

ÉTUDE COMPARATIVE DE MÉTHODES D'HOMOGENÉISATION DES SÉRIES DE PRÉCIPITATIONS

par

Claudie Beaulieu

Taha B.M.J. Ouarda

Ousmane Seidou

avec la collaboration de :

Gilles Boulet

Abderrahmane Yagouti

Chaire en hydrologie statistique Hydro-Québec/CRSNG
Chaire du Canada en estimation des variables hydrologiques

INRS-ETE

490, rue de la Couronne

Québec (Qc), Canada, G1K 9A9

Rapport de recherche N° R-849

Mars 2006

TABLE DES MATIÈRES

TABLE DES MATIÈRES	iii
LISTE DES FIGURES.....	v
LISTE DES TABLEAUX	vii
1. INTRODUCTION.....	1
2. MÉTHODOLOGIE.....	3
2.1 Sélection des méthodes d'homogénéisation	3
2.1.1 Méthode d'Alexandersson (snht)	3
2.1.2 Régression multiple (regm).....	4
2.1.3 Régression à deux phases (reg2)	5
2.1.4 Approche bivariée (bivt)	6
2.1.5 Test de Student séquentiel (stus).....	7
2.1.6 Test de Wilcoxon séquentiel (wils).....	8
2.1.7 Méthode de Jaruskova (jaru)	9
2.1.8 Méthode Bayésienne (rasm).....	9
2.2 Génération des séries synthétiques.....	11
2.2.1 Séries de base	12
2.2.2 Séries voisines	16
2.3 Critères d'évaluation des méthodes d'homogénéisation	17
2.3.1 Critère d'évaluation des méthodes sur les séries homogènes.....	17
2.3.2 Critère d'évaluation des méthodes sur les séries à un saut.....	18
2.3.3 Critère d'évaluation des méthodes sur les séries à plusieurs sauts.....	18
2.3.4 Critère d'évaluation des méthodes sur les séries avec un saut d'écart type	18
2.3.5 Critère d'évaluation des méthodes sur les séries avec une tendance	19
3. PERFORMANCES DES MÉTHODES ORIGINALES.....	21
3.1 Séries homogènes	21
3.2 Séries avec un saut de moyenne.....	25
3.3 Séries avec plusieurs sauts de moyenne	33
3.4 Séries avec un saut d'écart-type	35
3.5 Séries avec une tendance	37

4. PERFORMANCE DES DIFFÉRENTES MÉTHODES LORSQUE LES AMPLITUDES DES SAUTS SONT ESTIMÉES PAR DIFFÉRENCES DE MOYENNE.....	41
4.1 Séries homogènes	42
4.2 Séries avec un saut de moyenne.....	45
4.3 Séries avec plusieurs sauts de moyenne	53
4.3.1 Performance évaluée par sommes de carrés résiduelles.....	53
4.3.2 Performance évaluée par un critère de positionnement.....	54
5. DISCUSSION.....	61
6. CONCLUSION.....	67
7. BIBLIOGRAPHIE	69

LISTE DES FIGURES

Figure 1 :	Sauts faussement détectés sur les séries homogènes de 100 ans a) méthode de Jaruskova b) régression multiple c) approche bivariée d) méthode d'Alexandersson e) régression à deux phases f) test de Wilcoxon séquentiel g) test de Student séquentiel h) méthode Bayésienne.	24
Figure 2 :	Sauts identifiés (%) dans les séries avec un saut de moyenne de longueur 60 ans I) correctement identifiés II) bien identifiés a) méthode d'Alexandersson b) régression multiple c) régression à deux phases d) approche bivariée e) test de Student séquentiel f) test de Wilcoxon séquentiel g) méthode de Jaruskova h) méthode Bayésienne.	29
Figure 3 :	Sauts identifiés (%) dans les séries avec un saut de moyenne de longueur 100 ans I) correctement identifiés II) bien identifiés a) méthode d'Alexandersson b) régression multiple c) régression à deux phases d) approche bivariée e) test de Student séquentiel f) test de Wilcoxon séquentiel g) méthode de Jaruskova h) méthode Bayésienne.	32
Figure 4 :	Exemple de série de base contenant trois sauts.	41
Figure 5 :	Sauts faussement détectés sur les séries homogènes de 100 ans a) régression multiple b) méthode de Jaruskova c) approche bivariée d) méthode d'Alexandersson e) régression à deux phases f) test de Wilcoxon séquentiel g) test de Student séquentiel h) méthode Bayésienne.	45
Figure 6 :	Sauts identifiés (%) en estimant l'amplitude par différences de moyennes dans les séries avec un saut de moyenne de longueur 60 ans I) correctement identifiés II) bien identifiés a) méthode d'Alexandersson b) régression multiple c) régression à deux phases d) approche bivariée e) test de Student séquentiel f) test de Wilcoxon séquentiel g) méthode de Jaruskova h) méthode Bayésienne.	49
Figure 7 :	Sauts identifiés (%) en estimant l'amplitude par différences de moyennes dans les séries avec un saut de moyenne de longueur 100 ans I) correctement identifiés II) bien identifiés a) méthode d'Alexandersson b) régression multiple c) régression à deux phases d) approche bivariée e) test de Student séquentiel f) test de Wilcoxon séquentiel g) méthode de Jaruskova h) méthode Bayésienne.	52
Figure 8 :	Histogramme du critère de positionnement selon les huit méthodes des séries avec deux sauts.	56
Figure 9 :	Histogramme du critère de positionnement selon les huit méthodes des séries avec trois sauts.	57
Figure 10 :	Critères de positionnement moyens selon la distance entre les deux sauts.	58
Figure 11 :	Critères de positionnement moyens selon la suite des signes des sauts.	59

Figure 12 : Algorithme de la méthode hybride proposée.65

LISTE DES TABLEAUX

Tableau 1.	Sauts faussement détectés (%) sur les séries homogènes de 60 ans.	21
Tableau 2.	Sauts faussement détectés (%) sur les séries homogènes de 100 ans.	22
Tableau 3.	Sauts identifiés (%) sur les séries de 60 ans avec un saut de moyenne.	26
Tableau 4.	Sauts identifiés (%) sur les séries de 100 ans avec un saut de moyenne.	26
Tableau 5.	Rapport entre les sommes de carrés résiduelles espérées et observées après correction des séries avec deux sauts.	34
Tableau 6.	Rapport entre les sommes de carrés résiduelles espérées et observées après correction des séries avec trois sauts.	35
Tableau 7.	Sauts faussement détectés (%) sur les séries avec un saut d'écart-type de 100 ans.	36
Tableau 8.	Différence (%) des proportions entre les sauts faussement détectés sur les séries homogènes de 100 ans et sur les séries avec un saut d'écart-type.	37
Tableau 9.	Nombre de cas (%) pour lesquels un saut est détecté à l'intérieur de la tendance.	38
Tableau 10.	Nombre de cas (%) pour lesquels deux sauts ou plus sont détectés à l'intérieur de la tendance.	38
Tableau 11.	Sauts faussement détectés (%) sur les séries homogènes de 60 ans.	43
Tableau 12.	Sauts faussement détectés (%) sur les séries homogènes de 100 ans.	43
Tableau 13.	Sauts identifiés (%) sur les séries de 60 ans avec un saut de moyenne.	46
Tableau 14.	Sauts identifiés (%) sur les séries de 100 ans avec un saut de moyenne.	46
Tableau 15.	Rapport entre les sommes de carrés résiduelles espérées et observées après correction des séries avec deux sauts.	53
Tableau 16.	Rapport entre les sommes de carrés résiduelles espérées et observées après correction des séries avec trois sauts.	53
Tableau 17.	Statistiques descriptives du critère de positionnement des séries avec deux sauts.	55
Tableau 18.	Statistiques descriptives du critère de positionnement des séries avec trois sauts.	55
Tableau 19.	Résumé des forces et faiblesses des méthodes d'homogénéisation comparées.	63

1. Introduction

Les caractéristiques statistiques des enregistrements a 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. Par conséquent, il peut arriver que des décisions soient prises en se basant sur des données qui contiennent des erreurs.

Les données des réseaux de mesure de précipitations sont utilisées dans la plupart des modèles hydrologiques et climatiques. La fiabilité de ces données devrait être vérifiée avant de les inclure dans un modèle. En effet, un changement dans le climat régional d'une station de mesure pourrait se témoigner par l'observation d'une augmentation ou réduction des précipitations totales à une station de mesure, mais ce changement dans la série de précipitation pourrait également être le résultat d'une modification dans l'exposition de la station ou dans la prise de mesure. De ce fait, il est très important de développer des techniques robustes de détection de ce type de perturbations pour que les données utilisées se rapprochent le plus possible de la réalité. Le processus de détection et de correction des ruptures d'origine non climatique est appelé homogénéisation.

Ce rapport présente les résultats d'une étude comparative de méthodes d'homogénéisation sur des séries synthétiques de précipitations totales annuelles qui possèdent des caractéristiques statistiques typiques des précipitations totales observées au Québec. Le reste du rapport est

organisé comme suit : a) méthodes comparées et séries synthétiques générées; b) présentation des résultats; c) discussion des résultats et finalement d) conclusion.

2. Méthodologie

La méthodologie adoptée dans ce travail consiste à 1) sélectionner les méthodes d'homogénéisation qui semblent prometteuses pour les données de précipitation, 2) générer des séries synthétiques de différentes configurations et 3) estimer la performance des méthodes sélectionnées sur les séries synthétiques.

2.1 Sélection des méthodes d'homogénéisation

Dans le but d'identifier les méthodes qui semblent les plus prometteuses pour l'homogénéisation des précipitations, une revue de littérature des méthodes d'homogénéisation des données climatiques a été réalisée dans un premier temps (Beaulieu et al. 2005). Un certain nombre de méthodes ont été retenues sur la base de leur potentiel d'applicabilité aux séries de précipitations. Les méthodes retenues sont la méthode d'Alexandersson (Alexandersson, 1986), la régression multiple (Vincent, 1998), la régression à deux phases (Easterling & Peterson, 1995; Lund & Reeves, 2002), l'approche bivariée (Maronna & Yohai, 1978; Potter, 1981), le test de Student séquentiel (Gullett et al., 1990) et de Wilcoxon séquentiel (Karl & Williams, 1987; Ducré-Robitaille et al., 2003), la méthode développée par Jaruskova (1996) et une méthode Bayésienne (Rasmussen, 2001). Cette section présente les méthodes comparées.

2.1.1 Méthode d'Alexandersson (snht)

Une série de ratios est créée entre la série de base et les séries voisines:

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

La valeur de l'année i de la série de base est représentée par y_i , alors que x_{ij} dénote l'observation i de la série voisine j . Il y a k séries voisines et n observations dans chaque série. Le coefficient de corrélation entre la série de base et la série voisine j est noté par ρ_j . On teste l'hypothèse que les ratios standardisés suivent une distribution normale de moyenne nulle et de variance 1 alors que la contre-hypothèse est qu'il y a un changement de moyenne dans la série. Pour trouver l'emplacement de ce point de changement, une série de moyennes pondérées est créée :

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

où $\bar{z}_{1:i}$ et $\bar{z}_{i+1:n}$ sont les moyennes des ratios standardisés pour les segments de 1 à i et de $i+1$ jusqu'à n , le nombre total d'années dans la série. La statistique du test, $Q_p = \max_{i=1, \dots, n-1} \{|Q_i|\}$, est significative si elle dépasse sa valeur critique (Alexandersson 1986) et son indice correspond à la position du saut (p).

2.1.2 Régression multiple (regm)

Cette approche se base sur l'application de quatre modèles de régression multiple représentant divers types d'inhomogénéités (Vincent 1998). Pour les besoins de cette étude, deux modèles sont utilisés. Le premier sert à déterminer si la série de base est homogène. L'homogénéité de la série est vérifiée par un test d'indépendance des résidus (e.g. un intervalle de confiance sur l'autocorrélation d'ordre 1 des résidus). Lorsque les résidus sont indépendants, le modèle ajuste bien les données et la série semble homogène. Par contre, si les résidus sont autocorrélés, cela indique que le modèle n'ajuste pas bien les données et qu'il pourrait y avoir une inhomogénéité dans la série de base. On applique alors le modèle 3 décrivant un saut dans la série de base:

$$y_i = \begin{cases} \tau + \beta_{31}x_{i1} + \dots + \beta_{3k}x_{ik} + e_i & i = 1, \dots, p-1 \\ \tau + \delta_p + \beta_{31}x_{i1} + \dots + \beta_{3k}x_{ik} + e_i & i = p, \dots, n \end{cases} \quad (3)$$

où y_i représente l'observation i de la série de base, x_{ik} représente l'observation i de la série de référence k , τ et β_{1j} ($j=1, \dots, k$) sont les estimateurs aux moindres carrés. Il y a k séries voisines au total et n observations pour chacune. Les résidus, e_i , sont distribués selon une loi normale $N(0, \sigma_e^2)$. 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. Si les résidus sont indépendants, alors il semble y avoir un saut à la position p . Le test de Fisher comparant l'ajustement du modèle homogène avec l'ajustement du modèle avec un saut est utilisé pour voir si le modèle 3 ajuste mieux les données. Le seuil significatif du saut se calcule avec la statistique de Student. On recommence ensuite le processus sur chacun des segments jusqu'à ce que tous les segments soient homogènes.

2.1.3 Régression à deux phases (reg2)

Plusieurs versions de la régression à deux phases sont présentées dans la littérature (Solow 1987; Easterling et Peterson 1995 ; Lund et Reeves 2000 ; Wang 2002). Pour ce travail, le modèle de Lund et Reeves (2000) a été utilisé. Deux modèles de régression sont ajustés dans lesquels la variable explicative est le temps. Le premier modèle représente une série homogène. Le deuxième modèle représente une série dans laquelle il y a une discontinuité au temps p :

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

où y_i représente l'observation à l'année i de la série de base, τ_1 et τ_2 les ordonnées à l'origine avant et après le saut et λ_1 et λ_2 les pentes précédent et suivant le saut. L'emplacement du saut se trouve en ajustant le modèle pour toutes les valeurs possibles de p , en calculant les statistiques de Fisher comparant l'ajustement des deux modèles pour tous ces points et en choisissant la valeur maximale (F_{max}). Les valeurs critiques de cette statistique sont obtenus par simulation et fournis dans Lund et Reeves (2000). Itérativement, le même processus est répété jusqu'à ce que tous les segments semblent homogènes ou que les segments soient de longueur inférieure à 10. La méthode n'est pas appliquée directement sur la série de base, mais sur une série de différences entre la série de base et les séries voisines.

2.1.4 Approche bivariée (bivt)

Cette méthode a été développée par Maronna et Yohai (1978) et appliquée au problème de l'homogénéisation par Potter (1981). La technique repose sur le postulat que la série de base ($y_i, i = 1, \dots, n$) et une série de référence ($x_i, i = 1, \dots, n$) positivement corrélée appartiennent à une même distribution normale bivariée. On fait l'hypothèse qu'il y a un changement de moyenne dans la série de base qui ne se retrouve pas dans la série de référence. On calcule une série de rapports :

$$q_i = \frac{i(n-i)\delta_i^2 \left((S_x - (\sum_{j=1}^i x_j / i - \bar{x}_{1:n})^2 ni) / n - i \right)}{S_x S_y - S_{xy}^2}, \quad i = 1, \dots, n-1 \quad (5)$$

où

$$\delta_i = \frac{\left[S_x (\bar{y}_{1:n} - \sum_{j=1}^i y_j / i) - S_{xy} (\bar{x}_{1:n} - \sum_{j=1}^i x_j / i) \right] n}{(n-i)F_i}, \quad i = 1, \dots, n-1 \quad (6)$$

$$S_{xy} = \sum_{i=1}^n (x_i - \bar{x}_{1:n})(y_i - \bar{y}_{1:n}) \quad (7)$$

$$S_x = \sum_{i=1}^n (x_i - \bar{x}_{1:n})^2 \quad (8)$$

$$S_y = \sum_{i=1}^n (y_i - \bar{y}_{1:n})^2 \quad (9)$$

La statistique de ce test est donnée par $Q_p = \max_{i=1, \dots, n-1} \{|q_i|\}$. Les valeurs critiques de Q_p ont été obtenues par simulation (Maronna et Yohai, 1978). Lorsqu'elle est significative, cela signifie qu'il y a un saut à l'année p . Pour ce travail, la série de référence utilisée est la moyenne des séries voisines.

2.1.5 Test de Student séquentiel (stus)

Le test de Student séquentiel consiste à tester successivement l'égalité de deux moyennes pour identifier une rupture dans une série (Gullett et al. 1990). Ducré-Robitaille et al. (2003) suggérant d'augmenter la taille de la fenêtre mobile, on teste l'égalité des moyennes en utilisant 10 points avant et après chaque position potentielle :

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

où $\bar{q}_{i-10:i-1}$, \bar{q}_{i+9} , $s_{i-10:i-1}^2$ et s_{i+9}^2 sont les moyennes et les variances de la série de ratios des 10 observations avant et après i . La valeur maximale de la statistique de Student correspond à la position du saut le plus probable et elle est significative si elle dépasse la valeur critique de la distribution de Student avec 18 degrés de liberté. Dans ce cas, la série est séparée en deux segments et le processus reprend sur les deux segments jusqu'à ce que tous les sauts soient détectés. Cette procédure a été substituée à l'originale puisqu'en extrayant simultanément toutes

les statistiques significatives, un même saut est identifié plus d'une fois étant donné que les statistiques de Student successives sont très corrélées. De plus, étant donné que le test est appliqué sur la même série un grand nombre de fois, la probabilité de rencontrer l'erreur de type 1, c'est-à-dire le rejet de l'hypothèse nulle alors qu'elle est vraie, est augmentée. On a alors déterminé par simulation le seuil critique à utiliser pour que le seuil critique global soit de 5%. Enfin, puisque l'on travaille avec des précipitations, la méthode sera appliquée sur une série de ratios entre la série de base et les séries de référence comme avec la méthode d'Alexandersson.

2.1.6 Test de Wilcoxon séquentiel (wils)

Le test de Wilcoxon a été utilisé de plusieurs façons pour l'homogénéisation des données (Karl et Williams 1987; Lanzante 1990; Ducré-Robitaille et al. 2003). On a choisit la version plus récente qui consiste à calculer successivement les statistiques de Wilcoxon et d'évaluer leurs seuils significatifs avec une approximation normale :

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

où n représente la taille de la série testée, $R_i = \sum_{j=1}^i r_j$ et r_j sont les rangs de la première partie de la série. Le maximum de la série et sa position sont extraits ($Z_p = \max_{i=11, \dots, n-9} \{|Z_i|\}$). Si le test est significatif, alors il y a une rupture de moyenne à cet emplacement. On sépare alors la série en deux segments et on recommence le processus sur chacun des segments. On continue jusqu'à ce que tous les segments soient homogènes ou de longueur inférieure à 10. Comme avec la méthode du test de Student séquentiel, on a modifié les valeurs critiques pour avoir un seuil de 5%. Dans Ducré-Robitaille et al. (2003), le test est appliqué sur une série de différences entre la série de

base et les séries de référence. Pour ce travail, on a plutôt formé une série de ratios entre la série de base et les séries voisines comme dans la méthode d'Alexandersson.

2.1.7 Méthode de Jaruskova (jaru)

Cette méthode a été proposée pour détecter une rupture dans une série météorologique par Jaruskova (1996). Plusieurs approches y sont présentées, mais le modèle pour lequel le temps de changement est inconnu a été utilisé. On construit une série de différences entre la série de base et une série de référence et on pose l'hypothèse qu'il y a un changement de moyenne dans la série de différences. La statistique suivante est calculée pour toutes les positions de la série.

$$Q_i = \sqrt{\frac{(n-i)i}{n}} \frac{(\bar{d}_{1:i} - \bar{d}_{i+1:n})}{s_i}, \quad i=1, \dots, n-1 \quad (12)$$

où

$$s_i^2 = \frac{1}{n-2} \left[\sum_{j=1}^i (d_j - \bar{d}_{1:i})^2 + \sum_{j=i+1}^n (d_j - \bar{d}_{i+1:n})^2 \right] \quad (13)$$

et d_i représente la différence entre l'observation à la station de base et à une station de référence de l'année i et n est la longueur de la série de différences. On suppose que la série de différences est normalement distribuée. Le maximum de la série, $Q_p = \max_{i=1, \dots, n-1} \{|Q_i|\}$, est extrait et la rupture est significative si la statistique excède la valeur critique de la distribution (Jaruskova 1996). La série de référence est une moyenne des trois séries voisines synthétiques.

2.1.8 Méthode Bayésienne (rasm)

Cette méthode présentée par Rasmussen (2001) consiste à inférer les paramètres d'un modèle de régression linéaire par une approche Bayésienne analytique. Plusieurs modèles de

régression linéaire de ce type ont été considérés dans ses travaux, mais celui utilisé dans cette étude s'écrit :

$$y_i = \begin{cases} \tau_1 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + e_i & i = 1, \dots, p-1 \\ \tau_2 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + e_i & i = p, \dots, n \end{cases} \quad (14)$$

où y_i représente l'observation i de la série de base, x_{ik} représente l'observation i de la série de référence k , τ_1 , τ_2 et β_j ($j=1, \dots, 3$) représentent respectivement les ordonnées à l'origine avant et après le changement et les coefficients propres aux séries de référence. Le modèle repose sur les postulats de normalité et d'indépendance des données. Les densités de probabilités *a priori* sur les paramètres de régression et sur la position du saut sont non informatives (uniformes). La position du changement est prise comme étant le mode de la distribution *a posteriori*. Pour ce travail, l'analyse Bayésienne se fait sur la position du changement (p) et aussi sur le vecteur de paramètres $\theta = [\tau_1, \tau_2, \beta_1, \beta_2, \beta_3]^T$. La densité de probabilités *a posteriori* de la position est donné par :

$$pr(p | Y) = \left\{ |G_{p-1}^T G_{p-1}|^{-1/2} [Y^T Y - Y^T G_{p-1} (G_{p-1}^T G_{p-1})^{-1} G_{p-1}^T Y] \right\} \\ * \left\{ \sum_{i=1}^{n-1} |G_i^T G_i|^{-1/2} [Y^T Y - Y^T G_i (G_i^T G_i)^{-1} G_i^T Y]^{-(n-5)/2} \right\}^{-1} \quad (15)$$

où

$$G_{p-1}^T = \begin{pmatrix} 1 & 0 & x_{1,1} & x_{1,2} & x_{1,3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & x_{p-1,1} & x_{p-1,2} & x_{p-1,3} \\ 0 & 1 & x_{p,1} & x_{p,2} & x_{p,3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 1 & x_{n,1} & x_{n,2} & x_{n,3} \end{pmatrix} \quad (16)$$

et Y représente le vecteur d'observations de la série de base, n est la longueur du vecteur et x_i représente les observations de la série de référence. La densité *a posteriori* des paramètres du vecteur θ est donnée par :

$$pr(\theta | Y, p) = \left\{ \frac{\Gamma(\nu + 5/2) |G_{p-1}^T G_{p-1}|^{1/2}}{[\Gamma(1/2)]^5 \Gamma(\nu/2) [c\sqrt{\nu}]^5} \right\} * \left\{ 1 + \frac{(\theta - \hat{\theta})^T G_{p-1}^T G_{p-1} (\theta - \hat{\theta})}{\nu c^2} \right\}^{-(\nu+5)/2} \quad (17)$$

avec $\hat{\theta} = (G^T G)^{-1} G^T Y$ (18)

$$c^2 = \sum_{i=1}^n (y_i - \hat{y}_i) / (n - 2) \quad (19)$$

et ν est le nombre de degrés de liberté ($n-5$) et Γ est la fonction Gamma. Comme il n'existe pas d'expression analytique pour la distribution a posteriori de l'amplitude du changement, celle-ci a été calculée par simulations de Monte Carlo. L'intervalle de crédibilité Bayésien obtenu est utilisé pour vérifier si le saut est significatif.

2.2 Génération des séries synthétiques

Pour évaluer la performance des méthodes d'homogénéisation, il est intéressant de posséder des séries homogènes et inhomogènes de plusieurs types. Pour ce faire, on a généré des séries de précipitations synthétiques ayant le même comportement statistique (moyenne, écart-type, autocorrélation d'ordre 1 et corrélation spatiale) que les séries de précipitations totales annuelles typiques observées dans le sud du Québec. Le modèle utilisé a été proposé par Easterling et Peterson (1992) et repris par Vincent (1998) et par Ducre-Robitaille et al. (2003)

pour générer des séries synthétiques de température. Pour estimer les paramètres du modèle, on a choisi un ensemble de stations situées au sud du Québec et dans les provinces avoisinantes qui ont des enregistrements de qualité (de longues séries d'observations et peu de données manquantes) et les paramètres du modèle ont été estimés à partir de ces dernières.

Des séries synthétiques homogènes (moyenne et variance constante) et inhomogènes (un ou plusieurs sauts de moyenne, une tendance, un saut d'écart-type) ont été générées. Étant donné que la plupart des méthodes d'homogénéisation se basent sur une comparaison entre une série de base (celle qu'on veut homogénéiser) et une série de référence homogène (calculée à partir de stations voisines), des séries voisines corrélées avec la série de base ont également été générées.

2.2.1 Séries de base

1) HOMOGÈNES

Des séries de base homogènes ont été générées pour étudier la sensibilité des méthodes sur des séries qui ne comprennent aucune inhomogénéité. Les séries de base homogènes standardisées sont simulées selon le modèle suivant :

$$z_i = \phi_1 z_{i-1} + e_i \quad (20)$$

où z_i représente la précipitation totale de l'année i standardisée, ϕ_1 est l'autocorrélation entre les années et e_i est un résidu normalement distribué de moyenne 0 et de variance $1 - \phi_1^2$. Par la suite, on multiplie par l'écart-type et on ajoute la moyenne pour avoir y_i^* , la précipitation totale de l'année i . On a voulu reproduire une précipitation totale annuelle de 1089 mm, un écart-type de 142 mm et une autocorrélation d'ordre 1 de 0.02, moyenne des stations retenues. Au total, 10 000 séries homogènes (5000 de 60 et 5000 de 100 ans) ont été générées de cette façon.

2) SÉRIES AVEC UN SAUT DE MOYENNE

Des séries avec un changement de moyenne ont été générées pour étudier la puissance des méthodes à positionner et estimer l'amplitude d'un saut de moyenne. L'amplitude et la position du saut sont déterminées de façon aléatoire. Une série avec un saut se représente par :

$$y_i = \begin{cases} y_i^* - \delta_p \sigma, & i = 1, \dots, p-1 \\ y_i^*, & i = p, \dots, n \end{cases} \quad (21)$$

$$\text{avec} \quad \delta_p = \text{sign}(u - 1/2) * 3 * b, \quad u \square U(0,1), \quad b \square BETA(2,2) \quad (22)$$

$$p = 10 + ud, \quad ud \square DUNIF(n - 20) \quad (23)$$

où y_i représente l'observation i d'une série inhomogène de taille n , y_i^* l'observation i d'une série homogène, σ est l'écart-type du dernier segment de la série, δ_p dénote l'amplitude du saut et p sa position. L'amplitude varie aléatoirement entre 0 et 3 écarts-type. Pour simuler la position, on utilise une distribution uniforme discrète qu'on tronque pour éviter qu'il se produise un saut pendant les 10 premières ou les 10 dernières années de la série. Un nombre total de 50 000 séries (25 000 de taille 60 et 25 000 de taille 100) sont générées de cette façon.

3) SÉRIES AVEC PLUSIEURS SAUTS DE MOYENNE

Des séries avec des sauts multiples (2 et 3) sont également générées. Une série avec deux sauts se représente par :

$$y_i = \begin{cases} y_i^* - \delta_{p_1} * \sigma, & i = 1, \dots, p_1 - 1 \\ y_i^* - \delta_{p_2} * \sigma, & i = p_1, \dots, p_2 - 1 \\ y_i^*, & i = p_2, \dots, n \end{cases} \quad (24)$$

$$\text{avec } \delta_{p_1}, \delta_{p_2} = \text{sign}(u - 1/2) * 3 * b, \quad u \square U(0,1), \quad b \square BETA(2,2) \quad (25)$$

$$p_1 = 10 + ud, \quad ud \square DUNIF(n - 31) \quad (26)$$

$$p_2 = 10 + p_1 + ud, \quad ud \square DUNIF(n - 20 - p_1) \quad (27)$$

où y_i représente l'observation i d'une série inhomogène de taille n , y_i^* l'observation i d'une série homogène, σ est l'écart-type du dernier segment de la série, $\delta_{p_1}, \delta_{p_2}, p_1$ et p_2 dénotent respectivement l'amplitude et la position des deux sauts. L'amplitude est simulée de la même façon que pour un seul saut. Pour la position, on utilise toujours la loi uniforme discrète, mais le paramètre change selon le nombre de sauts générés. La distance entre les sauts de moyenne ne peut être inférieure à 10. Les séries avec trois sauts de moyenne sont générées selon le même principe. Au total, 15 000 séries avec deux sauts et 15 000 séries avec trois ruptures de moyenne ont été simulées.

4) SÉRIES AVEC UNE TENDANCE DANS LA MOYENNE

Lorsqu'une tendance est présente dans les données, il peut arriver que les méthodes d'homogénéisation l'interprètent comme un ou plusieurs sauts de moyenne consécutifs. On a simulé des séries avec une tendance dans la moyenne pour étudier le comportement des méthodes d'homogénéisation face à ce type d'inhomogénéité. Malgré le fait que les méthodes ne sont pas faites pour détecter des tendances, on s'intéresse quand même à leur performance à identifier un changement à l'intérieur de la tendance. Une série avec une tendance en moyenne se représente par :

$$y_i = \begin{cases} y_i^* - \delta_{p_1:p_2} \sigma, & i = 1, \dots, p_1 - 1 \\ y_i^* - \delta_{p_1:p_2} \sigma - mi, & i = p_1, \dots, p_2 - 1 \\ y_i^*, & i = p_2, \dots, n \end{cases} \quad (28)$$

avec $\delta_{p_1:p_2} = \text{signe}(u - 1/2) * 3 * b$, $u \square U(0,1)$, $b \square BETA(4,2)$ (29)

$$p_1 = 10 + ud, \quad ud \square DUNIF(n - 31) \quad (30)$$

$$p_2 = 10 + p_1 + ud, \quad ud \square DUNIF(n - 20 - p_1) \quad (31)$$

où y_i représente l'observation i d'une série inhomogène de taille n , y_i^* l'observation i d'une série homogène, σ est l'écart-type du dernier segment de la série, $\delta_{p_1:p_2}$ est l'amplitude de la tendance, m sa pente, p_1 et p_2 dénotent respectivement la position du début et de la fin de la tendance. L'amplitude se situe entre -3 et 3 écarts-type et les petites tendances sont peu fréquentes. La position du début et de la fin de la tendance est simulée de la même façon que pour une série à deux sauts. Un nombre total de 10 000 séries synthétiques avec une tendance aléatoire ont été générées.

5) SÉRIES AVEC UN SAUT D'ÉCART-TYPE

Des séries avec un saut de variance ont été générées pour déterminer les méthodes sensibles à ce type de changement. Les méthodes sélectionnées n'ont pas été développées pour détecter des changements de variance, mais il est intéressant de connaître le comportement des techniques en présence de ce type de rupture dans les données. La position du saut est déterminée de la même façon que pour un changement de moyenne. Par ailleurs, l'amplitude est tirée de la distribution BETA (8,2) auquel on ajoute un signe aléatoire et qu'on divise par 2. De cette façon,

les amplitudes simulées sont entre 0 et 50% de l'écart-type. Un nombre total de 10 000 séries avec un saut de variance avec une amplitude et une position aléatoire a été généré.

2.2.2 Séries voisines

Pour chaque série de base, trois séries voisines sont générées. On génère d'abord trois séries homogènes indépendantes de la série de base :

$$w_i = \phi_1 w_{i-1} + e_i \quad (32)$$

où ϕ_1 est l'autocorrélation entre les années et e_i est un résidu normalement distribué de moyenne 0 et de variance $1 - \phi_1^2$. Ensuite, on introduit une corrélation entre la série de base et chaque série voisine :

$$w_i = \psi z_i + w_i \quad (33)$$

où z_i représente la précipitation totale standardisée à la station de base de l'année i , w_i représente la précipitation totale standardisée à une station voisine de l'année i , ψ est une constante de corrélation entre les séries voisines et la série de base. Par la suite, la série w_i est restandardisée. Enfin, on multiplie par 142 et on ajoute 1089. Les séries voisines possèdent les mêmes caractéristiques statistiques que les séries de base. La constante de corrélation (0.7) a été déterminée par simulation sur 10 000 paires de séries base-référence pour reproduire une corrélation de 0.55. En effet, la corrélation moyenne entre les stations qui ont moins de 300 km de distance est de 0,55. En pratique, cette valeur varie énormément et peut affecter la performance des méthodes, elles performant probablement beaucoup mieux avec une forte corrélation base-voisines. Cependant, le réseau de stations du Québec est peu dense et parfois il

peut s'avérer difficile de trouver plusieurs voisines pour la même série de base. Dans le but de représenter le mieux possible cette réalité, le nombre de stations voisines a été limité.

2.3 Critères d'évaluation des méthodes d'homogénéisation

Les méthodes d'homogénéisation décrites ci-haut ont été appliquées successivement (au seuil critique de 5%) sur tous les ensembles de séries synthétiques jusqu'à ce que tous les segments soient déclarés homogènes ou de longueur inférieure à 10. La performance de chaque méthode sera évaluée de façon différente selon qu'elle est appliquée à une série homogène, une série à un saut, une série à plusieurs sauts, une série à un saut d'écart type ou une série avec une tendance. De plus, étant donné que les méthodes d'homogénéisation sélectionnées ont été développées pour détecter un seul saut dans une série, il est possible que leur performance soit dégradée en présence de sauts multiples. Pour éviter ceci, on peut estimer les amplitudes en terme de différences de moyenne des segments avant et après chaque saut, peu importe la méthode de détection utilisée. Les deux possibilités sont considérées dans la présente étude : le cas où la méthode est appliquée telle que publiée dans la littérature (section 3), et le cas où les méthodes sont utilisées pour positionner les sauts avant une estimation des amplitudes par différence de moyenne (section 4).

2.3.1 Critère d'évaluation des méthodes sur les séries homogènes

La performance des méthodes sur les séries homogènes sera estimée par la proportion d'erreur de type I (sauts faussement détectés) et aussi par l'amplitude et la position des sauts faussement détectés.

2.3.2 Critère d'évaluation des méthodes sur les séries à un saut

En ce qui concerne les séries à un saut, on s'intéresse particulièrement à la capacité de ces méthodes à positionner les inhomogénéités, et à en estimer l'amplitude. On considère qu'un saut est correctement identifié lorsque sa position est exacte et la différence entre l'amplitude estimée et l'amplitude réelle est au plus 20% de l'amplitude réelle. Un saut est bien identifié lorsque la position estimée se situe à 2 années ou moins de la position réelle et que l'erreur absolue de l'estimation de l'amplitude est inférieure ou égale à 50% de l'amplitude réelle. Les sauts bien positionnés sont situés à plus ou moins deux années de l'amplitude réelle. La performance de la méthode est évaluée par le pourcentage de sauts bien identifiées et bien positionnés.

2.3.3 Critère d'évaluation des méthodes sur les séries à plusieurs sauts

Pour les séries avec plusieurs sauts, on évalue la performance des méthodes à corriger les inhomogénéités détectées. On compare alors les sommes de carrés résiduelles des séries homogénéisées avec les sommes de carrés résiduelles espérées pour des séries bien homogénéisées. De plus, on évalue leur capacité à positionner tous les sauts imposés. On a donc défini un critère que l'on calcule à partir des distances entre les sauts réels et les sauts détectés. Le critère a été spécialement conçu pour tenir compte autant des détections de sauts fictifs que des échecs à détecter des sauts existants. On reviendra plus en détails sur la formulation mathématique de ce critère à la section 4.3.

2.3.4 Critère d'évaluation des méthodes sur les séries avec un saut d'écart type

Pour les séries avec un saut d'écart-type, on évalue le nombre de sauts faussement détectés et compare avec le pourcentage d'erreur de type I observés sur les séries homogènes. Les

techniques robustes à des sauts d'écart-type devraient donner approximativement le même pourcentage d'erreur de type I.

2.3.5 Critère d'évaluation des méthodes sur les séries avec une tendance

Enfin, pour les séries synthétiques avec une tendance, on compte le nombre de sauts détectés à l'intérieur de cette dernière. Les méthodes d'homogénéisation sélectionnées ne sont pas toutes en mesure de détecter des changements graduels. En effet, seuls les deux modèles de régression ont été développés pour identifier à la fois des sauts et des tendances. De ce fait, on s'intéresse aux résultats obtenus en présence d'une tendance afin de savoir si les méthodes sont sensibles à ce type d'inhomogénéité et si elles l'interprètent comme un ou plusieurs sauts consécutifs.

3. Performances des méthodes originales

Dans cette section, les méthodes sont appliquées telles qu'elles sont décrites dans la littérature, autant pour le positionnement des sauts que pour l'estimation de l'amplitude des sauts.

3.1 Séries homogènes

Les méthodes d'homogénéisation ont été appliquées à deux ensembles de séries homogènes (60 et 100 ans) pour évaluer le pourcentage d'erreur de type 1 commise. En d'autres termes, on a dénombré les cas pour lesquels on rejette l'hypothèse d'homogénéité (l'hypothèse alternative étant que la série de base contient un saut ou plus) alors qu'elle est vraie. Les tableaux 1 et 2 présentent les résultats obtenus sur les séries homogènes.

Tableau 1. Sauts faussement détectés (%) sur les séries homogènes de 60 ans.

Amplitude (en écart-type)	regm	jaru	bivt	snht	wils	reg2	stus	rasm
0-0.25	0.1	0.0	0.0	0.0	0.2	1.1	1.7	0.1
0.25-0.5	0.4	0.0	0.3	0.4	1.1	1.1	1.9	11.9
0.5-1	0.4	1.0	1.9	2.1	2.2	1.3	1.5	9.0
1-2	0.0	0.0	0.0	0.0	0.3	0.3	0.1	0.2
>2	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0
Total	0.9	1.0	2.2	2.5	3.8	4.0	5.2	21.2

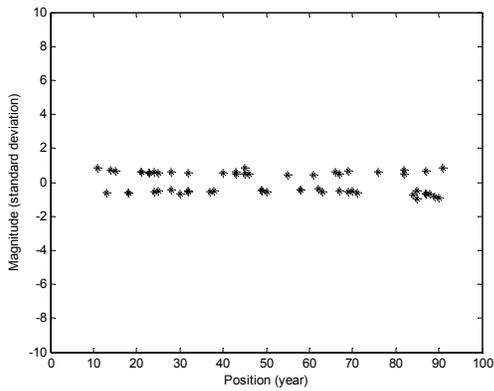
Tableau 2. Sauts faussement détectés (%) sur les séries homogènes de 100 ans.

Amplitude (en écart-type)	regm	jaru	bivt	snht	reg2	wils	stus	rasm
0-0.25	0.2	0.0	0.0	0.0	1.5	0.6	2.2	0.7
0.25-0.5	0.7	0.3	1.8	1.7	1.1	1.6	1.9	18.2
0.5-1	0.3	0.9	2.1	2.3	1.0	2.4	0.9	8.4
1-2	0.0	0.0	0.0	0.0	0.3	0.2	0.0	0.1
>2	0.0	0.0	0.0	0.0	0.5	0.0	0.0	0.0
Total	1.2	1.2	3.9	4.0	4.4	4.8	5.0	27.5

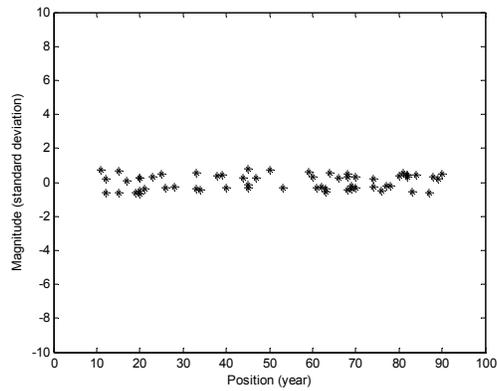
En général, toutes les méthodes performant bien pour identifier les séries homogènes à l'exception de la méthode Bayésienne. On peut remarquer que pour la plupart des méthodes l'amplitude des sauts faussement détectés dépasse rarement un écart-type. Le pourcentage de sauts faussement détectés varie entre environ 1% (Régression multiple, Méthode de Jaruskova) et 5% (Test de Student séquentiel) pour les méthodes classiques. La méthode Bayésienne donne un pourcentage de sauts faussement détectés nettement supérieur (plus que 20%). Dans une étude semblable effectuée sur les températures (Ducré-Robitaille et al., 2003), on a trouvé des seuils de fausse détection amplement supérieurs pour les méthodes de régression à deux phases et de Wilcoxon séquentiel. Ceci s'explique par le fait que dans notre modèle de régression à deux phases, on utilise la statistique de Fisher révisée par Lund & Reeves (2002) avec des valeurs critiques plus élevées qui font en sorte que ce test est plus conservatif. Ainsi, les taux de fausse détection sont améliorés avec l'utilisation de la statistique révisée. De plus, on a réduit le seuil critique des tests de Wilcoxon en raison du nombre élevé d'applications sur une même série de façon à ce que le seuil critique global se trouve aux environs de 5%. Les seuils critiques utilisés pour les méthodes de Wilcoxon et Student séquentiels ont été obtenus par simulation et diffèrent

selon la longueur de la série. La figure 1 présente l'amplitude des sauts faussement détectés des séries homogènes de 100 ans en fonction de leur position pour les méthodes comparées. Les figures des séries de 60 ans ne sont pas présentées puisqu'on retrouve les mêmes patrons. Les amplitudes sont calculées selon la description respective de chaque méthode.

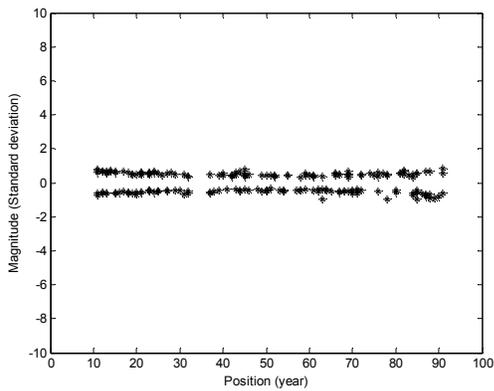
a)



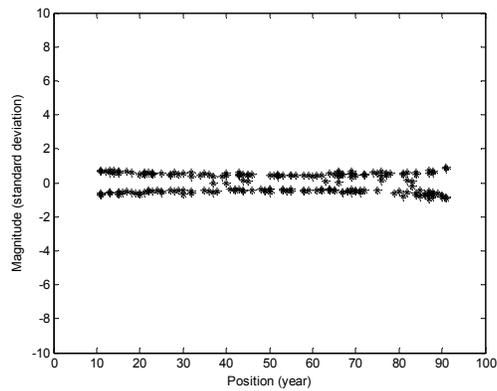
b)



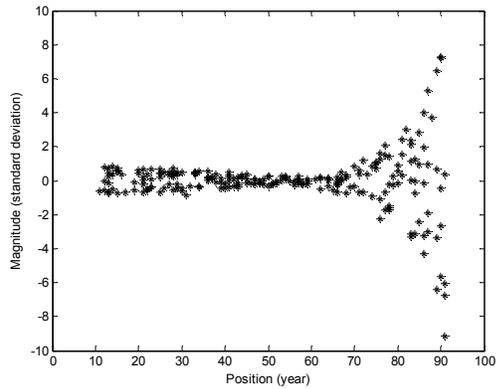
c)



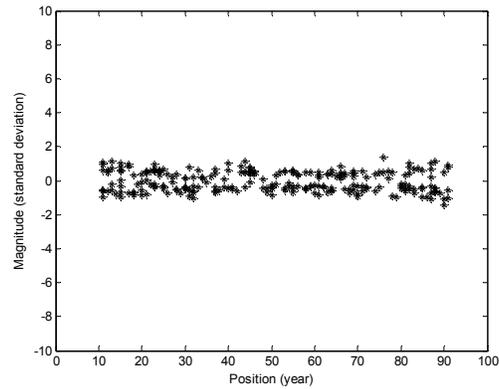
d)



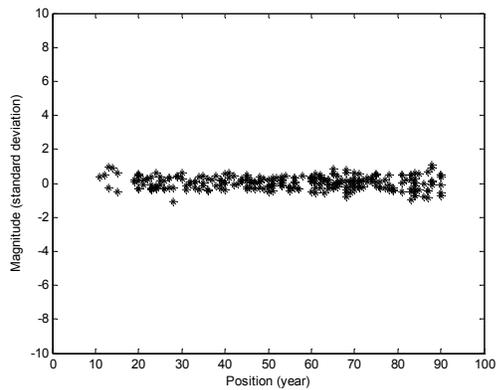
e)



f)



g)



h)

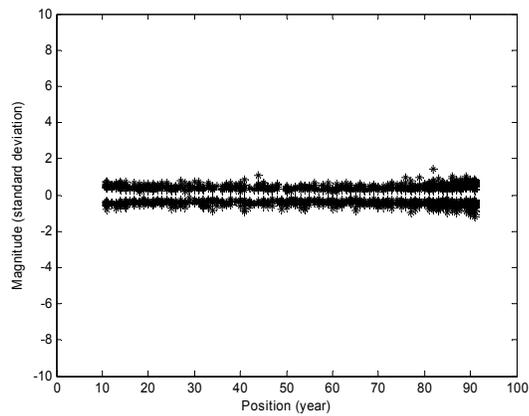


Figure 1 : Sauts faussement détectés sur les séries homogènes de 100 ans a) méthode de Jaruskova b) régression multiple c) approche bivariee d) méthode d'Alexandersson e) régression à deux phases f) test de Wilcoxon séquentiel g) test de Student séquentiel h) méthode Bayésienne.

On peut remarquer que la régression à deux phases tend à détecter des sauts d'amplitudes très élevées vers la fin de la série. En effet, on a obtenu des amplitudes allant jusqu'à huit écarts-types sur des séries homogènes alors que pour toutes les autres méthodes les sauts faussement détectés dépassent rarement un écart-type. L'amplitude a été estimée en comparant la moyenne du segment corrigé et du segment original. Le modèle proposé corrige pour les différences de pente avant et après un saut et lorsque le saut est situé vers la fin de la série, cette différence de pente peut devenir très grande. Le fait de corriger le premier segment de la série selon cette

différence de pente conduit à une amplitude erronée. Il serait important de faire un test sur le caractère significatif du changement de tendance. Par ailleurs, un autre modèle dans lequel on considère seulement les différences d'ordonnées à l'origine est présenté dans Lund et Reeves (2002) serait probablement plus appropriée pour des données qui ne contiennent pas de tendance.

Pour la méthode Bayésienne, le nombre élevé de sauts faussement détectés s'explique par le fait que cette méthode fait l'hypothèse implicite qu'il y a forcément un saut dans la série. En effet, la somme des probabilités a posteriori de la position du changement doit être égale à un, donc la méthode tend à positionner de toute façon un changement quelque part. De plus, en l'appliquant successivement, on augmente encore plus les chances de fausse détection. En général, dans le cas de séries homogènes, les probabilités de la position du changement sont concentrées vers les extrémités de la série. Cependant, en dépit de la décision de ne pas tenir compte des sauts proches des bords, on aboutit à un nombre relativement grand de changements faussement détectés. En utilisant cette méthode, il est fort probable que des sauts réels soient détectés, mais on risque fortement de détecter des sauts inexistantes.

3.2 Séries avec un saut de moyenne

Pour évaluer la puissance des méthodes à identifier la position et estimer l'amplitude d'un saut, les méthodes ont été appliquées à deux ensembles de séries synthétiques contenant un saut de moyenne de position et d'amplitude aléatoire. On a dénombré le nombre de sauts correctement identifiés, bien identifiés et bien positionnés. Les tableaux 3 et 4 présentent le pourcentage total de sauts correctement identifiés, bien identifiés et bien positionnés pour chacune des méthodes.

Tableau 3. Sauts identifiés (%) sur les séries de 60 ans avec un saut de moyenne.

	Snht	regm	reg2	bivt	stus	wils	jaru	rasm
Correctement identifié	51.5	48.7	39.0	52.3	32.1	57.3	53.0	57.3
Bien identifié	79.4	70.8	59.9	79.8	70.8	80.4	79.3	81.3
Bien positionné	81.6	72.0	64.8	82.2	71.2	80.9	81.4	83.3

Tableau 4. Sauts identifiés (%) sur les séries de 100 ans avec un saut de moyenne.

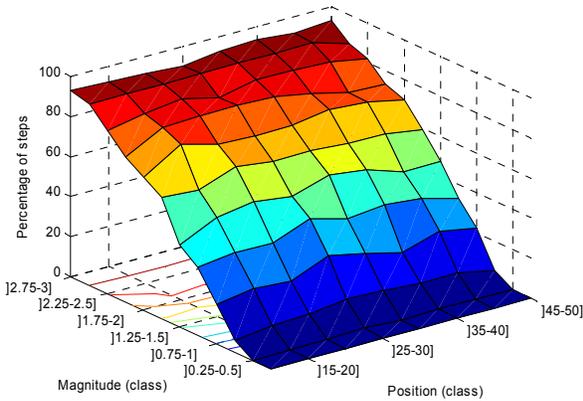
	Snht	regm	reg2	bivt	stus	wils	jaru	rasm
Correctement identifié	55.4	54.2	43.7	56.2	30.3	57.9	57.0	55.5
Bien identifié	82.4	76.2	66.2	82.8	67.4	82.4	82.4	82.6
Bien positionné	83.9	76.9	73.6	84.3	67.5	82.7	83.6	85.2

On peut voir que la plupart des méthodes performant bien à positionner les sauts. En effet, on retrouve un pourcentage de sauts bien positionnés de l'ordre de 75% et plus pour la plupart des méthodes excepté la régression à deux phases et le test de Student séquentiel. Par ailleurs, on peut voir que le test de Student séquentiel performe mieux pour les séries de 60 ans. Ceci est probablement dû au fait que les seuils critiques utilisés ont été déterminés par simulation et diffèrent pour les séries de 60 et 100 ans. En effet, en utilisant les mêmes seuils critiques, on observe une performance équivalente peu importe la longueur de la série.

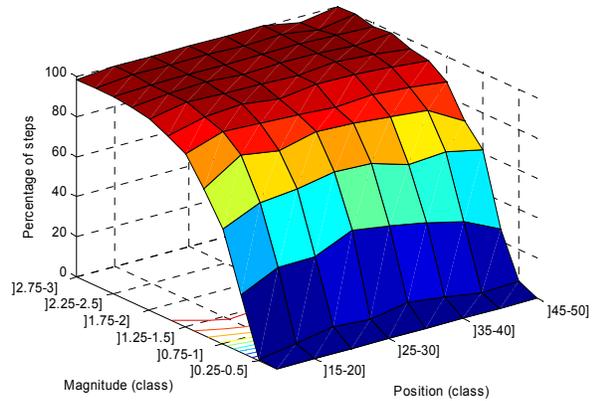
De manière générale, les méthodes d'homogénéisation performant assez bien à identifier approximativement l'emplacement et l'amplitude des sauts, mais une estimation correcte semble

plus problématique. En effet, la différence de pourcentage entre les sauts correctement identifiés et bien identifiés est supérieure à 20%. Donc, certaines méthodes arrivent à détecter l'emplacement des sauts, mais estiment moins bien l'amplitude. La régression à deux phases semble avoir ce problème (différence entre sauts bien identifiés et bien positionnés de 5% et plus). On verra dans la section suivante si la performance des méthodes est supérieure en estimant l'amplitude par différence de moyenne. On a également analysé le pourcentage de sauts correctement ou bien identifiés pour différentes classes d'amplitude et de position. Les figures 2 et 3 présentent les résultats.

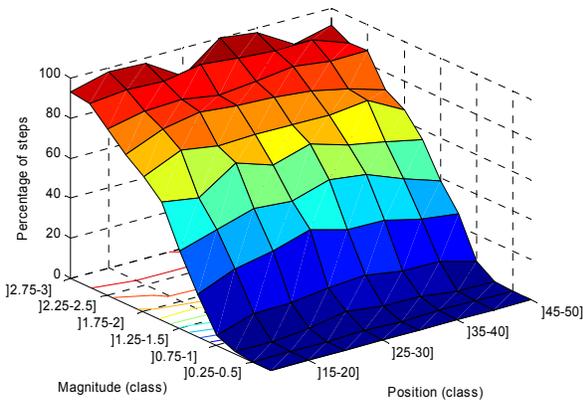
a)-I)



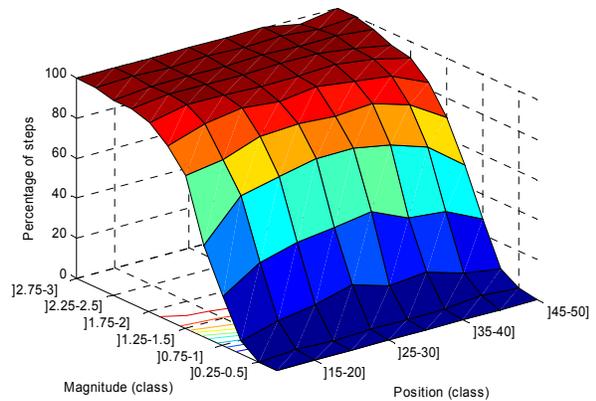
II)



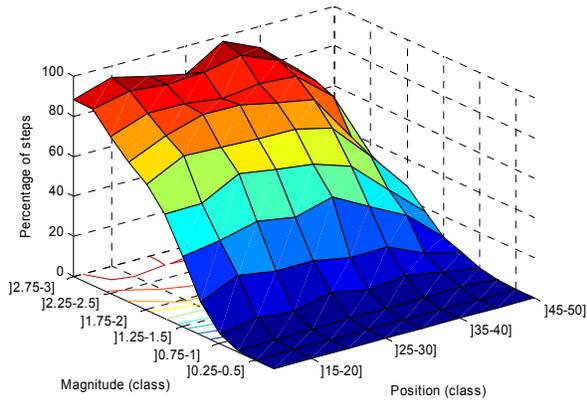
b)-I)



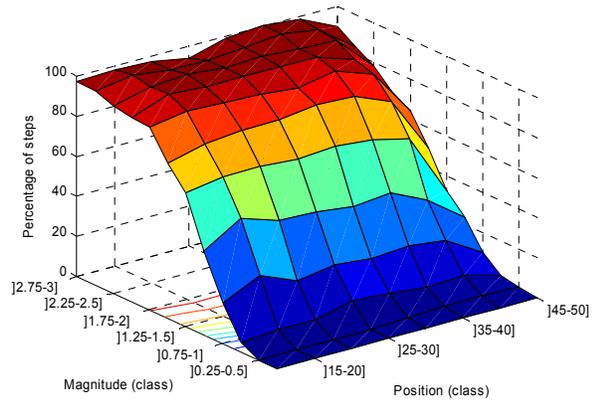
II)



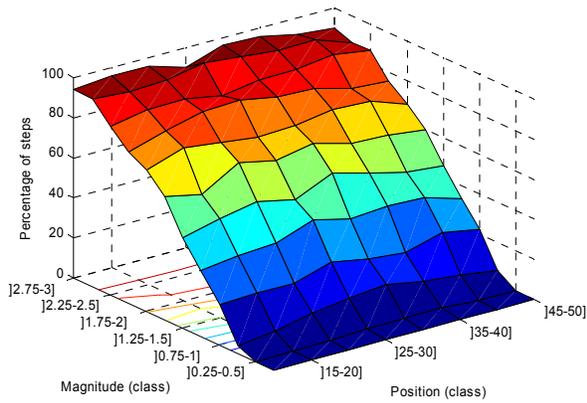
c)-I)



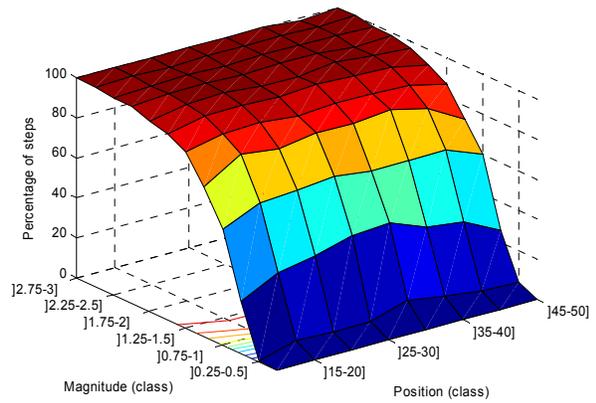
II)



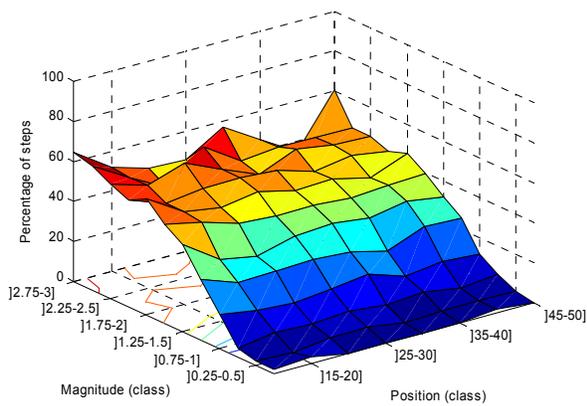
d)-I)



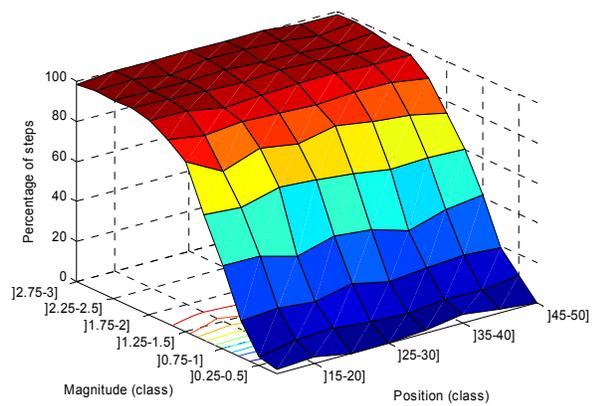
II)



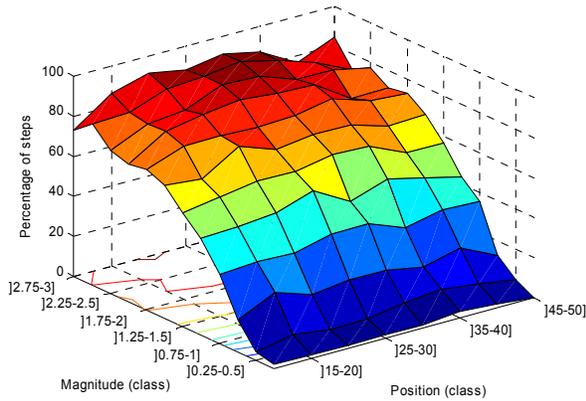
e)-I)



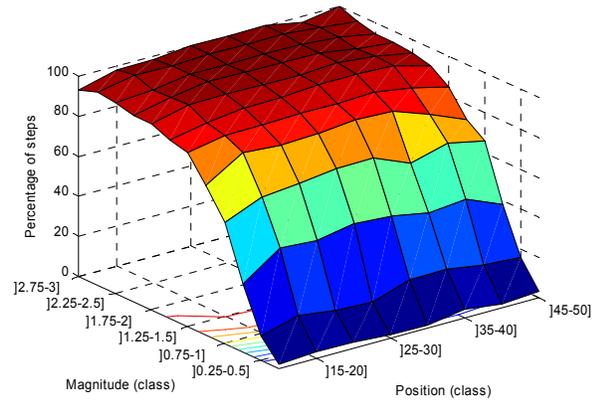
II)



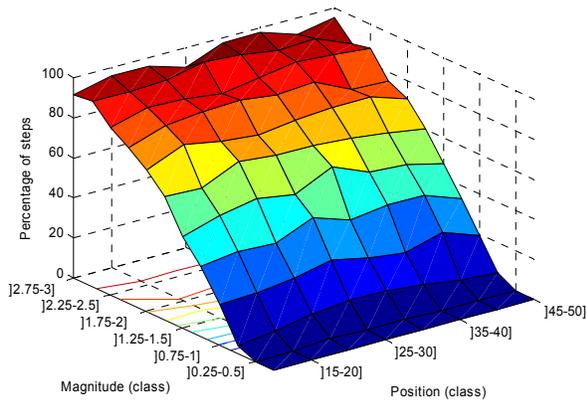
f)-I)



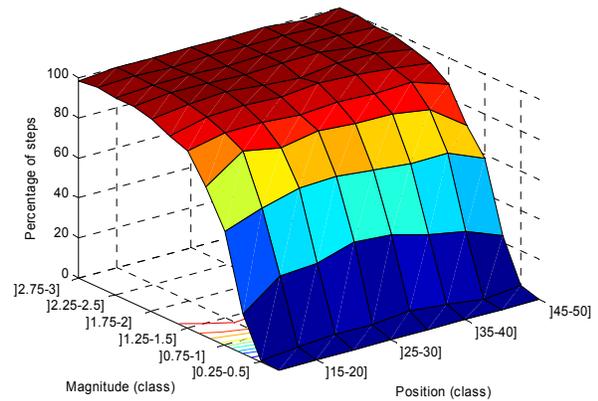
II)



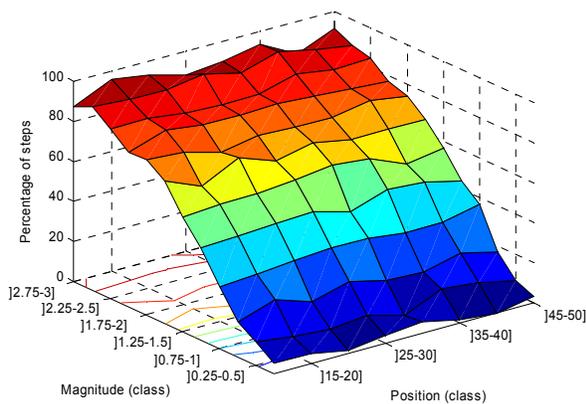
g)-I)



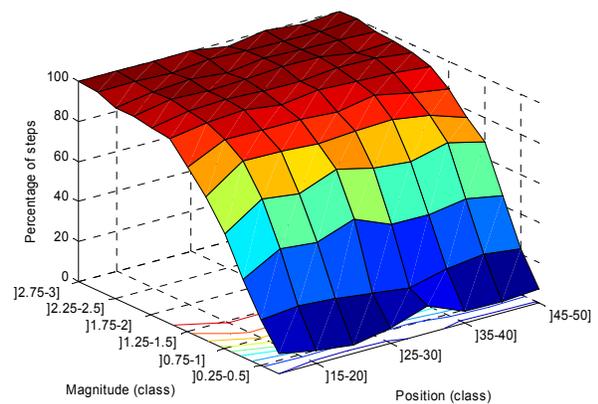
II)



h)-I)

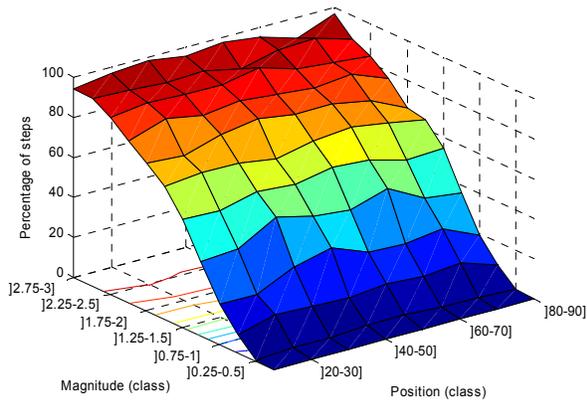


II)

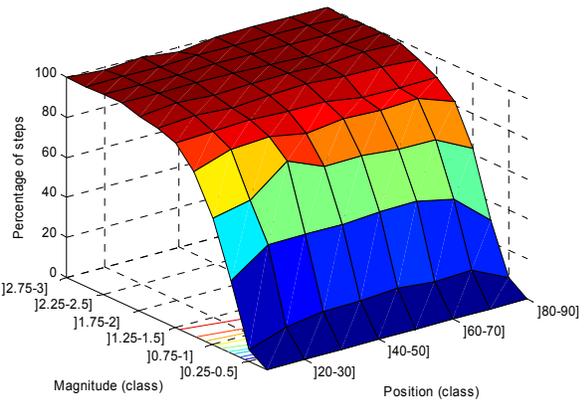


**Figure 2 : Sauts identifiés (%) dans les séries avec un saut de moyenne de longueur 60 ans
 I) correctement identifiés II) bien identifiés a) méthode d'Alexandersson b) régression multiple
 c) régression à deux phases d) approche bivariée e) test de Student séquentiel f) test de Wilcoxon
 séquentiel g) méthode de Jaruskova h) méthode Bayésienne.**

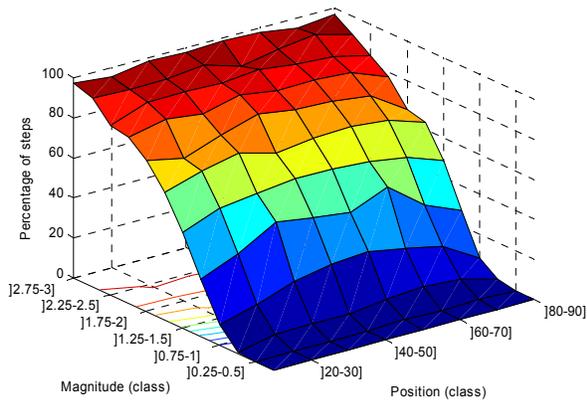
a)-I)



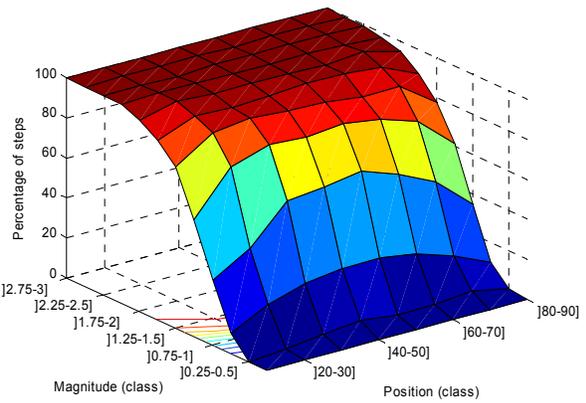
II)



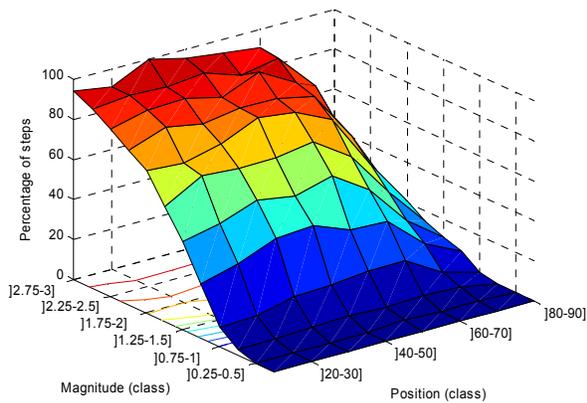
b)-I)



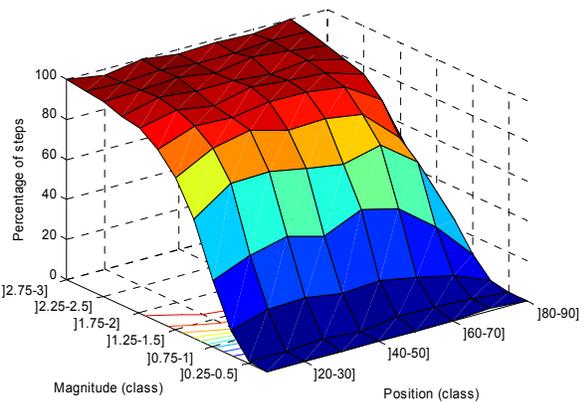
II)



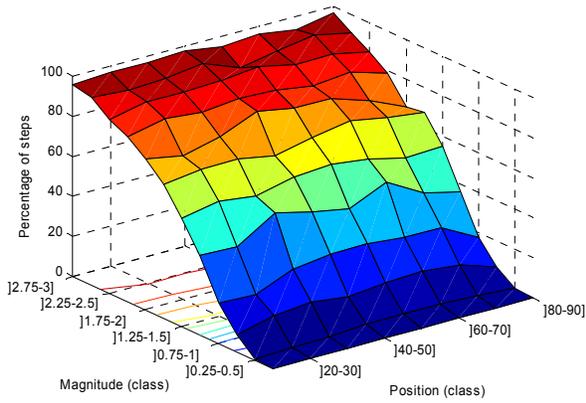
c)-I)



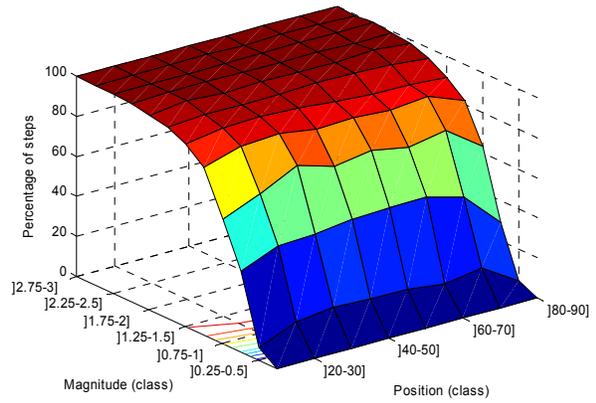
II)



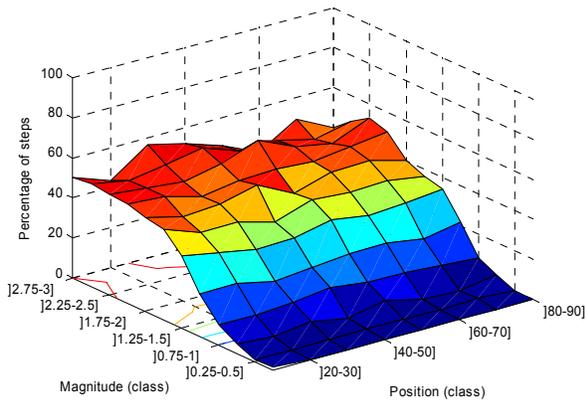
d)-I)



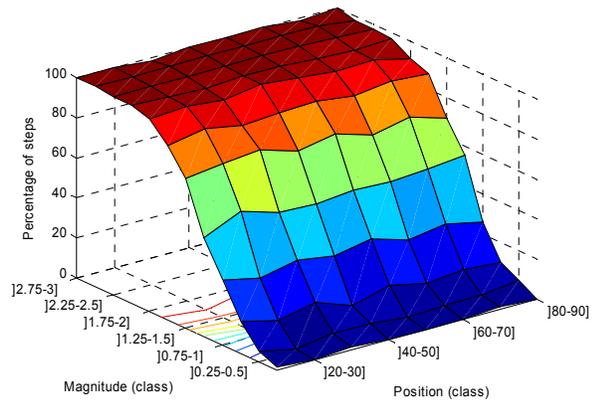
II)



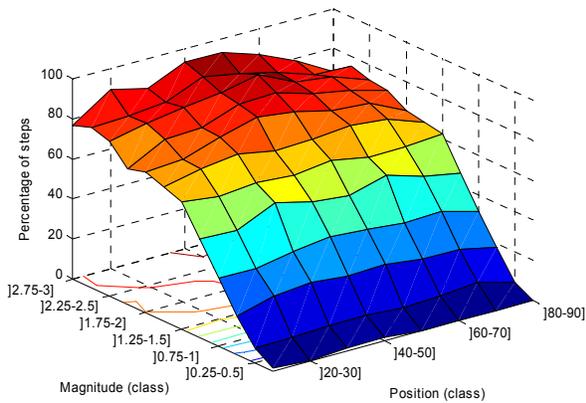
e)-I)



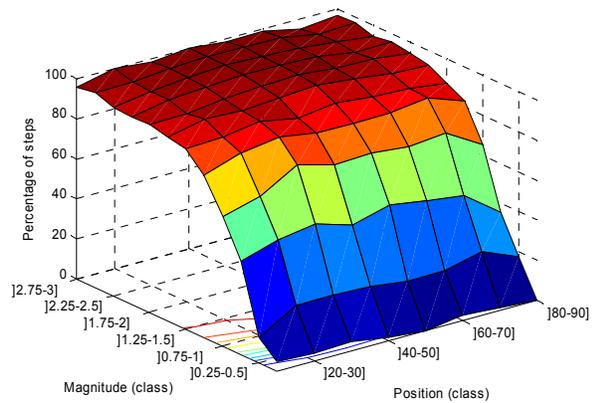
II)



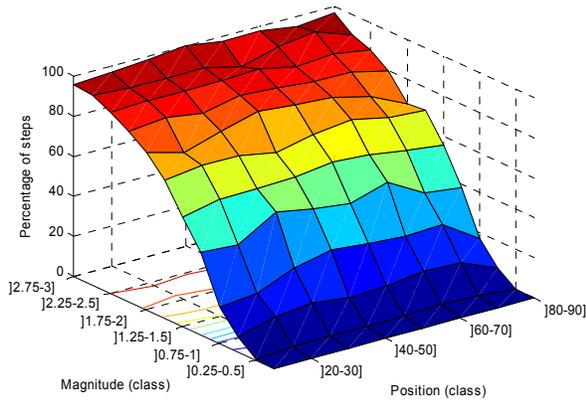
f)-I)



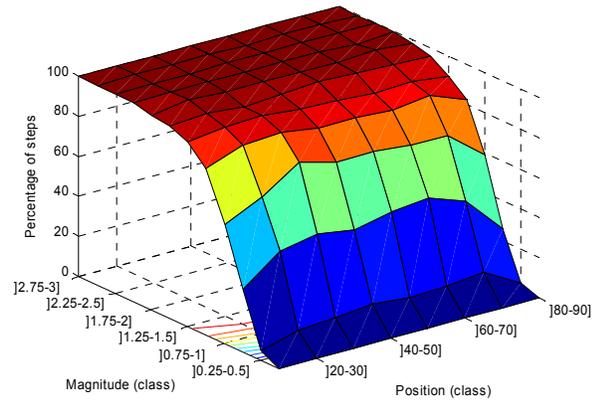
II)



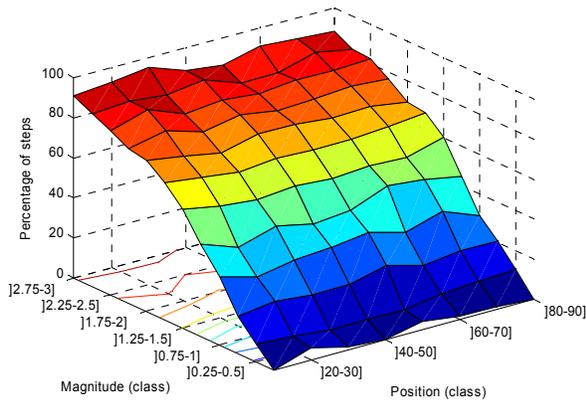
g)-I)



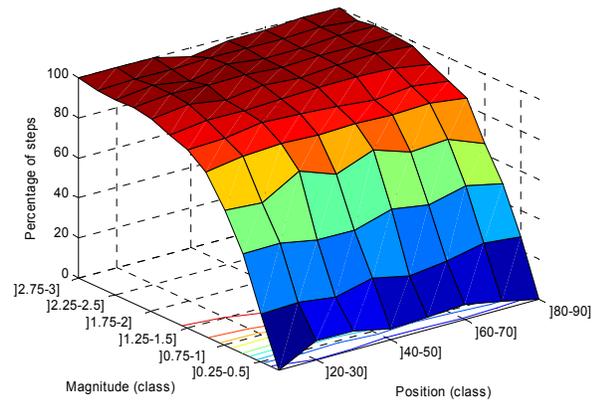
II)



h)-I)



II)



**Figure 3 : Sauts identifiés (%) dans les séries avec un saut de moyenne de longueur 100 ans
I) correctement identifiés II) bien identifiés a) méthode d'Alexandersson b) régression multiple
c) régression à deux phases d) approche bivariée e) test de Student séquentiel f) test de Wilcoxon
séquentiel g) méthode de Jaruskova h) méthode Bayésienne.**

On remarque que toutes les méthodes échouent à identifier les sauts de faible amplitude (0-0.5 écart-type). De plus, le pourcentage de sauts correctement ou bien identifiés augmente avec l'amplitude et les sauts sont bien détectés à partir de 1 ou 1.5 écart-type. Le test de Student séquentiel semble avoir beaucoup plus de difficulté que les autres techniques à correctement identifier un saut. Probablement que le fait d'utiliser une fenêtre mobile pénalise cette méthode. Une étude précédente est arrivée à cette conclusion avec une fenêtre mobile de 10 observations

(Ducré-Robitaille et al., 2003). Malgré qu'on ait augmenté la fenêtre mobile à 20 observations, cette méthode demeure peu performante. Par ailleurs, il ne semble pas vraiment y avoir de position problématique sauf pour la régression à deux phases. Encore une fois, on remarque la difficulté à estimer les amplitudes lorsque le saut est situé à la fin de la série. On s'attendait à ce que les sauts situés aux extrémités soient plus problématiques pour l'ensemble des méthodes. Probablement qu'en ayant ignoré tous les sauts détectés pendant les dix premières ou dix dernières observations, cet effet a été retiré.

3.3 Séries avec plusieurs sauts de moyenne

La plupart des méthodes comparées dans ce travail ne sont pas destinées à détecter des sauts multiples. En pratique, elles sont souvent utilisées à cette fin. On a donc comparé la performance des méthodes à identifier des sauts multiples sur des séries synthétiques qui contiennent deux ou trois sauts de moyenne. Les méthodes d'homogénéisation ont été appliquées successivement jusqu'à ce que tous les segments soient homogènes ou de longueur inférieure à 10 (on ne s'est pas limité au nombre de sauts réels présents dans la série). Pour chaque série, on a corrigé les inhomogénéités détectées en fonction du dernier segment de la série (considéré homogène) et ensuite calculé la somme de carrés résiduelle (Ducré-Robitaille et al., 2003) :

$$RSS_{obs} = \sum_{i=1}^n (y_i - \bar{y})^2 = Var(y_{1:n}) * (n-1) \quad (34)$$

où y_i représente l'observation i de la série de base et n sa taille. La somme de carrés résiduelle attendue d'une série homogène correspond donc à la variance du segment homogène multipliée par $n-1$:

$$RSS_{\text{exp}} = \text{Var}(y_{p_m:n}) * (n-1) \quad (35)$$

où p_m correspond à la position du dernier saut introduit dans la série de base. Puisque le dernier segment de la série constitue la base de la correction des séries, on peut s'attendre à retrouver approximativement la même valeur dans une série bien homogénéisée. Lorsque la somme de carrés résiduelle espérée est inférieure à la somme de carrés résiduelle observée, cela indique que la série n'a pas été assez corrigée parce que les inhomogénéités n'ont pas toutes été détectées et/ou que les amplitudes ont été mal estimées. Ainsi, la série est sous ajustée si $\frac{RSS_{\text{exp}}}{RSS_{\text{obs}}} < 1$ et sur ajustée si $\frac{RSS_{\text{exp}}}{RSS_{\text{obs}}} > 1$. Les méthodes d'homogénéisation ont été appliquées sur des séries synthétiques contenant 2 et 3 sauts et la somme de carrés résiduelle a été calculée. Pour chaque série, on a fait le rapport entre la somme de carrés résiduelle observée et attendue. Des statistiques descriptives sur les rapports de sommes de carrés résiduelles obtenues selon les différentes méthodes sont présentées dans les tableaux 5 et 6.

Tableau 5. Rapport entre les sommes de carrés résiduelles espérées et observées après correction des séries avec deux sauts.

$\frac{RSS_{\text{exp}}}{RSS_{\text{obs}}}$	Snht	regm	reg2	bivt	stus	wils	jaru	rasm
Moyenne	0.97	0.99	0.83	1.01	0.95	1.02	1.01	1.00
Médiane	0.91	0.98	0.87	1.00	0.93	1.00	0.99	0.99
Écart-type	0.39	0.28	0.33	0.28	0.28	0.29	0.29	0.28

Tableau 6. Rapport entre les sommes de carrés résiduelles espérées et observées après correction des séries avec trois sauts.

RSS_{exp} / RSS_{obs}	snht	regm	reg2	bivt	stus	wils	jaru	rasm
Moyenne	0.93	0.97	0.73	0.99	0.92	1.00	0.99	0.95
Médiane	0.86	0.94	0.72	0.96	0.89	0.96	0.95	0.92
Écart-type	0.40	0.34	0.37	0.36	0.32	0.35	0.35	0.34

Les méthodes qui performant le mieux à corriger des séries inhomogènes sont celles dont le rapport entre la somme de carrés résiduelle espérée et observée est près de 1. D'après les tableaux 5 et 6, on peut voir que les méthodes performant bien en général à corriger des sauts multiples sauf pour la régression à deux phases. En effet, la moyenne des rapports de sommes de carrés résiduelles est de 0.83 pour les séries à deux sauts et de 0.73 pour les séries à trois sauts. Dans la prochaine section, on pourra voir si le fait d'estimer l'amplitude en terme de différences de moyenne améliore les résultats de cette méthode. Il semble qu'en augmentant le nombre de sauts dans les séries, les méthodes tendent à sous ajuster les séries.

3.4 Séries avec un saut d'écart-type

Les méthodes comparées sont basées sur l'hypothèse que la variance est constante tout au long de la série. De ce fait, des variations à ce niveau pourraient affecter les résultats d'une procédure d'homogénéisation. Dans le but d'étudier la robustesse des méthodes quant au postulat d'égalité des variances, des séries synthétiques (10 000) avec un saut d'écart-type ont été générées. Le pourcentage de sauts de moyenne faussement détectés a été calculé. Le tableau 7 présente les sauts faussement détectés sur les séries synthétiques avec des sauts d'écart-type.

Tableau 7. Sauts faussement détectés (%) sur les séries avec un saut d'écart-type de 100 ans.

Méthode	jaru	regm	stus	wils	snht	bivt	reg2	rasm
% Rejet	2.4	2.9	4.7	5.1	5.4	5.4	7.7	27.7

On a comparé le taux de fausse détection sur les séries avec un saut d'écart-type avec celui obtenu des séries homogènes. En effet, une méthode robuste à un changement de variance devrait donner approximativement le même pourcentage de fausse détection en présence d'un changement d'écart-type. On remarque que le pourcentage de sauts faussement détectés a augmenté pour chacune des méthodes. Pour vérifier si un changement d'écart-type a un effet significatif sur le nombre de sauts détectés, on a fait un test de comparaison des proportions de sauts détectés sur les séries homogènes et sur les séries avec un saut de variance pour chacune des méthodes. Un intervalle de confiance a été calculé:

$$\begin{aligned} \hat{p}_1 - \hat{p}_2 - Z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} &\leq p_1 - p_2 \\ &\leq \hat{p}_1 - \hat{p}_2 + Z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \end{aligned} \quad (36)$$

dans lequel \hat{p}_1 et \hat{p}_2 représentent respectivement les proportions de sauts faussement détectés dans les séries homogènes de 100 ans et dans les séries avec un saut d'écart-type de 100 ans et n_1 et n_2 leurs tailles d'échantillons. Les résultats sont présentés dans le tableau 8.

Tableau 8. Différence (%) des proportions entre les sauts faussement détectés sur les séries homogènes de 100 ans et sur les séries avec un saut d'écart-type.

Méthode	Borne inférieure de l'IC (5%)	Différence	Borne supérieure de l'IC (5%)
reg2*	2.5	3.3	4.1
regm*	1.2	1.7	2.1
jaru*	0.8	1.2	1.6
bivt*	0.8	1.5	2.2
snht*	0.7	1.4	2.1
stus	-0.4	0.3	1
wils	-0.5	0.3	0.9
rasm	-1.3	0.2	1.7

* Test significatif au seuil de 5%

La plupart des différences de proportions sont significatives au seuil de 5% sauf pour les tests de Student et Wilcoxon séquentiel et pour la méthode Bayésienne. On s'y attendait pour le test de Wilcoxon séquentiel puisqu'il s'agit d'une méthode non paramétrique. Par contre, le test de Student séquentiel est basé sur l'égalité des variances alors ce résultat est inattendu. Encore une fois, ceci peut être dû à la fenêtre mobile étant donné que le changement d'écart-type a moins d'impact sur une petite portion de la série. Enfin, il semble qu'un changement au niveau de la variance dans la série de base augmente la probabilité de commettre l'erreur de type 1 pour la plupart des méthodes, mais cette augmentation est relativement faible. Bref, les sauts de variance ne semblent pas poser de problème quant au niveau de fausse détection.

3.5 Séries avec une tendance

Il peut arriver que des inhomogénéités sous forme plus graduelle se retrouvent dans les données. Toutefois, il est impossible de comparer les méthodes sélectionnées pour ce travail à

identifier des tendances puisque seules la régression multiple et la régression à deux phases ont la capacité de le faire. On a plutôt dénombré le nombre de sauts positionnés à l'intérieur de la tendance (-2 du début et +2 de la fin). Ceci a pour but de montrer que les inhomogénéités sous forme de tendance risquent d'être interprétées comme un ou plusieurs sauts consécutifs par la plupart des techniques. Les tableaux 9 et 10 présentent respectivement le nombre de cas pour lesquels un saut et deux sauts ou plus sont détectés à l'intérieur de la tendance.

Tableau 9. Nombre de cas (%) pour lesquels un saut est détecté à l'intérieur de la tendance.

Amplitude (en écart-type)	snht	regm	reg2	bivt	stus	wils	jaru	rasm
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

Tableau 10. Nombre de cas (%) pour lesquels deux sauts ou plus sont détectés à l'intérieur de la tendance.

Amplitude (en écart-type)	snht	regm	reg2	bivt	stus	wils	jaru	rasm
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

On remarque que dans la majorité des cas, la tendance est interprétée comme un saut de moyenne (tableau 9). Ceci se produit moins souvent lorsque la tendance est de faible amplitude (0-0.5 écart-type). Il arrive également que la tendance soit interprétée comme plusieurs sauts de moyenne consécutifs lorsqu'elle est de forte amplitude (tableau 10). Une façon d'éviter de se

méprendre sur le type de changement serait d'utiliser une méthode graphique combinée à une méthode objective. On pourra ainsi voir si le changement semble abrupt ou graduel.

4. Performance des différentes méthodes lorsque les amplitudes des sauts sont estimées par différences de moyenne

Les méthodes d'homogénéisation étudiées dans ce travail ne sont pas destinées à s'appliquer de façon séquentielle. Par conséquent, l'estimation de l'amplitude est problématique en présence de sauts multiples. Dans cette section, on calcule des différences de moyenne entre les segments respectifs pour estimer l'amplitude des sauts. Cette technique d'estimation vise à éviter l'interférence d'autres sauts éventuels dans le calcul de l'amplitude d'une inhomogénéité. D'abord, on applique une méthode d'homogénéisation successivement jusqu'à ce que tous les segments soient homogènes ou de longueur inférieure à 10. Par la suite, on calcule des différences de moyenne entre tous les segments consécutifs à partir du dernier segment.

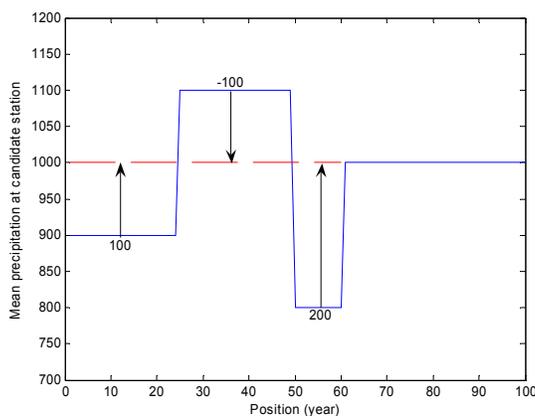


Figure 4 : Exemple de série de base contenant trois sauts.

Prenons par exemple la série de base représentée par la figure 4 qui contient trois sauts aux positions 25, 50 et 60. Les segments représentés sur la figure représentent les moyennes de

chaque portion de la série. Pour estimer les amplitudes, on considère que le dernier segment est homogène et on ajuste les segments précédents par rapport à celui-ci. On estime l'amplitude du dernier saut en calculant la différence des moyennes entre le dernier segment et celui des positions [50 : 60]. On obtient une amplitude positive de 200 et pour corriger on ajoute cette valeur aux données de ce segment de manière à le ramener à une précipitation totale moyenne de 1000. Par la suite, on estime l'amplitude du deuxième saut en calculant la différence de moyenne entre le segment corrigé de [50 : 100] et celui des positions [25 : 49]. La correction de ce segment le ramène également à une valeur moyenne de 1000. On refait la même chose pour le segment de 1 à 24. Les sauts détectés correspondent respectivement à des amplitudes de 100, -100 et 200. Toutes les amplitudes sont estimées avec cette technique dans la présente section.

Par ailleurs, comme l'analyse des séries avec un saut d'écart-type et avec une tendance ne change pas selon la technique employée pour estimer les amplitudes, les résultats des sections 3.4 et 3.5 ne sont pas repris.

4.1 Séries homogènes

Les résultats de la section 3.1 sont repris en exprimant l'amplitude en terme de différence de moyennes de la série de base avant et après le saut. Les tableaux 11 et 12 présentent les résultats obtenus sur les séries homogènes. La figure 5 présente l'amplitude des sauts faussement détectés en fonction de leur position.

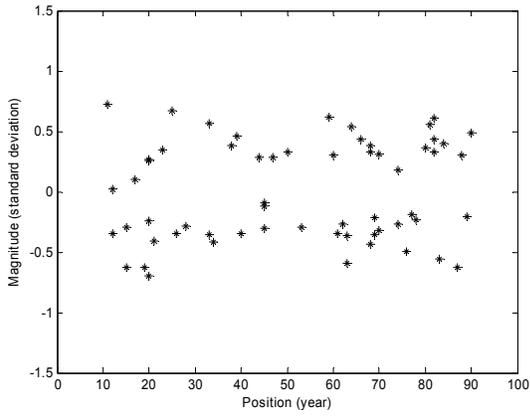
Tableau 11. Sauts faussement détectés (%) sur les séries homogènes de 60 ans.

Amplitude (en écart-type)	regm	jaru	bivt	snht	wils	reg2	stus	rasm
0-0.25	0.1	0.0	0.	0.1	0.3	1.5	1.7	0.7
0.25-0.5	0.4	0.1	0.4	0.4	1.1	1.2	1.9	11.8
0.5-1	0.4	0.9	1.7	2.0	2.1	1.2	1.5	8.5
1-2	0.0	0.0	0.1	0.0	0.3	0.1	0.1	0.2
>2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Total	0.9	1.0	2.2	2.5	3.8	4.0	5.2	21.2

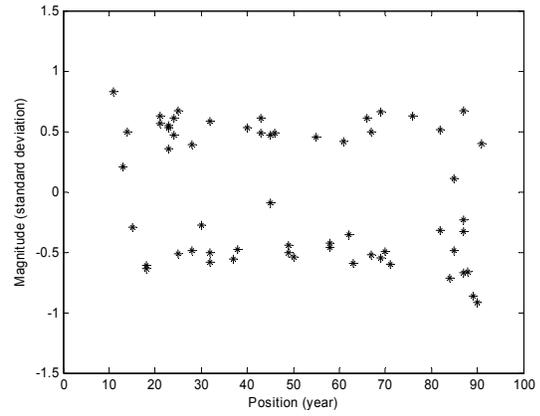
Tableau 12. Sauts faussement détectés (%) sur les séries homogènes de 100 ans.

Amplitude (en écart-type)	regm	jaru	bivt	snht	reg2	wils	stus	rasm
0-0.25	0.2	0.1	0.2	0.3	2.1	0.8	2.3	1.7
0.25-0.5	0.7	0.5	2.0	2.0	1.5	1.7	1.8	18.3
0.5-1	0.3	0.6	1.7	1.7	0.8	2.2	0.8	7.4
1-2	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.1
>2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Total	1.2	1.2	3.9	4.0	4.4	4.8	5.0	27.5

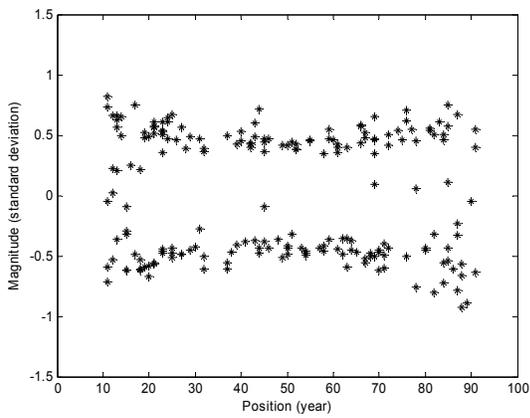
a)



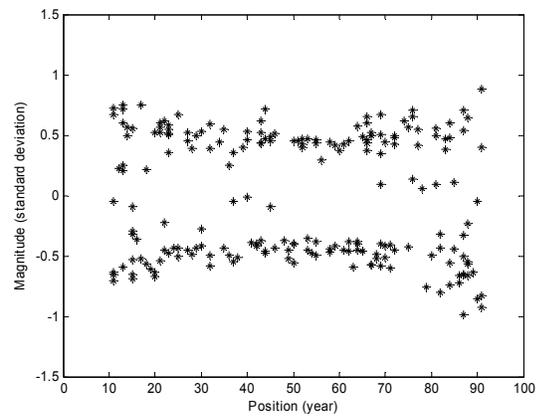
b)



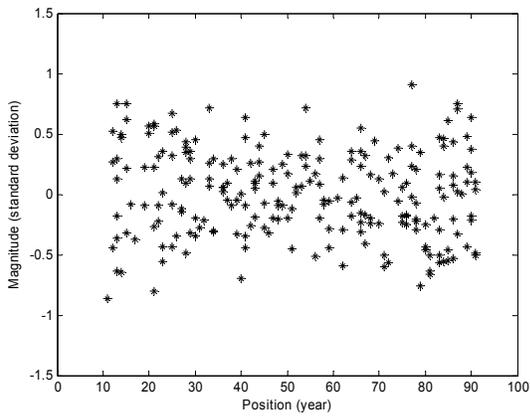
c)



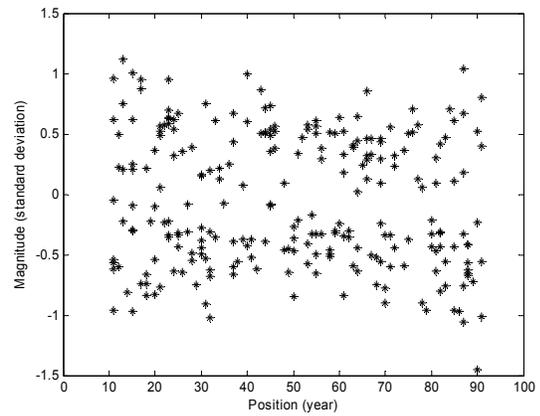
d)



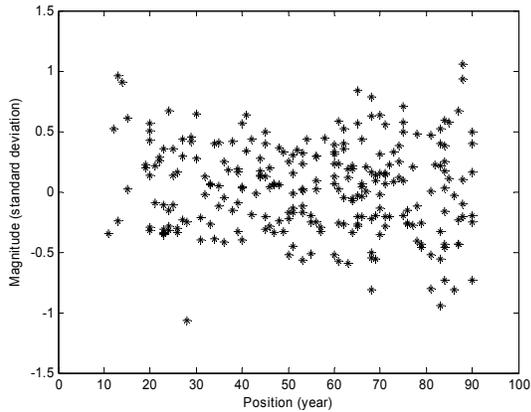
e)



f)



g)



h)

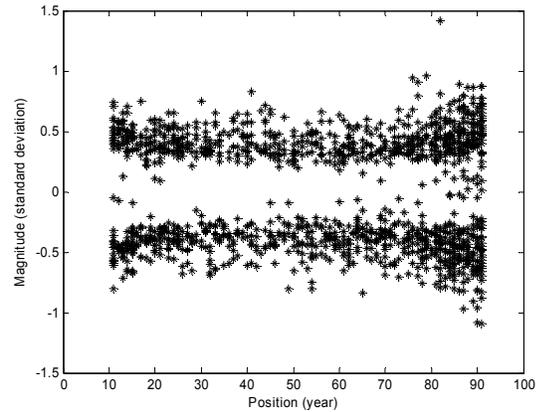


Figure 5 : Sauts faussement détectés sur les séries homogènes de 100 ans a) régression multiple b) méthode de Jaruskova c) approche bivariée d) méthode d’Alexandersson e) régression à deux phases f) test de Wilcoxon séquentiel g) test de Student séquentiel h) méthode Bayésienne.

On peut remarquer d’après la figure 5 qu’en estimant les amplitudes de cette façon, les sauts faussement détectés ont des amplitudes nettement inférieures pour la régression à deux phases (maximum de l’ordre de 1.5 écart-type). Lorsque cette méthode est appliquée, l’amplitude devrait être estimée en terme de différence de moyenne pour éviter de conclure à des sauts d’amplitude exagérée. Pour toutes les autres méthodes, la différence n’est pas marquée.

4.2 Séries avec un saut de moyenne

Dans cette section, l’analyse de la performance des méthodes sur les séries avec un saut est reprise, en estimant leur amplitude par différences de moyenne. Comme l’estimation des amplitudes par les méthodes de Student séquentiel et de Wilcoxon séquentiel se faisait déjà de cette façon, les résultats de ces deux méthodes demeurent inchangés. Les tableaux 13 et 14 présentent les pourcentages de sauts identifiés pour les séries synthétiques de 60 et 100 ans.

Tableau 13. Sauts identifiés (%) sur les séries de 60 ans avec un saut de moyenne.

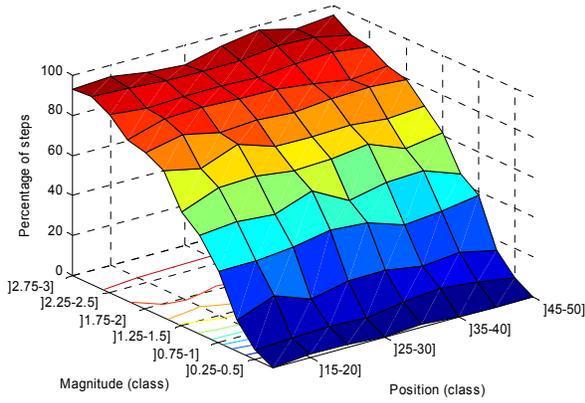
	snht	regm	reg2	bivt	stus	wils	jaru	rasm
Correctement identifié	60.0	56.0	51.9	61.6	32.1	57.3	61.7	58.6
Bien identifié	81.1	71.8	64.7	81.7	70.8	80.5	81.0	81.9
Bien positionné	81.6	72.0	64.8	82.2	71.2	80.9	81.4	83.3

Tableau 14. Sauts identifiés (%) sur les séries de 100 ans avec un saut de moyenne.

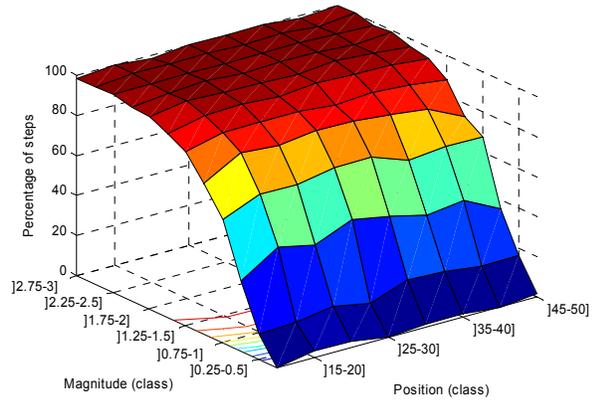
	snht	regm	reg2	bivt	stus	wils	jaru	rasm
Correctement identifié	61.5	59.3	57.4	62.7	30.3	57.9	63.0	57.8
Bien identifié	83.5	76.8	73.5	83.9	67.4	82.4	83.4	83.2
Bien positionné	83.9	76.9	73.6	84.3	67.5	82.7	83.6	85.2

On remarque que pour toutes les méthodes (sauf le test de Student séquentiel et le test de Wilcoxon séquentiel), cette façon d'estimer les amplitudes augmente le pourcentage de sauts correctement et bien identifiés. On remarque également cette amélioration dans les figures 6 et 7 qui présentent le pourcentage de sauts correctement et bien identifiés par rapport à l'amplitude et à la position des sauts.

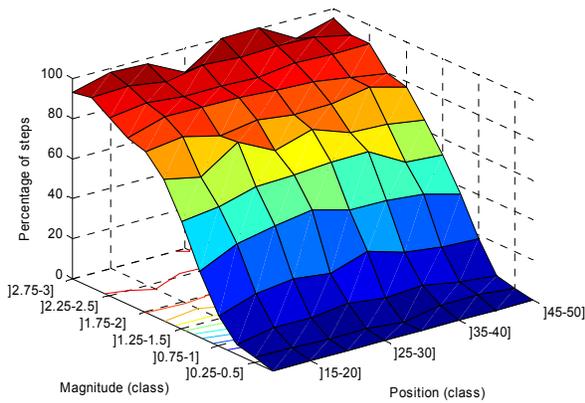
a)-I)



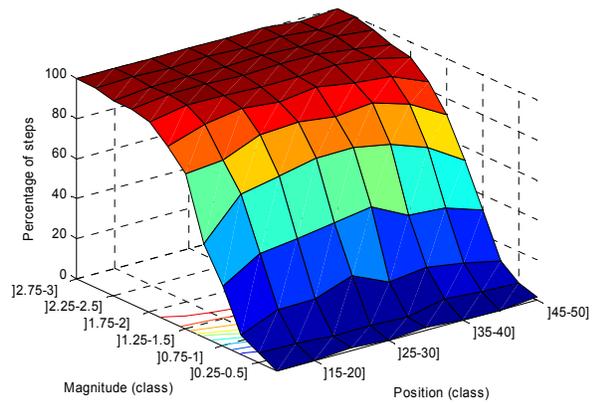
II)



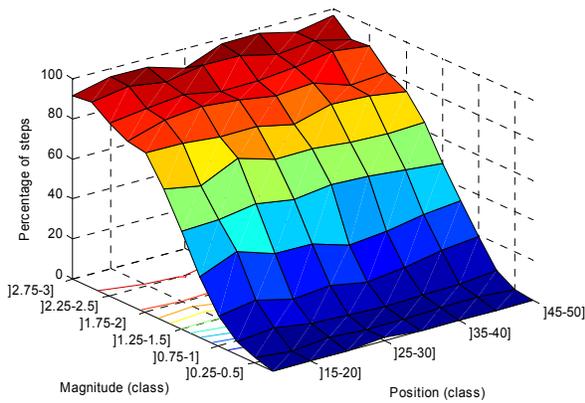
b)-I)



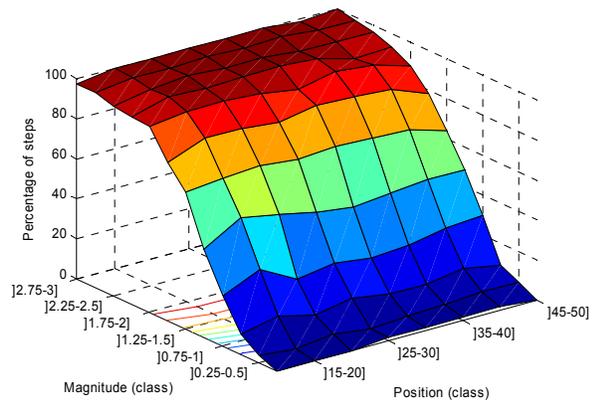
II)



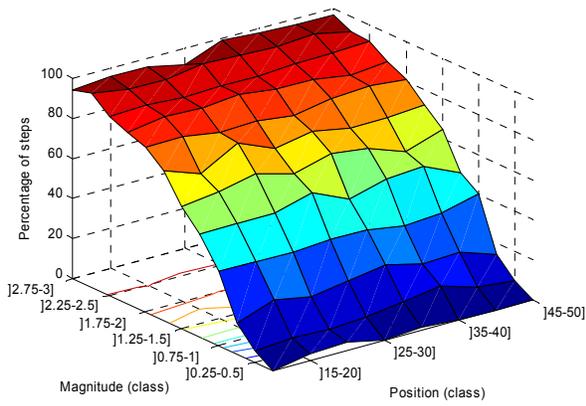
c)-I)



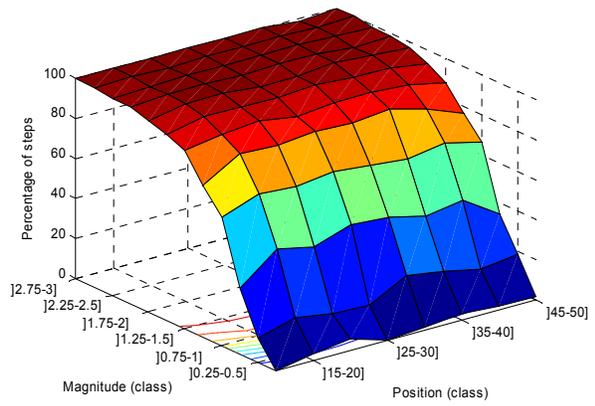
II)



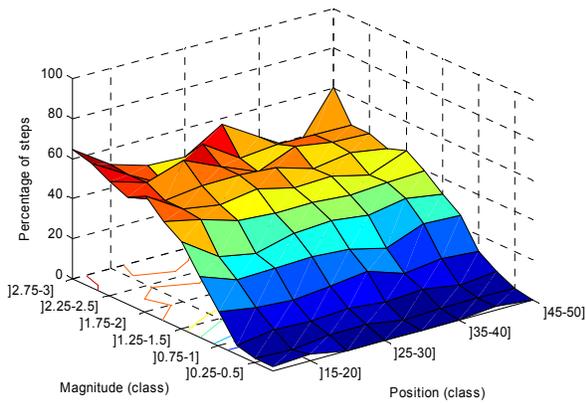
d)-I)



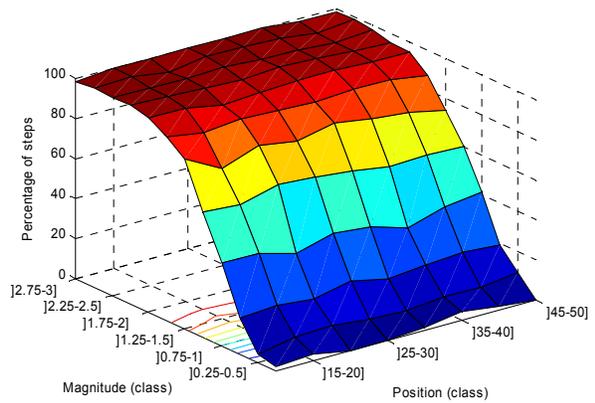
II)



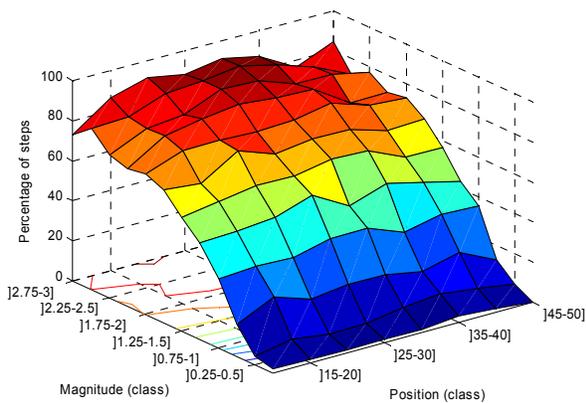
e)-I)



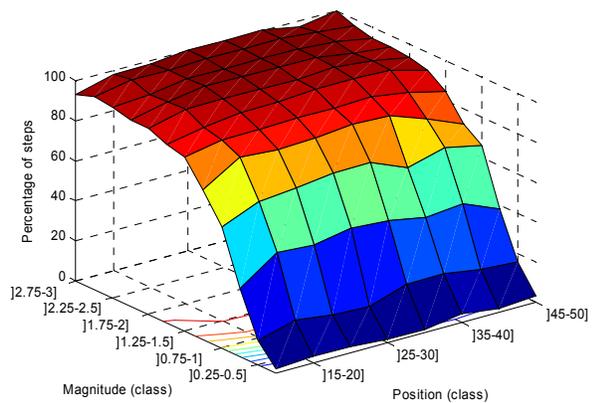
II)



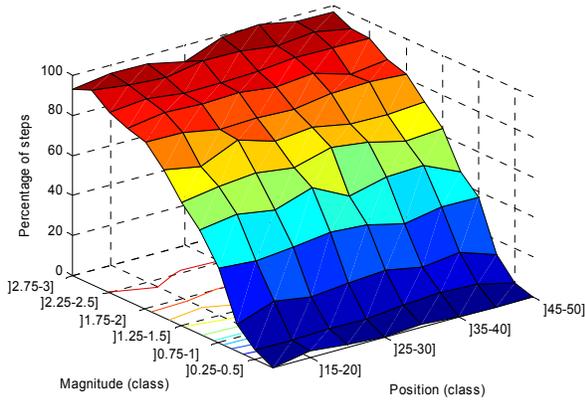
f)-I)



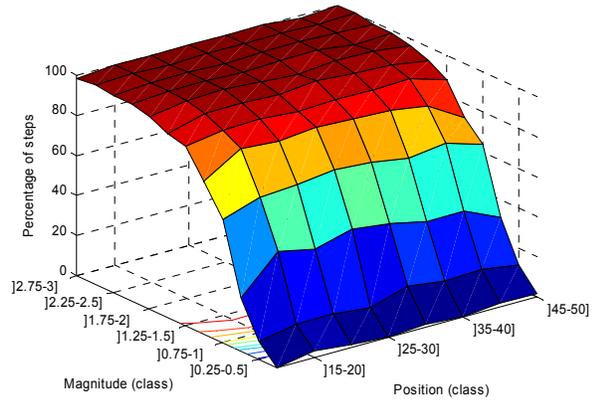
II)



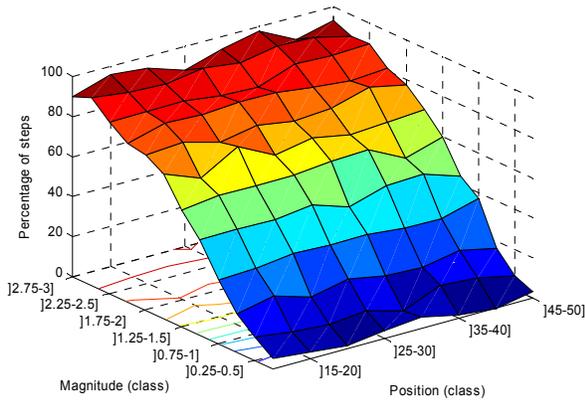
g)-I)



II)



h)-I)



II)

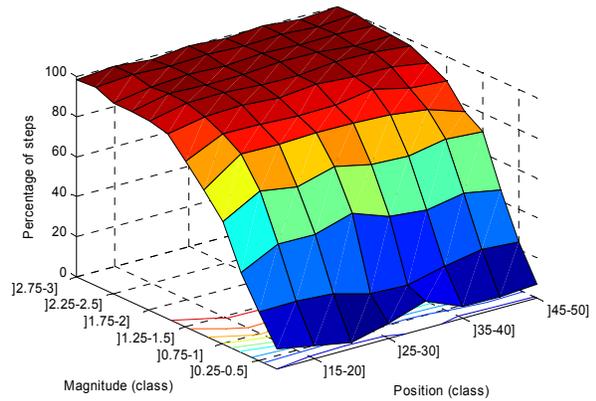
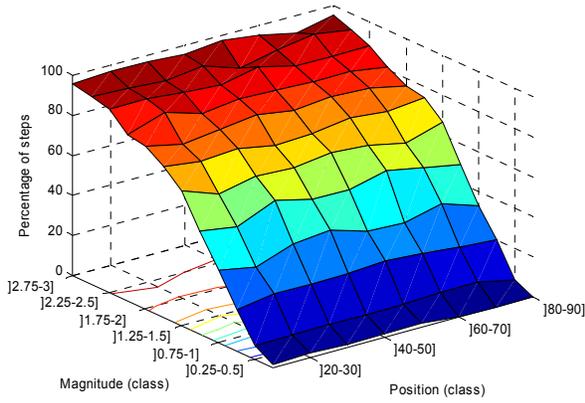
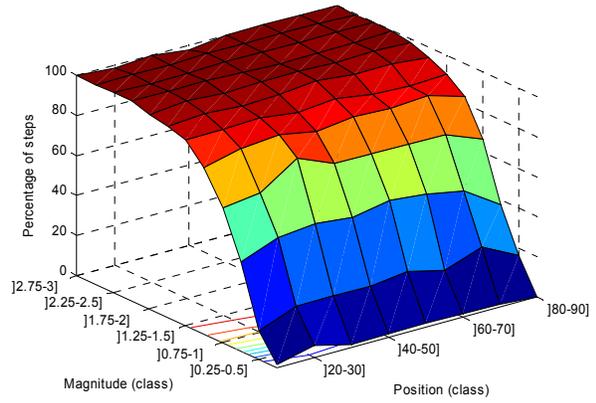


Figure 6 : Sauts identifiés (%) en estimant l'amplitude par différences de moyennes dans les séries avec un saut de moyenne de longueur 60 ans I) correctement identifiés II) bien identifiés a) méthode d'Alexandersson b) régression multiple c) régression à deux phases d) approche bivariée e) test de Student séquentiel f) test de Wilcoxon séquentiel g) méthode de Jaruskova h) méthode Bayésienne.

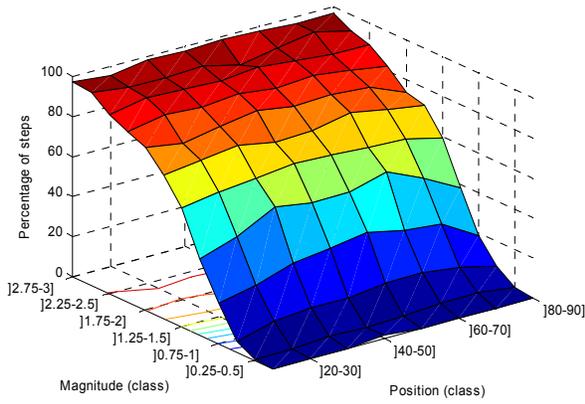
a)-I)



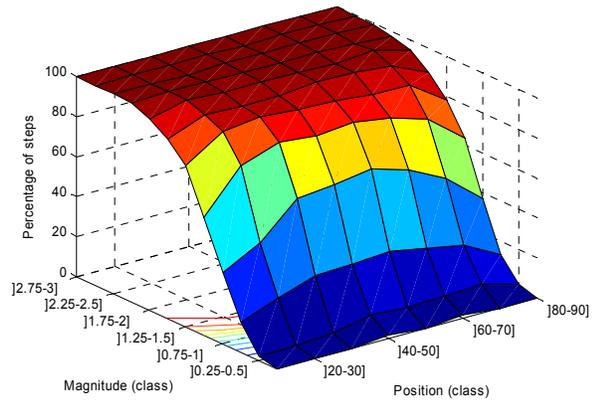
II)



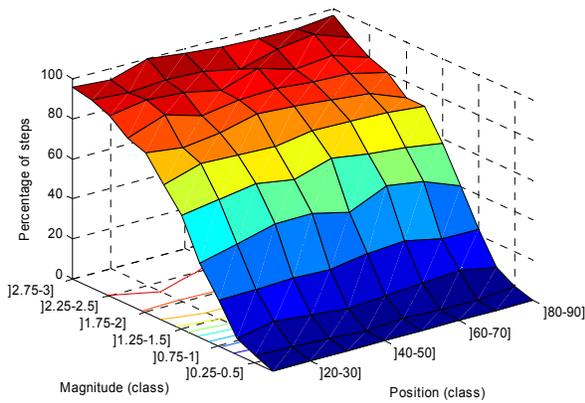
b)-I)



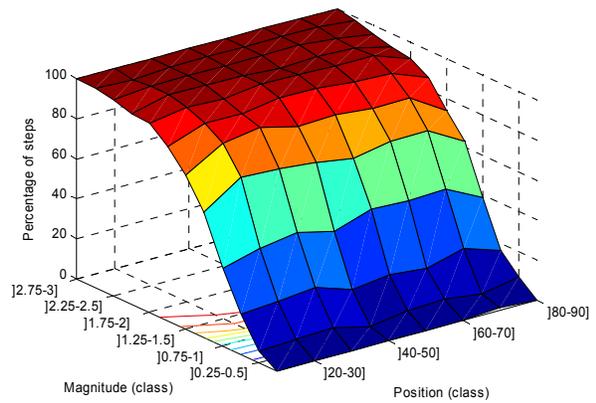
II)



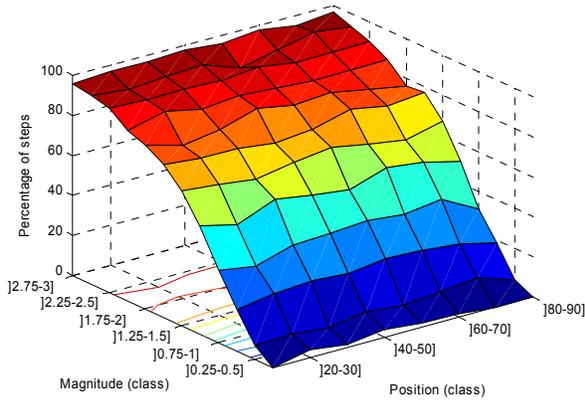
c)-I)



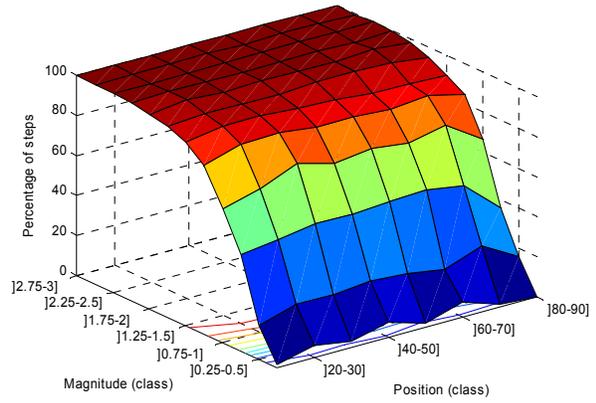
II)



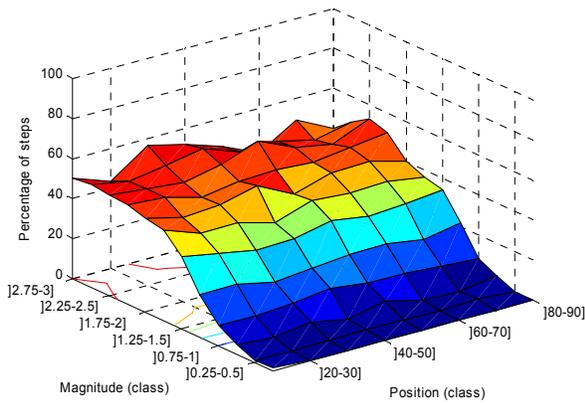
d)-I)



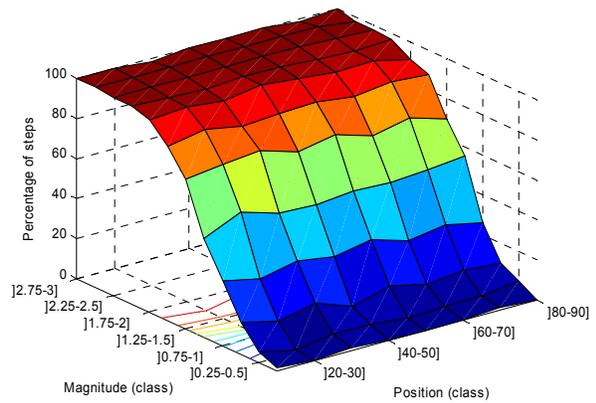
II)



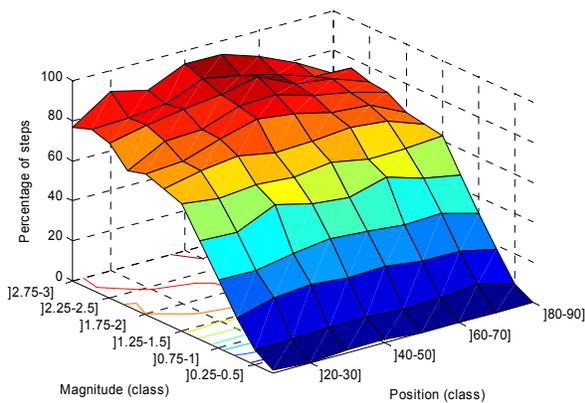
e)-I)



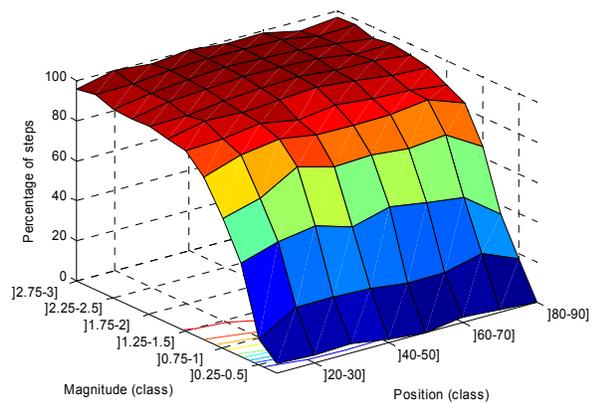
II)



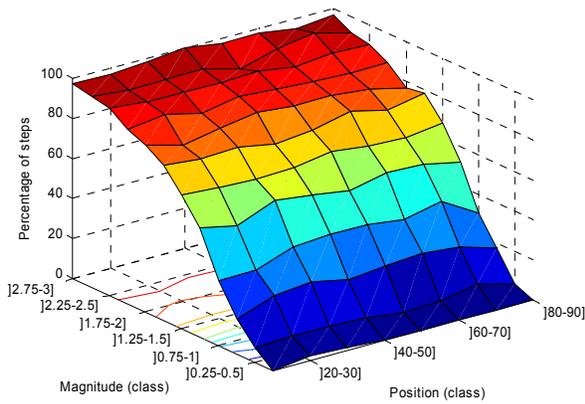
f)-I)



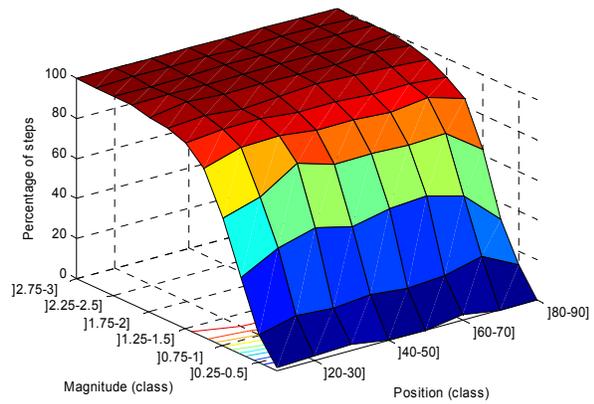
II)



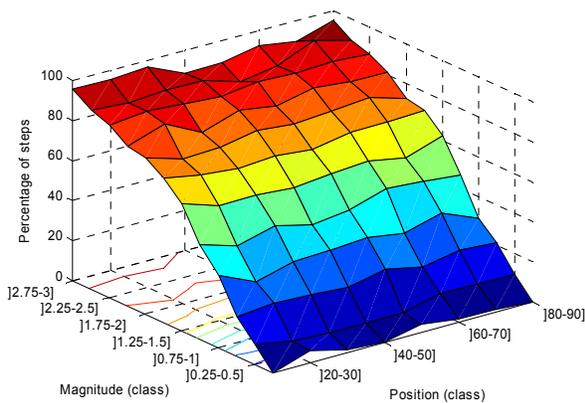
g)-I)



II)



h)-I)



II)

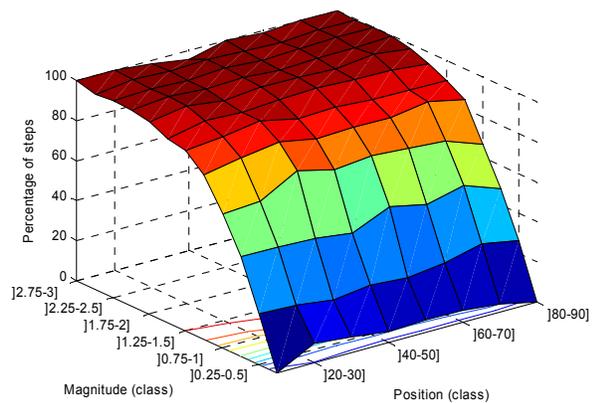


Figure 7 : Sauts identifiés (%) en estimant l'amplitude par différences de moyennes dans les séries avec un saut de moyenne de longueur 100 ans I) correctement identifiés II) bien identifiés a) méthode d'Alexandersson b) régression multiple c) régression à deux phases d) approche bivariée e) test de Student séquentiel f) test de Wilcoxon séquentiel g) méthode de Jaruskova h) méthode Bayésienne.

Toutes ces méthodes performant à bien identifier les sauts d'amplitude supérieure à deux écarts-type. Pour des sauts de 1 écart-type et moins, le pourcentage de sauts bien identifiés se détériore brusquement. Par ailleurs, la performance des méthodes de Student séquentiel et régression à deux phases se dégrade plutôt à partir de 1.5 écart-type et moins. Cependant, la position du saut ne pose plus de problème dans le cas de la régression à deux phases. Enfin, on

remarque que cette façon d'estimer les amplitudes augmente le pourcentage de sauts correctement et bien identifiés, plus particulièrement pour des sauts de faible amplitude. Il est donc préférable d'estimer les amplitudes de cette façon pour toutes les méthodes.

4.3 Séries avec plusieurs sauts de moyenne

4.3.1 Performance évaluée par sommes de carrés résiduelles

Le calcul des amplitudes des séries avec des sauts multiples a été repris et les séries ont été corrigées par la technique des différences de moyenne. On a recalculé les sommes de carrés résiduelles de chaque série et comparé avec les sommes de carrés résiduelles espérées. Les tableaux 15 et 16 présentent les résultats.

Tableau 15. Rapport entre les sommes de carrés résiduelles espérées et observées après correction des séries avec deux sauts.

	snht	regm	reg2	bivt	stus	wils	jaru	rasm
Moyenne	1.02	1.00	0.98	1.02	0.95	1.02	1.02	1.03
Médiane	1.01	0.99	0.97	1.01	0.93	1.00	1.01	1.01
Écart-type	0.29	0.29	0.29	0.29	0.28	0.29	0.29	0.29

Tableau 16. Rapport entre les sommes de carrés résiduelles espérées et observées après correction des séries avec trois sauts.

	snht	regm	reg2	bivt	stus	wils	jaru	rasm
Moyenne	1.00	0.98	0.95	1.01	0.92	1.00	1.00	0.99
Médiane	0.97	0.95	0.91	0.98	0.89	0.96	0.97	0.96
Écart-type	0.35	0.35	0.34	0.35	0.32	0.35	0.35	0.36

On remarque une grande amélioration de la performance pour la régression à deux phases en calculant les amplitudes par différences de moyenne. Pour les autres méthodes, l'amélioration ne semble pas très marquée. Étant donné que les rapports se situent tous près de 1, il est difficile d'évaluer les méthodes les plus appropriées pour détecter des sauts multiples.

4.3.2 Performance évaluée par un critère de positionnement

En estimant l'amplitude par la technique des différences de moyenne, des méthodes qui positionnent des sauts aux mêmes endroits évaluent les amplitudes de manière identique. Par conséquent, il est important de s'attarder aux positions détectées dans les séries avec sauts multiples. On a donc établi un critère de performance des méthodes à trouver l'emplacement de plusieurs sauts. Ce dernier mesure l'écart entre les positions des sauts réels et des sauts détectés. Soit $p_i^d, i = 1, \dots, m$ et $p_j^r, j = 1, \dots, l$ représentant respectivement les positions des sauts détectés et réels, le critère établi s'énonce comme suit :

$$C = \begin{cases} \frac{1}{m} \sum_{i=1}^m (p_i^d - p_i^r)^2, & l = m \\ \frac{1}{l} \left[\sum_{i=1}^m (p_i^d - p_i^r)^2 + |l - m|(n-1)^2 \right], & l > m \\ \frac{1}{m} \left[\sum_{j=1}^l (p_j^d - p_j^r)^2 + |l - m|(n-1)^2 \right], & l < m \end{cases} \quad (37)$$

où n représente la longueur des séries. Les paires (p_i^d, p_j^r) sont choisies de manière à minimiser le critère. Lorsque le nombre de sauts détectés et réels est le même, on calcule les sommes de carrés des différences entre les paires qui minimisent le critère. Lorsque le nombre de sauts détectés et réels diffère, on cherche la disposition des paires qui minimise le critère et on ajoute une pénalité de $(n-1)^2$ pour chaque saut faussement détecté ou non détecté. Cette valeur

correspond à la distance maximale possible entre deux sauts. Un critère de 0 signifie que tous les sauts ont été correctement positionnés. Un critère près de 0 indique que les sauts se situent près des positions réelles. Un critère très élevé désigne une série dans laquelle des sauts n'ont pas été détectés ou des sauts fictifs ont été détectés. Il a été calculé pour les ensembles de séries synthétiques avec deux et trois sauts. Les tableaux 17 et 18 présentent des statistiques descriptives sur les critères obtenus des différentes méthodes.

Tableau 17. Statistiques descriptives du critère de positionnement des séries avec deux sauts.

	snht	regm	reg2	bivt	stus	wils	jaru	rasm
Moyenne	1768	2757	3044	1730	3268	1812	1928	1813
Médiane	2	4901	4901	2	4901	3	2	18
Écart-Type	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

Tableau 18. Statistiques descriptives du critère de positionnement des séries avec trois sauts.

	snht	regm	reg2	bivt	stus	wils	jaru	rasm
Moyenne	2439	3291	3654	2378	3476	2565	2683	2216
Médiane	3267	3267	3267	3267	3267	3267	3267	2451
Écart-Type	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

Pour les séries avec deux sauts, l'approche bivariée a donné le critère moyen le moins élevé tandis que dans les séries avec trois sauts, c'est plutôt la méthode Bayésienne qui présente cette caractéristique. Une analyse de variance de Kruskal-Wallis a été réalisée au seuil significatif

de 5% afin d'arriver à cette conclusion. Néanmoins, dans les deux cas, l'approche bivariée, la méthode d'Alexandersson, le test de Wilcoxon séquentiel, la méthode de Jaruskova et la méthode Bayésienne ont des critères moyens moindres. De plus, dans le tableau 17, le critère médian pour ces cinq méthodes est très faible ce qui signifie que dans la moitié des séries synthétiques à deux sauts, ces méthodes ont très bien positionnés tous les sauts sans en détecter des non existants. Pour les séries avec trois sauts, les critères médians sont beaucoup plus élevés. En effet, lorsque le nombre de sauts réels s'accroît, il devient très ardu de tous les identifier. Le maximum obtenu pour toutes les méthodes est de 9801 et correspond au cas où tous les sauts réels ne sont pas détectés. Les figures 8 et 9 présentent un histogramme des critères obtenus sur les deux ensembles de séries synthétiques. Pour chaque classe de critère, huit bandes représentant les huit méthodes sont illustrées.

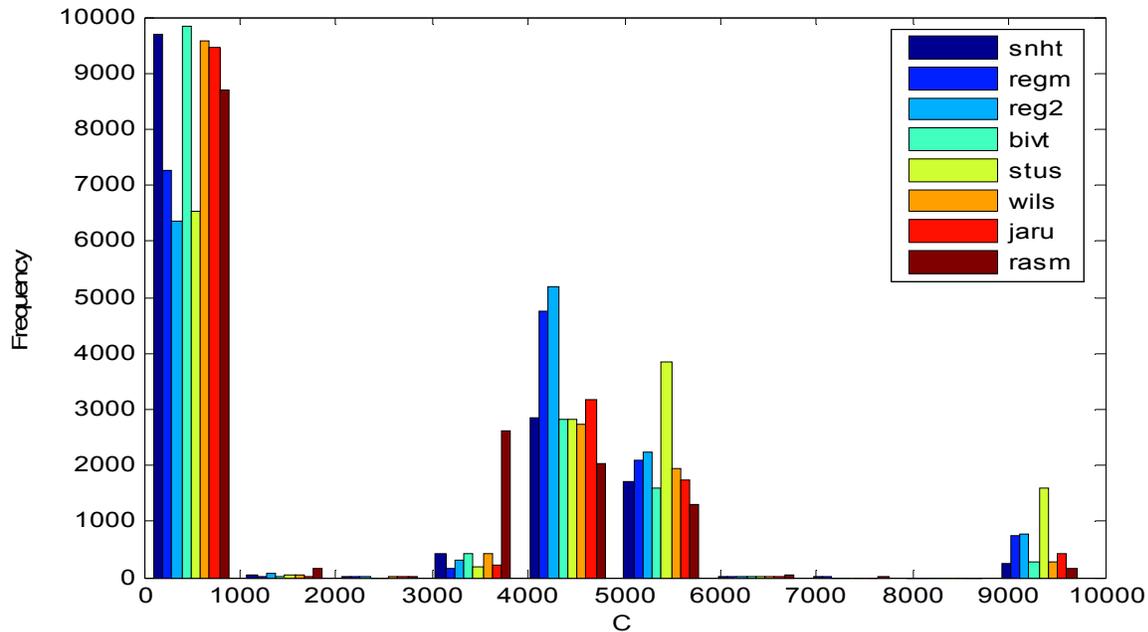


Figure 8 : Histogramme du critère de positionnement selon les huit méthodes des séries avec deux sauts.

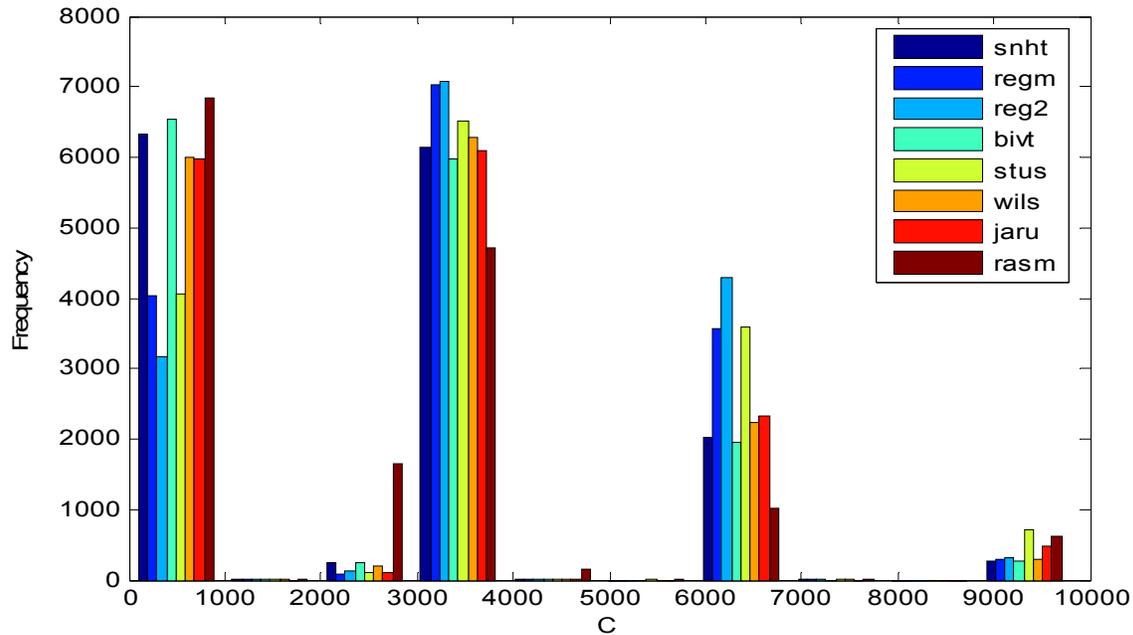


Figure 9 : Histogramme du critère de positionnement selon les huit méthodes des séries avec trois sauts.

Les méthodes les plus performantes sont celles dont les critères se retrouvent en forte majorité dans la première classe (critère compris entre 0 et 1000). On dénote cinq méthodes qui se démarquent dans la première classe, l'approche bivariée, la méthode d'Alexandersson, le test de Wilcoxon séquentiel, la méthode de Jaruskova et la méthode Bayésienne. La régression multiple, la régression à deux phases et le test de Student séquentiel ont des fréquences plus faibles dans la première classe et des fréquences élevées dans des classes subséquentes. Cela signifie qu'on retrouve plus de critères de positionnement élevé pour ces méthodes.

Par la suite, on s'est intéressé à la performance des différentes méthodes selon la distance séparant les deux sauts. Pour chaque méthode, on a calculé le critère de positionnement moyen selon différentes classes de distance entre les deux sauts. La figure 10 présente les résultats.

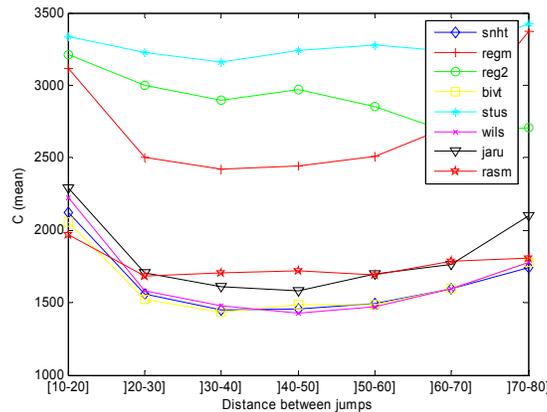


Figure 10 : Critères de positionnement moyens selon la distance entre les deux sauts.

On remarque effectivement que la distance entre les sauts semble influencer le critère de performance. Il semble que des sauts très rapprochés ou très éloignés sont plus difficilement identifiés que deux sauts de distance moyenne. Par ailleurs, pour le test de Student séquentiel la distance entre les sauts n'a aucun effet en raison de la fenêtre mobile. On utilise toujours 10 observations de part et d'autre de chaque saut. La régression à deux phases semble mieux performer à positionner deux sauts très éloignés que deux sauts rapprochés. Ensuite, on s'est intéressé à savoir s'il semble plus facile d'identifier des sauts de signes contraires (un saut positif suivi d'un saut négatif par exemple) ou deux sauts de même signe. On a analysé la performance des méthodes selon que les sauts successifs sont de signes contraires ou analogues. Pour chaque méthode, on a calculé le critère moyen selon les suites de signes possibles. La figure 11 présente les résultats.

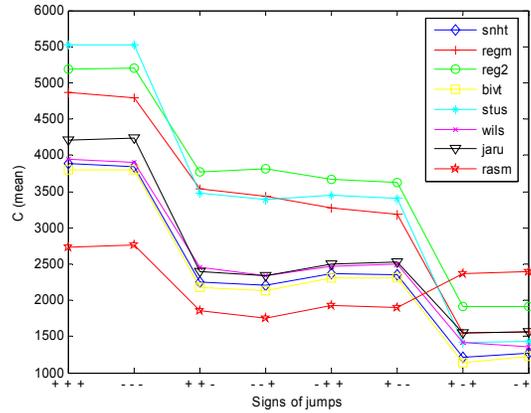


Figure 11 : Critères de positionnement moyens selon la suite des signes des sauts.

On remarque effectivement que la suite des signes semble influencer le critère de performance. En effet, dans les deux cas où les trois sauts sont du même signe, on obtient des critères moyens plus élevés que dans les autres cas. Lorsque que les sauts sont suivis d'un autre saut de signe opposé, la performance semble meilleure sauf pour la méthode Bayésienne. Enfin, les cas dans lesquels deux sauts de même signe sont suivis ou précédés d'un saut de signe opposé donnent une meilleure performance que trois sauts de même signe.

5. Discussion

Lorsqu'on évalue l'homogénéité d'une série, on ne sait pas préalablement si la série de base sera homogène ou si elle contiendra une ou plusieurs inhomogénéités. De ce fait, ce travail vise à identifier les méthodes qui possèdent la capacité de détecter à la fois des séries homogènes et des sauts de moyenne. De plus, les méthodes robustes à des changements de variance seraient préférables. Des séries synthétiques représentant les précipitations totales annuelles typiques du Québec ont été générées. Ces dernières ont été générées dans plusieurs ensembles (homogènes, un saut de moyenne, plusieurs sauts de moyenne, une tendance, un saut d'écart-type). On a présenté deux sections de résultats dans le but de comparer la performance des méthodes selon que l'on estime les amplitudes tel que présenté dans la littérature ou en calculant des différences de moyenne entre les segments respectifs de la série de base. On a remarqué que dans tous les cas (1, 2 ou 3 sauts), la méthode d'estimation des amplitudes par différences de moyenne donne de meilleurs résultats. Il est donc préférable, plus particulièrement en présence de sauts multiples, d'estimer les amplitudes de cette façon peu importe la méthode utilisée.

La performance des méthodes dans ce travail peut se résumer selon certains critères. Tout d'abord, il est très important que le nombre de sauts faussement détectés soit faible. Au-delà de 5%, cela devient très risqué. On pourrait introduire des inhomogénéités dans une série qui est homogène en réalité. Par la suite, une méthode d'homogénéisation doit avoir la capacité d'identifier un saut dans une série. Pour toutes amplitudes confondues, si la méthode peut positionner un saut au moins dans 75% des cas, on peut dire qu'il s'agit d'une bonne performance. Dans le cas des fortes amplitudes (2 écarts-types et plus), on s'attend à ce qu'elles soient identifiées presque 100% du temps. On s'attend également à observer un pourcentage de

sauts bien identifiés raisonnable à partir de 1 écart-type. Ensuite, on s'attend à une performance équivalente selon que l'on estime l'amplitude par différence de moyenne ou par chaque méthode respective dans le cas des séries avec un seul saut. Ces méthodes ont été développées pour détecter un seul saut, alors elles devraient être en mesure de bien l'estimer. Pour les sauts multiples, il est intéressant de déterminer des méthodes qui sont capables de positionner plusieurs sauts sans en négliger ni en détecter de fictifs. Enfin, les méthodes qui peuvent s'appliquer directement dans des conditions réelles sans modifications additionnelles sont avantageuses car elles permettent une économie de temps. Le tableau suivant présente un résumé des forces et faiblesses de chacune des méthodes. Cependant, il est à noter que les critères proposés sont subjectifs et qu'ils peuvent varier selon les besoins.

Tableau 19. Résumé des forces et faiblesses des méthodes d'homogénéisation comparées.

Méthode	Forces	Faiblesses
snht	<ul style="list-style-type: none"> - Faible pourcentage de fausse détection (<5%) - Pourcentage de sauts positionnés élevé (>80%) - Sauts de l'ordre de 1 écart-type bien identifiés (>70%) - Sauts de 2 écarts-type et plus bien identifiés (>90%) - Bonne capacité à positionner des sauts multiples 	<ul style="list-style-type: none"> - Amplitude mieux estimée par différences de moyenne dans le cas d'un seul saut
regm	<ul style="list-style-type: none"> - Faible pourcentage de fausse détection (<5%) - Sauts de l'ordre de 1 écart-type bien identifiés (>70%) - Sauts de 2 écarts-type et plus bien identifiés (>90%) 	<ul style="list-style-type: none"> - Difficulté à détecter des sauts multiples (C élevé) - Amplitude mieux estimée par différences de moyenne dans le cas d'un seul saut
reg2	<ul style="list-style-type: none"> - Faible pourcentage de fausse détection (<5%) - Sauts de 2 écarts-type et plus bien identifiés (>90%) 	<ul style="list-style-type: none"> - Pourcentage de sauts positionnés faible (<75%) - Sauts de l'ordre de 1 écart-type bien identifiés (<70%) - Difficulté à détecter des sauts multiples (C élevé) - Amplitude mieux estimée par différences de moyenne dans le cas d'un seul saut
bivt	<ul style="list-style-type: none"> - Faible pourcentage de fausse détection (<5%) - Pourcentage de sauts positionnés élevé (>80%) - Sauts de l'ordre de 1 écart-type bien identifiés (>70%) - Sauts de 2 écarts-type et plus bien identifiés (>90%) - Bonne capacité à positionner des sauts multiples 	<ul style="list-style-type: none"> - Amplitude mieux estimée par différences de moyenne dans le cas d'un seul saut
stus	<ul style="list-style-type: none"> - Faible pourcentage de fausse détection (<5%) - Sauts de 2 écarts-type et plus bien identifiés (>90%) 	<ul style="list-style-type: none"> - Pourcentage de sauts positionnés faible (<75%) - Sauts de l'ordre de 1 écart-type bien identifiés (<70%) - Difficulté à détecter des sauts multiples (C élevé) - Des seuils significatifs doivent être calculés pour appliquer sur des séries de longueur autre que 60 ou 100 ans.
wils	<ul style="list-style-type: none"> - Faible pourcentage de fausse détection (<5%) - Pourcentage de sauts positionnés élevé (>80%) - Sauts de l'ordre de 1 écart-type bien identifiés (>70%) - Sauts de 2 écarts-type et plus bien identifiés (>90%) - Bonne capacité à positionner des sauts multiples 	<ul style="list-style-type: none"> - Des seuils significatifs doivent être calculés pour appliquer sur des séries de longueur autre que 60 ou 100 ans.
jaru	<ul style="list-style-type: none"> - Faible pourcentage de fausse détection (<5%) - Pourcentage de sauts positionnés élevé (>80%) - Sauts de l'ordre de 1 écart-type bien identifiés (>70%) - Sauts de 2 écarts-type et plus bien identifiés (>90%) - Bonne capacité à positionner des sauts multiples 	<ul style="list-style-type: none"> - Amplitude mieux estimée par différences de moyenne dans le cas d'un seul saut
rasm	<ul style="list-style-type: none"> - Pourcentage de sauts positionnés élevé (>80%) - Sauts de l'ordre de 1 écart-type bien identifiés (>70%) - Sauts de 2 écarts-type et plus bien identifiés (>90%) - Bonne capacité à positionner des sauts multiples 	<ul style="list-style-type: none"> - Fort pourcentage de fausse détection (>20%) - Amplitude mieux estimée par différences de moyenne dans le cas d'un seul saut

L'application sur des séries qui contiennent des sauts d'écart-type a montré que la plupart des méthodes comparées dans ce travail sont peu sensibles à ce type d'inhomogénéité. On a donc omis cette caractéristique dans le tableau des forces et faiblesses. L'analyse des méthodes sur des séries synthétiques avec une tendance a montré qu'un changement plus graduel est souvent interprété comme un ou plusieurs sauts consécutifs à l'intérieur de la tendance. Il ne s'agit pas d'une faiblesse puisque les méthodes comparées ne sont pas destinées à cette fin. D'après les caractéristiques présentées dans ce tableau, une méthode hybride utilisant les forces de ces méthodes pourrait être utilisée. En se basant sur le fait qu'il est préférable d'omettre des inhomogénéités dans une série que d'en introduire des erronées, une possibilité de méthode hybride est illustrée dans la figure 12.

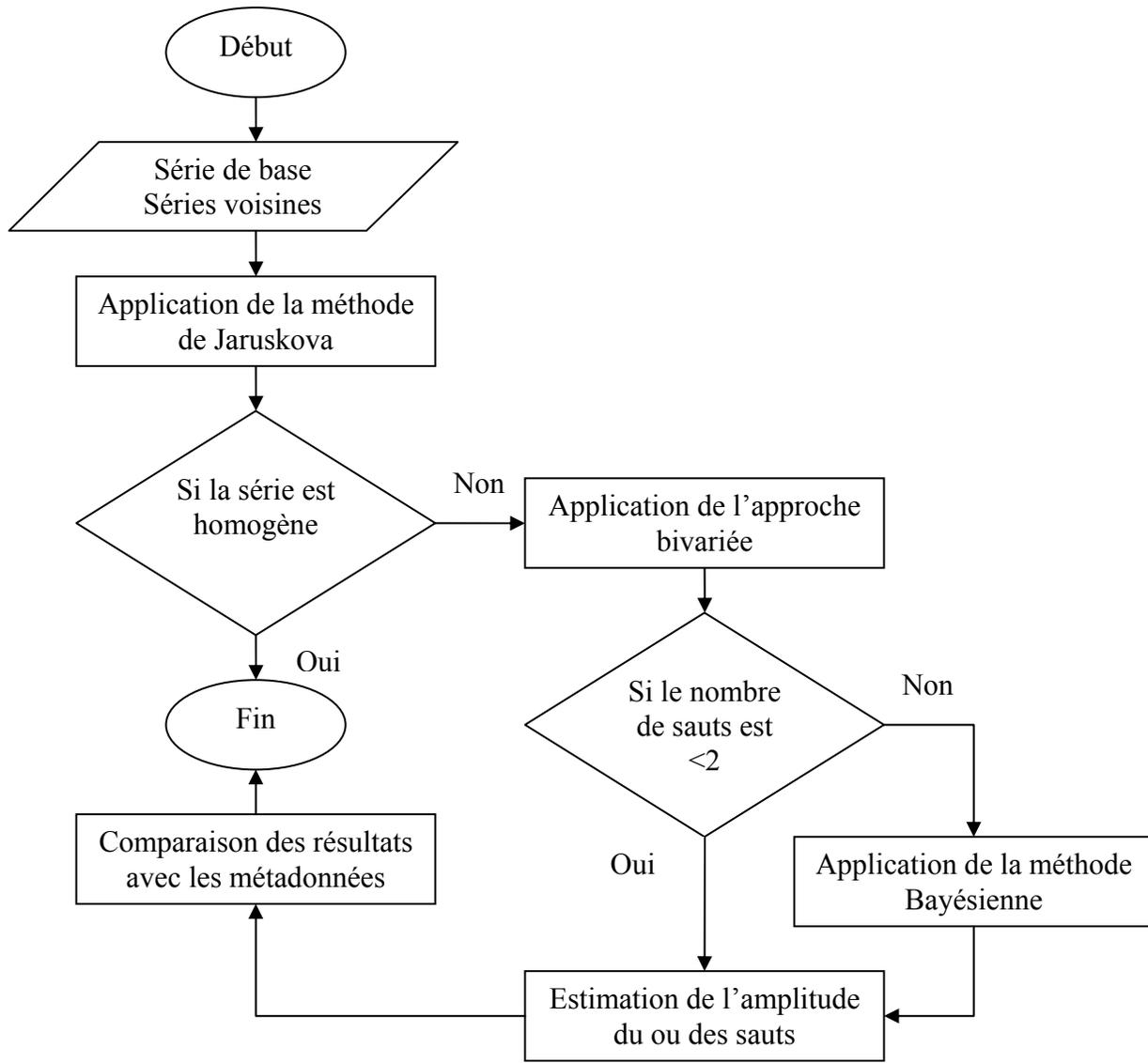


Figure 12 : Algorithme de la méthode hybride proposée.

On commence donc par appliquer la méthode de Jaruskova en raison de sa grande capacité à identifier des séries homogènes. La régression multiple donne une performance équivalente, mais la méthode de Jaruskova a été préférée parce qu'elle risque moins de déclarer des séries homogènes si elles ne le sont pas. En effet, sur les séries avec un saut, la méthode de Jaruskova se montre moins conservatrice que la régression multiple. Ensuite, si la méthode de

Jaruskova déclare la série inhomogène, on applique l'approche bivariée. En effet, sur les séries avec un saut, l'approche bivariée se situe en deuxième position pour le nombre de sauts positionnés. Étant donné que la méthode la plus performante sur les séries avec un saut (méthode Bayésienne) détecte aussi un grand nombre de sauts inexistants, on a préféré l'approche bivariée qui a un pourcentage de sauts faussement détecté faible sur les séries homogènes. De plus, on propose de déclarer un saut significatif seulement si l'explication se trouve dans les métadonnées. L'approche bivariée a également une très bonne capacité à détecter deux sauts et la méthode de Rasmussen performe mieux pour identifier trois sauts. Donc, à partir de 2 sauts et plus, on pourrait déclarer les sauts significatifs lorsqu'ils ont été détectés par l'approche bivariée ou par la méthode de Rasmussen et que les métadonnées peuvent expliquer ces inhomogénéités. Ainsi, on a une forte probabilité d'identifier tous les sauts existants et on évite les fausses détections grâce aux métadonnées. Cette démarche est assez conservatrice et se base sur l'idée qu'il vaut mieux ne pas détecter des sauts que d'appliquer des corrections pour des sauts inexistants.

6. Conclusion

On a observé des performances différentes pour les huit méthodes d'homogénéisation présentées dans ce travail. On retient quatre méthodes qui ont des performances à peu près équivalentes dans tous les cas (méthode d'Alexandersson, test bivarié, test de Wilcoxon séquentiel et méthode de Jaruskova). Ces méthodes peuvent être appliquées dans tous les cas. Par ailleurs, certaines méthodes ne peuvent s'appliquer sur tous types de séries de données. Par exemple, la régression multiple performe très bien dans l'identification des séries homogènes et de façon satisfaisante dans l'identification d'un seul saut. Cependant, en présence de sauts multiples, cette méthode est désavantagée. Ensuite, la méthode de Rasmussen performe très bien dans l'identification d'un seul saut ou de sauts multiples, mais tend à détecter des sauts non existants dans les séries homogènes. Enfin, le test de Student séquentiel et la régression à deux phases sont recommandables dans l'identification des séries homogènes, mais ont plus de difficulté à identifier un ou plusieurs sauts. Une méthode hybride utilisant les forces de plusieurs méthodes comme celle présentée dans la section 5 peut être utilisée.

Ce travail compare uniquement des méthodes objectives parce que l'application de méthodes graphiques sur des milliers de séries synthétiques constitue une tâche irréalisable. 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.

Les mathématiques des méthodes comparées dans cette étude ont été développées pour détecter un seul saut. L'application séquentielle de ces méthodes peut les désavantager puisqu'elles ne sont pas conçues pour détecter des sauts multiples. Cependant, la présence de plusieurs inhomogénéités dans une série de données réelles est un problème tout à fait concret. Le développement de méthodes capables d'identifier un ou plusieurs changements de plusieurs types (saut de moyenne, tendance, saut de variance) est nécessaire. De plus, on a testé la robustesse des méthodes face à un saut de variance, mais les méthodes d'homogénéisation sont basées sur d'autres postulats comme la normalité des données et l'homogénéité des séries voisines. La violation de ces derniers peut également altérer la performance des méthodes et il serait intéressant d'en évaluer l'impact. Ensuite, ces méthodes ont été appliquées dans des conditions spécifiques et les résultats de cette étude font partie d'un domaine de validité restreint. Donc, une application des méthodes jugées plus performantes par ce travail sur des données de précipitations totales annuelles réelles serait également intéressante. Enfin, les métadonnées permettent d'expliquer la provenance d'un changement et d'identifier si il s'agit d'un changement artificiel ou naturel. Idéalement, les réseaux d'observations seraient denses et bien documentés. Des travaux futurs dans ces directions seraient grandement appréciés.

7. Bibliographie

- Alexandersson H., 1986: A homogeneity test applied to precipitation data. *J. Climatol.*, **6**, 661-675.
- Beaulieu C., Ouarda, T.B.M.J. and O. Seidou, 2005 : Synthèse des techniques d'homogénéisation des séries climatiques et analyse d'applicabilité aux séries de précipitations. *Soumis au Journal des Sciences Hydrologiques*.
- Ducré-Robitaille J.-F., Boulet G. 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. Proceedings, *Fifth International Meeting on Statistical Climatology*, Toronto, ON, Atmospheric Environment Service, Environment Canada, J28-J32.
- 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., Vincent L.A. and P.J.F. Sajecki, 1990: Testing for homogeneity in temperature time series at Canadian climate stations. Canadian Climate Centre Rep. 90-4, Atmospheric Environment Service, Downsview, ON, 43 p.
- 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.
- 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.
- 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.
- 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.
- Slonosky V.C., Jones P.D. and T.D. Davies, 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.
- Vincent L.A., 1998: A technique for the identification of inhomogeneities in Canadian temperature series. *J. Climate*, **11**, 1094-1105.

Young K.C., 1993: Detecting and removing inhomogeneities from long-term monthly sea level pressure time series. *J. Climate*, **6**, 1205-1220.

Wang X.L., 2003: Comments on 'Detection of Undocumented Changepoints: A revision of the Two-Phase regression model'. *J. Climate* **16**, 3383-3385.