having a probability density $f(\mathbf{x} | \theta)$. The theorem allows to actualize the prior distribution $\pi(\theta)$ with the information contained in the observations. The choice of the prior distribution can be made in light of the different sources of information. The estimation of the parameter can be made with different Bayesian parametric estimators (posterior mean, posterior mode,...). For hypothesis testing, the posterior mode is often used (Robert, 1992).

3.2. Bayesian changepoint detection in natural exponential family

The technique presented in Lee (1998) allows the detection of a change in the parameter of a distribution belonging to the natural exponential family. A distribution belonging to the natural exponential family has a probability density function of the form:

$$f(x | \theta) = \exp\{\theta x + \varphi(\theta) + S(x)\} \quad (2)$$

where $\varphi(\theta)$ and $S(x)$ are two functions which depend on the form of the distribution. Suppose the following changepoint model:

$$x_i \sim \begin{cases} F_{\theta_1}, & i = 1, \dots, k \\ F_{\theta_2}, & i = k + 1, \dots, n \end{cases} \quad (3)$$

where the x_i 's are independent variables, F_{θ_1} and F_{θ_2} represents the distribution before and after the shift with parameters θ_1 and θ_2 respectively, k is the unknown position of the changepoint, and n is the length of the series. The joint distribution of $\mathbf{x} = (x_1, \dots, x_k, x_{k+1}, \dots, x_n)$ is given by:

$$f(\mathbf{x} | \theta) = \exp \left\{ \theta_1 \sum_{i=1}^k x_i + \theta_2 \sum_{i=k+1}^n x_i + k\varphi(\theta_1) + (n-k)\varphi(\theta_2) + \sum_{i=1}^n S(x_i) \right\} \quad (4)$$

The hypothesis that there is no change in the parameter of the distribution ($\theta_1 = \theta_2$) is tested against the hypothesis that there is a change in the parameter of the distribution beginning at the time $k + 1$:

$$\begin{aligned} H_0 : k = n, \\ H_1 : 1 \leq k \leq n - 1 \end{aligned} \quad (5)$$

The prior probability distribution for k is given by (Lee, 1998):

$$g(k) = \begin{cases} p, & k = n \\ \frac{1-p}{n-1}, & k \neq n \end{cases} \quad (6)$$

where p represents the prior probability of no-change ($0 \leq p \leq 1$). The case where no-change has occurred in the series corresponds to $k = n$ and the case where a change has occurred at the position k corresponds to $k \neq n$. $g(k)$ is uniformly distributed over $[1, \dots, n-1]$. For the parameters θ_1 and θ_2 , conjugate priors are used. The conjugate prior for the exponential family has the form:

$$\pi(\theta | m, c) \propto \exp \{ mc\theta + m\varphi(\theta) \} \quad (7)$$

where m and c are hyperparameters such that $m \geq 0$ and $c \in X$. X represents the sample space of the x_i 's. The prior information about θ_1 and θ_2 takes into account the position of the changepoint (Lee, 1998):

$$\begin{aligned}\pi(\theta_1 | k) &= \pi(\theta_1 | m_{1k}, c_{1k}), \quad k = n \\ \pi(\theta_1, \theta_2 | k) &= \prod_{i=1}^2 \pi(\theta_i | m_{ik}, c_{ik}), \quad k \neq n\end{aligned}\tag{8}$$

where $\pi(\theta_i | m_{ik}, c_{ik}) \propto \exp\{m_{ik}c_{ik}\theta_i + m_{ik}\varphi(\theta_i)\}$, $i = 1, 2$, $m_{ik} \geq 0$, $c_{ik} \in X$. The joint prior distribution of (k, θ_1, θ_2) is given by

$$\begin{aligned}\pi(k, \theta_1) &= p \frac{\exp\{m_{1k}c_{1k}\theta_1 + m_{1k}\varphi(\theta_1)\}}{\tilde{\pi}(\Theta | m_{1k}, c_{1k})}, \quad k = n \\ \pi(k, \theta_1, \theta_2) &= \frac{1-p}{n-1} \frac{\exp\left\{\sum_{i=1}^2 [m_{ik}c_{ik}\theta_i + m_{ik}\varphi(\theta_i)]\right\}}{\prod_{i=1}^2 \tilde{\pi}(\Theta | m_{ik}, c_{ik})}, \quad k \neq n\end{aligned}\tag{9}$$

where

$$\tilde{\pi}(\Theta | m_{ik}, c_{ik}) = \int_{\Theta} \exp\{m_{ik}c_{ik}\theta_i + m_{ik}\varphi(\theta_i)\} d\theta_i, \quad i = 1, 2\tag{10}$$

and Θ represents the parameter space of θ_1 and θ_2 . According to Lee (1998), the posterior distribution of the position of the changepoint k is:

$$\pi(k | \mathbf{x}) \propto \begin{cases} p \frac{\tilde{\pi}\left(\Theta | m_{1k} + n, \left(m_{1k}c_{1k} + \sum_{i=1}^n x_i\right) / m_{1k} + n\right)}{\tilde{\pi}\left(\Theta | m_{1k}, c_{1k}\right)}, & k = n \\ \frac{1-p}{n-1} \frac{\tilde{\pi}\left(\Theta | m_{1k} + k, \left(m_{1k}c_{1k} + \sum_{i=1}^k x_i\right) / m_{1k} + k\right)}{\tilde{\pi}\left(\Theta | m_{1k}, c_{1k}\right)} \\ \times \frac{\tilde{\pi}\left(\Theta | m_{2k} + (n-k), \left(m_{2k}c_{2k} + \sum_{i=k+1}^n x_i\right) / m_{2k} + (n-k)\right)}{\tilde{\pi}\left(\Theta | m_{2k}, c_{2k}\right)} \end{cases}, \quad k \neq n \quad (11)$$

The values of the hyperparameters m_{ik} can be thought of as the sample size of priors (Lee, 1998).

Thus, they can be set such that $m_{1k} = k$ and $m_{2k} = n - k$. The maximum likelihood type II (ML-

II) approach for prior selection (Berger, 1985) is used to estimate the hyperparameters c_{ik} . The

ML-II approach consists in a maximization of the marginal distribution of x , $m(\mathbf{x} | \pi)$, over the

hyperparameters c_{ik} , $i = 1, 2$, $k = 1, \dots, n$. Maximizing $m(\mathbf{x} | \pi)$ over π is equivalent to

maximizing the two following functions on c_{1k} and c_{2k} when $m_{1k} = k$ and $m_{2k} = n - k$ (Lee,

1998):

$$\frac{\tilde{\pi}\left(\Theta | 2k, (c_{1k} + \bar{x}_k) / 2\right)}{\tilde{\pi}\left(\Theta | k, c_{1k}\right)} \text{ or } \frac{\tilde{\pi}\left(\Theta | 2(n-k), (c_{2k} + \bar{x}_k^*) / 2\right)}{\tilde{\pi}\left(\Theta | n-k, c_{2k}\right)} \quad (12)$$

where $\bar{x}_k = \sum_{i=1}^k x_i / k$ and $\bar{x}_k^* = \sum_{i=k+1}^n x_i / (n - k)$. The logarithm of these functions gives an expression

that is similar to the function:

$$h(s) = \log\left(\tilde{\pi}\left(\Theta | 2k, (c + s) / 2\right)\right) - \log\left(\tilde{\pi}\left(\Theta | k, s\right)\right) \quad (13)$$

c is \bar{x}_k or \bar{x}_k^* and s represents the hyperparameters to maximize. The maximum values can be obtained with the derivative. $h(s)$ is concave downward and the ML-II hyperparameters are easily obtained for the normal distribution.

3.3. Bayesian Gamma Homogeneity Test (BGHT)

The gamma case of the test presented in Lee (1998) allows to test for a change in the scale parameter of a gamma series and to find its position. The shape parameter is assumed known and constant. A random variable x , having a gamma $G(\alpha, \eta)$ distribution has the following density function:

$$f(x) = \frac{x^{\alpha-1}}{\Gamma(\alpha)\eta^\alpha} \exp\{-x/\eta\}, \quad x > 0, \alpha > 0, \eta > 0 \quad (14)$$

where α is the shape parameter, η is the scale parameter and Γ represents the gamma function. The observations in the series are assumed independent. The changepoint model can be represented by:

$$x_i \sim \begin{cases} G(\alpha, \eta_1), & i = 1, \dots, k \\ G(\alpha, \eta_2), & i = k+1, \dots, n \end{cases} \quad (15)$$

where G represents the gamma distribution, η_1 and η_2 are the scale parameters of the series before and after the shift and α is the shape parameter of the series (assumed known).

The conjugate priors for the scale parameters before and after the change have a $G(m\alpha + 1, 1/mc)$ distribution. Lee (1998) found that the ML-II hyperparameters c and m are respectively $\hat{c}_{1k} = [1 + 1/k\alpha] \bar{x}_k$, $\hat{m}_{1k} = k$, $\hat{c}_{2k} = [1 + 1/(n-k)\alpha] \bar{x}_k^*$ and $\hat{m}_{2k} = n - k$ for the means before and after the shift. The posterior distribution of k given \mathbf{x} is:

$$\pi(k | \mathbf{x}) \propto p \cdot \frac{\Gamma(2n\alpha + 1) \cdot \left(1 + \frac{1}{n\alpha}\right)^{n\alpha+1}}{\Gamma(n\alpha + 1) \cdot \left(2 + \frac{1}{n\alpha}\right)^{2n\alpha+1} \cdot (n\bar{x}_n)^{n\alpha}}, \quad k = n \quad (16)$$

$$\pi(k | \mathbf{x}) \propto \begin{cases} \frac{1-p}{n-1} \cdot \frac{\Gamma(2k\alpha + 1) \cdot \left(1 + \frac{1}{k\alpha}\right)^{k\alpha+1}}{\Gamma(k\alpha + 1) \cdot \left(2 + \frac{1}{k\alpha}\right)^{2k\alpha+1} \cdot (k\bar{x}_k)^{k\alpha}} \\ \times \frac{\Gamma(2(n-k)\alpha + 1) \cdot \left(1 + \frac{1}{(n-k)\alpha}\right)^{(n-k)\alpha+1}}{\Gamma((n-k)\alpha + 1) \cdot \left(2 + \frac{1}{(n-k)\alpha}\right)^{2(n-k)\alpha+1} \cdot ((n-k)\bar{x}_k^*)^{(n-k)\alpha}} \end{cases}, k = 1, \dots, n-1 \quad (17)$$

The position of the shift can be estimated from the posterior distribution by taking the highest posterior probability (posterior mode). The mathematical details of the technique are presented in Lee (1998).

In homogenization, the test can be applied to series of precipitation ratios between the base series and the neighbor series. However, the test can also be applied to detect a change in the scale parameter of precipitation totals.

4. Case study

In Beaulieu *et al.* (2008a), an application of the Bayesian Normal Homogeneity Test (BNHT) to precipitation series is presented. The total annual precipitation series of Quebec City (station 7016294), Quebec, Canada, was tested for homogeneity. This station was chosen among a set of stations located in southern and central Quebec and nearby areas (Canada) with high-quality data (few missing values, long observation series) that were extracted from Environment Canada's database. It was also chosen because neighbor series and metadata are available for this station. The Quebec station is located at latitude 46.8, longitude -71.38 and at an altitude of 70 meters. Two neighbor stations, Shawinigan Falls (7018000) and La Pocatiere (7054095), were identified in the set of stations by considering the distance from the base station, the elevation

difference, the observation period, the correlations and the correlations computed from the first difference series (Beaulieu *et al.*, 2008a). In the metadata, no potential causes of inhomogeneities are documented for the two neighbor stations. However, the base station was relocated several times; and twice with a change in the elevation (1958 and 1977). These two relocations could have introduced an inhomogeneity in the base series, because precipitation series are especially sensible to changes in the instrument height (Heino, 1997). Figure 1 presents the precipitation series at the base and neighbor stations at the three different time scales.

The BNHT was applied to the series of ratios between the base series and the mean of the two neighbor series for the common period of observations: 1944-1982. A shift was detected in 1958, which means that the new regime begins in 1959. This shift could be due to a change in the elevation of the station that is documented in the metadata. At smaller time scales the use of the BNHT is not recommended as the data is not normal. However, as the gamma distribution is flexible, the BGHT can be used to perform a homogeneity analysis at different time scales. In this paper, the same case study is examined. The BGHT is used to detect inhomogeneities in total annual, seasonal and monthly precipitation at the Quebec station. For each case, the adequacy of the gamma and normal distributions to fit the ratio series is verified. Furthermore, the two distributions are compared according to the Akaike (1974) and Schwarz (1978) criteria. The Akaike (*AIC*) and Schwarz (*SIC*) information criteria can be expressed as follows:

$$AIC = 2d - 2\log(L) \tag{18}$$

$$SIC = d \cdot \log(n) - 2\log(L) \tag{19}$$

where d represents the number of parameters in the model, L is the maximized value of the likelihood function and n is the number of observations. The distribution which leads to a smaller criterion is considered to be the most appropriate. The BGHT is also based on the

independence assumption of the observations. In Beaulieu *et al.* (2008a), it was shown that the independence and goodness of fit tests can be biased in the presence of an inhomogeneity. This point is also discussed in Perreault *et al.* (1999). Consequently, for each time scale, the independence and goodness of fit tests are applied to the entire series of ratios and to the segments before and after the shift.

4.1. Annual scale

Figure 2 presents the time series and histogram of the ratio of annual precipitation totals of the base series over the mean of the two neighbors. It seems clear that there is a change in the mean around 1959. The BGHT was applied with different prior probabilities of no-change. The maximum likelihood shape parameter is 76. Figure 3 presents the posterior probability distributions for the shift's position in the series of ratios of total annual precipitation. With all values of the prior probability of no-change, the same shift is detected with BGHT and BNHT. Table 1 presents the results of the independence test, the goodness of fit of the gamma and normal distributions and the comparison of the two distributions with the AIC and SIC for the entire series and for the segments before and after the shift. The observations in each segment can be considered independent. The two distributions seem to provide an adequate fit to the entire series and to the two segments analyzed separately (Chi-square test, 5% critical level). The choice of the best distribution, according to the AIC and SIC, depends on the segment that is fitted. The identification of the best distribution to represent the ratios between the total annual precipitation in Quebec and its two neighbors needs hence to be carefully examined.

4.2. Seasonal scale

The homogeneity of the seasonal precipitation series in Quebec was assessed to determine if it is also affected by the same inhomogeneity and when exactly. The ratios of seasonal precipitation totals between the base series and the mean of the two neighbor series were computed and are presented in figure 4. The shift is also apparent at the seasonal level. Furthermore, the ratios seem positively skewed. The maximum likelihood shape parameter is 27. The BGHT was applied with different prior probabilities of no-change and the results are presented in figure 5. The mode of the posterior distribution is the summer 1959, which means that the new regime begins during the fall of 1959. This shift coincides with the change detected in the annual series. When tested separately, the independence hypothesis was accepted in each segment (table 1). The normal distribution does not seem adequate to fit the ratios, while the gamma distribution seems to fit the two segments separately. However, the gamma distribution seems to be a better choice than the normal distribution: the AIC and SIC are smaller in all cases.

4.3. Monthly scale

The same procedure was repeated to detect inhomogeneities in the monthly precipitation series. Figure 6 illustrates the ratios between the base series and the two neighbors and the associated histogram. The histogram has an asymmetric shape. The maximum likelihood shape parameter of the series of ratios is 11. The BGHT was applied with different prior probabilities of no-change and the results are presented in figure 7. With all prior probabilities of no-change, the mode of the posterior distribution is October 1959. Then, the shift detected in the monthly series coincides with the shifts detected in the seasonal and annual series. The gamma distribution seems more appropriate than the normal distribution to represent each segment (table1). The AIC and SIC are smaller with the gamma distribution, indicating that it is the most appropriate to represent the series of ratios. The autocorrelation is significant in the second segment. At the

monthly time scale, the presence of autocorrelation may be sometimes problematic. Furthermore, the presence of seasonality could be problematic. Figure 8 presents the boxplot of the ratios for each month. There is no evident seasonal pattern in the series of ratios.

5. Discussion

5.1. Seasonality and trends

Climatic time series usually have a strong seasonality, and this can affect changepoint analysis. The easiest way to handle the seasonality is to subtract from January's data the overall January average, from the February's data, the overall February average, and so on (Jaruskova, 1997). However, in the presence of a shift, the overall mean is biased, and then, the deseasoning can influence the changepoint detection, particularly if the shift is large (Lund *et al.*, 2007). However, it can be assumed that the seasonality in the base series is similar to the seasonal component of the neighbor series and that it is retrieved when producing the series of ratios between the base and neighbor series. The same hypothesis is usually made about the presence of a regional trend in the series.

5.2. Autocorrelation

For total annual precipitation series, the hypothesis of independence is realistic. At smaller time scales, the presence of autocorrelation is more likely to occur. In the case study, the independence was almost always respected, except in the last segment of the monthly series of ratios. This is problematic as the changepoint detection techniques which were developed for independent error series tend to detect more false shifts in the presence of autocorrelation (Lund *et al.*, 2007).

5.3. Distribution

In the case study, it was shown that the BGHT can be used to detect inhomogeneities in annual, seasonal and monthly precipitation totals. The gamma distribution is advantageous as it can accommodate a non null skewness. In the annual series, it was shown that both the normal and gamma distributions could be used to represent the series. The estimated shape parameter was 76. In Groisman and Easterling (1994), it is pointed out that a shape parameter exceeding 15 describes a gamma distribution that approximates a normal distribution. In the seasonal series, the shape parameter was 27 (higher than 15), but the gamma distribution led to smaller AIC and SIC. For the monthly series, the normal distribution seems less adequate (Chi-square test, 5% critical level). Furthermore, the shape parameter of the gamma distribution was 11, which indicates that the normal distribution is not appropriate.

When a series is not normally distributed, some transformations of the series (square root, cubic root, logarithm, etc.) can be applied to normalize it. However, these transformations are not efficient in all cases. Furthermore, the interpretation of a change in the parameters in the transformed series can cause problems (Jaruskova, 1997). As the gamma distribution can approximate the normal distribution, it gives a more general frame which does not necessitate any data transformations.

6. Concluding remarks

Hydrologic and climatic time series are often short. An advantage of reducing the time scale is that the observation series are longer and this can help to increase the power of detection of the test (Lund *et al.*, 2007). The BGHT test presented in this paper will be useful to detect inhomogeneities in precipitation series at different time scales. The BGHT is based on the Bayesian approach which allows the integration of different sources of information in the analysis. The information contained in the metadata can be incorporated in the prior distribution.

Even though metadata was available for the case study, the test was applied with different prior probabilities of no-change to study the sensitivity of the technique to the choice of prior.

In future work, the performance of the technique to detect inhomogeneities and the false detection rates should be assessed in a simulation study similar to the one presented in Beaulieu et al (2008a). The sensitivity of the test to the choice of the prior parameters should also be studied in detail. As the monthly series are likely to have a significant autocorrelation, its effect on the false detection rates and on the performance of the test to detect a shift should be verified.

The BGHT can also be applied to detect a change in the scale parameter of a precipitation series. This test could be applied to homogenize or to detect changes in other climatic series having a gamma distribution, such as wind speed for example. Furthermore, the posterior probabilities for a changepoint in a Poisson distribution following the methodology of Lee (1998) could be derived. Thus, the test could be applied to detect changes or inhomogeneities in the occurrences and amounts of daily precipitation series using a Poisson-gamma model.

Acknowledgements

The authors wish to thank the Fonds Québécois de Recherche sur la Nature et les Technologies (FQRNT), the National Sciences and Engineering Research Council of Canada (NSERC) and the Canada Research Chair Program for funding this research. The authors would also like to thank Mrs L. Vincent of the Meteorological Service of Canada for providing the metadata.

Notations

α	shape parameter of the gamma distribution
Γ	gamma function
η	scale parameter of the gamma distribution
η_1	scale parameter before the shift
η_2	scale parameter after the shift
Θ	parameter space for θ_1 and θ_2
θ	parameter of a distribution
θ_1	parameter of a distribution before the shift
θ_2	parameter of a distribution after the shift
$\pi(\theta)$	prior distribution of the parameter θ
$\pi(\theta_1 k)$	prior distribution of the parameter θ_1 given the position of the shift (under the no-change hypothesis)
$\pi(\theta_1, \theta_2 k)$	prior distribution of the parameters θ_1 and θ_2 given the position of the shift (under the change hypothesis)
$\pi(\theta m, c)$	prior distribution of the parameter θ with the hyperparameters m and c
$\pi(k \mathbf{x})$	posterior distribution of k given the observations \mathbf{x}
$\varphi(\theta)$	function of θ in the exponential family form
AIC	Akaike information criterion
c	hyperparameter for the prior distribution of the scale parameter
\hat{c}_{1k}	ML-II hyperparameter for the prior distribution of η_1
\hat{c}_{2k}	ML-II hyperparameter for the prior distribution of η_2

d	number of parameters of the distribution
F_{θ_1}	distribution before the shift
F_{θ_2}	distribution after the shift
$f(\mathbf{x} \theta)$	probability density of \mathbf{x} given the parameter θ
G	gamma distribution
$g(k)$	prior probability distribution of the shift's position
k	position of the shift
L	maximized likelihood function
m	hyperparameter for the prior distribution of the scale parameter
\hat{m}_{1k}	ML-II hyperparameter for the prior distribution of η_1
\hat{m}_{2j}	ML-II hyperparameter for the prior distribution of η_2
n	number of observations
p	prior probability of no-change
SIC	Schwarz information criterion
$S(x)$	function of x in the exponential family form
X	sample space for \mathbf{x}
\mathbf{x}	vector of observations
\bar{x}_k	mean of the first k observations
\bar{x}_k^*	mean of the last $n - k$ observations
\bar{x}_n	overall mean of the series

References

- Akaike, H. (1974), A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19, 716-723.
- Alexandersson, H. (1985), A simple stochastic model of the precipitation process. *J. Climate Appl. Meteor.*, 24, 1285-1295.
- Alexandersson, H. (1986), A homogeneity test applied to precipitation data, *J. Climatol.*, 6, 661-675.
- Beaulieu, C., T. B. M. J. Ouarda, and O. Seidou (2007), A review of homogenization techniques for precipitation data and their applicability to precipitation series (in French), *Hydrol. Sci. J.*, 52, 18-37.
- Beaulieu, C., T. B. M. J. Ouarda, O. Seidou and X. Zhang (2008a), A Bayesian Normal Homogeneity Test for the detection of artificial discontinuities in climatic series. To be *submitted*.
- Beaulieu, C., O. Seidou, T. B. M. J. Ouarda, X. Zhang, G. Boulet, and A. Yagouti (2008b), Intercomparison of homogenization techniques for precipitation data. *Water Resour. Res.*, 44, W02425.
- Berger, J. O. (1985), Statistical decision theory and Bayesian analysis. 2nd edition, Springer, New York, 617 p.
- Bobée, B. and F. Ashkar (1991), The gamma family and derived distributions in hydrology. Water Resources Publications, Littleton, Colorado, 203 p.

- Bridges, T. C. and C. T. Haan (1972), Reliability of precipitation probabilities estimated from the gamma distribution. *Mon. Wea. Rev.*, 100, 607-611.
- Briggs, W. M. and D. S. Wilks (1996), Estimating monthly and seasonal distributions of temperature and precipitation using the new CPC Long-Range forecasts. *J. Climate*, 9, 818-826.
- Buishand, T. A. (1989), Statistics of extremes in climatology. *Statistica Neerlandica*, 43, 1-29.
- Diaz, J. (1982), Bayesian detection of a change of scale parameter in sequences of independent gamma random variables. *J. Econometrics*, 19, 23-29.
- Ducré-Robitaille, J. F., G. Boulet, and L. A. Vincent (2003), Comparison of techniques for detection of discontinuities in temperature series. *Int. J. Climatol.*, 23, 1087-2003.
- Dunn, P. K. (2004), Occurrence and quantity of precipitation can be modelled simultaneously. *Int. J. Climatol.*, 24, 1231-1239.
- Easterling, D. R. and T. C. Peterson (1995), A new method for detecting undocumented discontinuities in climatological time series, *Int. J. Climatol.*, 15, 369-377.
- Ghorbanzadeh, G. and R. Lounes (2001), Bayesian analysis for detecting a change in exponential family. *Appl. Math. Comput.*, 124, 1-15.
- Groisman, P. Y. and D. R. Easterling (1994), Variability and trends of total precipitation and snowfall over the United States and Canada. *J. Climate*, 7, 184-205.
- Heino, R. (1997), Metadata and their role in homogenization. Proc. First Seminar for Homogenization of Surface Climate Data, Hungarian meteorological service, Budapest, Hungary.

- Hsu, A. (1979), Detecting shifts in gamma sequences with applications to stock price and air traffic flow analysis. *J. Am. Statist. Assoc.*, 74, 31-40.
- Ison, N. T., Feyerherm, A. M., and L. D. Bark (1971), Wet period precipitation and the gamma distribution. *J. Appl. Meteor.*, 10, 658-665.
- Jaruskova, D. (1997), Some problems with application of change-point detection methods to environmental data. *Environmetrics*, 8, 469-483.
- Juras, J. (1994), Some common features of probability distributions for precipitation. *Theor. Appl. Climatol.*, 49, 69-76.
- Kander, Z. and S. Zacks (1966), Tests procedures for possible changes in parameters of statistical distributions occurring at unknown time points. *The Ann. Math. Statist.*, 37, 1196-1210.
- Karl, T. R. and C. N. Jr. Williams (1987), An approach to adjusting climatological time series for discontinuous inhomogeneities, *J. Climate Appl. Meteor.*, 26, 1744-1763.
- Katz, R. W., Parlange, M. B. and P. Naveau (2002), Statistics of extremes in hydrology. *Adv. Water Res.*, 25, 1287-1304.
- Lanzante, J. R. (1996), Resistant, robust and non-parametric techniques for the analysis of climate data: theory and examples, including applications to historical radiosonde station data, *Int. J. Climatol.*, 16, 1197-1226.
- Lee, C. B. (1998), Bayesian analysis of a change-point in exponential families with applications. *Comp. Statist. Data Anal.*, 27, 195-208.
- Lund, R., X. L. Wang, Q. Lu, J. Reeves, C. Gallagher, and Y. Feng (2007), Changepoint detection in periodic and autocorrelated time series. *J. Climate*, 20, 5178-5190.

- Mestre, O. (2000), Méthodes statistiques pour l'homogénéisation des données climatiques. Ph.D. Thesis, Université Paul Sabatier, Toulouse, 230 p.
- Mooley, D.A. (1973), Gamma distribution probability model for asian summer monsoon monthly rainfall. *Mon. Wea. Rev.*, *101*, 160-176.
- Ozturk, A. (1982), On the study of a probability distribution for precipitation totals. *J. Appl. Meteor.*, *20*, 1499-1505.
- Perreault, L., M. Haché, M. Slivitzsky, and B. Bobée (1999), Detection of changes in precipitation and runoff over eastern Canada and U.S. using a Bayesian approach. *Stoch. Environ. Res. Risk Assess.*, *13*, 201-216.
- Peterson, T. C., D. R. Easterling, T. R. Karl, P. Groisman, N. Nicholls, N. Plummer, S. Torok, I. Auer, R. Boehm, D. Gullett, L. Vincent, R. Heino, H. Tuomenvirta, O. Mestre, T. Szentimrey, J. Salinger, E. J. Forland, I. Hanssen-Bauer, H. Alexandersson, P. Jones, and D. Parker (1998), Homogeneity adjustments of in situ atmospheric climate data: a review, *Int. J. Climatol.*, *18*, 1493-1517.
- Pettitt, A. N. (1979), A non-parametric approach to the change-point problem. *Appl. Statist.*, 126-135.
- Potter, K.W. (1981), Illustration of a new test for detecting a shift in mean in precipitation series, *Mon. Wea. Rev.*, *109*, 2040-2045.
- Reeves, J., J. Chen, X. L. Wang, R. Lund, and Q. Lu (2007), A review and comparison of changepoint detection techniques for climate data. *J. App. Meteor. Climatol.*, *46*, 900-915.

- Revfeim, K. J. A. (1985), A note on the comparison of theoretical and empirical quantiles for monthly rainfall totals. *Atmosphere-Ocean*, 23, 414-419.
- Robert, C. (1992), L'analyse statistique Bayésienne. Economica, Paris, 394 p.
- Ropelewski, C. F., Janowiak, J. E., and M. S. Halpert (1985), The analysis and display of real time surfaces climate data. *Mon. Wea. Rev.*, 113, 1101-1106.
- Schwarz, G. (1978), Estimating the dimension of a model. *Ann. Statist.*, 6, 461-464.
- Smith, A. F. M. (1975), A Bayesian approach to inference about a change-point in a sequence of random variables. *Biometrika*, 62, 407-416.
- Thom, H. C. S. (1958), A note on the gamma distribution. *Mon. Wea. Rev.*, 86, 117-122.
- Thompson, C. S. (1984), Homogeneity analysis of rainfall series: an application of the use of a realistic rainfall model. *J. Climatol.*, 4, 609-619.
- Vincent, L. A. (12 (2002), Future changes in the distribution of daily precipitation totals across North America. *Geophys. Res. Lett.*, 29, 1135.
- Wilks, D. S. (1990), Maximum likelihood estimation for the Gama distribution using data containing zeros. *J. Climate*, 3, 1495-1501.
- Wilks, D. S. (2006), Statistical methods in the atmospheric sciences. 2nd Edition, International Geophysics Series, Academic Press, London, 627 p.
- Wilks, D. S. and K. L. Eggleston (1992), Estimating monthly and seasonal precipitation distributions using the 30- and 90- day outlooks. *J. Climate*, 5, 252-259.

WMO, (2003), Guidelines on climate metadata and homogenization. *Report WMO-TD 1186*,
World Meteorological Organization, Geneva, Switzerland, 50 pp.

Worsley, K. J. (1986), Confidence regions and tests for a change-point in a sequence of
exponential family random variables. *Biometrika*, 73, 91-104.

Yue, S. and M. Hashino (2007), Probability distribution of annual, seasonal and monthly
precipitation in Japan. *Hydrol. Sci. J.*, 52, 863-877.

Table

Table 1. Independence (Wald-Wolfowitz test) and goodness of fit of the normal and gamma distributions (Chi-square test) to the series of ratios at the 5% critical level. Comparison of the two distributions with the AIC and SIC.

Time scale	Segment	Independence	Goodness of fit		AIC		SIC	
			G	N	G	N	G	N
Annual	1944-1982	R	A	A	-42.08	-43.81	-38.80	-40.53
	1944-1958	A	A	A	-18.93	-18.39	-17.65	-17.11
	1959-1982	A	A	A	-54.39	-54.71	-52.04	-52.36
Seasonal	Spring 1944-Fall 1982	R	R	R	-23.49	-23.13	-17.47	-17.11
	Spring 1944-Summer 1959	A	A	R	-20.74	-20.37	-16.65	-16.28
	Fall 1959-Fall 1982	A	A	R	-39.43	-35.80	-34.37	-30.73
Monthly	February 1944-January 1983	R	R	R	334.40	360.67	342.65	368.91
	February 1944-October 1959	A	A	A	85.60	95.05	91.95	101.40
	November 1959-January 1983	R	A	R	201.74	218.13	209.00	225.39

G: gamma distribution
 N: normal distribution
 R: hypothesis rejected
 A: hypothesis accepted

Figure captions

Figure 1. Precipitation totals of the base station and the two neighbor stations at different time scales.

Figure 2. Ratios of annual precipitation totals between the base series and the mean of the two neighbor series a) time series b) histogram.

Figure 3. Posterior probability distributions for the position of the change in the series of ratios (annual precipitation totals).

Figure 4. Ratios of seasonal precipitation totals between the base series and the mean of the two neighbor series a) time series b) histogram.

Figure 5. Posterior probability distributions for the position of the change in the series of ratios (seasonal precipitation totals).

Figure 6. Ratios of monthly precipitation totals between the base series and the mean of the two neighbor series a) time series b) histogram.

Figure 7. Posterior probability distributions for the position of the change in the series of ratios (monthly precipitation totals).

Figure 8. Boxplot of the ratios of monthly precipitation totals between the base series and the mean of the two neighbor series.

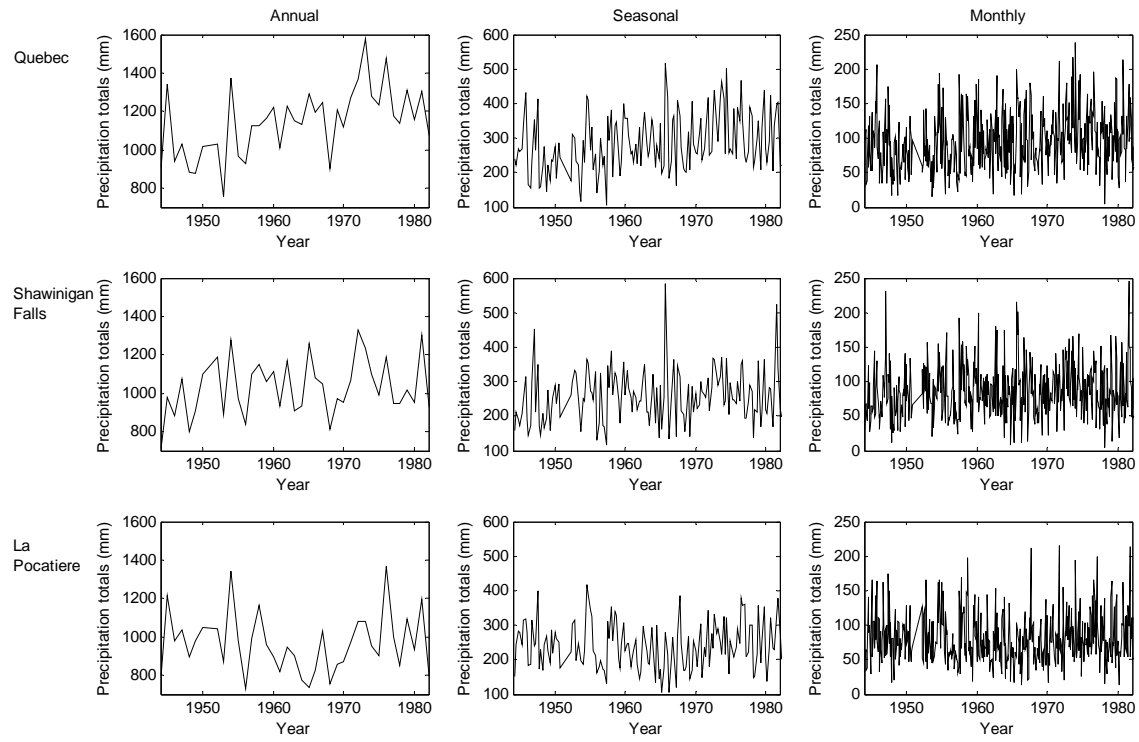


Figure 1.

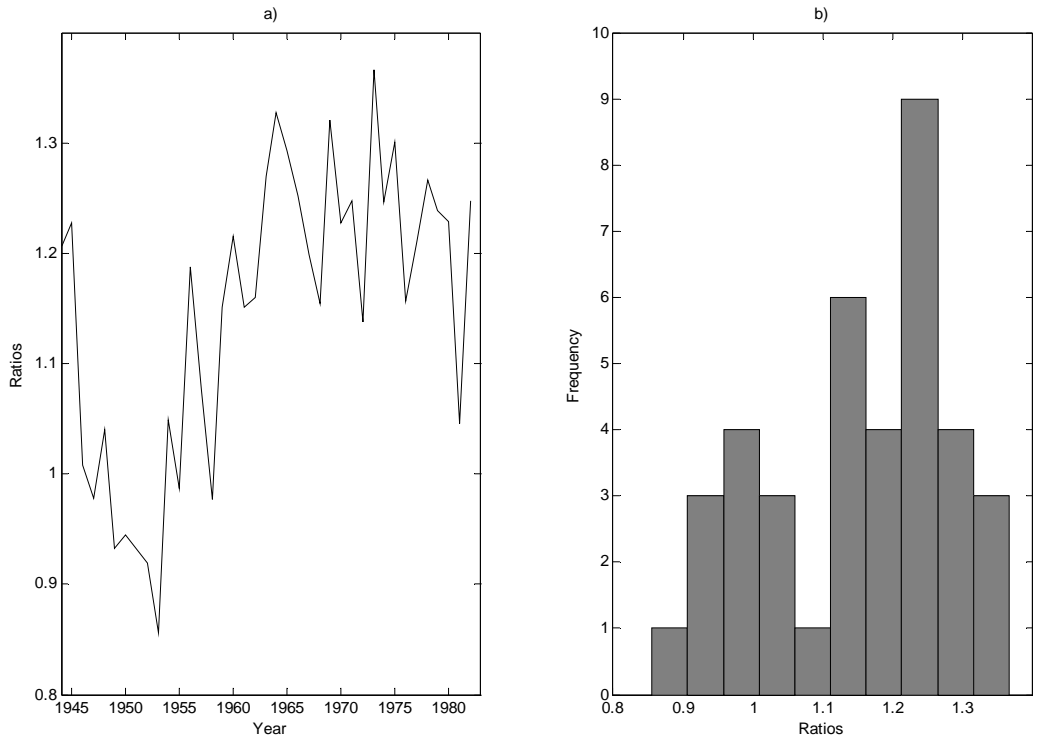


Figure 2.

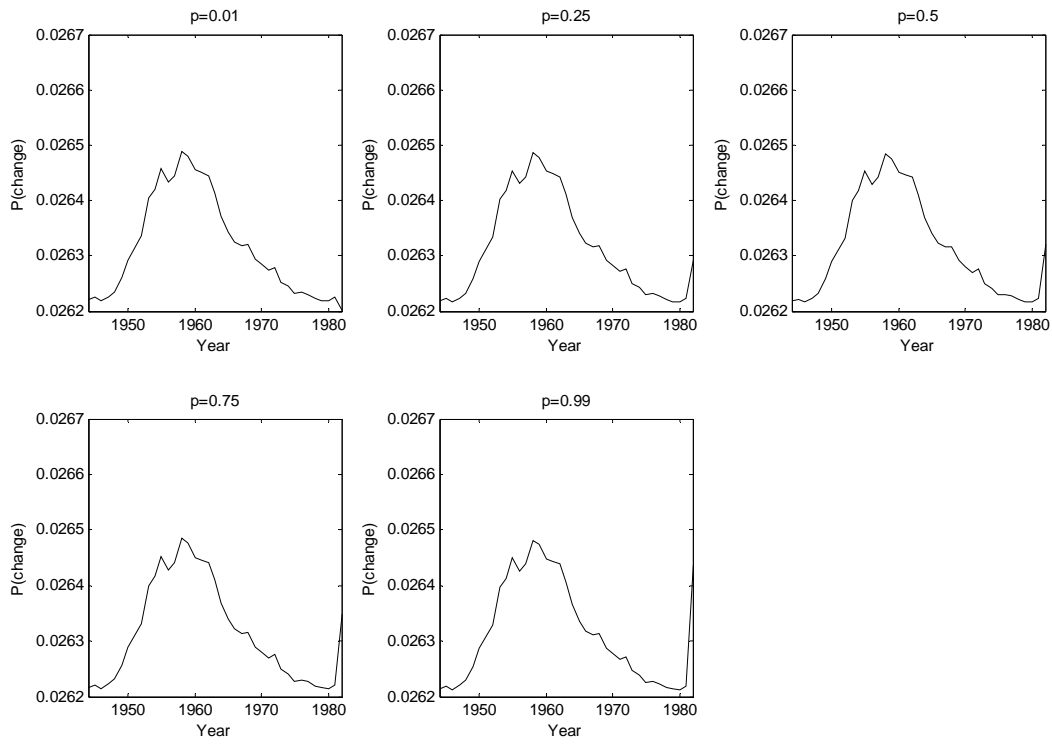


Figure 3.

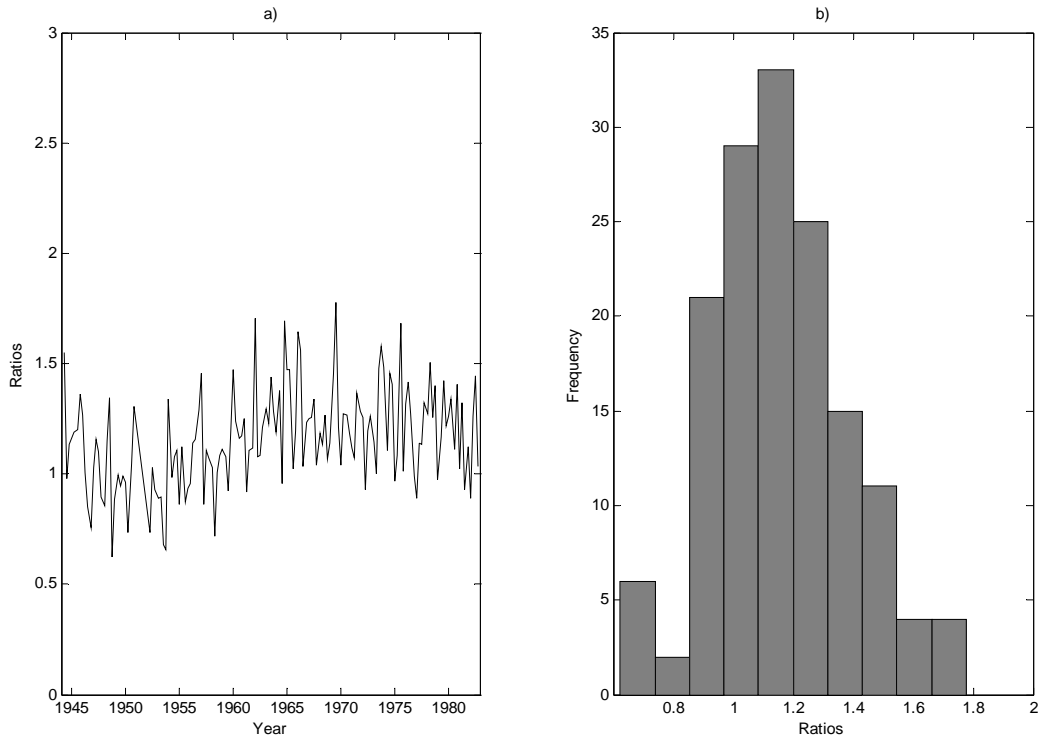


Figure 4.

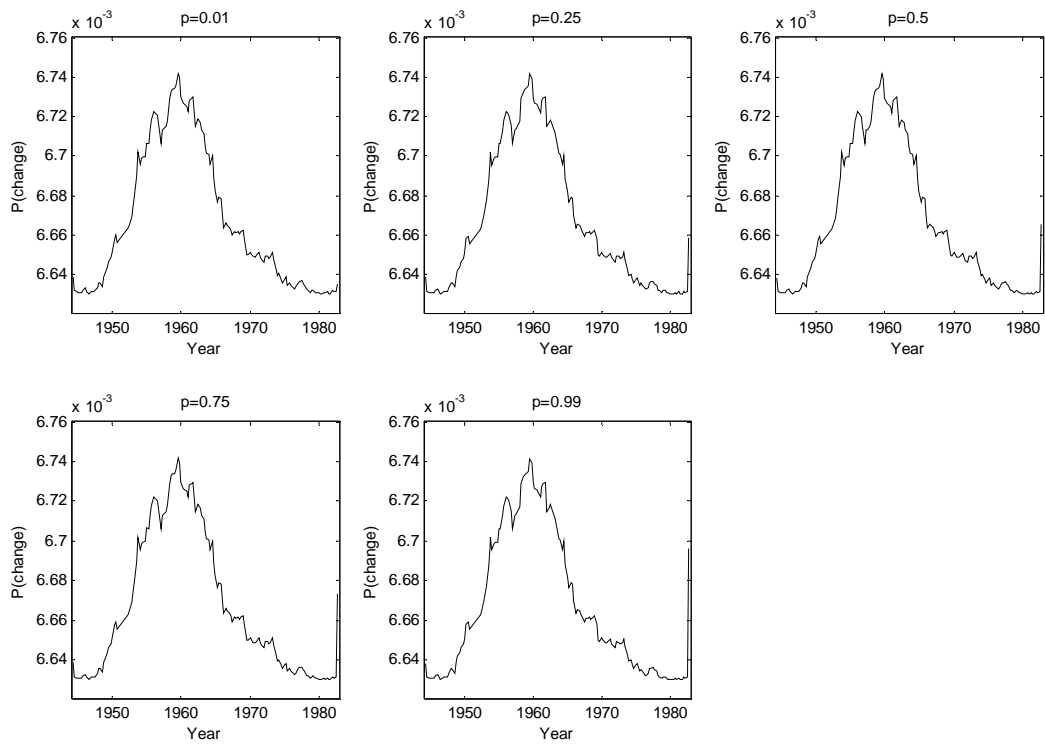


Figure 5.

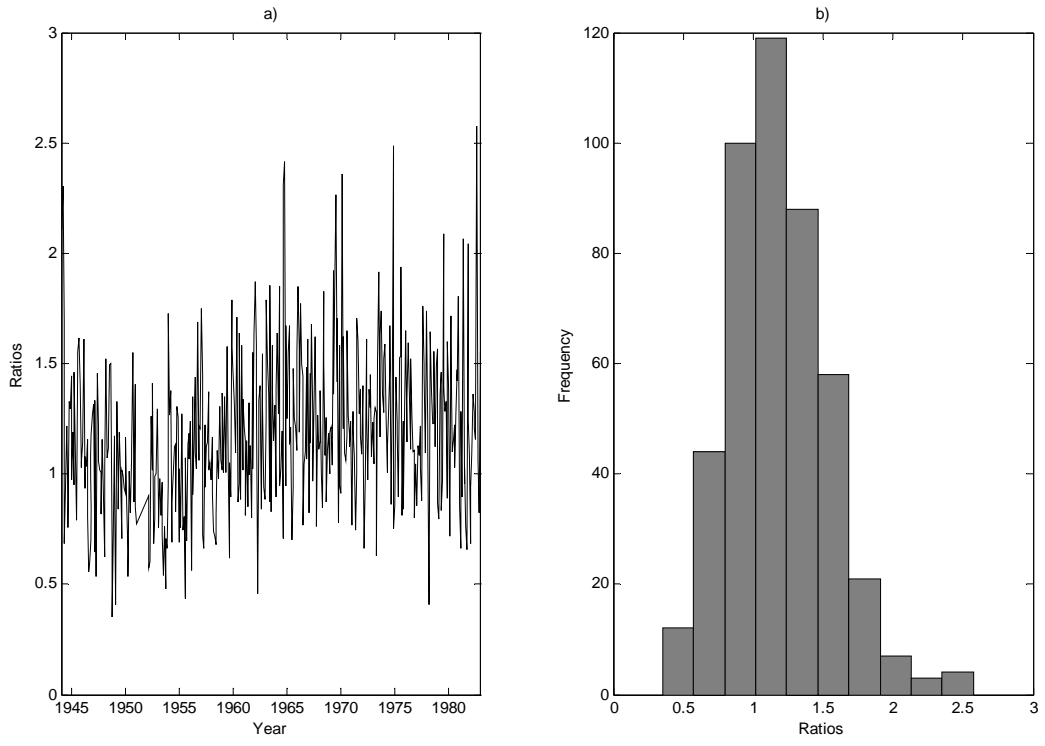


Figure 6.

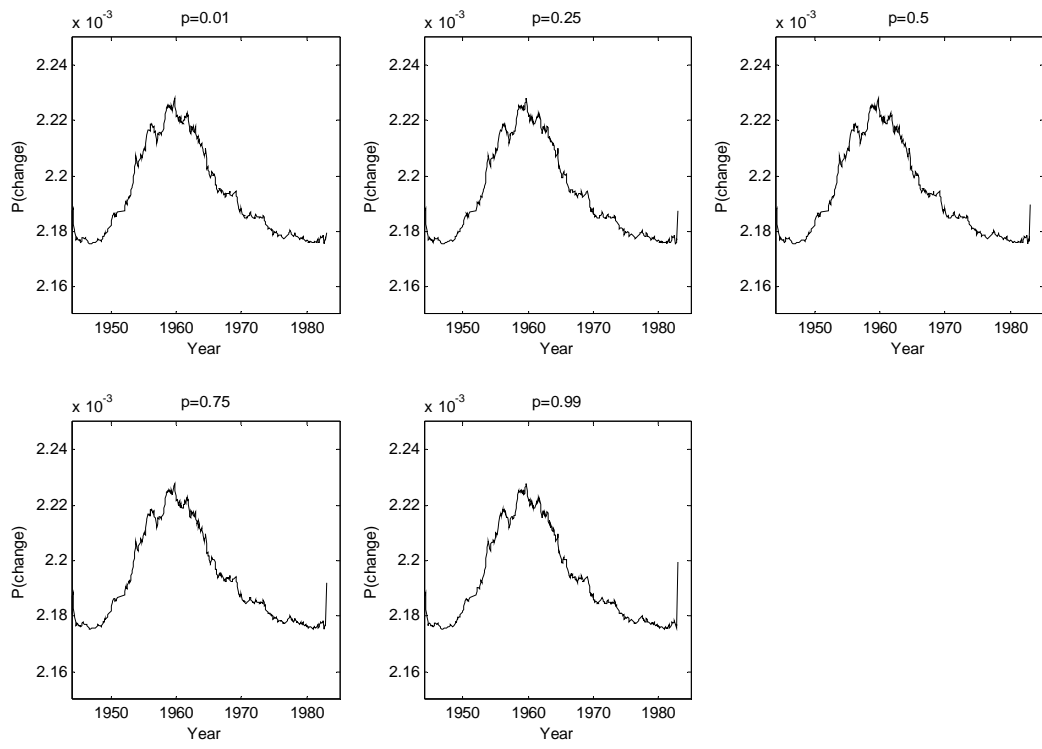


Figure 7.

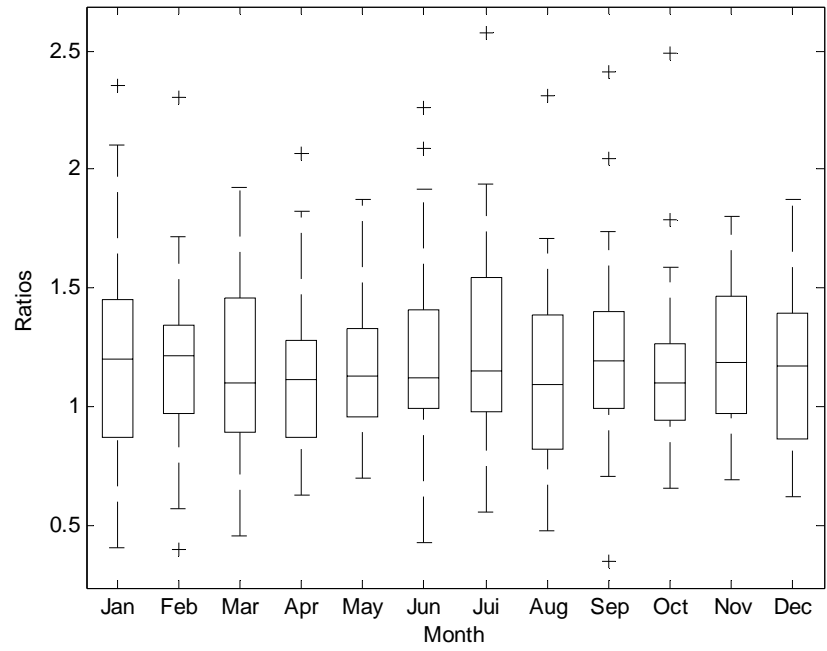


Figure 8

ANNEXES

11. Annexe A : Résultats non publiés

Dans cette annexe, quelques résultats qui n'ont pas été publiés sont présentés. Dans Beaulieu *et al.* (2008), l'effet de la distance entre deux sauts consécutifs et de la suite des signes consécutifs sur la performance de détection des différentes méthodes a été analysé. Dans les autres articles dans lesquels des techniques sont appliquées sur les mêmes séries synthétiques (Beaulieu *et al.*, 2009a, 2009b), ces analyses n'ont pas été présentées pour faire place à l'étude d'un cas pratique. Les figures A1 et A2 présentent ces résultats pour toutes les techniques comparées dans (Beaulieu *et al.*, 2008, 2009a, 2009b).

Dans Beaulieu *et al.* (2008), les techniques sont comparées sur des séries synthétiques avec un saut de variance et avec une tendance. L'amplitude et la position des sauts de variance sont déterminées de manière aléatoire : l'amplitude provient d'une distribution beta et varie entre 0 et 50% de l'écart-type et la position provient de la distribution uniforme discrète et varie entre 11 et 90. L'amplitude et la position du début et de la fin de la tendance sont également déterminées aléatoirement : l'amplitude provient d'une distribution beta et varie entre -3 et 3 écarts-types et la position du début et de la fin provient de la distribution uniforme discrète, varie entre 11 et 90 et elle est tronquée pour qu'il y ait au moins 10 observations entre le début et la fin de la tendance. Les résultats méthodes BARE, BAMS et BNHT ont été appliquées sur ces séries synthétiques, mais les résultats obtenus ne sont pas présentés dans Beaulieu *et al.* (2009a, 2009b). Les tableaux A1 et A2 présentent les résultats obtenus sur les séries avec un saut de variance pour toutes les techniques. Les tableaux A3 et A4 présentent les résultats obtenus sur les séries avec une tendance pour toutes les techniques.

Les méthodes bayésiennes BAMS, BARE et BNHT ont été appliquées sur des séries synthétiques qui contiennent différents niveaux d'autocorrélation (coefficients d'autocorrélation de 0, 0.2 et 0.4). Ces résultats sont présentés dans les figures A3 et A4.

I) Résultats sur les séries synthétiques avec plusieurs sauts

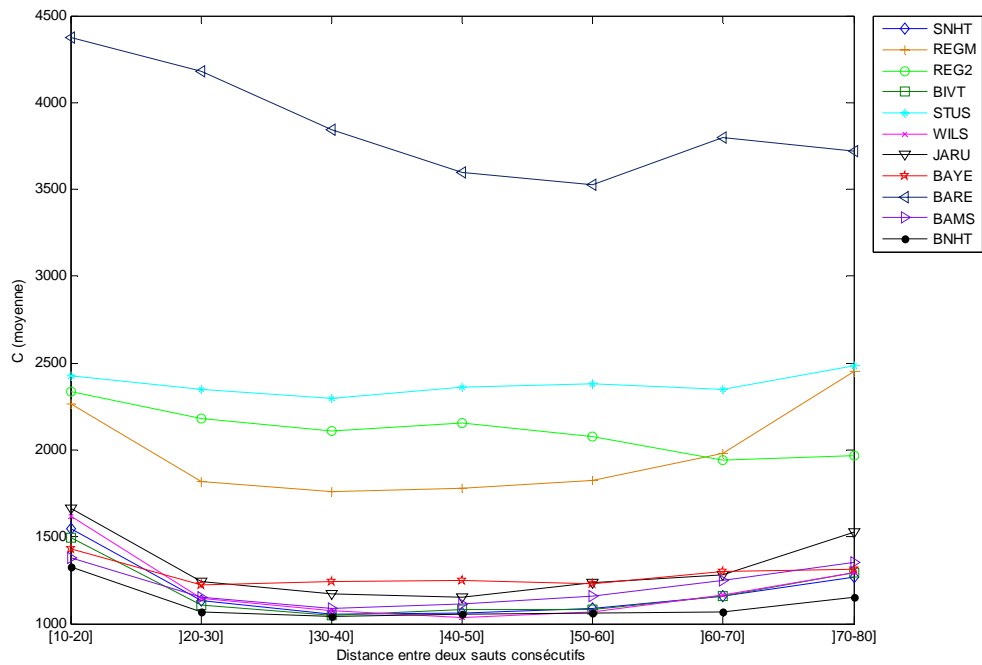


Figure A1 : Critère de performance moyen obtenu sur les séries avec deux sauts en fonction de la distance entre deux sauts consécutifs.

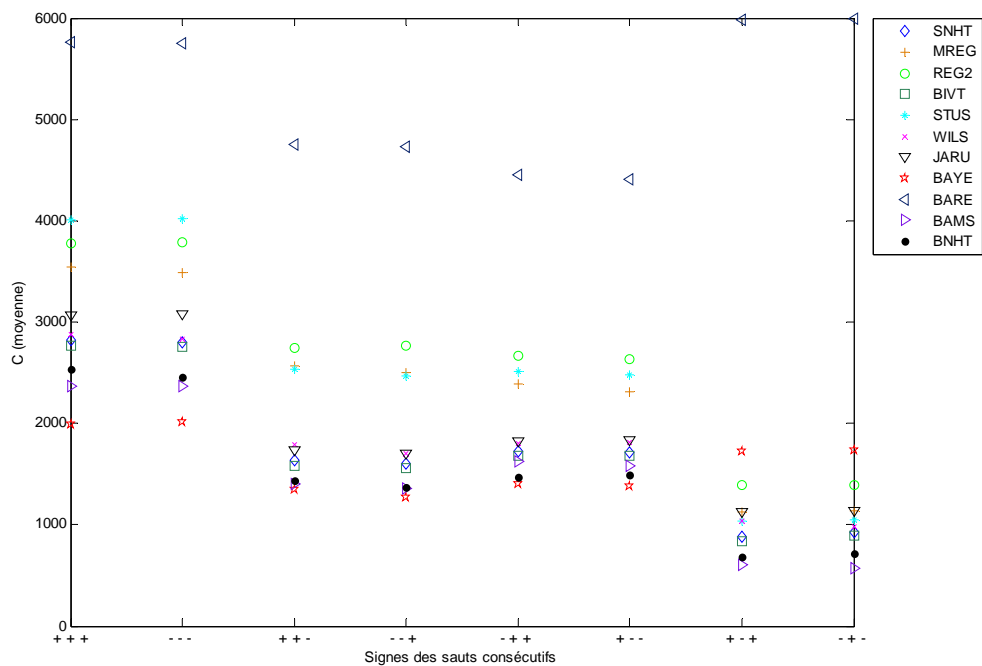


Figure A2 : Critère de performance moyen obtenu sur les séries avec trois sauts en fonction de l'ordre des signes des trois sauts consécutifs.

II) Résultats sur les séries synthétiques avec un saut de variance

Tableau A1 : Pourcentage de rejet de l'hypothèse d'homogénéité sur les séries avec un saut de variance.

Méthode	Rejet (%)
BAMS	40.8
BARE	7.5
BNHT	6.5
SNHT*	5.4
MREG*	2.9
REG2*	7.7
BIVT*	5.4
STUS*	4.7
WILS*	5.1
JARU*	2.4
BAYE*	27.7

*Résultat publié dans Beaulieu *et al.* (2008).

Tableau A2 : Différence entre les pourcentages de rejet de l'hypothèse d'homogénéité sur les séries avec un saut de variance et sur les séries homogènes.

Méthode	Borne inférieure de l'IC	Différence	Borne supérieure de l'IC
BAMS	37.2	38.3**	39.3
BARE	6.4	6.9**	7.5
BNHT	0.5	1.3**	2.1
SNHT*	0.7	1.4**	2.1
MREG*	1.2	1.7**	2.1
REG2*	2.5	3.3**	4.1
BIVT*	0.8	1.5**	2.2
STUS*	-1	-0.3	0.4
WILS*	-0.5	0.3	0.9
JARU*	0.8	1.2**	1.6
BAYE*	-1.3	0.2	1.7

*Résultat publié dans Beaulieu *et al.* (2008).

**Différence significative au seuil de 5%.

III) Résultats sur les séries synthétiques avec une tendance

Tableau A3 : Nombre de cas (%) pour lesquels un saut est détecté à l'intérieur de la tendance.

Amplitude (en écart-type)	BAMS	BARE	BNHT	SNHT*	MREG*	REG2*	BIVT*	STUS*	WILS*	JARU*	BAYE*
0-0.5	17.4	0.0	39.1	39.1	4.3	13.0	30.4	8.7	30.4	30.4	52.2
0.5-1	67.2	14.8	84.9	84.2	54.1	28.5	84.9	13.2	83.5	83.0	79.4
1-2	58.4	83.2	69.7	77.3	89.8	65.7	76.0	47.0	76.8	81.5	63.2
2-3	27.7	98.8	51.8	56.1	80.9	77.0	55.3	64.5	52.6	58.2	52.2

*Résultat publié dans Beaulieu *et al.* (2008).

Tableau A4 : Nombre de cas (%) pour lesquels deux sauts ou plus sont détectés à l'intérieur de la tendance.

Amplitude (en écart-type)	BAMS	BARE	BNHT	SNHT*	MREG*	REG2*	BIVT*	STUS*	WILS*	JARU*	BAYE*
0-0.5	4.3	0.0	0.0	0.0	0.0	0.0	4.3	0.0	0.0	0.0	4.3
0.5-1	7.2	0.0	4.5	2.6	1.2	0.0	2.6	0.0	4.3	1.4	13.6
1-2	40.1	0.0	29.8	21.9	6.1	1.8	23.4	2.7	22.3	17.7	36.5
2-3	72.3	0.2	48.2	43.9	19.0	6.4	44.7	12.3	47.3	41.7	47.8

*Résultat publié dans Beaulieu *et al.* (2008).

IV) Résultats sur les séries synthétiques autocorrélées

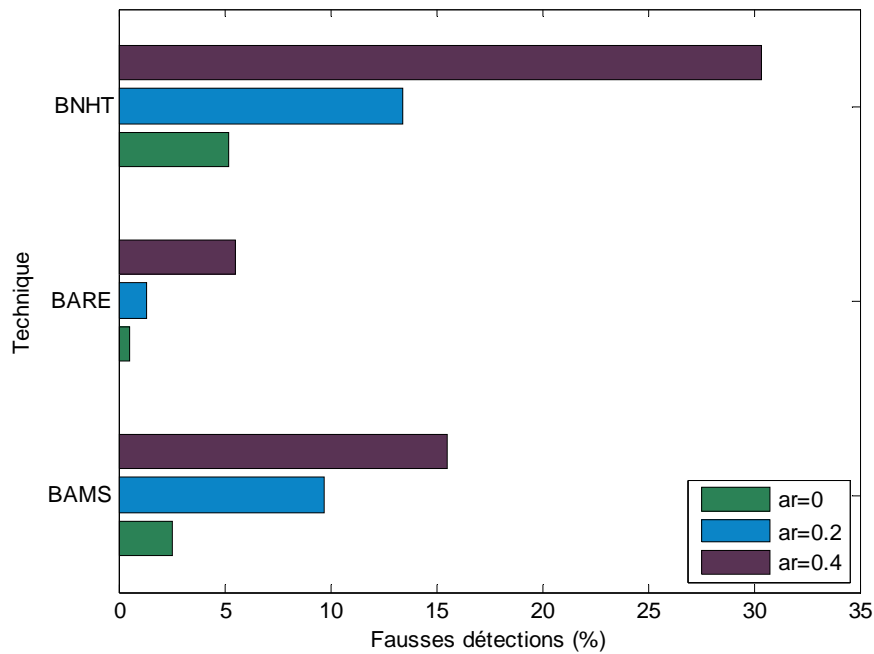


Figure A3 : Fausses détections (%) obtenues sur des séries homogènes qui contiennent différents niveaux d'autocorrélation (coefficient autorégressif d'ordre 1 de 0, 0.2 et 0.4).

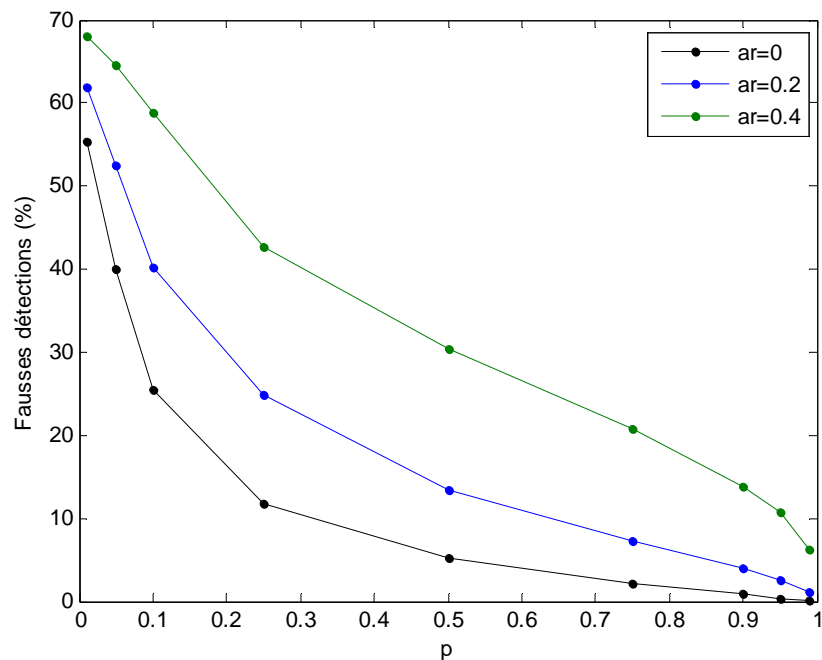


Figure A4 : Fausses détections (%) obtenues sur des séries homogènes qui contiennent différents niveaux d'autocorrélation (coefficient autorégressif d'ordre 1 de 0, 0.2 et 0.4) avec la technique BNHT appliquée avec différentes probabilités de non-changement a priori (p).

12. Annexe B : Pseudonymes utilisés pour désigner les techniques d'homogénéisation

Pseudonyme	Méthode	Référence(s)
AIC	Critère d'Akaike	-Akaike (1973)
BABI	Approche bayésienne bivariée pour un saut	-Asselin <i>et al.</i> (1999)
BAMS	Approche bayésienne multi-sauts	-Fearnhead (2006) -Seidou et Ouarda (2007)
BARE	Approche bayésienne régressive multi-sites	-Seidou <i>et al.</i> (2007)
BAUN	Approche bayésienne univariée pour un saut	-Lee et Heghinian (1977) -Ouarda <i>et al.</i> (1999) -Perreault <i>et al.</i> (1999)
BAYE	Méthode bayésienne univariée pour la détection d'un saut	-Rasmussen (2001)
BIVT	Approche bivariée	-Maronna & Yohai (1978) -Potter (1981)
BGHT	Test d'homogénéité bayésien (loi gamma)	-Lee (1998) -Beaulieu <i>et al.</i> (2009c)
BNHT	Test d'homogénéité bayésien (loi normale)	-Lee (1998) -Beaulieu <i>et al.</i> (2009b)
BUIT	Test sur les déviations cumulées	-Buishand (1982, 1984)
CAME	Méthode de Caussinus et Mestre	-Mestre (2000) -Caussinus et Mestre (2004)
CIC	Critère de Caussinus et Lyazrhi	-Caussinus et Lyazrhi (1997)
GAC2	Doubles accumulations	-Kohler (1949) -Bois (1986) -Brunet-Moret (1979)
GDCU	Analyse graphique des déviations cumulées	-Craddock (1979)
GDIF	Analyse graphique des séries de différences	-Jones <i>et al.</i> (1986)
GKZU	Filtre de Kolmogorov-Zurbenko	-Zurbenko <i>et al.</i> (1996)
GSCP	Sommes cumulatives parallèles	-Rhoades et Salinger (1993)
JARU	Méthode de Jaruskova	-Jaruskova (1996)
KEND	Test de tendance de Kendall	-Mann (1945) -Kendall (1975)
KWAL	Test de Kruskal-Wallis	-Kruskal et Wallis (1952)
NMETA	Test sur les quartiles	-Allen et DeGaetano (2000)
MASH	Méthode de Szentimrey	-Szentimrey (1996;1999)
ONDE	Méthode basée sur l'analyse des ondelettes	-Datsenko <i>et al.</i> (2002)
PMETA	Test de différences de moyennes local	-Karl et Williams (1987)
POIS	Homogénéisation par processus de Poisson	-Thompson (1984)

PSSH	Procédure de segmentation de série hydrométéorologique	-Hubert <i>et al.</i> (1989) -Kehagias <i>et al.</i> (2006)
PSTU	Test de Student pénalisé	-Wang <i>et al.</i> (2007)
PTTI	Test de Pettitt	-Pettitt (1979) -Boroneant et Tomozeiu (1999) -Wijngaard (2003)
REG2	Régression à deux phases	-Solow (1987) -Easterling et Peterson (1995) -Lund et Reeves (2002) -Wang (2003)
REGM	Régression multiple	-Vincent (1998)
REMU	Régression multiphase	-Gullett <i>et al.</i> (1991)
SIC	Critère de Schwarz	-Schwarz (1978)
SNHT	Test d'homogénéité standard normal (standard normal homogeneity test)	-Alexandersson (1986) -Alexandersson et Moberg (1997) -Khaliq et Ouarda (2007)
STUS	Test de Student séquentiel	-Gullett <i>et al.</i> (1990) -Ducré-Robitaille <i>et al.</i> (2003)
VNEU	Rapport de Von Neumann	-Von Neumann (1941)
WILS	Test de Wilcoxon séquentiel	-Lanzante (1996) -Ducré-Robitaille <i>et al.</i> (2003)
WWOL	Test de Wald-Wolfowitz	-Wald et Wolfowitz (1943)

13. Annexe C : Notation

Symbole	Définition
α	paramètre de forme
γ	coefficient d'asymétrie
β	coefficient de régression de la série de référence
β_j	coefficient de régression de la série voisine j
δ	amplitude du saut
η	paramètre d'échelle
λ_1	pende de la tendance avant le saut
λ_2	pende de la tendance après le saut
μ	moyenne
τ	ordonnée à l'origine
τ_1	ordonnée à l'origine avant le changement
τ_2	ordonnée à l'origine après le changement
e_i	résidus
k	nombre de séries voisines
n	longueur de la série de base
p	position du saut
q_i	série de comparaison entre la série de base et ses voisines au temps i
x_i	observation i de la série de référence
x_{ij}	observation i de la série voisine j
y_i	série de base au temps i

