

INRS-Eau
UNIVERSITÉ DU QUÉBEC

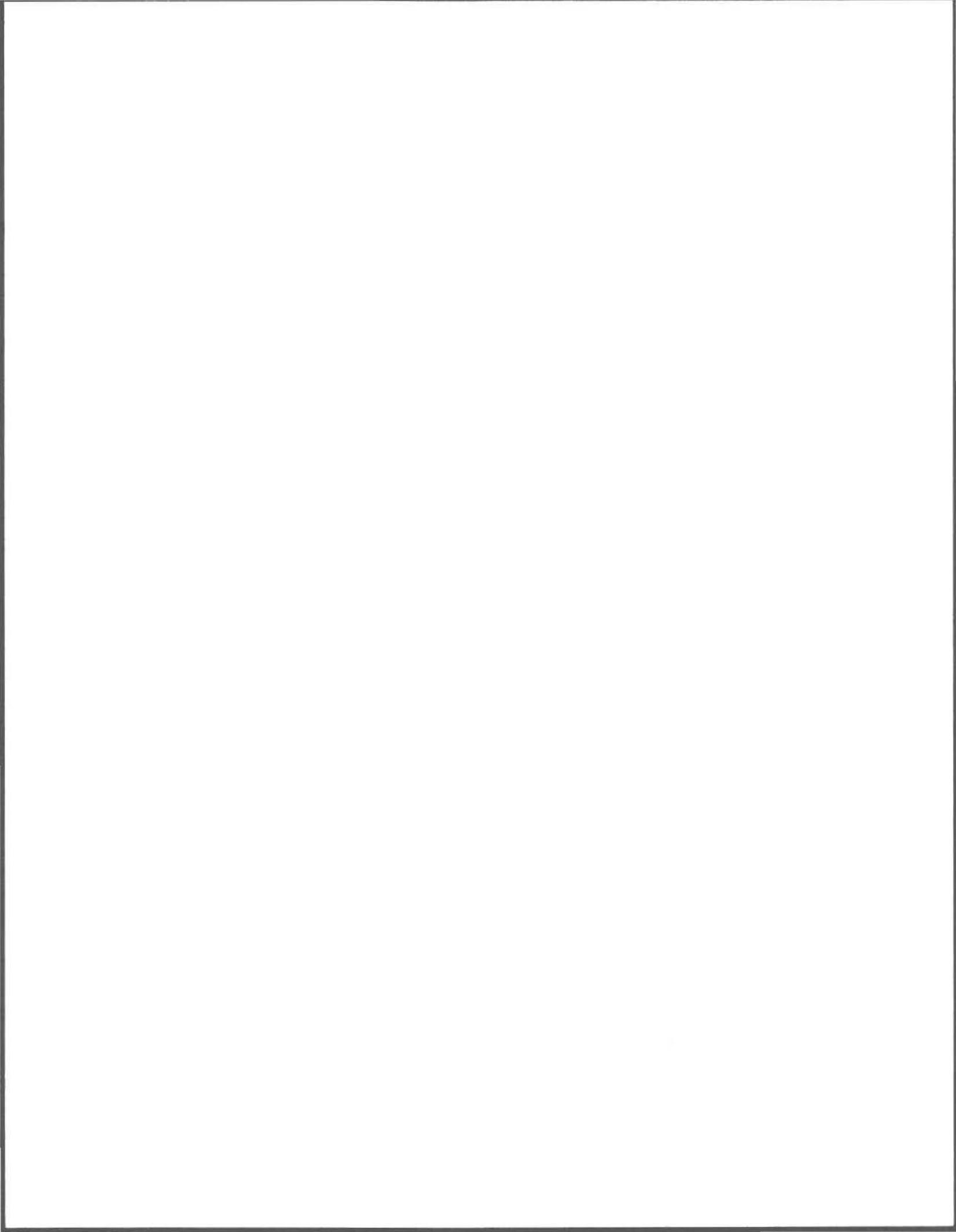
THÈSE
PRÉSENTÉE
COMME EXIGENCE PARTIELLE
DU DOCTORAT EN SCIENCES DE L'EAU

par

CLAUDE LABERGE

UTILISATION DE LA RÉGRESSION LINÉAIRE
POUR LA DÉTECTION DE TENDANCE DANS DES
SÉRIES CHRONOLOGIQUES ENVIRONNEMENTALES.

JUIN 1995



AVANT-PROPOS

L'idée de la présente recherche m'est venue à la suite de quelques années de travail dans le domaine de la détection de tendance dans des séries chronologiques reliées aux sciences de l'eau. Il m'a semblé que plusieurs chercheurs non-statisticiens auraient aimé utiliser la régression linéaire pour détecter des tendances mais ils avaient vaguement entendu dire que: 1) les résidus seraient autocorrélés et ils ne savaient pas quoi faire dans cette situation ; 2) la présence de saisons affectait le test et ils ne savaient pas non plus quoi faire dans cette situation et; 3) les valeurs aberrantes influençaient la régression linéaire mais ils ne savaient pas quoi faire pour réduire cette influence. Devant tant d'obstacles ces utilisateurs potentiels se tournaient généralement vers les tests non paramétriques.

La présente thèse doit être vue comme un document de référence **pratique** pour quiconque désire détecter des tendances à l'aide de la régression linéaire. L'accent mis sur le côté pratique devrait permettre une compréhension rapide pour les non statisticiens. Afin que cette thèse puisse être la plus utile possible, il fallait traiter de la majorité des caractéristiques restrictives (autocorrélation, valeurs aberrantes et saisonnalité) car il est très rare qu'une seule de ces caractéristiques soit présente et on reléguerait de nouveau la régression linéaire aux oubliettes aussitôt que les caractéristiques non traitées seraient présentes.

La volonté de traiter la majorité des caractéristiques restrictives nous oblige cependant à traiter chacune d'elles un peu superficiellement. Il aurait été impossible de résoudre complètement tous les problèmes reliés à toutes les situations associées aux caractéristiques restrictives. Malgré cette impossibilité d'aller dans les moindres détails, le présent travail permet de savoir comment utiliser la régression linéaire pour détecter des

tendances dans la majorité des séries chronologiques environnementales bien échantillonnées. Les problèmes non résolus devront faire l'objet de recherches ultérieures.

Finalement, il est important de noter que le but premier du présent travail est de développer un test de détection de tendance dont on connaît les propriétés, les avantages et les désavantages. Dans cette optique, les adaptations utilisées, pour l'autocorrélation par exemple, peuvent ne pas être optimales selon la théorie statistique (pour tenir compte de l'autocorrélation), mais si elles sont **plus simples d'utilisation** et qu'elles **n'affectent pas les propriétés du test** de détection de tendance, elles représentent quand même ici une solution adéquate.

RÉSUMÉ

Cette recherche s'attaque aux problèmes pratiques de la détection de tendance dans des séries chronologiques relativement bien échantillonnées (longues et équidistantes). Le but premier du présent travail est de fournir un document qui traite de la majorité des caractéristiques restrictives qui peuvent être rencontrées dans une étude de détection de tendance utilisant la régression linéaire. L'autocorrélation, les valeurs aberrantes et les saisonnalités sont traitées alors que la longueur et l'équidistance des séries sont supposées adéquates. On cherche alors des solutions simples et pratiques pour tenir compte des caractéristiques restrictives. Ces solutions sont validées à l'aide de simulations de Monte Carlo et les tests qui en résultent sont comparés aux tests non paramétriques généralement utilisés dans le domaine de l'eau.

L'utilisation de variables avec retard dans la régression a été sélectionnée pour tenir compte de la présence d'autocorrélation. À l'exception de la présence de très forte autocorrélation, les études de simulations ont montré que le seuil de 5% est conservé et que la puissance n'est pas significativement différente de la puissance théorique. On peut donc connaître les propriétés de ce test en présence d'autocorrélation sauf pour les cas de très forte autocorrélation. Lorsque l'ordre du processus autorégressif n'est pas connu, on a montré que l'utilisation d'une surparamétrisation permet d'effectuer le test de détection de tendance sans affecter le seuil et la puissance.

En présence de valeurs aberrantes seulement (pas d'autocorrélation), l'utilisation des M-Estimeurs permet une détection de tendance adéquate, puisque le seuil théorique est conservé et la puissance, bien que significativement plus faible que la puissance théorique, demeure stable pour des amplitudes et des emplacements différents des valeurs aberrantes. Les M-Estimeurs permettent donc un test beaucoup plus fiable que le test utilisant les estimateurs classiques.

En présence de IO (et d'autocorrélation), l'utilisation des M-Estimeurs permet également une détection de tendance adéquate: le seuil théorique est conservé et la puissance, encore significativement plus faible que la puissance théorique, demeure stable pour des amplitudes et des emplacements différents des valeurs aberrantes. L'introduction d'autocorrélation n'affecte donc pas la détection de tendance pourvu qu'on utilise une variable avec retard dans le modèle de régression multiple et que l'effet des valeurs aberrantes est propagé par la structure autorégressive (IO). Les M-Estimeurs permettent donc un test beaucoup plus fiable que le test utilisant les estimateurs classiques.

En présence de AO, les M-Estimeurs ne permettent pas une détection de tendance adéquate (le seuil est significativement plus grand que 5%). Il est nécessaire de trouver une

nouvelle méthode permettant d'annuler le biais important de l'estimateur du coefficient d'autorégression si l'on veut détecter adéquatement des tendances en présence de ce type de valeurs aberrantes.

En présence de saisonnalités seulement, l'utilisation des variables muettes permet d'obtenir un test de détection de tendance avec des propriétés semblables au cas théorique (régression linéaire simple sur des données $N(0,1)$ indépendantes à l'exception de la tendance). La principale difficulté de la méthode utilisée réside dans la modélisation des saisons. En effet, on a vu que la présence d'un décalage entre les saisons réelles et les saisons estimées peut amener le test de détection de tendance à présenter des propriétés indésirables.

En présence de saisonnalités et d'autocorrélation, l'utilisation des variables avec retard et muettes permet également d'obtenir un test de détection de tendance avec des propriétés semblables au cas théorique. Cependant, la présence d'un décalage entre les saisons réelles et les saisons estimées aura un effet encore plus important sur le test de détection de tendance. En présence de saisonnalités et d'autocorrélation, il faut donc être certain de la modélisation des saisons pour se fier au résultat du test de détection de tendance utilisant la régression linéaire.

Les comparaisons des adaptations de la régression linéaire avec les tests non paramétriques généralement utilisés dans le domaine de l'eau ont permis les conclusions suivantes:

- En présence de **IO de moyenne non nulle et d'autocorrélation**, le test de Spearman/Lettenmaier est fortement influencé par l'emplacement des IO mais pas par l'amplitude de ces dernières. Le test utilisant les estimateurs robustes est beaucoup moins influencé par l'emplacement des valeurs aberrantes tout en étant également peu influencé par l'amplitude des IO. Le test utilisant les estimateurs robustes est donc préférable si on ne connaît pas l'emplacement des valeurs aberrantes.
- En présence de **IO de moyenne nulle et d'autocorrélation**, le test de Spearman/Lettenmaier est préférable car sa puissance est plus grande que celle du test utilisant les estimateurs robustes et qu'il est peu influencé (pour la moyenne des simulations) par l'emplacement des valeurs aberrantes.
- En présence de **AO de moyenne nulle et d'autocorrélation**, l'utilisation de la régression autorégressive robuste permet un test adéquat. Le test de Spearman/Lettenmaier, utilisé dans le domaine de l'eau, présente des faiblesses évidentes à cause de l'influence de l'emplacement des valeurs aberrantes et du biais d'estimateurs non robustes pour le coefficient d'autorégression. Ce dernier test ne devrait donc pas être utilisé à moins d'obtenir une bonne estimation du coefficient d'autorégression et une bonne connaissance de l'emplacement des valeurs aberrantes.
- En présence de **AO de moyenne non nulle et d'autocorrélation**, aucun des tests étudiés ne présente des propriétés désirables, il est donc très hasardeux d'effectuer une détection de tendance dans cette situation. Il est évident que dans

cette situation on a besoin de recherches complémentaires pour effectuer une détection de tendance adéquate.

- En présence de **saisonnalités seulement**, le test de régression utilisant les variables muettes présente des propriétés équivalentes à celles du test de Kendall saisonnier. Le test de Kendall sur les moyennes saisonnières est moins puissant que les deux autres dans cette situation. Les seuils des tests de régression saisonnière et de Kendall saisonnier sont affectés par la présence de décalage systématique entre les saisons estimées et réelles. Les puissances des tests de régression et de Kendall saisonnier diminuent significativement avec l'introduction de décalage systématique entre les saisons estimées et réelles. Le test de Kendall sur les moyennes saisonnières n'est pas affecté par l'introduction de décalage systématique entre les saisons estimées et réelles.
- En présence de **saisonnalités et d'autocorrélation**, le test de régression autorégressive saisonnière est plus puissant que le test de Hirsch et Slack (sur les moyennes saisonnières). L'introduction d'un décalage systématique entre les saisons estimées et réelles influence cependant très fortement le test utilisant la régression alors que le test de Hirsch et Slack sur les moyennes saisonnières reste stable devant un tel décalage.

De façon générale, les comparaisons entre les tests paramétriques et non paramétriques montrent l'intérêt d'utiliser les deux approches dans des études de détection de tendance. La connaissance des forces et des faiblesses des deux approches permet alors des conclusions beaucoup plus éclairées. Les avantages de cette utilisation conjointe des méthodes paramétriques et non paramétriques sont également illustrés dans les applications sur des données réelles dans le dernier chapitre.

Dans ces applications, on a montré que:

- Les concentrations de sulfates et le pH ont augmenté de façon significative entre 1981 et 1987, au centre du lac Laflamme et à son exutoire.
- Le paramètre de minéralisation (Ca+Mg+Na) a diminué significativement, en concentration et en charge dans les précipitations, entre 1981 et 1987.
- Les charges de nitrates à l'exutoire du lac Laflamme ont diminué significativement entre 1981 et 1987.

Les conclusions précédentes sont obtenues avec les tests paramétriques et non paramétriques. Parmi les 18 séries traitées, on a cependant deux séries qui amènent une conclusion différente selon qu'on utilise les tests paramétriques ou les tests non paramétriques. En effet, pour les concentrations de nitrates au centre du lac Laflamme et à son exutoire, les tests paramétriques ne détectent pas de tendance alors que les tests non paramétriques concluent à une diminution significative entre 1981 et 1987.

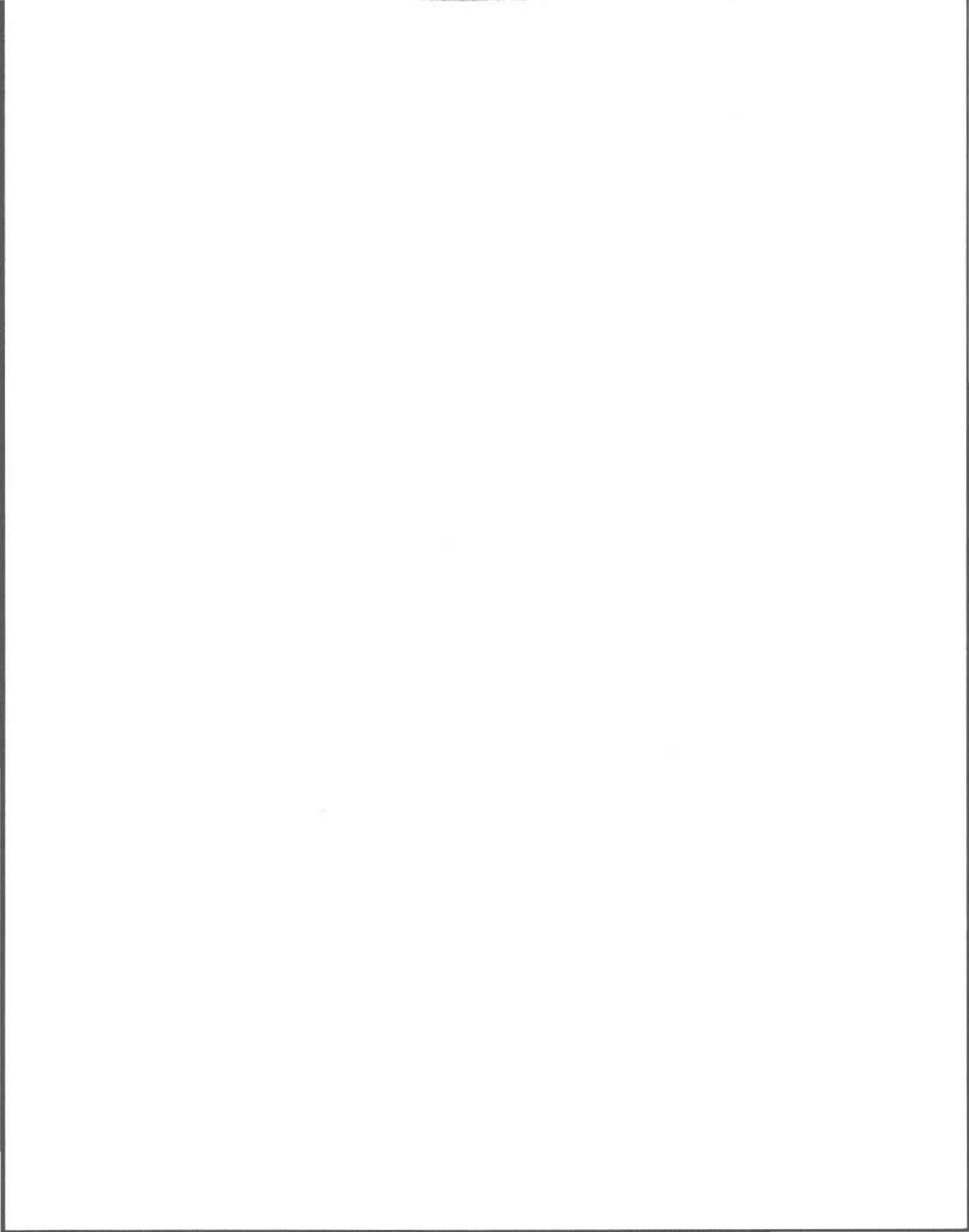


TABLE DES MATIÈRES

	Page
AVANT-PROPOS	iii
RÉSUMÉ	v
TABLE DES MATIÈRES	ix
LISTE DES TABLEAUX	xi
LISTE DES FIGURES.	xv
INTRODUCTION	1
CHAPITRE 1. MÉTHODES ET NOTATIONS UTILISÉES	5
1.1 Détection de tendance à l'aide de la régression	5
1.1.1 Variables indépendantes et types de tendance	6
1.1.2 Test de détection de tendance	7
1.1.3 Caractéristiques restrictives	10
1.1.4 Utilisation de la régression multiple.	13
1.2 Simulations.	13
1.3 Notations et définitions	20
CHAPITRE 2 : PRÉSENCE D'AUTOCORRÉLATION	23
2.1 Effets de l'autocorrélation et traitements possibles	24
2.2 Détection de tendance en présence d'autocorrélation	28
2.2.1 Étude de seuil lorsque p est connu	29
2.2.2 Étude de puissance lorsque p est connu.	31
2.2.3 Seuil et puissance lorsque p est inconnu	34
2.3 Type de tendance et présence d'autocorrélation	37
2.3.1 Détection de tendances par saut	38
2.3.2 Détection de tendances de type inconnu.	39
2.4 Discussion sur la présence d'autocorrélation.	41
CHAPITRE 3. PRÉSENCE DE VALEURS ABERRANTES	43
3.1 Méthodes robustes de Huber.	44
3.1.1 Présentation générale	45
3.1.2 Régression multiple robuste	46
3.2 Modèles de valeurs aberrantes	49
3.2.1 Valeurs aberrantes statiques (AO)	49
3.2.2 Valeurs aberrantes dynamiques (IO)	51
3.2.3 Emplacement des valeurs aberrantes	53
3.3 Détection de tendance en présence de valeurs aberrantes (VA).	54
3.3.1 Effet des VA sur les estimateurs classiques	54
3.3.2 Étude de seuil avec M-Estimeurs	57
3.3.3 Étude de puissance avec M-Estimeurs	59
3.4 Détection de tendance en présence de IO	62
3.4.1 Effet des IO sur les estimateurs classiques	63
3.4.2 Étude de seuil avec M-Estimeurs	65
3.4.3 Étude de puissance avec M-Estimeurs	67

3.5	Détection de tendance en présence de AO	69
3.5.1	Effet des AO sur les estimateurs classiques	70
3.5.2	Étude de seuil avec M-Estimateurs	72
3.5.3	Étude de puissance avec M-Estimateurs	74
3.6	Estimation robuste en absence de valeurs aberrantes	76
3.7	Discussion sur la présence de valeurs aberrantes	78
CHAPITRE 4. PRÉSENCE DE SAISONNALITÉS		81
4.1	Effet des saisonnalités sur les estimateurs classiques	81
4.2	Modélisation avec variables dichotomiques.	83
4.3	Détection de tendance en présence de saisonnalités.	85
4.3.1	Étude de seuil; saisons connues	85
4.3.2	Étude de puissance; saisons connues	87
4.3.3	Étude de seuil; présence de décalage	88
4.3.4	Étude de puissance; présence de décalage	89
4.3.5	Étude complémentaire en présence de décalage.	90
4.3.6	Étude complémentaire en absence de décalage	94
4.4	Détection de tendance en présence de saisonnalités et d'autocorrélation.	96
4.4.1	Étude de seuil et puissance; sans décalage	96
4.4.2	Étude de seuil et puissance; avec décalage	99
4.5	Utilisation des variables muettes en absence de saisonnalités	101
4.6	Discussion sur la présence de saisonnalités	104
CHAPITRE 5. COMPARAISON AVEC LES TESTS NON PARAMÉTRIQUES.		107
5.1	Test de Spearman/Lettenmaier	108
5.1.1	Comparaison des seuils en présence de IO	110
5.1.2	Comparaison des puissances en présence de IO.	112
5.1.3	Comparaison des seuils en présence de AO	115
5.1.4	Comparaison des puissances en présence de AO.	117
5.2	Tests non paramétriques saisonniers.	118
5.2.1	Comparaison des seuils en présence de décalage	121
5.2.2	Comparaison des puissances en présence de décalage.	123
5.3	Test de Hirsch et Slack.	126
5.3.1	Comparaison des seuils	128
5.3.2	Comparaison des puissances.	130
5.4	Synthèse des comparaisons	133
CHAPITRE 6. APPLICATION A DES SERIES ENVIRONNEMENTALES		137
6.1	Choix des paramètres à étudier	139
6.2	Étude des sulfates	141
6.3	Étude des nitrates	155
6.4	Étude du pH.	170
6.5	Étude du paramètre de minéralisation	176
6.6	Discussion sur les applications	190
CONCLUSION		195
BIBLIOGRAPHIE		199
APPENDICE A : Sous-routine FORTRAN pour le calcul des M-Estimateurs de Huber		203
APPENDICE B : Banque des données utilisées du lac Laflamme		207

LISTE DES TABLEAUX

	Page
Tableau 1.1 : Puissance ($1-\beta$) du test de détection de tendance et amplitudes relatives détectées (fonction de σ) pour différents paramètres de noncentralité (δ).	9
Tableau 1.2 : Générateur de nombre aléatoire $U(0,1)$; Langage FORTRAN.	14
Tableau 1.3 : Écarts types d'un seuil ou d'une puissance empirique selon le nombre de séries simulées (S) et la probabilité de rejet (p).	19
Tableau 2.1 : Résultats numériques de la simulation 1; Seuil de la régression non autorégressive en présence d'autocorrélation.	25
Tableau 2.2 : Résultats numériques de la simulation 2; Seuil de la régression autorégressive en présence d'autocorrélation.	30
Tableau 2.3 : Résultats numériques de la simulation 2b; Seuil de la régression autorégressive en présence d'autocorrélation $\phi=0,25$.	31
Tableau 2.4 : Résultats numériques de la simulation 2c; Seuil de la régression autorégressive en présence d'autocorrélation $\phi=0,95$.	31
Tableau 2.5 : Résultats numériques de la simulation 3; Puissance de la régression autorégressive en présence d'autocorrélation.	32
Tableau 2.6 : Résultats numériques de la simulation 3b; Puissance de la régression autorégressive en présence d'autocorrélation (tendance : $\Delta_m=0,000667$).	33
Tableau 2.7 : Résultats numériques de la simulation 3c; Puissance de la régression autorégressive en présence d'autocorrélation (tendance : $\Delta_m=0,001333$).	34
Tableau 2.8 : Résultats numériques de la simulation 4; Seuil de la régression autorégressive avec surparamétrisation en présence d'autocorrélation.	35
Tableau 2.9 : Résultats numériques de la simulation 5; Puissance de la régression autorégressive avec surparamétrisation en présence d'autocorrélation.	36
Tableau 2.10: Résultats numériques de la simulation 6; Puissance de la régression autorégressive en présence d'une tendance par saut et d'autocorrélation.	39
Tableau 2.11: Résultats numériques de la simulation 7; Puissance de la régression autorégressive avec mauvaise modélisation du type de tendance.	40
Tableau 3.1 : Moyennes et écart types des différents processus en présence de valeurs aberrantes statiques; pour les valeurs de ϕ , μ_{AO} et σ_{AO} utilisées dans les simulations.	51
Tableau 3.2 : Moyennes et écart types des différents processus en présence de valeurs aberrantes dynamiques; pour les valeurs de ϕ , μ_{IO} et σ_{IO} utilisées dans les simulations.	52
Tableau 3.3 : Résultats numériques de la simulation 8; Seuil de la régression non robuste en présence de valeurs aberrantes.	55
Tableau 3.4 : Résultats numériques de la simulation 9; Puissance de la régression non robuste en présence de valeurs aberrantes.	57
Tableau 3.5 : Résultats numériques de la simulation 10; Seuil de la régression robuste en présence de valeurs aberrantes.	58

	Page
Tableau 3.6 : Résultats numériques de la simulation 11; Puissance de la régression robuste en présence de valeurs aberrantes: $\mu_{VA}=5, \sigma_{VA}=2$.	60
Tableau 3.7 : Résultats numériques de la simulation 11b; Puissance de la régression robuste en présence de valeurs aberrantes: $\mu_{VA}=10, \sigma_{VA}=3$.	60
Tableau 3.8 : Résultats numériques de la simulation 11c; Puissance de la régression robuste en présence de valeurs aberrantes: $\mu_{VA}=0, \sigma_{VA}=3$.	60
Tableau 3.9 : Résultats numériques de la simulation 12; Seuil de la régression non robuste en présence de IO.	64
Tableau 3.10: Résultats numériques de la simulation 12; Puissance de la régression non robuste en présence de IO.	65
Tableau 3.11: Résultats numériques de la simulation 13; Seuil de la régression robuste en présence de IO.	66
Tableau 3.12: Résultats numériques de la simulation 14; Puissance de la régression robuste en présence de IO: $\mu_{IO}=5, \sigma_{IO}=2$.	68
Tableau 3.13: Résultats numériques de la simulation 14b; Puissance de la régression robuste en présence de IO : $\mu_{IO}=10, \sigma_{IO}=3$.	68
Tableau 3.14: Résultats numériques de la simulation 14c; Puissance de la régression robuste en présence de IO : $\mu_{IO}=0, \sigma_{IO}=3$.	68
Tableau 3.15: Résultats numériques de la simulation 15; Seuil de la régression non robuste en présence de AO.	71
Tableau 3.16: Résultats numériques de la simulation 15; Puissance de la régression non robuste en présence de AO.	72
Tableau 3.17: Résultats numériques de la simulation 16; Seuil de la régression robuste en présence de AO: $\mu_{AO}=5, \sigma_{AO}=2$.	73
Tableau 3.18: Résultats numériques de la simulation 16b; Seuil de la régression robuste en présence de AO de moyenne nulle: $\mu_{AO}=0, \sigma_{AO}=3$.	74
Tableau 3.19: Résultats numériques de la simulation 17; Puissance de la régression robuste en présence de AO : $\mu_{AO}=0, \sigma_{AO}=3$.	75
Tableau 3.20: Résultats numériques de la simulation 18; Seuil de la régression robuste en absence de valeurs aberrantes.	77
Tableau 3.21: Résultats numériques de la simulation 18; Puissance de la régression robuste en absence de valeurs aberrantes.	77
Tableau 4.1 : Résultats numériques de la simulation 19; Seuil de la régression saisonnière en présence de saisons connues.	86
Tableau 4.2 : Résultats numériques de la simulation 20; Puissance de la régression saisonnière en présence de saisons connues.	87
Tableau 4.3 : Résultats numériques de la simulation 19; Seuil de la régression saisonnière en présence de saisons et d'un décalage.	89
Tableau 4.4 : Résultats numériques de la simulation 20; Puissance de la régression saisonnière en présence de saisons et d'un décalage.	90
Tableau 4.5 : Résultats numériques de la simulation 21; Seuil de la régression saisonnière (12 saisons) en présence de 4 saisons et d'un décalage.	92
Tableau 4.6 : Résultats numériques de la simulation 22; Puissance de la régression saisonnière (12 saisons) en présence de 4 saisons et d'un décalage.	93

	Page
Tableau 4.7 : Résultats numériques de la simulation 21; Seuil de la régression saisonnière (12 saisons) en présence de 4 saisons mais sans décalage.	95
Tableau 4.8 : Résultats numériques de la simulation 22; Puissance de la régression saisonnière (12 saisons) en présence de 4 saisons mais sans décalage.	95
Tableau 4.9 : Résultats numériques de la simulation 23; Seuil de la régression autorégressive saisonnière en présence d'autocorrélation et de saisons.	97
Tableau 4.10: Résultats numériques de la simulation 24; Puissance de la régression autorégressive saisonnière en présence d'autocorrélation et de saisons.	98
Tableau 4.11: Résultats de la simulation 23; Seuil de la régression autorégressive saisonnière en présence d'autocorrélation, de saisons et de décalage.	99
Tableau 4.12: Résultats de la simulation 24; Puissance de la régression autorégressive saisonnière en présence d'autocorrélation, de saisons et de décalage.	100
Tableau 4.13: Résultats de la simulation 25; Seuil de la régression saisonnière en absence de saisons.	103
Tableau 4.14: Résultats de la simulation 26; Puissance de la régression saisonnière en absence de saisons.	103
Tableau 4.15: Effets sur les séries de résidus d'une mauvaise estimation des saisons en présence d'autocorrélation.	105
Tableau 5.1 : Comparaison des seuils des différents tests de détection de tendance en présence de IO et d'autocorrélation.	111
Tableau 5.2 : Comparaison des puissances des différents tests de détection de tendance en présence de IO et d'autocorrélation.	113
Tableau 5.3 : Comparaison des seuils des différents tests de détection de tendance en présence de AO et d'autocorrélation.	115
Tableau 5.4 : Comparaison des puissances des différents tests de détection de tendance en présence de AO de moyenne nulle et d'autocorrélation.	117
Tableau 5.5 : Comparaison des seuils du test de la régression saisonnière et des tests de Kendall saisonniers.	122
Tableau 5.6 : Comparaison des puissances de la régression saisonnière et des tests de Kendall saisonniers .	124
Tableau 5.7 : Comparaison des seuils de la régression autorégressive saisonnière et du test de Hirsch et Slack .	129
Tableau 6.1 : Résultats des tests de détection de tendance pour la série de concentrations de sulfates dans les précipitations.	143
Tableau 6.2 : Résultats des tests de détection de tendance pour la série de charges de sulfates dans les précipitations.	146
Tableau 6.3 : Résultats des tests de détection de tendance pour la série de concentrations de sulfates au centre du lac.	147
Tableau 6.4 : Résultats des tests de détection de tendance pour la série de concentrations de sulfates à l'exutoire.	150
Tableau 6.5 : Résultats des tests de détection de tendance pour la série de charges de sulfates à l'exutoire.	154
Tableau 6.6 : Résultats des tests de détection de tendance pour la série de concentrations de nitrates dans les précipitations.	156
Tableau 6.7 : Résultats des tests de détection de tendance pour la série de charges de nitrates dans les précipitations.	161
Tableau 6.8 : Résultats des tests de détection de tendance pour la série de concentrations de nitrates au centre du lac.	162

Tableau 6.9 : Résultats des tests de détection de tendance pour la série de concentrations de nitrates à l'exutoire.	165
Tableau 6.10: Résultats des tests de détection de tendance pour la série de charges de nitrates à l'exutoire.	169
Tableau 6.11: Résultats des tests de détection de tendance pour la série de pH dans les précipitations.	172
Tableau 6.12: Résultats des tests de détection de tendance pour la série de pH dans le lac.	174
Tableau 6.13: Résultats des tests de détection de tendance pour la série de pH à l'exutoire.	176
Tableau 6.14: Résultats des tests de détection de tendance pour la série de concentrations du paramètre de minéralisation dans les précipitations.	177
Tableau 6.15: Résultats des tests de détection de tendance pour la série de charges du paramètre de minéralisation dans les précipitations.	180
Tableau 6.16: Résultats des tests de détection de tendance pour la série de concentrations du paramètre de minéralisation au centre du lac.	183
Tableau 6.17: Résultats des tests de détection de tendance pour la série de concentrations du paramètre de minéralisation à l'exutoire.	187
Tableau 6.18: Résultats des tests de détection de tendance pour la série de charges du paramètre de minéralisation à l'exutoire.	188
Tableau 6.19: Résumé de l'étude de détection de tendance utilisant la régression linéaire.	191
Tableau 6.20: Résumé de l'étude de détection de tendance utilisant les tests non paramétriques.	192

LISTE DES FIGURES

	Page
Figure 6.1 : Série des concentrations de sulfates dans les précipitations à la station Montmorency.	142
Figure 6.2 : Résidus standardisés de la régression autorégressive classique; série des concentrations de sulfates dans les précipitations.	142
Figure 6.3 : Série des charges de sulfates dans les précipitations à la station Montmorency.	145
Figure 6.4 : Résidus standardisés de la régression saisonnière classique; série des charges de sulfates dans les précipitations.	145
Figure 6.5 : Série des concentrations de sulfates au centre du lac Laflamme.	148
Figure 6.6 : Résidus standardisés de la régression autorégressive classique; série des concentrations de sulfates au centre du lac.	148
Figure 6.7 : Série des concentrations de sulfates à l'exutoire du lac Laflamme.	151
Figure 6.8 : Résidus standardisés de la régression autorégressive classique; série des concentrations de sulfates à l'exutoire.	151
Figure 6.9 : Série des charges de sulfates à l'exutoire du lac Laflamme.	153
Figure 6.10: Résidus standardisés de la régression saisonnière autorégressive classique; série des charges de sulfates à l'exutoire.	153
Figure 6.11: Série des concentrations de nitrates dans les précipitations à la station Montmorency.	158
Figure 6.12: Résidus standardisés de la régression classique; série des concentrations de nitrates dans les précipitations.	158
Figure 6.13: Série des charges de nitrates dans les précipitations à la station Montmorency.	160
Figure 6.14: Résidus standardisés de la régression classique; série des charges de nitrates dans les précipitations.	160
Figure 6.15: Série des concentrations de nitrates au centre du lac Laflamme.	164
Figure 6.16: Résidus standardisés de la régression saisonnière autorégressive classique; série des concentrations de nitrates au centre du lac.	164
Figure 6.17: Série des concentrations de nitrates à l'exutoire du lac Laflamme.	166
Figure 6.18: Résidus standardisés de la régression saisonnière autorégressive classique; série des concentrations de nitrates à l'exutoire.	166
Figure 6.19: Série des charges de nitrates à l'exutoire du lac Laflamme.	168
Figure 6.20: Résidus standardisés de la régression saisonnière autorégressive classique; série des charges de nitrates à l'exutoire.	168
Figure 6.21: Série des pH dans les précipitations à la station Montmorency.	171
Figure 6.22: Résidus standardisés de la régression autorégressive classique; série des pH dans les précipitations.	171
Figure 6.23: Série des pH au centre du lac Laflamme.	173
Figure 6.24: Résidus standardisés de la régression autorégressive classique; série des pH au centre du lac.	173
Figure 6.25: Série des pH à l'exutoire du lac Laflamme.	175

	Page
Figure 6.26: Résidus standardisés de la régression autorégressive classique; série des pH à l'exutoire.	175
Figure 6.27: Série des concentrations du paramètre de minéralisation dans les précipitations à la station Montmorency.	178
Figure 6.28: Résidus standardisés de la régression classique; série des concentrations du paramètre de minéralisation dans les précipitations.	178
Figure 6.29: Série des charges du paramètre de minéralisation dans les précipitations à la station Montmorency.	181
Figure 6.30: Résidus standardisés de la régression classique; série des charges du paramètre de minéralisation dans les précipitations.	181
Figure 6.31: Série des concentrations du paramètre de minéralisation au centre du lac Laflamme.	184
Figure 6.32: Résidus standardisés de la régression saisonnière autorégressive classique; série des concentrations du paramètre de minéralisation au centre du lac.	184
Figure 6.33: Série des concentrations du paramètre de minéralisation à l'exutoire du lac Laflamme.	186
Figure 6.34: Résidus standardisés de la régression saisonnière autorégressive classique; série des concentrations du paramètre de minéralisation à l'exutoire.	186
Figure 6.35: Série des charges du paramètre de minéralisation à l'exutoire du lac Laflamme.	189
Figure 6.36: Résidus standardisés de la régression saisonnière autorégressive classique; série des charges du paramètre de minéralisation à l'exutoire.	189

INTRODUCTION

Depuis une vingtaine d'années, plusieurs chercheurs se sont penchés sur le problème de la détection de tendance dans les séries de qualité de l'eau. Certaines caractéristiques restrictives présentes dans ces séries (autocorrélation, saisonnalités, mauvais échantillonnage, valeurs aberrantes, valeurs tronquées, etc.) ont cependant rendu les tests classiques (test t, test de KENDALL, test de Spearman, etc.) inutilisables dans la majorité des applications. Devant ces restrictions, des adaptations de tests non paramétriques ont été développées afin de permettre la détection de tendance dans les séries chronologiques de qualité de l'eau. Lettenmaier (1976) a étudié l'utilisation du test de Spearman en présence d'autocorrélation, Hirsch, Slack et Smith (1982) ont construit un test saisonnier utilisant le test de Mann-Kendall (Mann, 1945; Kendall, 1975), alors que Hirsch et Slack (1984) ont également adapté le test de Mann-Kendall pour les cas où des saisonnalités et de l'autocorrélation sont présentes. Ces tests permettent de couvrir adéquatement presque toutes les situations pouvant être rencontrées pour la détection de tendance dans les séries chronologiques du domaine de l'eau. Pour cette raison, ils sont devenus très populaires dans le domaine de l'environnement et, en particulier, dans le domaine de l'eau.

Cependant, ces tests non paramétriques ne sont pas une panacée. Parmi les limites de leur utilisation, notons :

1. L'absence de valeurs critiques théoriques pour le test de Spearman/Lettenmaier. Les valeurs critiques sont actuellement obtenues par simulation.
2. Ces tests sont adaptés à la présence d'une tendance linéaire, on a donc une perte de puissance lorsqu'ils sont utilisés en présence d'une tendance par saut. L'adaptation

du test de Mann-Whitney pourrait pallier à ce problème, mais la validation d'une telle adaptation reste à faire.

3. Ces tests, bien connus dans le domaine de l'eau, le sont beaucoup moins dans d'autres domaines. Ils sont donc rarement disponibles dans les logiciels statistiques classiques. Cluis, Laberge et Houle (1987) ont développé un logiciel (DETECT) utilisant ces tests, mais les possibilités de ce logiciel sont limitées et sa distribution se fait surtout de bouche à oreille.
4. Ces tests ne permettent pas une estimation du niveau de la série et de l'amplitude de la tendance. Certains estimateurs non paramétriques peuvent combler ce manque, mais leur disponibilité sur support informatique est très limitée.
5. Les différents tests ayant été développés séparément, il y a peu de points communs entre eux et l'exécution d'une analyse complète (plusieurs séries de structures différentes) de détection de tendance devient alors ardue.

Les tests non paramétriques demeurent cependant la seule porte de sortie, lorsque les ensembles de données contiennent la majorité des caractéristiques restrictives, en particulier quand on est en présence d'un mauvais échantillonnage (faible nombre d'échantillons et observations non équidistantes).

Le besoin de plus en plus important, pour les décideurs publics d'effectuer un suivi environnemental adéquat, a conduit à meilleure structuration des banques de données. Des plans d'échantillonnage mieux structurés génèrent maintenant des séries chronologiques beaucoup mieux échantillonnées (longues et équidistantes). Evidemment, beaucoup d'autres caractéristiques (autocorrélation, saisonnalités, valeurs aberrantes, etc.) demeurent restreignantes dans l'utilisation de tests de détection de tendance. Cependant, l'existence de séries de qualité de l'eau bien échantillonnées semble maintenant permettre l'utilisation de méthodes paramétriques. L'avantage important des tests paramétriques sur les tests non paramétriques dans les études environnementales de détection de tendance est l'estimation directe de paramètres tels l'amplitude de la tendance et le niveau de la série. Ces estimations permettent de porter un jugement complémentaire au simple résultat d'absence ou de présence de tendance. En effet, en plus de savoir qu'une tendance est présente, il est important de savoir à quel niveau se situe la série, ce résultat peut ensuite être comparé à

une norme d'usage ou un critère de qualité. De plus, en présence d'une tendance monotone, l'estimation de la pente de la tendance permet de prévoir les niveaux futurs des séries de qualité en supposant que cette tendance se poursuive.

Le présent travail cherche à mettre sur pied une procédure simple et efficace pour la détection de tendance à l'aide d'une méthode paramétrique: la régression linéaire. Le choix de la régression linéaire repose sur les avantages suivants:

1. facilement adaptable à la détection de tendance;
2. possibilités d'adaptation pour tenir compte de l'autocorrélation, des saisonnalités et des valeurs aberrantes;
3. grande disponibilité sur supports informatiques et;
4. facilité d'utilisation pour le non-statisticien.

La procédure qui sera développée suppose cependant que les séries chronologiques sont bien échantillonnées (séries assez longues et équidistantes). Des adaptations seront développées pour tenir compte de la présence des caractéristiques restrictives suivantes: autocorrélation, valeurs aberrantes et saisonnalité. Ces adaptations seront validées (seuil et puissance) et comparées aux méthodes non paramétriques usuelles à l'aide de simulations de Monte-Carlo.

Le but principal du présent travail est donc de regrouper les solutions des problèmes d'utilisation de la régression linéaire pour la détection de tendance dans des séries chronologiques environnementales. Afin de permettre une compréhension rapide aux chercheurs de différentes disciplines reliées au domaine de l'eau (biologistes, chimistes, géographes, etc) l'utilisation de développements statistiques théoriques est réduite au maximum. Bien que ce travail vise en premier lieu le traitement de séries chronologiques provenant du domaine de l'eau, les développements peuvent être repris pour toute série chronologique bien échantillonnée dans laquelle on désire détecter une tendance (économie, environnement, etc).

Ce travail se divise en 6 chapitres. Le premier chapitre présente l'utilisation de la régression linéaire pour la détection de tendance, les développements associés aux simulations et les notations qui seront utilisées tout au long du travail. Le chapitre 2 présente les problèmes introduits par la présence d'autocorrélation et les solutions permettant une détection de tendance en présence de cette caractéristique restrictive. Les chapitres 3 et 4 traitent respectivement les problèmes et les solutions associés à la présence de valeurs aberrantes et de saisonnalités. Dans ces trois derniers chapitres, les simulations permettent de valider le seuil et la puissance du test de détection de tendance en présence de ces caractéristiques restrictives.

Le chapitre 5 présente une comparaison entre les adaptations des chapitres 2, 3 et 4 et les tests non paramétriques de détection de tendance (Spearman/Lettenmaier, Mann-Kendall Saisonnier et Hirsch et Slack) généralement utilisés dans le domaine de l'eau. On s'attarde alors au seuil et à la puissance des différents tests. Cette comparaison permet de faire ressortir les avantages et les désavantages des deux approches. Une synthèse détaillée des comparaisons est présentée à la section 5.4.

Finalement, le chapitre 6 permet de mettre en pratique les différentes adaptations à l'aide d'applications sur des séries réelles. Les séries étudiées consistent en des mesures hebdomadaires de concentrations et de charges au Lac Laflamme, à son exutoire ainsi que dans les précipitations sur ce lac. Les séries sont d'une durée de 6 ans débutant en décembre 1981 pour se terminer en décembre 1987. On traite les paramètres chimiques suivants: sulfates, nitrates, pH et paramètre de minéralisation.

CHAPITRE 1

MÉTHODES ET NOTATIONS UTILISÉES

Ce premier chapitre présente les résultats, associés à la régression linéaire, nécessaires aux développements des chapitres suivants, la façon de procéder pour les simulations et différentes notations et définitions permettant une meilleure compréhension des développements des chapitres suivants.

1.1 Détection de tendance à l'aide de la régression.

La détection de tendance temporelle à l'aide de la régression linéaire est en théorie assez simple: l'utilisation d'une variable explicative associée au temps et au type de tendance que l'on veut détecter, ainsi que le modèle (1.1) permettent de détecter des tendances dans une série chronologique : $Y_t, t = 1, \dots, T$.

$$Y_t = \beta_0 + \beta_1 X_t + e_t \quad (1.1)$$

Les différentes composantes de l'équation (1.1) sont:

Y_t : la série chronologique mesurée. Le but des études de détection de tendance est de chercher la présence ou non d'une tendance à l'aide de cette série.

β_0 : l'ordonnée à l'origine dans le modèle de régression linéaire. L'étude de ce paramètre est généralement de peu d'intérêt pour les études de détection de tendance sauf pour estimer le niveau initial de la série.

β_1 : l'amplitude (ou la pente) de la tendance. L'étude de ce paramètre constitue la pierre angulaire de la détection de tendance à l'aide de la régression linéaire.

X_t : une variable non stochastique qui identifie le type de tendance que l'on cherche à détecter.

e_t : la série de résidus.

1.1.1 Variables indépendantes et types de tendance.

Dans le présent travail, deux variables non stochastiques (X_t) sont utilisées comme variable indépendante pour la détection de tendance à l'aide de la régression. Elles sont associées soit à une tendance par saut, soit à une tendance monotone.

Tendance par saut

Ce type de tendance est caractérisé par un changement instantané du niveau de la série avec des niveaux stationnaires avant et après le saut. La variable X_t la mieux adaptée à ce type de tendance est:

$$X_t = \begin{cases} 0 & \text{si } t \leq t_0 \\ 1 & \text{si } t > t_0 \end{cases} \quad (1.2)$$

où t_0 est la date du saut.

Tendance monotone

Ce type de tendance est caractérisé par un changement graduel du niveau de la série. Une variable X_t bien adaptée à ce type de tendance est:

$$X_t = t, \quad t = 1, \dots, T \quad (1.3)$$

Dans ce travail, l'hypothèse d'équidistance des observations implique que l'intervalle d'échantillonnage est constant dans le temps. La variable explicative définie en (1.3) permet donc d'obtenir la pente de la tendance associée à l'intervalle d'échantillonnage. Une fois la variable X_t choisie, la détection de tendance se fait à l'aide des tests classiques associés à la régression linéaire simple.

1.1.2 Test de détection de tendance.

La détection de tendance à l'aide de la régression linéaire se fait en testant la signification du paramètre β_1 défini à l'équation (1.1). À moins d'avis contraire, tous les tests de détection de tendance discutés dans ce travail sont donc définis par les deux hypothèses suivantes:

$$\begin{aligned} H_0 : \beta_1 = 0 & : \text{pas de tendance} \\ H_1 : \beta_1 \neq 0 & : \text{tendance} \end{aligned}$$

Les développements théoriques pour la régression linéaire simple montrent que pour le test sur β_1 avec un seuil de 5%, on accepte H_0 si :

$$|t^*| = \left| \frac{b_1}{s(b_1)} \right| \leq t_{1-\alpha/2, T-2} \quad (1.4)$$

alors que l'on rejette H_0 si :

$$|t^*| = \left| \frac{b_1}{s(b_1)} \right| > t_{1-\alpha/2, T-2} \quad (1.5)$$

où b_1 est l'estimateur de β_1 obtenu à l'aide de la régression linéaire, $s(b_1)$ est l'écart type de cet estimateur et où $t_{1-\alpha/2, T-2}$ est le $(1 - \alpha/2)$ percentile de la loi de Student (t) avec $T-2$ degrés de liberté. Dans le cas de la régression linéaire simple,

$$s^2(b_1) = \frac{EQM}{\sum_{t=1}^T (X_t - \bar{X})^2} = \frac{\sum_{t=1}^T (\hat{Y}_t - \bar{Y})^2 / (T-2)}{\sum_{t=1}^T (X_t - \bar{X})^2} \quad (1.6)$$

où $\hat{Y}_t = b_0 + b_1 X_t$. Le lecteur intéressé à la justification théorique de ce test peut consulter Neter et Wasserman (1974, Chapitre 3) ou Bickel et Doksum (1977, Chapitre 7).

Puissance du test de détection de tendance

Comme la détection de tendance à l'aide de la régression linéaire ne suppose qu'une variable explicative particulière, la puissance du test sur β_1 pour le modèle (1.1) correspond à la puissance du test de détection de tendance. Les résultats classiques de la régression linéaire simple montrent que la puissance de ce test est obtenue à l'aide de la loi de Student non centrée avec paramètre de noncentralité δ défini par:

$$\delta^2 = \frac{\Delta^2}{\sigma^2} \sum_{t=1}^T (X_t - \bar{X})^2 \quad (1.7)$$

où Δ est l'amplitude de la tendance, soit la pente pour une tendance monotone ou le changement de niveau pour une tendance par saut.

Pour les variables X_t associées aux tendances monotone et par saut, respectivement équations (1.2) et (1.3), on peut développer le terme $\sum_{t=1}^T (X_t - \bar{X})^2$ pour obtenir:

$$\delta^2 = \frac{\Delta^2}{\sigma^2} \frac{t_0(T-t_0)}{T} \quad (1.8)$$

pour une tendance par saut avec variable X_t définie à l'équation (1.2). Alors que pour une tendance monotone avec $X_t = t$ on obtient :

$$\delta^2 = \frac{\Delta^2}{\sigma^2} \frac{T(T+1)(T-1)}{12} \quad (1.9)$$

Le tableau 1.1 présente les puissances pour différentes valeurs de δ . Ce tableau montre également les amplitudes relatives (fonction de σ) associées à chaque valeur de Δ , ainsi que pour chaque type de tendance. Tous les résultats sont obtenus pour $T=300$, dans le cas des tendances par saut, plusieurs valeurs de t_0 sont présentées. Le choix de $T=300$ correspond à l'ordre de grandeur des séries qui seront traitées dans le chapitre des applications (chapitre 6).

Tableau 1.1
Puissance $(1-\beta)$ du test de détection de tendance et amplitudes relatives détectées (fonction de σ) pour différents paramètres de noncentralité (δ).

δ	$(1-\beta)$	Δ_m	Δ_s $t_0=150$	Δ_s $t_0=100$	Δ_s $t_0=50$	Δ_s $t_0=25$
0,0	5%	$0,0\sigma$	$0,0\sigma$	$0,0\sigma$	$0,0\sigma$	$0,0\sigma$
0,5	8%	$0,0003\bar{\sigma}$	$0,0577\sigma$	$0,0612\sigma$	$0,0775\sigma$	$0,1044\sigma$
1,0	17%	$0,0006\bar{\sigma}$	$0,1155\sigma$	$0,1225\sigma$	$0,1549\sigma$	$0,2089\sigma$
1,5	33%	$0,001\sigma$	$0,1732\sigma$	$0,1837\sigma$	$0,2324\sigma$	$0,3133\sigma$
2,0	50%	$0,0013\bar{\sigma}$	$0,2309\sigma$	$0,2449\sigma$	$0,3098\sigma$	$0,4178\sigma$
2,5	70%	$0,0016\bar{\sigma}$	$0,2887\sigma$	$0,3062\sigma$	$0,3873\sigma$	$0,5222\sigma$
3,0	85%	$0,002\sigma$	$0,3464\sigma$	$0,3674\sigma$	$0,4647\sigma$	$0,6266\sigma$
3,5	93%	$0,0023\bar{\sigma}$	$0,4041\sigma$	$0,4286\sigma$	$0,5422\sigma$	$0,7311\sigma$
4,0	98%	$0,0026\bar{\sigma}$	$0,4619\sigma$	$0,4899\sigma$	$0,6196\sigma$	$0,8355\sigma$
4,5	99,2%	$0,003\sigma$	$0,5196\sigma$	$0,5511\sigma$	$0,6971\sigma$	$0,9400\sigma$
5,0	99,9%	$0,0033\bar{\sigma}$	$0,5774\sigma$	$0,6124\sigma$	$0,7745\sigma$	$1,0444\sigma$

Les puissances $(1-\beta)$ sont obtenues à l'aide du tableau A-5 présenté par Neter et Wasserman (1975, p 814). Comme les développements du présent travail supposent des séries longues ($T>300$), seule la courbe de puissance avec degré de liberté infini est nécessaire; les autres courbes de puissances présentées par Neter et Wasserman sont

donc sans intérêt pour le présent travail. Il faut noter que les puissances asymptotiques théoriques, à un certain Δ_0 , peuvent également être obtenues en utilisant :

$$p(\Delta_0) = P(|N(0, 1) + \delta| > 2) = 1 - \Phi(2 - \delta) + \Phi(-2 - \delta)$$

où Φ représente la distribution normale standardisée et δ est tel que défini aux équations 1.7, 1.8 ou 1.9.

On peut utiliser la deuxième colonne du tableau 1.1 pour trouver la puissance $(1 - \beta)$ d'un test (en fonction de δ) pour toutes les valeurs de T suffisamment grandes (avec $T > 50$, la précision sera suffisante pour la grande majorité des applications).

Il faut cependant noter que, pour des valeurs de T différentes de 300, les amplitudes relatives (colonnes 3 à 7) changent. Les colonnes associées aux amplitudes relatives doivent donc être utilisées seulement dans le cas où $T = 300$. Pour des valeurs près de 300, la puissance ne varie cependant pas beaucoup. Par exemple, pour $T=285$, on détectera, 85 fois sur 100, une tendance monotone de $0,00216\sigma$ (comparativement à $0,002\sigma$ pour $T = 300$) et on détectera, 79 fois sur 100, (vs 85% pour $T = 300$) une tendance de $0,002\sigma$. De même si l'on a $T = 315$, on détectera, 85 fois sur 100, une tendance monotone de $0,00185\sigma$ et 89 fois sur 100 une tendance monotone de $0,002\sigma$.

Les puissances présentées au tableau 1.1 seront appelées puissances théoriques. Ce sera donc à ces puissances que l'on référera dans les prochains chapitres lorsqu'il sera question d'une comparaison avec la puissance théorique. Il faut cependant noter que les puissances $(1 - \beta)$ présentées au tableau 1.1 ne sont pas des valeurs exactes. La lecture de ces puissances au tableau A-5 de Neter et Wasserman peut amener certaines erreurs (maximum 2%), en particulier pour les puissances entre 20% et 80%. Cette légère imprécision doit être prise en compte lors de la comparaison des puissances empiriques avec les puissances du tableau 1.1.

1.1.3 Caractéristiques restrictives

Les développements théoriques présentés jusqu'ici pour la détection de tendance reposent sur certaines hypothèses de base. Ainsi, pour obtenir des tests adéquats sur β_1 , les résidus (e_t) doivent être indépendants et identiquement distribués $\mathcal{N}(0, \sigma^2)$. Comme on travaille avec des séries chronologiques, ces dernières hypothèses sont rarement vérifiées. De plus, les séries rencontrées dans le domaine de l'eau présentent d'autres caractéristiques restreignant considérablement l'utilisation des méthodes statistiques classiques.

Les caractéristiques restrictives généralement présentes dans les séries chronologiques du domaine de l'eau seront maintenant discutées. Pour chacune d'elles, une brève description sera suivie de la façon dont elle est traitée dans le présent travail.

Autocorrélation

La présence d'autocorrélation dans les séries chronologiques du domaine de l'eau est associée à la "mémoire" des processus hydrologiques; i.e. une dépendance à court terme des observations. L'autocorrélation amène une violation de l'hypothèse d'indépendance des observations qui se répercute généralement dans une dépendance des résidus. Le traitement en présence d'autocorrélation fera l'objet du chapitre 2. Cette caractéristique restrictive est donc largement étudiée dans ce travail.

Valeurs aberrantes

La présence de valeurs aberrantes amène généralement une violation de l'hypothèse de normalité des résidus. Le traitement en présence de valeurs aberrantes fera l'objet du chapitre 3. Cette caractéristique restrictive est également largement étudiée dans ce travail.

Saisonnalités

La présence de saisonnalités est associée à différents facteurs cycliques: température, utilisation biologique (nitrates), recharge hivernale par écoulement souterrain (calcium), etc. Les saisonnalités amènent une violation de l'hypothèse d'indépendance des observations et de normalité des résidus. Le traitement en présence de saisonnalités fera

l'objet du chapitre 4. Avec les valeurs aberrantes et l'autocorrélation, les saisonnalités constitue donc une caractéristique restrictive largement considérée dans ce travail.

Valeurs tronquées

La présence de valeurs tronquées est associée à des limitations dans les méthodes de mesures: seuil de détection des appareils (amène des valeurs notées "inférieures à", par exemple : <0.001), dilutions pour les coliformes (amène des valeurs notées "supérieures à", par exemple : >5000), etc. Les valeurs tronquées amènent des problèmes dans l'estimation des différents paramètres: niveau de la série, moyenne de la série, pente de la tendance, etc. Le traitement en présence de valeurs tronquées n'est pas discuté dans le présent travail, parce que les séries étudiées ne contiennent pas de telles valeurs et que le traitement des valeurs tronquées constitue un sujet complexe qui dépasse les limites du présent travail. La présence de valeurs tronquées n'en demeure pas moins un problème épineux dont il faut tenir compte dans certaines situations, si l'on désire une détection de tendance adéquate.

Mauvais échantillonnage

La présence de mauvais échantillonnage (séries courtes et non équidistantes) est généralement associée aux limitations budgétaires des organismes ayant le mandat de recueillir l'information dans le domaine de l'eau. Dans le présent travail, on suppose l'absence de cette caractéristique restrictive, c'est-à-dire que les séries sont longues et que les observations sont équidistantes. Bien qu'il soit illusoire de croire que de telles séries forment la majorité des séries du domaine de l'eau, il est quand même de plus en plus possible de trouver des séries bien échantillonnées.

Dans un contexte de détection de tendance à l'aide de la régression linéaire, les prochains chapitres expliquent les problèmes inhérents à l'autocorrélation, les valeurs aberrantes et les saisonnalités et présentent des solutions à ces problèmes, afin de permettre des tests adéquats pour les tendances dans des séries bien échantillonnées.

1.1.4 Utilisation de la régression multiple.

La section 1.1.2 a présenté l'utilisation de la régression linéaire simple pour la détection de tendance. La présence de caractéristiques restrictives (réf. section 1.1.3) oblige l'utilisation de la régression multiple puisque des variables indépendantes supplémentaires doivent être introduites, afin d'exclure l'effet de ces caractéristiques sur le test de détection de tendance. Comme, en pratique, les variables indépendantes ne sont pas toujours parfaitement indépendantes, l'équation (1.6) ne peut plus être utilisée pour obtenir la variance de l'estimateur de β_1 . La variance de cet estimateur est plutôt obtenue à l'aide de la matrice de variance-covariance:

$$s^2(b) = EQM(X'X)^{-1}$$

$s(b_1)$ est alors un élément de la diagonale principale de cette matrice de variance-covariance.

La présence de variables explicatives non indépendantes amène également des complications au niveau de l'interprétation des résultats. En effet, deux variables explicatives indépendantes auront les mêmes estimateurs que l'on utilise deux régressions linéaires simples ou une régression multiple à deux variables (pour un exemple voir Neter et Wasserman 1974 p. 255). L'interprétation en présence de variables explicatives non indépendantes n'est pas aussi simple, puisque l'introduction de nouvelles variables dans le modèle change les estimateurs des paramètres déjà présents dans le modèle. Les implications de la dépendance des variables explicatives sur la détection de tendance seront discutées dans les chapitres où la régression multiple est utilisée.

1.2 Simulations

L'utilisation de simulations de Monte-Carlo permettra de mettre en évidence les caractéristiques des tests de détection de tendance (seuil et puissance) et de certains estimateurs. La section 1.2 discute différents points associés à ces simulations: générateur

de nombre aléatoire, transformation des séries simulées, conditions expérimentales, présentation des résultats et précision des résultats.

Générateur de nombres aléatoires

Le générateur de nombres aléatoires utilisé est présenté au tableau 1.2; écrit en langage FORTRAN 77, il permet de générer des nombres de distribution uniforme entre 0 et 1. Ce générateur est une combinaison de deux MLCG (Multiplicative Linear Congruential Generator) et sa période est de 2.30584×10^{18} (L'Ecuyer 1986). Comme chaque simulation utilise environ 1.5×10^6 nombres pseudo-aléatoires, on peut donc effectuer environ 1.5×10^{12} simulations avant de retomber sur les mêmes nombres pseudo-aléatoires!

Tableau 1.2
Générateur de nombre aléatoire $U(0,1)$; Langage FORTRAN.

```

SUBROUTINE UNIFORM(V)
  INTEGER*4 N1,N2,K,J
  COMMON/VAR/N1,N2
  K=N1/53668
  N1=40014*(N1-K*53668)-K*12211
  IF (N1.LT.0) N1=N1+2147483563
  K=N2/52774
  N2=40692*(N2-K*52774)-K*3791
  IF (N2.LT.0) N2=N2+2147483399
  J=(N1-2147483563)+N2
  IF (J.LT.1) J=J+2147483562
  V=J*4.656613E-10
  RETURN
  END

```

Une série de nombres pseudo-aléatoires $N(0,1)$ indépendants est ensuite obtenue en utilisant la méthode de BOX et MULLER (1958). Pour chaque paire de nombres $U(0,1)$: u_{2t} et u_{2t+1} on crée deux nombres pseudo-aléatoires $N(0,1)$: e_{2t} et e_{2t+1} en utilisant les équations suivantes:

$$e_{2t} = \sqrt{-2 \ln u_{2t}} \cos(2\pi u_{2t+1})$$

$$e_{2t+1} = \sqrt{-2 \ln u_{2t}} \sin(2\pi u_{2t+1})$$

Ces séries de nombres pseudo-aléatoires $N(0,1)$ sont ensuite utilisées pour construire des séries chronologiques avec une ou plusieurs caractéristiques restrictives comparables aux caractéristiques présentes dans les séries chronologiques environnementales.

Transformation en séries chronologiques.

Les séries de nombres $N(0,1)$ sont ensuite transformées afin de contenir des caractéristiques restrictives (autocorrélation, saisonnalité, valeurs aberrantes) avec des valeurs connues pour les paramètres associées à ces caractéristiques. Les séries avec des caractéristiques et des valeurs de paramètre connues permettent : a) d'évaluer les effets des caractéristiques restrictives sur les différents estimateurs ainsi que sur le seuil et la puissance du test de détection de tendance et; b) de valider les adaptations utilisées pour tenir compte des caractéristiques restrictives.

Comme le test de détection de tendance (test sur β_1) constitue la pierre angulaire du présent travail, les études de son seuil et de sa puissance seront les résultats les plus importants à discuter.

Dans une **étude de seuil**, on créera des séries contenant la(les) caractéristique(s) restrictive(s) mais ne contenant pas de tendance. On estimera le seuil en calculant le pourcentage de séries simulées où H_0 est rejeté (détection de tendance en absence de tendance réelle).

Dans une **étude de puissance**, on créera des séries contenant la(les) caractéristique(s) restrictive(s) en plus de contenir une tendance. On estimera la puissance en calculant le pourcentage de séries simulées où H_0 est rejeté (détection de tendance en présence de tendance).

Les résultats des simulations associés aux estimateurs ponctuels seront présentés afin de vérifier leur convergence et permettre d'expliquer un seuil ou une puissance différents des résultats théoriques (section 1.1.2).

Conditions expérimentales

Pour chacune des simulations, on présentera sous forme de tableau les différentes conditions expérimentales. On donne d'abord un titre permettant de situer rapidement le but de la simulation. On présente ensuite les différents estimateurs étudiés.

La longueur des séries et le nombre de séries simulées suivent. En général on simulera 500 séries de 300 valeurs. La longueur (300 valeurs) est choisie pour sa concordance avec les séries qui seront traitées dans le chapitre sur les applications, alors que le choix de 500 séries simulées permet une précision adéquate (discutée à la section 1.2.5) tout en n'exagérant pas inutilement le temps d'exécution des simulations.

Les dernières conditions expérimentales présentées sont : le modèle sous-jacent (série simulée) et le modèle de régression utilisé pour obtenir les estimateurs nécessaires au test de détection de tendance.

Présentation des résultats des simulations.

Pour chaque simulation on présentera des résultats associés aux estimateurs ponctuels ainsi qu'un seuil ou une puissance empirique. Pour les estimateurs ponctuels des paramètres de la régression on présente un tableau de la forme suivante:

Estimateur	Moyenne	Ecart type	Précision
$REQM$	\overline{REQM}	$E.T.(REQM)$	$1.96 \frac{E.T.(REQM)}{\sqrt{S}}$
b_1	$\overline{b_1}$	$E.T.(b_1)$	$1.96 \frac{E.T.(b_1)}{\sqrt{S}}$
d_1	$\overline{d_1}$	$E.T.(d_1)$	$1.96 \frac{E.T.(d_1)}{\sqrt{S}}$
l_1	$\overline{l_1}$	$E.T.(l_1)$	$1.96 \frac{E.T.(l_1)}{\sqrt{S}}$

Sur la première ligne on aura donc la moyenne des REQM pour l'ensemble des S séries simulées, l'écart type empirique des REQM également obtenu à l'aide des S séries simulées et finalement une mesure de précision de la moyenne des REQM qui équivaut à la demi-étendue d'un intervalle de confiance de cette moyenne au niveau de confiance de 95%. Sur les lignes suivantes, on présente les moyennes, écarts types et précisions de la moyenne des paramètres associés aux différentes variables explicatives introduites dans le modèle. Seul l'estimateur associé à la tendance (b_1) sera toujours présent. Les estimateurs des coefficients d'autorégression (d) et des coefficients saisonniers (l) seront présentés lorsque les paramètres correspondants sont dans le modèle de régression utilisé.

Le seul autre résultat présenté dépend de l'objectif de la simulation. Dans le cas d'une étude de seuil, on présentera le seuil empirique qui correspond au nombre de séries simulées pour lesquelles on rejette l'hypothèse nulle ($H_0 : \beta_1 = 0$: Pas de tendance) divisé par le nombre total de séries simulées.

$$\hat{\alpha} = \text{Seuil empirique} = \frac{\# \text{ de rejet de } H_0}{\# \text{ de séries simulées}}$$

Dans le cas d'une étude de puissance on présente la puissance empirique qui est obtenue à l'aide du même calcul. La seule différence vient de l'utilisation d'un modèle sous-jacent avec une tendance (H_0 fausse).

$$(1 - \hat{\beta}) = \text{Puissance empirique} = \frac{\# \text{ de rejet de } H_0}{\# \text{ de séries simulées}}$$

Précision des simulations.

Afin de pouvoir porter un jugement quantitatif sur les résultats des simulations, il est important de connaître la précision des estimateurs obtenus. Dans le cas de seuil et de de

puissance obtenus par simulation, on peut évaluer leur précision à l'aide des développements suivants.

On peut d'abord supposer que, pour chaque série simulée, le résultat du test (accepter ou rejeter H_0) correspond à une épreuve de Bernoulli $B(p)$ où p est la probabilité de rejet de H_0 et $(1-p)$ est la probabilité d'acceptation de H_0 . Dans le cas d'une étude de seuil (pas de tendance présente, i.e. H_0 vraie) on a $\alpha = \text{Seuil} = p$ alors que dans le cas d'une étude de puissance (tendance présente, i.e. H_1 vraie) on a $(1 - \beta) = \text{Puissance} = p$.

La construction des séries simulées amène des séries indépendantes. Cette indépendance des séries implique une indépendance des résultats du test d'une série à l'autre. Finalement, l'indépendance des résultats des tests amène des épreuves de Bernoulli indépendantes, ce qui implique que les résultats de l'ensemble des séries simulées seront associés à une loi binomiale $B(S, p)$ où S est le nombre de séries simulées et où p est toujours la probabilité de rejeter H_0 .

La variance d'un seuil empirique ($\hat{\alpha}$) est alors équivalente à la variance d'une loi binomiale $B(S, \alpha)$:

$$\text{Var}(\hat{\alpha}) = \frac{\alpha(1-\alpha)}{S}$$

où le seuil théorique (α) est connu puisque les séries sont simulées. On utilise $\alpha = 0.05$ pour l'ensemble des simulations.

De même, la variance d'une puissance empirique sera:

$$\text{Var}(1 - \hat{\beta}) = \frac{\beta(1-\beta)}{S}$$

La puissance théorique est connue (tableau 1.1). Dans certains cas, l'introduction de variables dépendantes dans le modèle de régression peut amener un seuil empirique ou une

puissance empirique significativement différents des valeurs théoriques. Dans ces situations, les variances ne peuvent être calculées à l'aide des valeurs théoriques de α et β .

Le tableau 1.3 présente les écarts types d'un seuil ou d'une puissance empirique pour différentes valeurs de S et de p. En multipliant ces écarts types par 1.96 on obtient une mesure de la précision des seuils et des puissances obtenus par simulations de Monte Carlo. En fait, ces mesures de précision représentent la demi-étendue d'intervalles de confiance à 95% pour α et $(1 - \beta)$.

Cette mesure de précision peut servir pour porter un jugement quantitatif sur la correspondance des seuils ou puissances empiriques avec les valeurs théoriques connues. De plus, elle permet de choisir le nombre de chiffres significatifs à utiliser pour présenter les résultats, il serait inutile de présenter un seuil empirique de .05678 lorsque l'écart type de cette valeur empirique est .02. Dans le présent travail, les seuils et puissances empiriques seront donc généralement présentés avec 3 chiffres après le point.

Tableau 1.3
Écarts types d'un seuil ou d'une puissance empirique selon
le nombre de séries simulées (S) et la probabilité de rejet (p).

S	p=0,05	p=0,10	p=0,25	p=0,40	p=0,50
	ou p=0,95	ou p=0,90	ou p=0,75	ou p=0,60	
100	0,022	0,030	0,043	0,490	0,050
250	0,014	0,019	0,027	0,031	0,032
300	0,013	0,017	0,025	0,028	0,029
400	0,011	0,015	0,022	0,024	0,025
500	0,010	0,013	0,019	0,022	0,022
1000	0,007	0,009	0,014	0,015	0,016

En ce qui concerne les autres estimateurs, on peut estimer leur précision à l'aide de l'estimation de leur variance. En effet, comme les résultats des simulations sont indépendants, la variance de la moyenne d'un estimateur b_1 sera:

$$Var(\bar{b}_1) = \frac{Var(b_1)}{S}$$

et la mesure de la précision est alors : $1.96Var(\bar{b}_1)$.

1.3 Notations et définitions

Notations générales

S	: Nombre de séries simulées.
T	: Nombre d'observations dans la série.
Y_t	: Série étudiée (dans laquelle on recherche une tendance).
X_t	: Variable explicative (non stochastique) associée à la tendance.
y_t	: Observation au temps t de la série étudiée.
x_t	: Valeur au temps t de la variable explicative.
σ^2	: Variance des erreurs e_t .
$REQM$: Racine carrée de l'erreur quadratique moyenne (estimateur de σ).
$REQM_i$: $REQM$ de la i_e série simulée.
\overline{REQM}	: Moyenne des $REQM_i$.
$s(REQM)$: Estimateur de l'écart type des $REQM$

Notations associées aux tendances

Δ	: Amplitude (connue) dans le modèle sous-jacent, on distingue Δ_m : tendance monotone, de Δ_s : tendance par saut.
β_1	: Paramètre associé à la tendance dans le modèle de régression.
$b_{1,i}$: Estimateur de β_1 pour la i_e série simulée.
\bar{b}_1	: Moyenne des estimateurs $b_{1,i}$.
$s(b_1)$: Estimateur de l'écart type de l'estimateur de β_1 .

 Notations associées à l'autocorrélation

- Y_{t-k} : Série avec retard de k unités de temps.
 ϕ_k : Paramètre, dans le modèle de régression, associé à la série avec retard k.
 $d_{k,i}$: estimateur de ϕ_k , pour la i_o série simulée.
 \bar{d}_k : Moyenne des estimateurs $d_{k,i}$.
 $s(d_k)$: Estimateur de l'écart type de l'estimateur de ϕ_k .
-

 Notations associées aux saisonnalités

- X_{sz} : Variable indépendante muette associée à la saison "z".
 λ_z : Paramètre, dans le modèle de régression, associé à la saison "z".
 $l_{z,i}$: Estimateur de λ_z , pour la i_o série simulée.
 \bar{l}_z : Moyenne des estimateurs $l_{z,i}$.
 $s(l_z)$: Estimateur de l'écart type de l'estimateur de λ_z .
-

 Notations associées aux estimateurs robustes et aux valeurs aberrantes

- AO : Se rapporte aux "Additive Outliers" de la littérature. En français l'auteur a choisit le terme Valeurs Aberrantes Statiques.
 IO : Se rapporte aux "Innovation Outliers" de la littérature. En français l'auteur a choisit le terme Valeurs Aberrantes Dynamiques.

On parlera donc de μ_{AO} pour la moyenne des valeurs aberrantes statiques, etc.

- M : Se rapporte aux M-Estimeurs. On parlera donc de b_{1M} pour le M-Estimeur de β_1 , etc.
-

Notations associées aux tests

α : Seuil d'un test. Probabilité de rejeter H_0 (pas de tendance) lorsque H_0 est vraie.

$(1 - \beta)$: Puissance d'un test. Probabilité de rejeter H_0 lorsque H_0 est fausse.

$\hat{\alpha}$: Seuil empirique estimé à partir d'une simulation.

$(1 - \hat{\beta})$: Puissance empirique estimée à partir d'une simulation.

Simulation : une étude empirique sur un sujet particulier.

Séries simulées : série de nombres aléatoires construite pour une étude de simulation. Il y a donc plusieurs séries simulées dans une simulation.

CHAPITRE 2

PRÉSENCE D'AUTOCORRÉLATION

L'autocorrélation représente une dépendance à court terme des observations. Dans une série chronologique, on dira que les observations sont autocorrélées lorsque les observations successives sont corrélées. On peut alors parler d'autocorrélation avec retard de 1, 2, ..., k unités de temps. En effet, une autocorrélation avec retard de k unités représente une corrélation entre les observations des séries Y_t et Y_{t+k} (équivalente à la corrélation entre les observations des séries Y_t et Y_{t+k}).

L'amplitude de l'autocorrélation est déterminée par les coefficients d'autocorrélation. Le coefficient d'autocorrélation d'ordre k d'une série chronologique Y_t , $t=1, \dots, T$ est estimé par:

$$r_k = \frac{\sum_{t=1}^{T-k} (y_t - \bar{y})(y_{t+k} - \bar{y})}{\sum_{t=1}^T (y_t - \bar{y})^2}$$

Des tests sur ces coefficients d'autocorrélation permettent de porter un jugement sur la présence d'une autocorrélation significative. Dans l'éventualité d'une autocorrélation significative, les modèles de Box et Jenkins (1976) ont été développés afin de permettre une

caractérisation du type d'autocorrélation. Dans le domaine des sciences de l'eau, les modèles qui nous intéressent tout particulièrement sont les modèles autorégressifs AR(p), qui permettent de bien modéliser la majorité des séries disponibles. Les modèles de moyennes mobiles MA(q), bien qu'important dans d'autres domaines, présentent un intérêt beaucoup moindre dans le domaine de l'eau; ces modèles ne seront pas discutés dans ce travail.

Un modèle autorégressif d'ordre p est défini par:

$$Y_t = \sum_{i=1}^p \phi_i Y_{t-i} + e_t$$

où ϕ_i est le coefficient d'autorégression d'ordre i, Y_{t-i} est la série avec retard de i unités et e_t est une série de bruits blancs (indépendants et identiquement distribués) de moyenne 0 et de variance σ^2 . On note alors que la réalisation au temps t de la série Y_t dépend des réalisations aux temps $t-1, \dots, t-p$. Pour cette raison, les processus autorégressifs sont souvent identifiés comme des processus avec mémoire.

2.1 Effets de l'autocorrélation et traitements possibles.

Certains auteurs dont Wonnacott et Wonnacott (1980) distinguent le traitement en présence de résidus autocorrélés du traitement en présence d'une variable dépendante autocorrélée. Bien qu'élégante en théorie, cette distinction est très difficile à faire en pratique. De plus, Wonnacott et Wonnacott (1980, p247) concluent que pour un échantillon suffisamment grand, les estimateurs obtenus dans les deux situations seront, à toute fin pratique, équivalents. Devant ces résultats et l'hypothèse de ce travail d'utiliser des séries longues, on s'attardera peu à distinguer ces deux situations.

La présence d'autocorrélation amène la violation d'au moins une hypothèse de base de la régression linéaire: soit l'indépendance des résidus ou l'indépendance des observations (y_t). Il faut cependant noter que, dans les deux cas, on retrouvera des résidus autocorrélés

en utilisant une régression non autorégressive (réf. modèle (1.1), p. 5). Illustrons à l'aide d'une simulation, les problèmes introduits par la présence d'autocorrélation lors de l'utilisation de ce modèle.

SIMULATION 1

CONDITIONS EXPERIMENTALES

Titre : Utilisation du modèle (1.1) en présence d'autocorrélation (seuil).

Estimateurs étudiés : b_1 , $REQM$, seuil.

Nombre de séries simulées : 500 X **Longueur des séries :** 300.

Modèle sous-jacent : AR(1) avec $\phi = 0,75$, $\Delta = 0$ et $\sigma^2 = 1,0$.

Modèle de régression utilisé : $Y_t = \beta_0 + \beta_1 X_t + e_t$; $X_t = t$.

Comme aucune tendance n'est présente ($\Delta = 0$), on peut estimer le seuil du test de détection de tendance en calculant le pourcentage de rejet de l'hypothèse H_0 . Voyons d'abord, au tableau 2.1, les résultats associés aux estimateurs des paramètres de localisation et d'échelle.

Tableau 2.1

Résultats numériques pour les estimateurs de la simulation 1;
Seuil de la régression non autorégressive en présence d'autocorrélation.

Estimateur	Moyenne	Ecart type	Précision
$REQM$	1,47	0,11	0,0096
b_1	-0,0002	0,0025	0,0002

$$\hat{\alpha} = \text{Seuil empirique} = 0,46$$

Un coup d'oeil rapide permet de conclure que le $REQM$ est un estimateur biaisé de $\sigma = 1$. On surestime donc significativement la variance des erreurs. Cette surestimation amène une tendance à rejeter moins souvent l'hypothèse nulle d'absence de tendance.

L'étude de l'estimateur b_1 montre qu'il n'est pas biaisé (résultat démontré théoriquement dans KMENTA 1971). Le problème principal vient plutôt de la très grande variabilité de b_1 à cause de l'introduction de l'autocorrélation. En effet, l'autocorrélation amène des séries d'observations successives du même côté de la moyenne, ces séries ont tendance à amener des estimateurs de β_1 souvent plus grand en valeur absolue. La comparaison des différents écarts types permet de comprendre ce qui se passe. La valeur théorique de $s(b_1)$ peut être déterminée à partir de l'équation 1.6 (avec $REQM=1$ et $X_t=t$) : on obtient alors $s(b_1)=0,00067$. L'estimateur de $s(b_1)$ obtenu à l'aide des simulations est 0,00099 qui correspond à $1,47 \times 0,00067$; le biais étant donc introduit par la surestimation de σ en utilisant le $REQM$. Cet estimateur est utilisé comme écart type de b_1 , mais le tableau 2.1 montre que l'écart type des estimateurs $b_{1,i}$ est en moyenne de 0,0025. En utilisant l'estimateur de $s(b_1)$ on sous-estime donc significativement l'écart type des estimateurs $b_{1,i}$, ce qui amène un trop grand nombre de rejet. On obtient donc un seuil empirique de 46% qui est beaucoup plus grand que le seuil théorique de 5%.

On conclut donc que l'utilisation du modèle (1.1), en présence d'autocorrélation, amène la détection trop fréquente de tendances "inexistantes". Ce résultat montre l'obligation de tenir compte de l'autocorrélation lorsque l'on travaille avec la régression linéaire; une utilisation pour la détection de tendance ne fait pas exception à cette règle.

Plusieurs possibilités de correction de la présence d'autocorrélation sont présentées dans la littérature. L'économétrie est un domaine particulièrement tourné vers la résolution de ce problème d'utilisation de la régression linéaire. Kmenta (1971) ainsi que Wonnacott et Wonnacott (1980) présentent différentes méthodes de correction pour tenir compte de la présence d'autocorrélation: les premières différences et les différences généralisées.

Méthodes des premières différences.

Cette première méthode suppose que l'autocorrélation se trouve dans les erreurs et que la structure de l'autocorrélation est :

$$e_t = e_{t-1} + v_t$$

où les v_t sont indépendants et identiquement distribués $N(0, \sigma^2)$. En utilisant le modèle (1.1) on peut soustraire la réalisation au temps "t" de la réalisation au temps "t-1" pour obtenir:

$$(Y_t - Y_{t-1}) = \beta_1(X_t - X_{t-1}) + (e_t - e_{t-1})$$

Les erreurs $(e_t - e_{t-1})$ ont alors les propriétés requises pour l'utilisation des moindres carrés. On peut donc estimer β_1 avec une régression de $(Y_t - Y_{t-1})$ sur $(X_t - X_{t-1})$. Cette première méthode repose cependant sur un modèle d'erreur peu général.

Méthodes des différences généralisées.

Cette méthode suppose également que l'autocorrélation se trouve dans les erreurs, mais utilise un modèle plus général pour les erreurs:

$$e_t = \phi e_{t-1} + v_t$$

avec $|\phi| < 1$. En utilisant de nouveau le modèle (1.1), on peut soustraire la réalisation au temps "t-1" multipliée par ϕ de la réalisation au temps "t", pour obtenir:

$$Y_t - \phi Y_{t-1} = \beta_0(1 - \phi) + \beta_1(X_t - \phi X_{t-1}) + (e_t - \phi e_{t-1})$$

Les erreurs $(e_t - \phi e_{t-1})$ ont alors les propriétés requises pour l'utilisation des moindres carrés. Lorsque ϕ est connu, on peut donc estimer β_1 avec une régression de $(Y_t - \phi Y_{t-1})$ sur $(X_t - \phi X_{t-1})$. Certains ajustements sont cependant nécessaires (Wonnacott et Wonnacott, 1980, p.231) pour les observations au temps "1" puisque Y_0 et X_0 sont alors indéfinis.

Lorsque ϕ est inconnu, on doit alors procéder en deux étapes. Deux approches sont présentées. La première, souvent associée à Durbin, consiste à effectuer la régression :

$$Y_t = \phi Y_{t-1} + \beta_0(1 - \phi) + \beta_1(X_t - \phi X_{t-1}) + (e_t - \phi e_{t-1})$$

pour obtenir un estimateur de ϕ , alors que la deuxième consiste à faire une régression avec le modèle (1.1) et à utiliser les résidus pour obtenir un premier estimateur du coefficient d'autorégression:

$$\hat{\phi} = \frac{\sum \hat{e}_t \sum \hat{e}_{t-1}}{\sum \hat{e}_{t-1}^2}$$

Dans les deux cas, on utilise l'estimateur de ϕ pour construire les variables :

$$\tau Y = (Y_t - \hat{\phi} Y_{t-1})$$

$$\tau X = (X_t - \hat{\phi} X_{t-1})$$

et on effectue la régression:

$$\tau Y_t = \beta_0(1 - \phi) + \beta_1 \tau X_t + v_t$$

Il est important de noter que toutes ces méthodes supposent que l'autocorrélation suit une structure autorégressive d'ordre 1 : AR(1). Bien que cette structure permette de bien modéliser la majorité des situations dans le domaine de l'eau, il peut arriver que certaines applications demandent une structure plus complexe. La section 2.2 présente une autre méthode, plus flexible à cet égard, qui sera utilisée dans le présent travail pour tenir compte de l'autocorrélation.

2.2 Détection de tendance en présence d'autocorrélation.

En présence d'autocorrélation, une dépendance à court terme est introduite dans la série des résidus. Afin d'extraire cette dépendance, on remplacera le modèle (1.1) par :

$$Y_t = \beta_0 + \beta_1 X_t + \sum_{i=1}^p \phi_i Y_{t-i} + e_t \quad (2.1)$$

où Y_t , β_0 , β_1 , X_t et e_t sont identiques à l'équation (1.1). On introduit donc, du côté des variables explicatives, les variables avec retard $Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}$. Ces nouvelles

variables explicatives permettent d'introduire, à l'intérieur du modèle, les corrélations entre les observations successives et, donc, de les exclure de la série des résidus.

Cette méthode qui sera appelée **régression autorégressive** doit cependant être validée avant d'être utilisée dans les applications. On doit donc répondre à quelques interrogations avant de pouvoir utiliser le modèle (2.1) systématiquement pour la détection de tendance en présence d'autocorrélation: Est-ce que le seuil du test de détection de tendance est égal au seuil théorique? Comment la puissance du test de détection de tendance se compare-t-elle avec la puissance théorique? Combien de variables explicatives avec retard doit-on inclure dans le modèle (2.1)?

Wonnacott et Wonnacott (1980, p244) abordent brièvement le sujet en associant cette méthode à la présence d'autocorrélation dans la variable dépendante. Les auteurs concluent que si Y_1 est connu les estimateurs des moindres carrés sont les estimateurs du maximum de vraisemblance. De plus, les auteurs stipulent que l'effet de la restriction " Y_1 connu" a tendance à diminuer à mesure que le nombre d'observations augmente. Les prochaines sections permettront de juger de l'applicabilité de cette méthode avec des séries de 300 observations.

2.2.1 Etude du seuil lorsque p est connu.

Lorsque l'ordre du processus autorégressif (p) est connu, le choix du nombre de variables avec retard est immédiat. En effet, en présence d'un modèle AR(1), une seule variable avec retard permettra d'éliminer l'autocorrélation présente dans les résidus. Plus généralement, p variables avec retard permettront de tenir compte de l'autocorrélation introduite par un modèle AR(p).

La simulation 2 permet l'étude du seuil du test de détection de tendance en utilisant la régression autorégressive en présence d'autocorrélation. Comme aucune tendance n'est présente dans ces séries, le pourcentage de rejet de H_0 représente le seuil empirique du test.

Le tableau 2.2 présente les résultats numériques associés aux estimateurs ponctuels. On remarque que l'introduction d'une variable avec retard permet un estimateur de σ non biaisé et que l'estimateur de β est de nouveau non biaisé.

SIMULATION 2

CONDITIONS EXPERIMENTALES

Titre : Utilisation du modèle (2.1) en présence d'autocorrélation (seuil).

Estimateurs étudiés : b_1 , d_1 , *REQM*, seuil.

Nombre de séries simulées : 500 X **Longueur des séries :** 300.

Modèle sous-jacent : AR(1) avec $\phi_1 = 0,75$, $\Delta = 0$, $\sigma^2 = 1,0$.

Modèle de régression utilisé : $Y_t = \beta_0 + \beta_1 X_t + \phi_1 Y_{t-1} + e_t, X_t = t$

Tableau 2.2

Résultats numériques pour les estimateurs de la simulation 2;
Seuil de la régression autorégressive en présence d'autocorrélation.

Estimateur	Moyenne	Ecart type	Précision
<i>REQM</i>	1,001	0,04	0,0036
b_1	0,00002	0,0007	0,00006
d_1	0,733	0,04	0,0036

$$\hat{\alpha} = \text{Seuil empirique} = 0,056$$

On note finalement que le seuil empirique n'est pas significativement différent de la valeur théorique. Afin de compléter cette étude de seuil, on a effectué les simulations 2b et 2c qui utilisent les mêmes conditions expérimentales que la simulation 2, sauf pour les valeurs de ϕ_1 fixées à 0,25 et 0,95. Les résultats numériques pour les estimateurs ponctuels et les seuils empiriques sont présentés aux tableaux 2.3 et 2.4.

Tableau 2.3
 Résultats numériques de la simulation 2b; Seuil de la
 régression autorégressive en présence
 d'autocorrélation avec $\phi=0,25$.

Estimateur	Moyenne	Ecart type	Précision
<i>REQM</i>	1,002	0,04	0,0036
b_1	-0,00001	0,0007	0,00006
d_1	0,237	0,06	0,005

$$\hat{\alpha} = \text{Seuil empirique} = 0,064$$

Tableau 2.4
 Résultats numériques de la simulation 2c; Seuil de la
 régression autorégressive en présence
 d'autocorrélation $\phi=0,95$.

Estimateur	Moyenne	Ecart type	Précision
<i>REQM</i>	0,998	0,04	0,0037
b_1	-0,00006	0,0007	0,00006
d_1	0,927	0,03	0,002

$$\hat{\alpha} = \text{Seuil empirique} = 0,102$$

Ces résultats permettent de constater que, pour une autocorrélation faible, le seuil observé n'est pas significativement différent du seuil théorique, alors que, pour une autocorrélation fortement élevée, le seuil observé est plus grand que le seuil théorique.

Il apparaît donc que, pour une autocorrélation allant jusqu'à 0,75, la méthode présentée pour tenir compte de l'autocorrélation est tout à fait adéquate en ce qui concerne le seuil du test de détection de tendance. Pour une autocorrélation très forte, il faut savoir que le test a un seuil légèrement plus grand que le seuil théorique: on détectera donc trop souvent des tendances inexistantes.

2.2.2 Etude de la puissance lorsque p est connu.

On cherche ici à connaître la puissance du test sur β_1 à détecter une tendance linéaire en présence d'autocorrélation et en utilisant le modèle (2.1). Il est important de

noter que, dans les séries simulées, la tendance est introduite dans la série e_t et que l'autocorrélation est introduite en utilisant la série e_t avec tendance. Il faut alors noter, qu'en utilisant le modèle 2.1, l'estimateur de β_1 (b_1) est associé à la tendance introduite dans les erreurs (Δ_e) et que cette tendance correspond à une tendance $(1-\phi)\Delta_y$, où Δ_y est la tendance réelle dans la série Y_t . En présence d'autocorrélation, l'estimateur b_1 est donc associé à la tendance dans les erreurs.

La simulation 3 étudie donc la puissance du test de détection de tendance alors que les résultats numériques des estimateurs ponctuels et de la puissance sont présentés au tableau 2.5.

SIMULATION 3

CONDITIONS EXPERIMENTALES

Titre : Utilisation du modèle (2.1) en présence d'autocorrélation (puissance).

Estimateurs étudiés : b_1 , d_1 , *REQM*, puissance.

Nombre de séries simulées : 500 X **Longueur des séries :** 300.

Modèle sous-jacent : AR(1) avec $\phi_1 = 0,75$, $\Delta_m = 0,002667$, $\sigma^2 = 1$.

Modèle de régression utilisé : $Y_t = \beta_0 + \beta_1 X_t + \phi_1 Y_{t-1} + e_t$, $X_t = t$.

Tableau 2.5

Résultats numériques de la simulation 3; Puissance de la régression autorégressive en présence d'autocorrélation.

Estimateur	Moyenne	Ecart type	Précision
<i>REQM</i>	1,002	0,04	0,0035
b_1	0,00286	0,0008	0,00007
d_1	0,732	0,04	0,0035

Puissance empirique = 0,980

Tout comme, dans l'étude de seuil, l'estimateur de σ n'est pas significativement différent de la valeur théorique. On remarque cependant que l'estimateur de β_1 (0,00286) apparaît biaisé par rapport à la valeur théorique (0,00267) étant donné la précision de l'estimateur (0,00007). Ce biais est cependant introduit par la sous-estimation de ϕ_1 qui affecte l'estimation de β_1 . Pour appuyer ce point, on montre que :

$$0,00267 = \frac{(1-0,75)}{(1-0,732)} 0,00286$$

On rappelle que la puissance empirique du test correspond ici au pourcentage des b_1 standardisés qui sont supérieurs à 1,96 ou inférieurs à -1,96 (rejet de H_0). Pour la simulation 3, la puissance empirique est de 98%. La puissance théorique, telle que définie à la section 1.1.2, est de 0,98 (tableau 1.1) dans des conditions semblables. Il y a donc concordance entre les puissances empirique et théorique pour cette simulation. Les faibles biais des estimateurs de β_1 et de ϕ_1 n'affectent donc pas la puissance du test de détection de tendance. L'effet inverse des biais de ces deux estimateurs explique également leur effet négligeable sinon nul sur la puissance du test de détection de tendance.

Les simulations 3b et 3c permettent de confirmer ces observations. On utilise alors les mêmes conditions expérimentales que pour la simulation 3 sauf pour $\Delta_m = 0,000667$ et $\Delta_m = 0,001333$. Les tableaux 2.6 et 2.7 présentent les résultats numériques de ces deux simulations complémentaires.

Tableau 2.6

Résultats numériques de la simulation 3b; Puissance de la régression autorégressive en présence d'autocorrélation (tendance : $\Delta_m = 0,000667$).

Estimateur	Moyenne	Ecart type	Précision
<i>REQM</i>	0,999	0,04	0,0036
b_1	0,00074	0,0008	0,00007
d_1	0,732	0,04	0,0035

Puissance empirique = 0,208

Tableau 2.7
 Résultats numériques de la simulation 3c; Puissance de la
 régression autorégressive en présence d'autocorrélation
 (tendance : $\Delta_m = 0,001333$).

Estimateur	Moyenne	Ecart type	Précision
<i>REQM</i>	1,001	0,04	0,0036
b_1	0,00144	0,0008	0,00007
d_1	0,729	0,04	0,0037

Puissance empirique = 0,528

On remarque de nouveau un biais dans l'estimation du coefficient d'autorégression qui se répercute dans l'estimation de la pente de la tendance. Cependant, comme la précision des estimateurs demeure la même, mais que l'amplitude du biais dépend de l'amplitude de la tendance, l'estimateur de β_1 n'est pas toujours significativement différent de la valeur théorique. En tenant compte de la précision d'une proportion (tableau 1.3) et de la précision des entrées sur le tableau 1.1, on conclut de nouveau que les puissances empiriques ne sont pas différentes des puissances théoriques.

On peut aussi généraliser ce résultat à tous les cas d'autocorrélation inférieure à 0.75 (p.e. une simulation avec $\phi_1 = 0,25$ et $\Delta_m = 0,002667$ a donné une puissance de 98,4 %). Cependant, on ne peut porter de jugement pour les niveaux d'autocorrélation très élevés puisque le seuil n'est pas conservé dans cette situation.

2.2.3 Seuil et puissance lorsque p est inconnu.

Généralement, il est impossible de connaître l'ordre exact du processus autorégressif présent dans la structure des données. Cependant, certains résultats théoriques permettent de croire que cette situation n'est pas incontrôlable. Parmi différents résultats présentés par Tsay et Tiao (1984), on note qu'une surparamétrisation dans un modèle de régression de la structure autorégressive d'un processus purement autorégressif (utilisation de k séries avec retard pour estimer les paramètres d'un $AR(p)$, $k > p$) amène des estimateurs de ϕ_1, \dots, ϕ_p convergents tout comme les estimateurs de $\phi_{p+1}, \dots, \phi_k$ qui convergent vers 0.

Afin d'illustrer ces résultats et de s'assurer qu'une surparamétrisation n'amène aucun effet sur l'estimation de β_1 , la simulation 4 a été effectuée. Le tableau 2.8 présente les résultats numériques associés à cette simulation. Ces résultats permettent de constater que la surparamétrisation à l'aide de variables avec retard n'affecte pas significativement la détection de tendance. En effet, les estimateurs de β_1 , σ et ϕ_1 , ainsi que le seuil empirique ne sont pas significativement différents des valeurs théoriques ni des valeurs de la simulation 2 où il n'y a pas de surparamétrisation. Les simulations 4b et 4c complètent cette étude de seuil; les conditions expérimentales sont alors les mêmes que pour la simulation 4 sauf pour ϕ_1 qui est remplacé par les valeurs 0,25 et 0,95. Les seuils empiriques obtenus sont : $\hat{\alpha} = 0,062$ pour $\phi_1=0,25$ ($d_1 =0,244$) et $\hat{\alpha} = 0,090$ pour $\phi_1=0,95$ ($d_1 =0,938$).

SIMULATION 4

CONDITIONS EXPERIMENTALES

Titre : Effet de la surparamétrisation sur le seuil.

Estimateurs étudiés : b_1 , d_1 , d_2 , d_3 , *REQM*, seuil.

Nombre de séries simulées : 500 X **Longueur des séries :** 300.

Modèle sous-jacent : AR(1) avec $\phi_1 = 0,75$, $\phi_2 = \phi_3 = 0$, $\Delta_m = 0,002667$, $\sigma^2 = 1,0$

Modèle de régression utilisé : $Y_t = \beta_0 + \beta_1 X_t + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-3} + e_t$, $X_t = t$.

Tableau 2.8

Résultats numériques de la simulation 4; Seuil de la régression autorégressive avec surparamétrisation en présence d'autocorrélation.

Estimateur	Moyenne	Ecart type	Précision
<i>REQM</i>	1,000	0,04	0,0037
b_1	0,00003	0,0007	0,00006
d_1	0,739	0,06	0,005
d_2	0,001	0,07	0,007
d_3	-0,010	0,06	0,005

$\hat{\alpha} = \text{Seuil empirique} = 0,062$

Comme les seuils empiriques obtenus dans les simulations 4, 4b et 4c ne sont pas significativement différents des seuils empiriques des simulations 2, 2b et 2c, on peut conclure que :

1. La surparamétrisation de l'autocorrélation n'a pas d'influence sur le seuil de la détection de tendance (test sur β_1).
2. On note de nouveau un seuil empirique significativement plus grand que le seuil théorique lorsque l'autocorrélation est de l'ordre de 0,95.

La simulation 5 étudie l'effet de la surparamétrisation sur la puissance du test de détection de tendance.

SIMULATION 5

CONDITIONS EXPERIMENTALES

Titre : Effet de la surparamétrisation sur la puissance.

Estimateurs étudiés : b_1 , d_1 , d_2 , d_3 , *REQM*, seuil.

Nombre de séries simulées : 500 X **Longueur des séries :** 300.

Modèle sous-jacent : AR(1) avec $\phi_1 = 0,75$, $\phi_2 = \phi_3 = 0$, $\Delta_m = 0,002667$, $\sigma^2 = 1,0$.

Modèle de régression utilisé : $Y_t = \beta_0 + \beta_1 X_t + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-3} + e_t$, $X_t = t$.

Tableau 2.9

Résultats numériques de la simulation 5; Puissance de la régression autorégressive avec surparamétrisation en présence d'autocorrélation.

Estimateur	Moyenne	Ecart type	Précision
<i>REQM</i>	0,999	0,04	0,0035
b_1	0,00295	0,0008	0,00007
d_1	0,740	0,06	0,005
d_2	-0,002	0,07	0,006
d_3	-0,009	0,06	0,005

Puissance empirique = 0,970

Ce résultat permet de constater que la surparamétrisation de variables avec retard n'amène pas une puissance différente de la puissance théorique. On obtient de nouveau des estimateurs de β_1, σ et ϕ_1 qui ne sont pas significativement différents des valeurs théoriques et une puissance empirique non significativement différente de la puissance théorique. Les simulations 5b et 5c complètent cette étude de puissance; les conditions expérimentales sont alors les mêmes que pour la simulation 5, sauf pour Δ_m qui est remplacé par les valeurs 0,000667 et 0,001333: les puissances empiriques obtenues sont alors respectivement de 0,202 et de 0,484. Ces puissances empiriques ne sont pas significativement différentes des puissances théoriques (tableau 1.1): soit respectivement 0,17 et 0,52.

Comme la surparamétrisation de la régression autorégressive n'affecte pas l'estimation des paramètres étudiés, il peut être intéressant d'effectuer une première régression avec plusieurs variables avec retard pour ensuite faire une autre régression en éliminant les variables avec retard associées à des paramètres non significativement différents de 0. Les résultats de la dernière régression seraient alors utilisés pour la détection de tendance. Bien que les résultats des deux régressions ne soient pas significativement différents, il est plus élégant de laisser de côté les variables non significatives dans le modèle final (parcimonie).

2.3 Type de tendance et présence d'autocorrélation.

Jusqu'ici, les simulations effectuées n'ont traité que des tendances monotones linéaires. Dans la présente section, deux questions supplémentaires sont discutées:

1. les résultats obtenus pour des tendances monotones sont-ils valables dans le cas de tendances par saut? et;
2. comment se comportent seuil et puissance lorsqu'un modèle construit pour une tendance monotone est utilisé en présence de tendance par saut?

2.3.1 Détection de tendances par saut.

Étant donné que seule la variable X_t change en présence de tendance par saut et que de plus, la variable :

$$X_t = \begin{cases} 0 & \text{si } t \leq t_0 \\ 1 & \text{si } t > t_0 \end{cases}$$

est tout à fait adaptée à la détection de tendance par saut, on doit s'attendre à ce que les résultats obtenus dans le cas de tendances monotones soient valables pour la détection de tendances par saut en utilisant la variable X_t définie plus haut. La simulation 6 présente une étude de puissance qui permet d'appuyer ce point.

SIMULATION 6

CONDITIONS EXPERIMENTALES

Titre : Utilisation du modèle (2.1) en présence d'autocorrélation (puissance).

Estimateurs étudiés : b_1 , d_1 , $REQM$, puissance.

Nombre de séries simulées : 500 X **Longueur des séries :** 300.

Modèle sous-jacent : AR(1) avec $\phi_1 = 0,75$, $\Delta_s = 0,3464$, $\sigma^2 = 1,0$.

Modèle de régression utilisé : $Y_t = \beta_0 + \beta_1 X_t + \phi_1 Y_{t-1} + e_t$, $X_t = 1$ si $t > t_0$

Le tableau 2.10 présente les résultats numériques associés à cette simulation. Ces résultats sont tout à fait semblables à ceux obtenus avec des tendances monotones. On remarque une légère sous-estimation du coefficient d'autorégression qui amène une légère surestimation de l'amplitude du saut. Ces biais n'affectent cependant pas la puissance du

test puisque cette dernière n'est pas significativement différente de la puissance théorique de 0,85 (réf. tableau 1.1).

Tableau 2.10

Résultats numériques de la simulation 6; Puissance de la régression autorégressive en présence de tendance par saut et d'autocorrélation.

Estimateur	Moyenne	Ecart type	Précision
<i>REQM</i>	1,001	0,04	0,0035
b_1	0,365	0,129	0,011
d_1	0,733	0,04	0,0034

Puissance empirique = 0,874

2.3.2 Détection de tendance de type inconnu.

Dans certaines situations, le choix de la variable X_t peut être dicté par des facteurs externes. Par exemple, lors d'une étude sur l'effet de la mise en marche d'une usine d'assainissement sur la qualité des eaux d'une rivière, on utilisera un test associé à une tendance par saut. Le type de tendance présente dans la série est alors secondaire puisque le but de l'étude est de savoir s'il existe une tendance par saut à une date connue.

Cependant, en pratique, on connaît très rarement le type de tendance présente dans une série chronologique. Au mieux, on peut identifier le type de tendance le plus probable à l'aide de méthodes telles les courbes CUSUM.

Devant de telles difficultés, on doit se demander, comment se comportent seuil et puissance du test de détection de tendance lorsque la variable X_t n'est pas exactement adaptée au type de tendance présente dans la série. Comme le seuil n'a pas de sens en présence de tendance (H_0 vraie), le choix de la variable X_t est sans importance au niveau de cette caractéristique. En effet, quelle que soit la variable X_t utilisée, le seuil du test de détection de tendance doit être égal au seuil théorique.

La simulation 7 permet l'étude de la puissance lorsque la variable X_t utilisée ne correspond pas exactement au type de tendance, alors que le tableau 2.11 présente les résultats numériques de cette simulation.

SIMULATION 7

CONDITIONS EXPERIMENTALES

Titre : Effet sur la puissance d'une variable X_t mal adaptée.

Estimateurs étudiés : b_1 , d_1 , $REQM$, puissance.

Nombre de séries simulées : 500 X **Longueur des séries :** 300.

Modèle sous-jacent : AR(1) avec $\phi_1 = 0,75$, $\sigma^2 = 1$ et tendance par saut $\Delta_s = 0,1155$

Modèle de régression utilisé : $Y_t = \beta_0 + \beta_1 X_t + \phi_1 Y_{t-1} + e_t$, $X_t = t$.

Tableau 2.11

Résultats numériques de la simulation 7; Puissance de la régression autorégressive avec mauvaise modélisation du type de tendance.

Estimateur	Moyenne	Ecart type	Précision
<i>REQM</i>	1,000	0,04	0,0036
b_1	0,00058	0,0007	0,00006
d_1	0,730	0,04	0,0036

Puissance empirique = 0,142

Ces résultats numériques permettent de conclure que les estimateurs de σ et de ϕ_1 ne sont pas biaisés, alors que l'estimateur de la pente de la tendance monotone ($b_1=0,00058$) ne correspond pas à l'amplitude du saut introduit dans la simulation ($\Delta_s=0,1155$). En effet, b_1 représente la pente d'une droite passant par les deux nuages de points constitués avant et après le saut. En ce qui concerne la puissance, elle est à la limite d'être significativement plus petite que la puissance théorique de 0,17 (réf. tableau 1.1).

Les simulations 7b et 7c permettent de compléter cette étude de puissance; les conditions expérimentales sont les mêmes que dans la simulation 7, sauf pour la valeur de Δ_s , fixée à 0,2309 et 0,4619. Les puissances empiriques obtenues sont respectivement de 0,42 et de 0,87. Ces puissances empiriques sont significativement plus petites que les puissances théoriques de 0,50 et de 0,98. On en conclut que le choix d'une variable X_t mal adaptée au type de tendance amène une perte de puissance. Il faut quand même noter que cette perte de puissance n'est pas drastique. De plus, une tendance présente dans une série ne sera jamais parfaitement linéaire ou parfaitement par saut, on peut donc croire qu'en pratique, la perte de puissance est moins importante que celle obtenue dans les simulations 7b et 7c.

2.4 Discussion sur la présence d'autocorrélation.

Ce chapitre a permis de valider l'utilisation de la régression autorégressive pour tenir compte de la présence d'autocorrélation. À l'exception de la présence de très forte autocorrélation, les études de simulations ont montré que le seuil de 5% est conservé et que la puissance empirique n'est pas significativement différente de la puissance théorique (réf. section 1.1.2). On peut donc connaître les propriétés de ce test en présence d'autocorrélation sauf pour les cas de très forte autocorrélation.

Il faut cependant rappeler que les résultats présentés dans le présent chapitre sont associés aux amplitudes des tendances au niveau des erreurs. Lorsque l'intérêt premier est la tendance (Δ_y) dans la série Y_t , il est possible de l'estimer en utilisant la formule suivante :

$$\Delta_y = b_1 / (1-d_1)$$

où b_1 est l'estimateur de la pente de la tendance dans les erreurs et d_1 est l'estimateur du coefficient d'autorégression. Cette distinction entre la tendance dans les erreurs et la tendance dans la série Y_t permet de conclure que :

1. pour une amplitude donnée au niveau des erreurs, le test utilisant la régression autorégressive aura la même puissance quelque soit le niveau du coefficient d'autorégression;
2. pour une puissance donnée, l'amplitude de la tendance dans la série Y_t qui est détectée par la régression autorégressive augmente à mesure que le niveau d'autocorrélation augmente; par exemple, on détectera 85% du temps une tendance (dans la série Y_t) de $0,002\sigma$ si $\phi=0$, de $0,008\sigma$ si $\phi=0,75$ et de $0,020\sigma$ si $\phi=0,90$. Il faut noter que, dans tous les cas de l'exemple, l'amplitude de la tendance au niveau des erreurs est de $0,002\sigma$.

Lorsque l'ordre du processus autorégressif qui amène l'autocorrélation n'est pas connu, on a montré que l'utilisation d'une surparamétrisation permet d'effectuer le test de détection de tendance sans affecter le seuil et la puissance.

Lors de l'application de cette méthode sur des données réelles, on propose la procédure suivante :

1. Effectuer un test de Durbin-Watson sur les résidus d'une régression n'utilisant pas de variable avec retard. Si le test ne conclut pas qu'une autocorrélation significative est présente, on oublie l'autocorrélation comme caractéristique restrictive.
2. Si le test de Durbin-Watson conclut qu'une autocorrélation significative est présente introduire une variable avec retard si le modèle AR(1) est approprié, ou plusieurs variables avec retard si l'ordre du processus engendrant l'autocorrélation est inconnu.

L'utilisation du test de Durbin-Watson n'est cependant pas nécessaire, puisque les résultats du présent chapitre montre que la surparamétrisation peut être utilisée. Une alternative serait donc d'utiliser un certain nombre de variables avec retard (3 variables avec retard apparaît un nombre approprié pour la majorité des applications dans le domaine de l'eau) et de juger de la présence d'autocorrélation à partir de la signification des estimateurs associés aux variables avec retard.

CHAPITRE 3

PRÉSENCE DE VALEURS ABERRANTES

La présence de valeurs douteuses est un sujet largement étudié dans le domaine statistique, puisque toutes les méthodes reposant sur l'hypothèse de normalité (*t*-test, analyse de variance, régression, etc.) sont affectées par la présence d'observations aberrantes. De façon générale, l'analyse statistique en présence de valeurs aberrantes se fait en deux étapes: Dans un premier temps, on a l'**identification** des valeurs aberrantes (Barnett et Lewis, 1978, Hawkins 1980) qui se fait généralement à l'aide de méthodes graphiques ou à l'aide d'identification d'écart importants entre une observation et la moyenne ou entre une valeur prédite et une valeur mesurée. Dans la deuxième étape, le **traitement** statistique des valeurs aberrantes se fait souvent en exécutant une analyse avec et sans les valeurs identifiées comme aberrantes et en comparant les résultats afin de mettre en évidence l'effet des valeurs douteuses sur les conclusions de l'analyse. Lorsque le nombre de valeurs aberrantes est assez grand, un tel traitement devient rapidement très encombrant. Il est alors préférable d'avoir recours à des méthodes qui sont peu influencées par la présence de valeurs aberrantes comme les méthodes non paramétriques et les méthodes robustes d'Huber.

L'utilisation des méthodes non paramétriques permet un traitement statistique adéquat en présence de valeurs aberrantes, car on utilise généralement les rangs des observations au lieu des observations brutes. Une telle transformation permet de limiter l'effet des valeurs extrêmes, puisque les valeurs minimales et maximales auront toujours les rangs 1 et T quels que soient leurs écarts avec le reste des observations. Jusqu'à présent l'utilisation des méthodes non paramétriques représente presque la seule option pour la détection de tendance dans le domaine de l'eau. L'utilisation des tests de Spearman, de Mann-Whitney et de Mann-Kendall, ainsi que des adaptations de ces trois tests pour tenir compte de l'autocorrélation et/ou des saisonnalités (Lettenmaier 1976, Hirsch, Slack et Smith 1982, Hirsch et Slack 1984) permet une détection de tendance adéquate pour les séries chronologiques contenant des valeurs douteuses.

L'utilisation des méthodes robustes d'Huber représente une option supplémentaire, encore peu connue et très peu utilisée pour la détection de tendance. Pourtant les développements associés à Huber (1981) peuvent être d'une grande utilité pour la détection de tendance en présence de valeurs aberrantes et peuvent facilement être incorporés dans des études utilisant la régression linéaire. Le présent chapitre discute des méthodes d'estimation robuste d'Huber (1981) et présente leur utilisation conjointe avec la régression linéaire pour détecter des tendances dans les séries chronologiques.

3.1 Méthodes robustes d'Huber.

3.1.1 Présentation générale

Les estimateurs qui seront utilisés dans le présent chapitre pour diminuer l'influence des valeurs aberrantes sont les M-Estimeurs. Ces estimateurs sont de type maximum de vraisemblance, i.e. qu'ils minimisent une somme du type :

$$\sum \rho(x_i; T_n)$$

ou encore qu'ils sont la solution d'une équation telle:

$$\sum \psi(x_i; T_n) = 0$$

où ρ est une fonction arbitraire et où $\psi(x; \theta) = (\partial/\partial\theta)\rho(x; \theta)$.

Le choix $\rho(x; \theta) = -\log f(x; \theta)$ donne l'estimateur classique du maximum de vraisemblance. Cependant, il est bien connu que ces estimateurs sont fortement influencés par la présence de valeurs aberrantes. Afin de diminuer l'influence de ces valeurs atypiques, Huber a proposé la fonction :

$$\Psi_{H,k}(v) = \begin{cases} v, & \text{si } |v| < k \\ k \operatorname{sign}(v) & \text{si } |v| \geq k \end{cases}$$

L'utilisation de cette fonction permet de borner l'influence d'une observation sur les estimateurs. Ces M-Estimeurs sont souvent utilisés pour estimer le paramètre de localisation (Huber 1981). Ils ont également été utilisés pour l'estimation des paramètres des séries chronologiques : Le premier ouvrage majeur Denby et Martin (1979) présente l'utilisation des M-Estimeurs pour estimer le coefficient d'autocorrélation d'ordre 1. Les ouvrages subséquents Martin (1980), Martin et Yohai (1985), Bustos et Yohai (1986) s'attardent à l'estimation de tous les coefficients des modèles ARMA (coefficients autorégressifs et de moyennes mobiles).

Laberge (1988) a étudié l'estimation robuste des paramètres de localisation et d'échelle dans les séries chronologiques. L'estimation de ces deux paramètres est primordiale dans une optique de détection de tendance, alors que l'estimation des paramètres des modèles ARMA devient plutôt secondaire et vise surtout à éliminer l'effet d'autocorrélation qui amène la non indépendance des observations. Bien qu'en présence de valeurs aberrantes seulement, l'estimation robuste de ces deux paramètres permette

d'effectuer une détection de tendance par saut en utilisant un test semblable au test t , elle ne permet pas une détection de tendance en présence de saisonnalités et d'autocorrélation.

Afin de pouvoir détecter des tendances en présence de valeurs aberrantes, d'autocorrélation et/ou de saisonnalités, les M-Estimeurs doivent être utilisés pour estimer les paramètres des modèles de régression développés aux chapitres 2 et 4. Les développements associés à l'utilisation des M-Estimeurs avec la régression linéaire sont présentés dans Huber (1981). La prochaine section présente les développements utilisés pour rendre robuste la détection de tendance à l'aide de la régression linéaire.

3.1.2 Régression multiple robuste.

L'estimation robuste des paramètres de la régression linéaire se fait en deux étapes (Huber 1981, p179): 1) l'**étape échelle** où l'on estime la variance des résidus et; 2) l'**étape localisation** où l'on estime les paramètres du modèle de régression. Pour l'étape localisation, deux approches sont proposées: la première utilise les résidus modifiés ("Modified Residuals") alors que la seconde utilise les poids modifiés ("Modified Weights"). La première approche est utilisée pour les développements du présent chapitre.

L'algorithme de calcul des différents estimateurs sera maintenant présenté alors que la sous-routine FORTRAN utilisée pour les calculs des M-estimateurs est présentée à l'appendice A.

1. Calcul des estimateurs préliminaires.

Comme les M-Estimeurs sont la solution d'un système d'équations non linéaires, ils sont obtenus à l'aide de méthodes numériques itératives convergeant vers cette solution. Afin de débiter les itérations, il est nécessaire d'avoir de bons estimateurs préliminaires qui

devraient être assez proche de la solution. Les estimateurs classiques du maximum de vraisemblance satisfont à ce critère, ils sont donc utilisés comme estimateurs préliminaires.

2. Étape échelle

À chaque itération, on ajuste l'estimation du paramètre d'échelle à l'aide de l'équation :

$$(\sigma^{(m+1)})^2 = \frac{1}{(T-p)\beta(k)} \sum \psi\left(\frac{r_i}{\sigma^{(m)}}\right)^2 (\sigma^{(m)})^2$$

où $\sigma^{(k)}$ est l'estimateur du paramètre d'échelle à la k e itération, T est le nombre d'observations, p est le nombre de paramètres, $\beta(k) = E(\psi^2(Z))$ est l'espérance sous l'hypothèse de normalité (i.e. Z est une variable aléatoire $N(0,1)$) et r_i est la série de résidus de la régression linéaire. Il apparaît donc que $(\sigma^{(m+1)})^2$ est un estimateur ordinaire de la variance mais qui est calculé à l'aide des résidus winsorisés obtenus avec la fonction d'Huber et corrigé pour le biais à l'aide du facteur $\beta(k)$. Afin d'illustrer cette remarque notons simplement que :

$$\psi\left(\frac{r_i}{\sigma^{(m)}}\right)(\sigma^{(m)}) = \begin{cases} r_i & \text{si } |r_i| < k\sigma^{(m)} \\ k \text{sign}(r_i)\sigma^{(m)} & \text{si } |r_i| \geq k\sigma^{(m)} \end{cases}$$

le résidu reste donc inchangé s'il est plus petit que k écarts types, mais sa valeur est limitée à k écarts types (en gardant le signe du résidu) lorsqu'il est plus grand que cette dernière valeur. Il faut bien noter ici que les grands résidus ne sont pas éliminés des calculs mais on limite plutôt leurs amplitudes afin de diminuer leur influence sur les estimateurs. Il existe une alternative intéressante pour l'étape échelle dans Rivest (1988).

3. Étape localisation.

Une fois l'estimateur d'échelle obtenu, on peut recalculer à chaque itération l'estimateur du paramètre de localisation à l'aide du système d'équations suivant:

$$\mathbf{b}^{(m+1)} = \mathbf{b}^{(m)} + (X'X)^{-1}X'\mathbf{r}^{*(m+1)}$$

où $\mathbf{b}^{(k)}$ est un vecteur de dimension $p \times 1$ contenant les estimateurs de β pour la k e itération, X est la matrice "design" et \mathbf{r}^* est le vecteur des résidus winsorisés à l'aide de la fonction de Huber. Donc

$$r_i^{*(m+1)} = \psi\left(\frac{r_i^{(m)}}{\sigma^{(m+1)}}\right)(\sigma^{(m+1)}) = \begin{cases} r_i^{(m)} & \text{si } |r_i^{(m)}| < k\sigma^{(m+1)} \\ k \text{sign}(r_i^{(m)})\sigma^{(m+1)} & \text{si } |r_i^{(m)}| \geq k\sigma^{(m+1)} \end{cases}$$

et les estimateurs sont obtenus à l'aide de la méthode de calcul associée à Gauss-Newton pour résoudre les systèmes d'équations non linéaires.

4. Test sur la convergence des estimateurs.

Maintenant que les estimateurs sont obtenus pour une itération, on teste la convergence des estimateurs. Dans un premier temps, on s'assure que le paramètre d'échelle a convergé: i.e.

$$|\sigma^{(m+1)} - \sigma^{(m)}| < \varepsilon\sigma^{(m+1)}$$

où ε est le critère de convergence sélectionné par l'utilisateur. On utilise généralement des valeurs de l'ordre de 0,01 ou 0,001. Dans les simulations, $\sigma^{(m+1)}$ est de l'ordre de 1 et on utilise $\varepsilon=0,001$, la précision numérique des estimateurs de l'écart type est donc de l'ordre du millièmes. Dans un second temps, après la convergence de l'estimateur du paramètre d'échelle, on teste la convergence de tous les paramètres du modèle de régression. Les estimateurs ont convergé lorsque :

$$|b_i^{(m+1)} - b_i^{(m)}| < \varepsilon s(b_i) \quad , \quad \forall i.$$

où $s(b_i)$ est l'écart type de l'estimateur b_i .

Si tous les estimateurs ont convergé, $\sigma^{(m+1)}$ et $\mathbf{b}^{(m+1)}$ représentent les estimateurs robustes du paramètre d'échelle et des paramètres du modèle de régression sinon, ces mêmes estimateurs sont utilisés dans l'itération suivante.

3.2 Modèles de valeurs aberrantes.

L'étude des valeurs aberrantes dans les séries chronologiques est différente des études où les observations sont indépendantes, puisque la structure autorégressive des séries chronologiques peut amener deux types distincts de valeurs aberrantes. Ainsi, Fox (1972) a distingué les valeurs aberrantes statiques ("Additive Outliers") qui ne sont pas propagées par la structure autorégressive et les valeurs aberrantes dynamiques ("Innovation Outliers") qui sont propagées par cette même structure de dépendance. Afin de garder une certaine homogénéité avec la littérature, majoritairement anglophone, les valeurs aberrantes statiques et dynamiques seront respectivement notées AO et IO.

Dans les sections suivantes il sera toujours important de distinguer trois processus: 1) le **processus sous-jacent** que l'on cherche à étudier; 2) le **processus mesuré**, le seul sur lequel on a de l'information et; 3) le **processus aberrant** qui masque le processus sous-jacent.

3.2.1 Valeurs aberrantes statiques (AO).

Les valeurs aberrantes statiques représentent des impulsions atypiques qui ne sont pas propagées par la structure autorégressive de la série chronologique. Il s'agit donc généralement d'un effet exogène au phénomène étudié. Une erreur de transcription est un bon exemple de valeur aberrante statique.

Dans un processus purement autorégressif d'ordre q , ce type de valeur aberrante peut donc s'écrire de la façon suivante:

$$x_t = \sum_{i=1}^q \phi_i x_{t-i} + e_t$$

$$y_t = x_t + v_t$$

où $\{y_t\}_{t=1}^T$ est le processus observé et où les impulsions e_t sont indépendantes et identiquement distribuées $N(0,1)$. Les v_t constituent le processus aberrant, ils sont supposés indépendants entre eux, indépendants des x_t et identiquement distribués avec distribution:

$$f_v(\cdot; \gamma, \mu_{AO}, \sigma_{AO}^2) = (1 - \gamma) \delta(\cdot) + \gamma N(\cdot; \mu_{AO}, \sigma_{AO}^2)$$

où $\delta(\cdot)$ est une loi dégénérée à 0, μ_{AO} et σ_{AO}^2 sont respectivement le moyenne et la variance des valeurs aberrantes et où $\Pr(v_t = 0) = (1 - \gamma)$ et $\Pr(v_t \neq 0) = \gamma$. Notons que pour ce type de valeurs aberrantes, il est assez simple de reconnaître les processus sous-jacent (x_t), mesuré (y_t) et aberrant (v_t). Notons qu'en général, dans les publications sur l'estimation robuste des paramètres des modèles de séries chronologiques, on travaille avec des valeurs aberrantes statiques de moyenne nulle: $\mu_{AO} = 0$. Cependant, la structure particulière de certaines séries chronologiques dans le domaine de l'eau nous amène à considérer le cas d'AO de moyenne non nulle: $\mu_{AO} \neq 0$.

Laberge (1988) présente les moyennes et écarts types du processus mesuré en présence d'AO. Il montre alors que :

$$E(y_t) = \mu_y = \gamma \mu_{AO}$$

$$Var(y_t) = \sigma_y^2 = \sigma^2 + \gamma \sigma_{AO}^2 + \gamma(1 - \gamma) \mu_{AO}^2$$

où σ^2 est la variance du processus sous-jacent. Dans le cas d'un modèle AR(1) avec coefficient d'autorégression ϕ , on a :

$$\sigma^2 = \frac{\sigma_e^2}{(1-\phi^2)}$$

où σ_e^2 est la variance des impulsions du processus sous-jacent. Le tableau 3.1 présente les valeurs théoriques des moyennes et écarts types des différents processus pour les valeurs de ϕ , μ_{AO} et σ_{AO} utilisées dans les simulations. N.B. $\gamma=0,002$ ce qui correspond à environ une valeur aberrante par année.

Tableau 3.1
Moyennes et écarts types des différents processus en présence de valeurs aberrantes statiques; pour les valeurs de ϕ , μ_{AO} et σ_{AO} utilisées dans les simulations.

ϕ	μ_{AO}	σ_{AO}	μ	σ	μ_y	σ_y
0,75	0	3	0	1,51	0	1,57
0,75	5	2	0	1,51	0,1	1,69
0,75	10	3	0	1,51	0,2	2,11

3.2.2 Valeurs aberrantes dynamiques (IO).

Les valeurs aberrantes dynamiques représentent des impulsions atypiques, propagées par la structure autorégressive de la série chronologique. Il s'agit donc généralement d'un effet endogène au phénomène étudié. Une crue de printemps et un rejet industriel sont de bons exemples de valeurs aberrantes dynamiques.

Dans un processus purement autorégressif d'ordre q , ce type de valeur aberrante peut donc s'écrire de la façon suivante:

$$x_t = \sum_{i=1}^q \phi_i x_{t-i} + e_t$$

$$y_t = x_t$$

où y_t est le processus observé et où les impulsions e_t sont indépendantes et identiquement distribuées, mais avec une distribution caractérisée par des queues plus grandes que la loi

normale. Dans le présent travail, les fonctions de densité des impulsions de processus avec IO seront des lois normales contaminées:

$$f_e(\cdot; \gamma, \sigma^2) = (1 - \gamma)N(\cdot; 0, \sigma^2) + \gamma N(\cdot; \mu_{IO}, \sigma_{IO}^2)$$

où μ_{IO} et σ_{IO}^2 sont respectivement le moyenne et la variance des valeurs aberrantes et où γ représente le pourcentage d'impulsions aberrantes dans le processus. Notons que, pour ce type de valeurs aberrantes, les processus sous-jacent et aberrant sont tous les deux associés aux impulsions de la distribution f_e , il est donc plus difficile de les distinguer dans les équations. Tout comme pour les valeurs aberrantes statiques, la structure particulière de certaines séries chronologiques dans le domaine de l'eau, nous amène à considérer le cas de valeurs aberrantes dynamiques de moyenne non nulle: $\mu_{IO} \neq 0$.

Laberge (1988) présente les moyennes et écarts types du processus mesuré en présence d'IO. Il montre alors qu'avec un processus sous-jacent AR(1) avec coefficient d'autocorrélation ϕ :

$$E(y_t) = \frac{\gamma \mu_{IO}}{(1-\phi)}$$

$$Var(y_t) = \frac{(1-\gamma)\sigma^2 + \gamma\sigma_{IO}^2 + \gamma(1-\gamma)\mu_{IO}^2}{(1-\phi^2)}$$

Le tableau 3.2 présente les valeurs théoriques des moyennes et écarts types des différents processus pour les valeurs de ϕ , μ_{IO} et σ_{IO} utilisées dans les simulations. N.B. $\gamma = 0,002$ ce qui correspond à environ une valeur aberrante par année.

Tableau 3.2

Moyennes et écarts types des différents processus en présence de valeurs aberrantes dynamiques; pour les valeurs de ϕ , μ_{IO} et σ_{IO} utilisées dans les simulations.

ϕ	μ_{IO}	σ_{IO}	μ	σ	μ_y	σ_y
0,75	0	3	0	1	0	1,08
0,75	5	2	0	1	0,4	1,24
0,75	10	3	0	1	0,8	1,77

3.2.3 Emplacement des valeurs aberrantes.

Dans les séries chronologiques, il faut noter que l'emplacement temporel des valeurs aberrantes a également un effet sur l'estimation des paramètres et par le fait même sur la détection des tendances. En effet, si des valeurs aberrantes positives se trouvent au début de la série, on détectera beaucoup de tendances négatives (diminutions du niveau de la série), à l'inverse, si des valeurs aberrantes positives se trouvent à la fin de la série, on détectera beaucoup de tendances positives (augmentations du niveau de la série). Pour cette raison, la situation temporelle des valeurs aberrantes sera également discutée dans les différentes simulations.

Les valeurs aberrantes seront introduites de façon systématique; Il est important de noter que le fait d'introduire les valeurs aberrantes de façon aléatoire ne changerait pas les résultats associés aux méthodes robustes. À moins d'avis contraire, une valeur aberrante est introduite à toutes les 48 observations. On aura donc environ 2% de valeurs aberrantes soit 6 par séries simulées de 300 observations.

On peut alors voir le rapprochement avec les effets annuels comme la crue de printemps, la fonte des neiges, etc, qui affectent une ou deux observations hebdomadaires à chaque année. Il est parfois contesté de traiter ces observations comme valeurs aberrantes puisqu'elles ne sont pas aberrantes au sens strict du terme. Cependant, lorsque l'intérêt de l'étude de détection de tendance se porte sur les 98% d'information autre que les crues ou la fonte, il est important d'exclure au maximum l'effet de ces facteurs printaniers. On voit donc que les développements des sections suivantes s'appliquent à toutes les situations où le processus mesuré est composé de deux sous processus et que l'on est intéressé à détecter des tendances dans le sous processus comprenant la majorité de l'information sans être affecté par le sous-processus minoritaire (valeurs aberrantes, crues, fontes, etc).

3.3 Détection de tendance en présence de valeurs aberrantes (VA).

En absence d'autocorrélation, les deux types de valeurs aberrantes sont équivalents puisque la propagation des IO par la structure autorégressive est inexistante. Dans les deux cas (IO et AO) on se retrouve donc en présence de valeurs aberrantes dans un bruit blanc (observations indépendantes). Avant d'approfondir l'effet conjoint des valeurs aberrantes et de l'autocorrélation, la présente section étudie: 1) l'effet des valeurs aberrantes sur les estimateurs classiques et; 2) l'utilisation des M-Estimeurs pour contrer cet effet. Il est important de toujours garder en mémoire que notre intérêt est de détecter adéquatement les tendances présentes dans un **processus sous-jacent** sans être affecté par la présence de **valeurs aberrantes** quand la seule information disponible est le **processus mesuré**.

3.3.1 Effets des VA sur les estimateurs classiques.

Les effets des valeurs aberrantes sur les estimateurs classiques du maximum de vraisemblance ont été largement discutés dans la littérature. L'effet principal de la présence de valeurs aberrantes est d'augmenter considérablement la variance des processus mesurés. Comme les estimateurs classiques du maximum de vraisemblance estiment cette variance, ils sont donc biaisés par rapport à la variance du processus sous-jacent. Les valeurs aberrantes de moyenne nulle n'affectent pas les estimateurs classiques de la moyenne, cependant, lorsque les valeurs aberrantes ont une moyenne non nulle, la moyenne du processus mesuré est différente de la moyenne du processus sous-jacent. Encore une fois, comme les estimateurs classiques du maximum de vraisemblance estiment la moyenne du processus mesuré, ils sont biaisés par rapport à la moyenne du processus sous-jacent.

Voyons maintenant comment ces caractéristiques des estimateurs ponctuels affectent le test de détection de tendance à l'aide de la régression linéaire. La simulation 8, illustre l'effet des valeurs aberrantes (VA) sur les estimateurs classiques du maximum de

vraisemblance ainsi que sur le seuil du test de détection de tendance. Le tableau 3.3 présente les résultats numériques de cette simulation.

SIMULATION 8

CONDITIONS EXPERIMENTALES

Titre : Étude de seuil en présence de valeurs aberrantes (VA).

Estimateurs étudiés : b_1 , *REQM*, seuil.

Nombre de séries simulées : 500 X **Longueur des séries :** 300.

Modèle sous-jacent : $N(0, 1)$, $\Delta_m = 0$.

Modèle de VA : $\mu_{VA} = 5$, $\sigma_{VA} = 2$, $t_{VA} = 2, 50, 98, 146, 194, 242$.

Modèle de régression utilisé : $Y_t = \beta_0 + \beta_1 X_t + e_t$; $X_t = t$.

Tableau 3.3

Résultats numériques de la simulation 8; Seuil de la régression non robuste en présence de valeurs aberrantes.

Estimateur	Moyenne	Ecart type	Précision
<i>REQM</i>	1,25	0,07	0,0065
b_1	-0,0004	0,0007	0,0001

$$\hat{\alpha} = \text{Seuil empirique} = 0,042$$

Notons d'abord qu'on retrouve une valeur aberrante au début de la série (2^e observation), alors qu'il n'y en a aucune à la fin (dans les 50 dernières observations). Cet emplacement des VA explique la valeur moyenne de b_1 , significativement plus petite que 0. On remarque également que le *REQM* moyen n'est pas significativement différent de l'écart type du processus mesuré 1,24 (réf. tableau 3.2) tout en étant biaisé par rapport à l'écart type du processus sous-jacent. En ce qui concerne le seuil empirique, il n'est pas significativement différent de 5%. Cependant, le peu de valeurs aberrantes et l'amplitude moyenne des valeurs aberrantes peuvent expliquer l'égalité statistique de ces deux seuils. De plus, l'effet inverse des biais des deux estimateurs ponctuels peut également expliquer

cette égalité. Le biais sur la pente amène à détecter plus de tendances, alors que le biais sur la variance amène à détecter moins de tendances.

La simulation 8b, identique à la simulation 8 sauf pour $\mu_{VA} = 10$ et $\sigma_{VA} = 3$, permet de compléter cette étude de seuil, on note alors que : 1) le b_1 moyen (-0,0007) est de nouveau négatif et significatif; 2) le *REQM* moyen (1,78) n'est pas significativement différent de l'écart type du processus mesuré (1,77) et; 3) le seuil empirique (0,018) est significativement plus petit que le seuil théorique (5%). Le faible seuil empirique implique que l'on ne rejette pas assez souvent l'hypothèse H_0 , ce résultat montre que la surestimation de la variance prend ici plus d'importance que l'effet de l'emplacement des VA.

La simulation 9 présente une étude de puissance du test de détection de tendance avec les estimateurs classiques en présence de valeurs aberrantes (VA), alors que le tableau 3.4 présente les résultats numériques de cette simulation. On note alors que: 1) le *REQM* moyen est significativement différent de l'écart type du processus sous-jacent, puisqu'il converge vers l'écart type du processus mesuré (1,24, réf. tableau 3.2); 2) le b_1 moyen (0,0016) est significativement plus petit que la tendance introduite dans les séries (0,002) et; 3) la puissance empirique (0,462) est beaucoup plus faible que la puissance théorique (0,85, réf. tableau 1.1).

SIMULATION 9

CONDITIONS EXPERIMENTALES

Titre : Étude de puissance en présence de valeurs aberrantes (VA).

Estimateurs étudiés : b_1 , *REQM*, puissance.

Nombre de séries simulées : 500 X **Longueur des séries :** 300.

Modèle sous-jacent : $N(0, 1), \Delta_m = 0,002$.

Modèle de VA : $\mu_{VA} = 5, \sigma_{VA} = 2, t_{VA} = 2, 50, 98, 146, 194, 242$.

Modèle de régression utilisé : $Y_t = \beta_0 + \beta_1 X_t + e_t, X_t = t$.

Tableau 3.4

Résultats numériques de la simulation 9; Puissance de la régression non robuste en présence de valeurs aberrantes.

Estimateur	Moyenne	Ecart type	Précision
<i>REQM</i>	1,27	0,08	0,0067
b_1	0,0016	0,0007	0,00006

Puissance empirique = 0,462

Tout comme dans l'étude de seuil, l'emplacement des valeurs aberrantes joue un rôle important sur les résultats de cette simulation. En effet, comme la tendance introduite est positive $\Delta_m = 0,002$ et qu'une valeur aberrante positive se trouve au début de la série, cette dernière élimine partiellement la tendance et amène le plus faible pourcentage de rejets de H_0 observé au tableau 3.4.

En conclusion, on voit que la présence de valeurs aberrantes affecte grandement les tests de détection de tendance lorsqu'on utilise les estimateurs classiques du maximum de vraisemblance. Il est donc primordial d'utiliser une méthode statistique beaucoup moins influencée par ces valeurs atypiques. Les prochaines simulations étudient les caractéristiques (seuil et puissance) des tests de détection de tendance avec l'utilisation des M-Estimeurs.

3.3.2 Étude de seuil avec l'utilisation des M-Estimeurs.

La section précédente a permis de mettre en évidence la forte influence des valeurs aberrantes et de leur emplacement sur le seuil du test de détection de tendance. La simulation 10 étudie ce même seuil, mais avec l'utilisation des M-Estimeurs dans l'estimation des différents paramètres de la régression multiple. Le tableau 3.5 présente les résultats de cette simulation.

SIMULATION 10

CONDITIONS EXPERIMENTALES

Titre : Étude de seuil, en présence de VA, avec utilisation des M-Estimeurs.

Estimateurs étudiés : b_{1M} , $\hat{\sigma}_M$, seuil.

Nombre de séries simulées : 500 X **Longueur des séries :** 300.

Modèle sous-jacent : $N(0,1)$, $\Delta_m = 0$.

Modèle de VA : $\mu_{VA} = 5$, $\sigma_{VA} = 2$, $t_{VA} = 2, 50, 98, 146, 194, 242$

Modèle de régression utilisé : $Y_t = \beta_0 + \beta_1 X_t + e_t$, $X_t = t$.

Tableau 3.5

Résultats numériques de la simulation 10; Seuil de la régression robuste en présence de valeurs aberrantes.

Estimateur	Moyenne	Ecart type	Précision
b_{1M}	-0,00013	0,0008	0,00007
$\hat{\sigma}_M$	1,035	0,073	0,006

$$\hat{\alpha} = \text{Seuil empirique} = 0,076$$

En comparant ces résultats, avec ceux du tableau 3.3, on remarque que : 1) le b_{1M} moyen (-0,00013) est plus près de 0 que b_1 moyen (-0,0004); 2) le $\hat{\sigma}_M$ moyen (1,035) est beaucoup plus près de 1 (valeur théorique de l'écart type) que le *REQM* moyen (1,25) et ; 3) dans les deux cas, le seuil empirique n'est pas significativement différent de 5%.

Cependant, la simulation 10b identique à la simulation 10 sauf pour $\mu_{VA} = 10$ et $\sigma_{VA} = 3$, montre également un seuil empirique (0,058) qui n'est pas significativement différent du seuil théorique (5%). Avec de telles valeurs aberrantes, la simulation 8b a montré que le seuil empirique du test de détection de tendance utilisant les estimateurs classiques (0,018) était significativement plus petit que 5%.

Ces résultats permettent de conclure que l'utilisation d'estimateurs robustes d'Huber permet de diminuer significativement l'influence des valeurs aberrantes: 1) Biais significatif pour l'estimateur robuste de la pente, mais plus faible que pour l'estimateur classique et; 2) Seuil empirique non significativement différent de 5% avec l'utilisation des M-Estimeurs et significativement plus faible que 5% pour les estimateurs classiques en présence de valeurs aberrantes de grande amplitude.

3.3.3 Étude de puissance avec l'utilisation des M-Estimeurs.

Les simulations 11, 11b et 11c étudient la puissance du test de détection de tendance en présence de valeurs aberrantes avec l'utilisation des M-Estimeurs. La présentation des résultats de ces trois simulations permet d'avoir une idée d'ensemble de la puissance du test de détection de tendance avec les M-Estimeurs. La différence entre les trois simulations réside au niveau de la moyenne et de la variance des VA. Pour les simulations 11, 11b et 11c les μ_{VA} sont respectivement 10, 5 et 0, alors que les σ_{VA} sont respectivement 3,2 et 3. Les tableaux 3.6, 3.7 et 3.8 présentent les résultats numériques associés à ces trois simulations, on peut alors conclure que :

1. La puissance empirique avec l'utilisation des M-Estimeurs est significativement plus faible que la puissance théorique (estimateurs classiques dans le cas d'observations indépendantes et identiquement distribuées $N(0,1)$);
2. En présence de VA de moyenne positive au début de la série, la puissance empirique avec l'utilisation des M-Estimeurs est significativement plus élevée que la puissance empirique avec l'utilisation des estimateurs classiques du maximum de vraisemblance. Lorsqu'on a pour les M-Estimeurs des puissances empiriques 0,708 et 0,746, les estimateurs classiques présentent des puissances empiriques de 0,462 et 0,072 dans les mêmes situations.

SIMULATION 11

CONDITIONS EXPERIMENTALES

Titre : Étude de puissance, en présence de VA, avec l'utilisation des M-Estimeurs.

Estimateurs étudiés : b_{1M} , $\hat{\sigma}_M$, puissance.

Nombre de séries simulées : 500 X **Longueur des séries :** 300.

Modèle sous-jacent : $N(0,1)$ avec $\Delta_m = 0,002$.

Modèle de VA : $\mu_{VA} = 5$, $\sigma_{VA} = 2$, $t_{VA} = 2, 50, 98, 146, 194, 242$

Modèle de régression utilisé : $Y_t = \beta_0 + \beta_1 X_t + e_t$, $X_t = t$.

Tableau 3.6

Résultats numériques de la simulation 11; Puissance de la régression robuste en présence de valeurs aberrantes: $\mu_{VA}=5$, $\sigma_{VA}=2$.

Estimateur	Moyenne	Ecart type	Précision
b_{1M}	0,00188	0,0007	0,00007
σ_M	1,029	0,074	0,007

Puissance empirique = 0,708

Tableau 3.7

Résultats numériques de la simulation 11b; Puissance de la régression robuste en présence de valeurs aberrantes: $\mu_{VA}=10$, $\sigma_{VA}=3$.

Estimateur	Moyenne	Ecart type	Précision
b_{1M}	0,00187	0,0007	0,00006
σ_M	1,034	0,076	0,007

Puissance empirique = 0,746

Tableau 3.8

Résultats numériques de la simulation 10c; Puissance de la régression robuste en présence de valeurs aberrantes: $\mu_{VA}=0$, $\sigma_{VA}=3$.

Estimateur	Moyenne	Ecart type	Précision
b_{1M}	0,00202	0,0007	0,00006
σ_M	1,020	0,076	0,007

Puissance empirique = 0,802

3. En présence de VA de moyenne nulle, la puissance empirique avec l'utilisation des M-Estimeurs (0,802) n'est pas significativement différente de la puissance empirique avec l'utilisation des estimateurs classiques du maximum de vraisemblance (0,828).

4. En présence de VA de moyenne nulle, l'estimateur robuste de la pente de la tendance n'est pas significativement différent de la valeur introduite dans les séries simulées (0,002). En présence de VA de moyenne non nulle, les estimateurs robustes de la pente sont significativement plus petits que 0,002. Cet écart significatif est introduit par l'emplacement des VA; on remarque cependant que le biais est beaucoup moins grand que celui des estimateurs classiques.

5. Finalement, l'estimation robuste de σ donne des estimateurs significativement différents de la valeur théorique (1,0). Cependant, les valeurs obtenues 1,029, 1,034 et 1,020 sont beaucoup plus près de 1 que les estimateurs classiques qui sont respectivement 1,273, 1,821 et 1,080. Rappelons que, dans ce dernier cas, les estimateurs classiques estiment plutôt l'écart type du processus mesuré.

En présence de valeurs aberrantes, il apparaît donc que les estimateurs robustes permettent une détection de tendanc tout à fait adéquate, puisque le seuil est conversé (5%) et la puissance, bien que significativement plus faible que la puissance théorique (estimateurs classiques en présence d'observations indépendantes et identiquement distribuées $N(0,1)$), demeure près de cette dernière. On peut donc avoir une bonne idée de la puissance du test de détection de tendance avec les M-Estimeurs, puisqu'elle varie peu selon l'amplitude et l'emplacement des valeurs aberrantes, ce qui n'est pas le cas des estimateurs classiques du maximum de vraisemblance.

En ce qui concerne les estimateurs robustes de la pente de la tendance, ils sont significativement influencés par l'emplacement des valeurs aberrantes, mais ils le sont beaucoup moins que les estimateurs classiques. Ces estimateurs peuvent donc être très utiles dans les applications où l'on s'intéresse à la tendance du processus sous-jacent; par exemple, lorsqu'on s'intéresse à la tendance d'une série de charges de sulfates transportées dans une rivière, mais qu'on désire que les crues printanières n'influencent pas notre détection de tendance.

3.4 Détection de tendance en présence d'IO.

Dans la section précédente, il a été question de la présence de valeurs aberrantes dans des bruits blancs (absence d'autocorrélation). En présence de valeurs aberrantes et d'autocorrélation, il faut alors distinguer les deux types de valeurs aberrantes discutés à la section 3.2, IO et AO. Il faut donc remarquer que la discussion de ces deux types de valeurs aberrantes implique la discussion de la présence d'autocorrélation.

En présence d'IO, plusieurs études ont montré que les M-Estimeurs sont de bons estimateurs des paramètres des séries chronologiques (Denby et Martin 1979, Martin et Yohai 1985, Laberge 1988). L'utilisation de ces estimateurs robustes devrait donc permettre une détection de tendance adéquate en présence de valeurs aberrantes dynamiques. Les simulations de la présente section permettent de valider cette affirmation tout en illustrant les avantages par rapport aux estimateurs classiques du maximum de vraisemblance.

3.4.1 Effet des IO sur les estimateurs classiques.

L'effet des IO sur les estimateurs classiques du maximum de vraisemblance est présenté dans Laberge (1988): L'effet principal de la présence de valeurs aberrantes dynamiques est d'augmenter considérablement la variance des processus mesurés. Les estimateurs classiques du maximum de vraisemblance estiment cette variance, ils sont donc biaisés par rapport à la variance du processus sous-jacent.

Les IO de moyenne nulle n'affectent pas les estimateurs classiques de la moyenne. Cependant, lorsque les IO ont une moyenne non nulle, la moyenne du processus mesuré est différente de la moyenne du processus sous-jacent, ce qui amène un biais par rapport à la moyenne du processus sous-jacent. Finalement, l'estimateur classique de ϕ , le coefficient d'autorégression, est peu affecté par la présence de IO. Voyons maintenant comment ces caractéristiques des estimateurs ponctuels affectent le test de détection de tendance à l'aide de la régression linéaire. La simulation 12 et le tableau 3.9 étudient le seuil de ce test, en présence d'IO.

SIMULATION 12

CONDITIONS EXPERIMENTALES

Titre : Étude de seuil en présence d'IO et d'autocorrélation.

Estimateurs étudiés : b_1 , d_1 , *REQM*, seuil.

Nombre de séries simulées : 500 X **Longueur des séries :** 300.

Modèle sous-jacent : AR(1) avec $\phi = 0,75$, $\Delta_m = 0$, $\sigma^2 = 1$.

Modèle de IO : $\mu_{IO} = 5$, $\sigma_{IO} = 2$, $t_{IO} = 45, 93, 141, 189, 237, 285$.

Modèle de régression utilisé : $Y_t = \beta_0 + \beta_1 X_t + \phi_1 Y_{t-1} + e_t$, $X_t = t$.

Tableau 3.9
 Résultats numériques de la simulation 12; Seuil de la
 régression non robuste en présence d'IO.

Estimateur	Moyenne	Ecart type	Précision
<i>REQM</i>	1,252	0,07	0,0062
b_1	0,00023	0,0008	0,00007
d_1	0,726	0,04	0,003

$$\hat{\alpha} = \text{Seuil empirique} = 0,046$$

Les estimateurs moyens des paramètres de localisation (0,00023) et d'échelle (1,252) sont significativement différents des valeurs associées au processus sous-jacent (respectivement 0 et 1), mais le seuil empirique n'est pas significativement différent de 5%. Une seconde simulation avec des valeurs aberrantes dynamiques de plus grande amplitude ($\mu_{IO} = 10$ et $\sigma_{IO} = 3$) montre une différence significative entre le seuil empirique (0,016) et le seuil théorique (0,05), alors qu'une autre simulation avec $\mu_{IO} = 10$ et $\sigma_{IO} = 3$ mais avec la première valeur aberrante au temps $t_0=2$ ne montre pas de différence significative entre le seuil empirique (0,04) et le seuil théorique. Ces résultats sont tout à fait conformes aux résultats obtenus en présence de valeurs aberrantes seulement (réf. section 3.3.1):

1. Les estimateurs classiques estiment les paramètres du processus mesuré. Ils sont donc significativement différents des valeurs des paramètres du processus sous-jacent.
2. Le seuil dépend de l'amplitude et de l'emplacement des valeurs aberrantes. Certaines caractéristiques amènent plus de rejet (p.e. localisation des IO aux extrémités) alors que d'autres amènent moins de rejet (p.e. surestimation de la variance), les effets inverses s'annulent donc pour amener, **parfois**, un seuil égal au seuil théorique.

Donc, comme dans le cas de valeurs aberrantes seulement, il est hasardeux d'utiliser les estimateurs classiques pour la détection de tendance.

Le tableau 3.10 étudie la puissance de ce test dans les mêmes conditions expérimentales que la simulation 12, sauf pour la présence d'une tendance monotone d'amplitude $\Delta_m=0,002$.

Tableau 3.10
Résultats numériques de la simulation 12; Puissance de la régression non robuste en présence d'IO.

Estimateur	Moyenne	Ecart type	Précision
<i>REQM</i>	1,248	0,07	0,0066
b_1	0,00238	0,0009	0,00008
d_1	0,723	0,04	0,0035

Puissance empirique = 0,772

La comparaison de ces résultats avec ceux de la section 3.3.1 montre que l'introduction d'autocorrélation en plus de la présence d'IO n'amène pas de problème majeur supplémentaire aux estimateurs classiques. Les estimateurs classiques n'en demeurent pas moins inutiles pour la détection de tendance en présence de valeurs aberrantes. L'augmentation de la puissance empirique (0,772) comparativement à la puissance présentée à la section 3.3.1 (0,462) est exclusivement due à l'emplacement des valeurs aberrantes. En effet, à la section 3.3.1 les valeurs aberrantes sont au début des séries (moins de rejets de H_0) alors que dans la simulation 11 les valeurs aberrantes sont à la fin des séries (plus de rejets de H_0).

Les sections 3.4.2 et 3.4.3 étudient maintenant la détection de tendance, à l'aide des estimateurs robustes, en présence d'IO.

3.4.2 Étude de seuil avec l'utilisation des M-Estimeurs.

La section 3.3.2 a montré qu'en présence de valeurs aberrantes, l'utilisation des M-Estimeurs permet de conserver le seuil théorique de 5% pour la détection de tendance, ce qui n'est pas le cas des estimateurs classiques qui amènent un seuil variant selon l'amplitude et l'emplacement des valeurs aberrantes. La simulation 13 étudie le seuil du test

utilisant les estimateurs robustes en présence d'IO. Afin de tenir compte de l'autocorrélation, les développements du chapitre 2 sont mis à profit puisqu'on utilise une variable avec retard Y_{t-1} . Le tableau 3.11 présente les résultats numériques associés à cette simulation.

SIMULATION 13

CONDITIONS EXPERIMENTALES

Titre : Étude de seuil en présence d'IO (M-Estimeurs).

Estimateurs étudiés : b_{1M} , $\hat{\sigma}_{1M}$, d_{1M} , *seuil*.

Nombre de séries simulées : 500 X **Longueur des séries :** 300.

Modèle sous-jacent : AR(1) avec $\phi_1 = 0,75, \sigma^2 = 1$.

Modèle de IO : $\mu_{IO} = 5, \sigma_{IO} = 2, t_{IO} = 45, 93, 141, 189, 237, 285$.

Modèle de régression utilisé : $Y_t = \beta_0 + \beta_1 X_t + \phi_1 Y_{t-1} + e_t, X_t = t$.

Tableau 3.11

Résultats numériques de la simulation 13; Seuil de la régression robuste en présence d'IO.

Estimateur	Moyenne	Ecart type	Précision
b_{1M}	0,00007	0,0008	0,00007
$\hat{\sigma}_M$	1,032	0,074	0,006
d_{1M}	0,737	0,038	0,0033

$$\hat{\alpha} = \text{Seuil empirique} = 0,068$$

L'estimateur robuste de la pente n'est donc pas significativement différent de 0, l'estimateur du paramètre d'échelle est légèrement plus grand que 1 et le seuil n'est pas significativement différent du seuil théorique de 5%. Ces résultats sont confirmés par des simulations utilisant d'autres amplitudes et emplacements de valeurs aberrantes. Le seuil est donc conservé lorsqu'on utilise la régression linéaire, une variable avec retard et les

M-Estimeurs pour détecter des tendances en présence d'IO. La section 5.1 sur la comparaison avec le test de Spearman présente les autres résultats de simulations qui permettent d'appuyer cette affirmation.

3.4.3 Étude de puissance avec l'utilisation des M-Estimeurs.

De nouveau, on cherche à savoir si l'introduction d'autocorrélation (en plus de la présence de valeurs aberrantes dynamiques) affecte les résultats du test de détection de tendance utilisant la régression linéaire et les M-Estimeurs, mais cette fois on s'attarde à la puissance du test. La section 3.3.3 avait montré qu'en présence de valeurs aberrantes seulement, la puissance de ce test était légèrement plus faible que la puissance théorique. Les simulations 14, 14b et 14c permettent une vue d'ensemble de la puissance du test utilisant les estimateurs robustes en présence d'IO.

SIMULATION 14

CONDITIONS EXPÉRIMENTALES

Titre : Étude de puissance en présence d'IO (M-Estimeurs).

Estimateurs étudiés : b_{1M} , $\hat{\sigma}_M$, d_{1M} , *puissance*.

Nombre de séries simulées : 500 X **Longueur des séries :** 300.

Modèle sous-jacent : AR(1) avec $\phi = 0,75$, $\Delta_m = 0,002$, $\sigma^2 = 1$.

Modèle de IO : $\mu_{IO} = 5$, $\sigma_{IO} = 2$, $t_{IO} = 45, 93, 141, 189, 237, 285$

Modèle de régression utilisé : $Y_t = \beta_0 + \beta_1 X_t + \phi_1 Y_{t-1} + e_t$, $X_t = t$

Les simulations 14b et 14c ont les mêmes conditions expérimentales que la simulation 14 sauf pour les modèles des IO qui ont alors $\mu_{IO} = 10$ et $\sigma_{IO} = 3$, pour la simulation 14b et $\mu_{IO} = 0$ et $\sigma_{IO} = 3$, pour la simulation 14c. Les tableaux 3.12, 3.13 et 3.14 présentent les résultats de ces trois simulations.

Tableau 3.12

Résultats numériques de la simulation 14; Puissance de la régression robuste en présence d'IO, $\mu_{IO}=5$, $\sigma_{IO}=2$.

Estimateur	Moyenne	Ecart type	Précision
b_{1M}	0,00215	0,0009	0,00008
$\hat{\sigma}_M$	1,036	0,076	0,007
d_{1M}	0,734	0,036	0,0032

Puissance empirique = 0,754

Tableau 3.13

Résultats numériques de la simulation 14b; Puissance de la régression robuste en présence d'IO, $\mu_{IO}=10$, $\sigma_{IO}=3$.

Estimateur	Moyenne	Ecart type	Précision
b_{1M}	0,00215	0,0008	0,00007
$\hat{\sigma}_M$	1,037	0,076	0,007
d_{1M}	0,741	0,026	0,0023

Puissance empirique = 0,790

Tableau 3.14

Résultats numériques de la simulation 14c; Puissance de la régression robuste en présence d'IO, $\mu_{IO}=10$, $\sigma_{IO}=3$.

Estimateur	Moyenne	Ecart type	Précision
b_{1M}	0,00214	0,0008	0,00007
$\hat{\sigma}_M$	1,025	0,074	0,007
d_{1M}	0,735	0,038	0,0034

Puissance empirique = 0,766

Les estimateurs de la pente sont légèrement surestimés à cause de la sous-estimation du coefficient d'autorégression et de l'emplacement des valeurs aberrantes. Les estimateurs du paramètre d'échelle sont également surestimés; cette surestimation est cependant

causée par la plus grande variation des valeurs aberrantes. Dans le cas des coefficients d'autorégression, l'estimateur robuste est significativement plus petit que 0,75.

Les puissances empiriques obtenues dans la section 3.3.3 en présence de valeurs aberrantes seulement sont respectivement de 0,708, 0,746 et 0,802. En présence d'IO de moyenne non nulle, les puissances empiriques sont donc un peu plus élevées dans la présente section. Cette légère augmentation est introduite par l'emplacement des IO à la fin des séries (implique un plus grand nombre de rejets de H_0) comparativement au début des séries dans la section 3.3.3. Dans le cas d'IO de moyenne nulle (simulation 14c), la puissance est légèrement plus faible en présence d'IO comparativement à la section 3.3.3. Dans ce dernier cas ($\mu_{IO} = 0$), l'emplacement des valeurs aberrantes a moins d'importance, puisqu'elles ne sont pas obligatoirement positives, c'est ce qui explique l'absence d'augmentation de puissance comme dans le cas des IO de moyenne non nulle.

Il apparaît donc que le test de détection de tendance utilisant la régression linéaire et les M-Estimeurs présente une puissance plus faible que la puissance théorique. Cependant, cette puissance théorique peut être impossible à atteindre en présence d'IO, il sera donc beaucoup plus intéressant de comparer la puissance de ce nouveau test avec la puissance des tests généralement utilisés dans le domaine de l'eau. La comparaison avec la puissance du test de Spearman/Lettenmaier est effectuée au chapitre 5. Il est aussi apparu que la puissance du test utilisant les M-Estimeurs est sensiblement affectée par l'emplacement des valeurs aberrantes. La comparaison du chapitre 5, permettra également de voir si le test de Spearman/Lettenmaier réagit de la même façon à la localisation des valeurs aberrantes.

3.5 Détection de tendance en présence de AO.

Comme dans le cas des IO, il est bon de rappeler que la présence d'AO implique la présence d'autocorrélation. Il est donc nécessaire d'inclure à nouveau une variable avec retard dans les modèles de régression afin de tenir compte de cette autocorrélation.

En présence d'AO, des études ont montré que l'estimation des paramètres des séries chronologiques à l'aide des M-Estimeurs présentait certaines difficultés (Denby et Martin 1979, Martin et Yohai 1985, Laberge 1988). Le principal problème est la sous-estimation, parfois importante, du coefficient d'autorégression. L'utilisation des M-Estimeurs ne permet donc pas nécessairement une détection de tendance adéquate en présence de valeurs aberrantes statiques. Les simulations de la présente section permettent de juger de la valeur de l'utilisation des M-Estimeurs, tout en illustrant les avantages par rapport aux estimateurs classiques du maximum de vraisemblance.

3.5.1 Effet des AO sur les estimateurs classiques.

L'effet des AO sur les estimateurs classiques du maximum de vraisemblance est présenté dans Laberge (1988); cet effet est assez semblable à l'effet des IO. Les AO augmentent considérablement la variance des processus mesurés. Les estimateurs classiques du maximum de vraisemblance estimant cette variance, ils sont donc biaisés par rapport à la variance du processus sous-jacent. Les AO de moyenne nulle n'affectent pas les estimateurs classiques de la moyenne. Lorsque les AO ont une moyenne non nulle, la moyenne du processus mesuré est différente de la moyenne du processus sous-jacent, ce qui amène un estimateur biaisé par rapport à la moyenne du processus sous-jacent.

Finalement, la différence importante avec la présence d'IO se retrouve au niveau de l'estimateur classique du coefficient d'autorégression qui est beaucoup plus affecté par la présence d'AO. Le fait que les AO ne suivent pas la structure autorégressive rend l'estimation du coefficient d'autorégression plus difficile.

Voyons maintenant comment les caractéristiques de ces estimateurs ponctuels (classiques) affectent le test de détection de tendance à l'aide de la régression linéaire. La simulation 15 étudie le seuil de ce test, en présence d'AO. Les résultats au tableau 3.15 illustrent tous les problèmes des estimateurs classiques en présence d'AO: i.e. surestimation de la variance du processus sous-jacent, sous-estimation du coefficient d'autorégression et estimation de la pente de la tendance affectée par l'emplacement des valeurs aberrantes. Comme certains effets inverses s'annulent, le seuil empirique n'est pas très loin du seuil théorique de 5%. Cependant, il est évident que dans certaines situations le seuil empirique sera très loin de 5%.

SIMULATION 15

CONDITIONS EXPERIMENTALES

Titre : Étude de seuil en présence d'AO (Estimateurs classiques).

Estimateurs étudiés : b_1 , d_1 , *REQM*, seuil.

Nombre de séries simulées : 500 X **Longueur des séries :** 300.

Modèle sous-jacent : AR(1) avec $\phi_1 = 0,75$

Modèle de AO : $\mu_{AO} = 5, \sigma_{AO} = 2, t_{AO} = 2, 50, 98, 146, 194, 242$

Modèle de régression utilisé : $Y_t = \beta_0 + \beta_1 X_t + \phi Y_{t-1} + e_t, X_t = t$

Tableau 3.15

Résultats numériques de la simulation 15; Seuil de la régression non robuste en présence d'AO.

Estimateur	Moyenne	Ecart type	Précision
<i>REQM</i>	1,356	0,10	0,0086
b_1	-0,00018	0,0011	0,00009
d_1	0,590	0,06	0,005

$$\hat{\alpha} = \text{Seuil empirique} = 0,094$$

Le tableau 3.16 étudie la puissance de ce test dans les mêmes conditions expérimentales que la simulation 15, sauf pour la présence d'une tendance monotone d'amplitude $\Delta_m = 0,002$ dans les impulsions. La surestimation du paramètre d'échelle amène moins de rejets de H_0 alors que la surestimation de la pente de la tendance et la sous-estimation du coefficient d'autorégression amènent plus de rejets de H_0 . La puissance empirique, obtenue à l'aide de ces mauvaises estimations ponctuelles, est assez près de la puissance théorique. Cependant, on peut difficilement croire que la puissance empirique est égale à la puissance théorique, puisque le seuil n'est pas conservé et que tous les estimateurs ponctuels, utilisés pour le test, sont fortement biaisés par rapport aux paramètres du processus sous-jacent.

Tableau 3.16
Résultats numériques de la simulation 15; Puissance de la régression non robuste en présence d'AO

Estimateur	Moyenne	Ecart type	Précision
<i>REQM</i>	1,469	0,104	0,0091
b_1	0,00352	0,0013	0,00011
d_1	0,543	0,055	0,0048

Puissance empirique = 0,882

Les sections 3.5.2 et 3.5.3 étudient l'utilisation des M-Estimeurs en présence d'AO afin de voir les améliorations par rapport aux estimateurs classiques et les limitations causées par la sous-estimation du coefficient d'autorégression.

3.5.2 Étude de seuil avec M-Estimeurs.

La simulation 16 présente une étude de seuil pour le test utilisant les M-Estimeurs en présence d'AO et d'autocorrélation. Le tableau 3.17 présente ensuite les résultats numériques associés à cette simulation. En comparant ces résultats numériques aux résultats de l'estimation classique (tableau 3.15), il est clair que les estimateurs robustes sont beaucoup plus près des paramètres du processus sous-jacent: l'estimateur robuste de

la pente (-0,00002) n'est pas significativement différent de 0, l'estimateur robuste du paramètre d'échelle (1,063) est plus près de 1 que l'estimateur classique (1,356) et l'estimateur robuste du coefficient d'autorégression (0,669) est plus près de 0,75 que l'estimateur classique (0,590). Malgré toutes ces améliorations au niveau des estimateurs ponctuels, il est surprenant de voir que le seuil obtenu à l'aide des estimateurs robustes est plus grand que le seuil obtenu à l'aide des estimateurs classiques, ce dernier étant déjà significativement plus grand que le seuil théorique de 5%. L'utilisation des M-Estimeurs ne permet donc pas un test adéquat en présence d'AO de moyenne non nulle.

SIMULATION 16

CONDITIONS EXPERIMENTALES

Titre : Étude de seuil en présence d'AO (M-Estimeurs) .

Estimateurs étudiés : b_{1M} , $\hat{\sigma}_M$, d_{1M} , *seuil*.

Nombre de séries simulées : 500 X **Longueur des séries :** 300.

Modèle sous-jacent : AR(1) avec $\phi_1 = 0,75$, $\sigma^2 = 1$.

Modèle de AO : $\mu_{AO} = 5$, $\sigma_{AO} = 2$, $t_{AO} = 2, 50, 98, 146, 194, 242$

Modèle de régression utilisé : $Y_t = \beta_0 + \beta_1 X_t + \phi Y_{t-1} + e_t$, $X_t = t$

Tableau 3.17

Résultats numériques de la simulation 16; Seuil de la régression robuste en présence d'AO: $\mu_{AO}=5$, $\sigma_{AO}=2$.

Estimateur	Moyenne	Ecart type	Précision
b_{1M}	-0,00002	0,0009	0,00008
$\hat{\sigma}_M$	1,063	0,076	0,007
d_{1M}	0,669	0,052	0,0045

$$\hat{\alpha} = \text{Seuil empirique} = 0,104$$

Le tableau 3.18 présente les résultats numériques de la simulation 16b, identique à la simulation 16 sauf pour la présence d'AO de moyenne nulle ($\mu_{AO} = 0$ et $\sigma_{AO} = 3$). On note alors que les M-Estimeurs sont beaucoup plus fiables qu'en présence d'AO de moyenne non nulle. En comparant les tableaux 3.17 et 3.18, on voit que l'estimateur de la pente est toujours non significativement différent de 0, alors que les estimateurs du paramètre d'échelle et du coefficient d'autorégression sont plus près des paramètres du modèle sous-jacent lorsque les AO ont une moyenne nulle. Finalement, le seuil empirique (0,069) n'est pas significativement différent du seuil théorique (0,05).

Tableau 3.18

Résultats numériques de la simulation 16b: Seuil de la régression robuste en présence d'AO de moyenne nulle: $\mu_{AO}=0$, $\sigma_{AO}=3$.

Estimateur	Moyenne	Ecart type	Précision
b_{1M}	-0,00002	0,0008	0,00007
$\hat{\sigma}_M$	1,040	0,073	0,006
d_{1M}	0,700	0,049	0,0043

$$\hat{\alpha} = \text{Seuil empirique} = 0,069$$

3.5.3 Étude de puissance avec M-Estimeurs.

La simulation 17 présente une étude de puissance pour le test utilisant les M-Estimeurs en présence d'AO. Comme les M-Estimeurs ne permettent pas de conserver le seuil théorique en présence d'AO de moyenne non nulle, la simulation 17 étudie la puissance en présence de moyenne nulle seulement. Le tableau 3.19 présente les résultats numériques associés à cette simulation.

L'estimateur robuste de la pente est significativement différent de la pente théorique (0.002). Cette différence importante est introduite par la sous-estimation du coefficient d'autocorrélation puisque la construction des séries se fait avec $\phi_1=0,75$ et l'introduction des

AO amène un estimateur de l'ordre de 0,70; la dépendance des deux estimateurs (b_{1M} et d_{1M}) amène l'estimateur de la pente à converger vers :

$$\frac{(1-d_{1M})}{(1-\phi)} \Delta_m = \frac{0,308}{0,25}(0,002) = 0,00246$$

Cette relation entre la pente théorique et la pente estimée selon la sous-estimation du coefficient d'autorégression est valable pour les estimateurs classiques également. Cette relation s'applique de plus dans le cas des études de seuil mais, comme la pente théorique est alors de 0, on s'explique facilement que la pente estimée n'est pas significativement différente de 0.

SIMULATION 17

CONDITIONS EXPERIMENTALES

Titre : Étude de puissance en présence d'AO (M-Estimeurs) .

Estimateurs étudiés : b_{1M} , $\hat{\sigma}_M$, d_{1M} , *puissance*.

Nombre de séries simulées : 500 X **Longueur des séries :** 300.

Modèle sous-jacent : AR(1) avec $\phi_1 = 0,75$ et $\Delta_m = 0,002$.

Modèle de AO : $\mu_{AO} = 0$, $\sigma_{AO} = 3$, $t_{AO} = 2, 50, 98, 146, 194, 242$

Modèle de régression utilisé : $Y_t = \beta_0 + \beta_1 X_t + \phi Y_{t-1} + e_t$, $X_t = t$

Tableau 3.19

Résultats numériques de la simulation 17; Puissance de la régression robuste en présence d'AO: $\mu_{AO}=0$, $\sigma_{AO}=3$.

Estimateur	Moyenne	Ecart type	Précision
b_{1M}	0,00251	0,0010	0,00009
$\hat{\sigma}_M$	1,040	0,077	0,007
d_{1M}	0,692	0,050	0,0044

Puissance empirique = 0,838

En ce qui concerne l'estimation du paramètre d'échelle et du coefficient d'autorégression, les estimateurs ponctuels présentent les mêmes caractéristiques que dans l'étude de seuil (légère surestimation du paramètre d'échelle et légère sous-estimation du coefficient d'autorégression).

Finalement, la puissance empirique du test (0,838) n'est pas significativement différente de la puissance théorique (0,85), il faut cependant être prudent avant de généraliser à cause des biais des estimateurs ponctuels. La section 5.1 sur la comparaison avec le test de Spearman/Lettenmaier permet de voir la puissance de ce dernier test dans les mêmes conditions.

3.6 Estimation robuste en absence de valeurs aberrantes.

Les sections précédentes montrent que l'utilisation des M-Estimeurs en présence de valeurs aberrantes améliore énormément la détection de tendance à l'aide de la régression multiple. Cependant, une utilisation abusive de ces estimateurs peut avoir des effets négatifs dans les études de détection de tendance. Afin d'illustrer ce phénomène la simulation 18 étudie le seuil et la puissance du test de détection de tendance avec les M-Estimeurs lorsqu'aucune valeur aberrante n'est présente dans les séries. Le tableau 3.20 présente les résultats numériques de cette simulation.

L'estimateur robuste de la pente converge vers la valeur théorique alors que l'estimateur robuste du paramètre d'échelle (1,005) n'est pas significativement différent de la valeur théorique (1). On note finalement que le seuil empirique (0,038) n'est pas significativement différent de la valeur théorique (0,05); une autre simulation dans les mêmes conditions expérimentales a donné un seuil empirique (0,056), encore une fois non significativement différent de la valeur théorique. L'utilisation des M-Estimeurs en absence de valeurs aberrantes n'affecte donc pas le seuil du test de détection de tendance.

SIMULATION 18

CONDITIONS EXPERIMENTALES

Titre : Étude de seuil : utilisation des M-Estimeurs en absence de VA .

Estimateurs étudiés : $b_{1M}, \hat{\sigma}_M$, *seuil*.

Nombre de séries simulées : 500 X **Longueur des séries** : 300.

Modèle sous-jacent : $N(0,1)$.

Modèle de VA : Aucun

Modèle de régression utilisé : $Y_t = \beta_0 + \beta_1 X_t + e_t$, $X_t = t$

Tableau 3.20

Résultats numériques de la simulation 18; Seuil de la régression robuste en absence de valeurs aberrantes.

Estimateur	Moyenne	Ecart type	Précision
b_{1M}	-0,00001	0,0007	0,00006
σ_M	1,005	0,076	0,007

$$\hat{\alpha} = \text{Seuil empirique} = 0,038$$

Le tableau 3.21 présente les résultats numériques d'une simulation dans les mêmes conditions que la simulation 18, sauf pour l'introduction d'une tendance monotone de pente 0,002.

Tableau 3.21

Résultats numériques de la simulation 18; Puissance de la régression robuste en absence de valeurs aberrantes.

Estimateur	Moyenne	Ecart type	Précision
b_{1M}	0,00196	0,0007	0,00006
$\hat{\sigma}_M$	1,005	0,074	0,007

$$\text{Puissance empirique} = 0,786$$

Cette étude de puissance permet de conclure que les estimateurs robustes de la pente de la tendance (0,00196) et du paramètre d'échelle (1,005) ne sont pas significativement différents des valeurs théoriques (0,002 et 1). On remarque cependant une légère diminution significative de la puissance empirique (0,786) du test de détection de tendance par rapport à la puissance théorique. Cette puissance théorique est la puissance du test utilisant les estimateurs classiques dans les conditions expérimentales de la simulation 18. On remarque cependant que la puissance du test utilisant les M-Estimeurs en absence de valeurs aberrantes est supérieure ou égale à la puissance du même test en présence de valeurs aberrantes.

Donc, l'utilisation des M-Estimeurs en absence de valeurs aberrantes n'amène pas une diminution de puissance par rapport à l'utilisation en présence de valeurs aberrantes. Cependant, en absence de valeurs aberrantes, les estimateurs classiques donnent un test plus puissant.

3.7 Discussion sur la présence de valeurs aberrantes.

En présence de valeurs aberrantes seulement (pas d'autocorrélation), l'utilisation des M-Estimeurs permet une détection de tendance adéquate, puisque le seuil théorique est conservé et la puissance, bien que significativement plus faible que la puissance théorique, demeure stable pour des amplitudes et des localisations différentes des valeurs aberrantes. Les M-Estimeurs permettent donc un test beaucoup plus fiable que le test utilisant les estimateurs classiques.

En présence d'IO (et d'autocorrélation), l'utilisation des M-Estimeurs permet également une détection de tendance adéquate: le seuil théorique est conservé et la puissance, encore significativement plus faible que la puissance théorique, demeure stable pour des amplitudes et des localisations différentes des valeurs aberrantes. L'introduction d'autocorrélation n'affecte donc pas la détection de tendance pourvu qu'on utilise une variable avec retard dans le modèle de régression multiple et que l'effet des valeurs

aberrantes est propagé par la structure autorégressive (IO). Les M-Estimeurs permettent donc un test beaucoup plus fiable que le test utilisant les estimateurs classiques. La comparaison du chapitre 5 permet de voir les différences avec le test de Spearman/Lettenmaier.

En présence d'AO, les M-Estimeurs ne permettent pas une détection de tendance adéquate (le seuil est significativement plus grand que 5%). Il est nécessaire de trouver une nouvelle méthode permettant d'annuler le biais important de l'estimateur du coefficient d'autorégression si l'on veut détecter adéquatement des tendances en présence de ce type de valeurs aberrantes. La section 5.1 sur la comparaison avec le test de Spearman/Lettenmaier permettra de voir si ce dernier test a davantage de succès en présence d'AO.

Il est généralement recommandé d'utiliser conjointement les estimateurs robustes et les estimateurs classiques. Les conclusions qu'on peut alors tirer sont beaucoup plus éclairées. En effet, lorsque les deux méthodes présentent des estimateurs ponctuels équivalents, ainsi que des conclusions semblables pour la détection de tendance, on peut alors conclure que les valeurs extrêmes ne sont pas aberrantes et qu'elles n'affectent pas la détection de tendance. Si les deux méthodes n'amènent pas la même conclusion pour la détection de tendance, le bon jugement permet généralement de choisir la bonne conclusion. On fera alors confiance aux estimateurs robustes si la présence de valeurs aberrantes amène des estimateurs ponctuels très différents d'une méthode à l'autre, alors qu'on fera confiance aux estimateurs classiques si les valeurs extrêmes n'affectent pas les estimateurs ponctuels (i.e. estimateurs robustes et classiques non significativement différents), les conclusions différentes seraient alors introduites par la plus faible puissance du test utilisant les estimateurs robustes.

L'identification du type de valeur aberrante (AO ou IO) peut se faire à l'aide des méthodes développées par Abraham et Chuang (1989) ou Chang, Tiao et Chen (1988).

Cependant ces méthodes sont assez encombrantes à utiliser lorsque le nombre d'observations aberrantes est grand, ou lorsque plusieurs paramètres sont étudiés. Il est cependant possible d'avoir une bonne idée du type de valeur aberrante en se basant sur les observations suivantes:

1. La présence d'AO amène généralement une différence significative entre les estimateurs classiques et robustes du coefficient d'autorégression.
2. Dans la série des résidus de la régression linéaire classique utilisant une variable avec retard, les IO amènent généralement un seul résidu aberrant suivi d'un résidu normal, alors que les AO amènent généralement deux résidus aberrants consécutifs de signes opposés. Ce résultat vient de la propagation ou non des valeurs aberrantes par la structure autorégressive. Pour les IO, la propagation par la structure autorégressive amène l'observation suivant la valeur aberrante à demeurer environ au même niveau que la valeur aberrante, ce qui amène un résidu normal. Pour les AO, la non propagation par la structure autorégressive amène l'observation suivant la valeur aberrante à retourner au niveau du reste de la série amenant un grand résidu de signe opposé.

Lorsqu'on a une bonne idée du type de valeur aberrante présente dans la série, la connaissance des propriétés des tests utilisant les estimateurs classiques et robustes permet des conclusions éclairées. Le chapitre 5 permet également de juger de la valeur du test de Spearman/Lettenmaier selon le type de valeur aberrante. Cependant, lorsque le type de valeur aberrante est complètement incertain, il faut être très prudent avec les conclusions de l'étude à cause de l'absence de test valable en présence d'AO.

CHAPITRE 4

PRÉSENCE DE SAISONNALITÉS

4.1 Effet des saisonnalités et traitements possibles.

L'étude de séries chronologiques en présence de saisonnalités est particulièrement importante dans le domaine de l'économie. On a qu'à penser aux différents indicateurs économiques (taux d'inflation, taux de chômage, ventes au détail, etc) qui contiennent une composante saisonnière. Cette composante doit être prise en compte afin de pouvoir porter un jugement valable sur l'évolution du niveau de ces séries. En effet, la présence de saisonnalités amène des changements de niveaux fréquents et de courte durée; ces fluctuations à court terme doivent être prises en compte pour permettre une étude valable de l'évolution à long terme (tendance) d'une série. De façon générale, les saisonnalités peuvent masquer la présence de tendance à long terme ou de valeurs aberrantes, elles peuvent également introduire des tendances fictives en amenant, par exemple, des hautes valeurs au début d'une série et des valeurs basses à la fin de la même série.

Lorsque les causes de saisonnalités et les dates de changement des conditions sont connues (p.e. ventes de jouet le mois précédent Noël, taux de chômage dans les Maritimes à la fin de la période de pêche, etc), l'extraction de la composante saisonnière est une tâche

assez simple. Lorsque les causes de saisonnalités sont connues mais que les dates de changement des conditions ne sont pas connues (p.e. crues printanières des rivières et la majorité des phénomènes naturels), l'extraction des saisonnalités devient plus compliquée bien qu'en général on puisse avoir une bonne idée du début et de la fin des cycles saisonniers. Finalement, lorsque ni les causes de saisonnalités, ni les dates de changement des conditions ne sont connues, l'extraction des saisonnalités devient une tâche assez complexe (p.e. nombre de meurtres par mois, (McKIE, 1985)).

Pierce (1980) présente un recensement des nouveaux développements pour l'extraction des saisonnalités. Bien que cet article ne soit pas très récent, il discute de la majorité des méthodes encore utilisées pour l'extraction des saisonnalités soit: 1) l'ajustement à l'aide des modèles ARIMA (Box et Jenkins 1976); 2) l'ajustement à l'aide de moyennes mobiles (p.e. Census X-11, Shiskin et al. 1967) et; 3) l'extraction des saisonnalités à l'aide de variables muettes dichotomiques.

Dans les volumes de référence en économétrie (p.e. Wonnacott et Wonnacott 1980, Kmenta, J., 1971) on discute brièvement des méthodes d'extraction des saisonnalités. On propose alors l'utilisation de variables muettes ("dummy") dichotomiques et/ou l'utilisation des filtres de moyennes mobiles.

Finalement, dans le domaine de l'eau, où le but principal derrière l'extraction des saisonnalités est généralement la détection de tendance. Hirsch, Slack et Smith (1982) ont développé un test non paramétrique saisonnier qui permet de tenir compte de la présence de saisonnalités. Ces auteurs utilisent le test de Mann-Kendall pour chacune des saisons pour ensuite faire la somme des statistiques obtenues afin d'obtenir un test pour l'ensemble de la série. Comme chaque statistique mensuelle a une distribution qui tend vers la loi normale, on peut alors facilement déterminer la distribution limite de la somme des statistiques. Étant donné ces avantages importants en présence de caractéristiques restrictives des séries

chronologiques dans le domaine de l'eau, ce test est largement utilisé pour la détection de tendance en environnement.

Cependant, ce test non paramétrique ne permet pas de porter un jugement sur la valeur des différents paramètres de localisation (p.e. niveau de la série étudiée et amplitude des tendances). En effet, la seule conclusion que permet ce test est l'absence ou non de saisonnalité. Bien que des estimateurs paramétriques et nonparamétriques puissent être obtenus à posteriori pour les paramètres de localisation, un tel traitement en deux étapes demeure encombrant dans des études contenant plusieurs séries sur lesquelles on désire détecter des tendances.

En continuité avec les méthodes utilisées dans les autres chapitres, la méthode qui apparaît la plus appropriée, pour le traitement qui nous intéresse, est l'utilisation de variables muettes dichotomiques pour l'extraction des saisonnalités dans les modèles de régression multiples. Afin de valider cette utilisation dans le contexte de détection de tendance de séries environnementales, le présent chapitre étudie le seuil et la puissance du test de détection de tendance : 1) En présence de saisonnalités seulement et 2) En présence de saisonnalités et d'autocorrélation. Dans ce dernier cas, l'utilisation de variables avec retard permettra de tenir compte de l'autocorrélation.

En présence de saisonnalités seulement, une comparaison aux niveaux du seuil et de la puissance avec le test de Mann-Kendall Saisonnier sera effectuée à la section 5.2, alors qu'en présence de saisonnalités et d'autocorrélation, une comparaison avec le test d'Hirsch et Slack sera effectuée à la section 5.3.

4.2 Modélisation avec variables muettes dichotomiques.

Wonnacott et Wonnacott (1980) présentent l'utilisation de variables muettes dichotomiques pour extraire la composante saisonnière dans un modèle de régression

linéaire. Supposons la présence de "k" saisons; on utilise alors "k-1" variables dichotomiques pour modéliser les "k" saisons: les "k-1" estimateurs associés aux variables dichotomiques représentent alors l'écart avec la saison non associée à une variable muette.

Afin d'illustrer cette méthode de modélisation, prenons l'exemple de quatre saisons : hiver, printemps, été et automne. Les trois variables muettes dichotomiques suivantes permettent d'extraire les composantes saisonnières dans une régression linéaire:

$$X_{sE} = \begin{cases} 1 & \text{si l'observation est mesurée à l'été} \\ 0 & \text{sinon} \end{cases}$$

$$X_{sA} = \begin{cases} 1 & \text{si l'observation est mesurée à l'automne} \\ 0 & \text{sinon} \end{cases}$$

$$X_{sP} = \begin{cases} 1 & \text{si l'observation est mesurée au printemps} \\ 0 & \text{sinon} \end{cases}$$

Les estimateurs associés à ces variables dichotomiques permettent alors d'évaluer l'écart entre la moyenne de chacune de ces saisons et la moyenne de l'hiver. On peut facilement introduire ces variables dans le modèle de régression linéaire présenté au chapitre 2, pour extraire la dépendance dans la série des résidus (éq. 2.1), et ainsi obtenir le modèle suivant:

$$Y_t = \beta_0 + \beta_1 X_t + \sum_{i=1}^p \phi_i Y_{t-i} + \lambda_P X_{sP} + \lambda_E X_{sE} + \lambda_A X_{sA} + \theta_t \quad (4.1)$$

Dans d'autres situations on peut être intéressé à travailler avec douze saisons (une par mois), on peut alors utiliser onze variables muettes dichotomiques (une variable associée à chacun des mois entre février et décembres inclus) qui nous donneront l'écart entre la moyenne du mois en question et la moyenne de mois de janvier.

4.3 Détection de tendance en présence de saisonnalités.

Avant de procéder à la validation (seuil et puissance) de la régression saisonnière (utilisation des variables muettes), en présence de saisonnalités, discutons d'abord de l'effet des saisonnalités sur les estimateurs classiques des paramètres d'échelle et de localisation. Étant donnée la grande similarité entre l'effet des valeurs aberrantes et l'effet des saisonnalités (changement de niveau de courte durée à l'intérieur d'une série), aucune simulation ne sera effectuée pour illustrer l'effet de la présence des saisonnalités. En étudiant simplement les résultats du chapitre 3, on peut conclure que la présence de saisonnalités aura les effets suivants sur les estimateurs classiques, si elle n'est pas prise en compte:

1. Comme dans le cas des valeurs aberrantes, l'estimateur du paramètre d'échelle sera surévalué à cause des changements de niveau qui amènent des écarts importants par rapport à la moyenne générale;
2. L'estimateur du coefficient d'autorégression est également biaisé par la présence de saisonnalités puisque, dans les cas où l'on a plusieurs observations consécutives par saison, on aura une autocorrélation plus importante amenée par la présence de groupes d'observations consécutives à des niveaux différents.
3. Finalement, le seuil et la puissance des tests de détection de tendance seront fortement affectés par les biais décrits aux points 1 et 2, en plus d'être affectés par l'emplacement des saisons: par exemple, même en absence de tendance, une saison de niveau élevé au début d'une série amènera souvent le test associé au modèle (1.1) à conclure à la présence d'une tendance décroissante; cette erreur sera encore plus fréquente si la saison à la fin de la série présente un niveau faible.

4.3.1 Étude de seuil; saisons connues.

La simulation 19 permet d'étudier le seuil du test de détection de tendance en présence de saisonnalités avec l'utilisation de la régression saisonnière. Le tableau 4.1

présente les résultats de cette simulation. Cette simulation montre clairement que le seuil (5%) est conservé et que tous les estimateurs convergent vers les valeurs introduites dans les séries simulées. Cependant, nous sommes ici en présence de saisonnalités dont on connaît les dates de changement de saison, ce qui en pratique n'est généralement pas le cas. La section 4.3.3 présente une étude de seuil lorsque les dates de changement de saison ne sont pas connues.

SIMULATION 19

CONDITIONS EXPERIMENTALES

Titre : Étude de seuil : Utilisation des variables muettes en présence de saison.

Estimateurs étudiés : b_1, l_1, l_2, l_3 REQM, *seuil*.

Nombre de séries simulées : 500 X **Longueur des séries :** 300.

Modèle sous-jacent : $N(0,1), \Delta_m = 0$.

Modèle de Saisons : $n_1 = 12$ et $\mu_1 = 0, n_2 = 13$ et $\mu_2 = 5, n_3 = 12$ et $\mu_3 = 3, n_4 = 13$ et $\mu_4 = 1$.

Modèle de régression utilisé : $Y_t = \beta_0 + \beta_1 X_t + \lambda_1 X_{s1} + \lambda_2 X_{s2} + \lambda_3 X_{s3} + e_t, X_t = t$.

Tableau 4.1

Résultats numériques de la simulation 19; Seuil de la régression saisonnière en présence de saisons connues.

Estimateur	Moyenne	Ecart type	Précision
b_1	0,00000	0,0007	0,00006
REQM	0,9998	0,044	0,004
l_1	4,997	0,168	0,015
l_2	2,991	0,177	0,015
l_3	0,986	0,178	0,016

$$\hat{\alpha} = \text{Seuil empirique} = 0,046$$

4.3.2 Étude de puissance; saisons connues.

La simulation 20 permet d'étudier la puissance du test de détection de tendance en présence de saisonnalités avec l'utilisation des variables muettes. Le tableau 4.2 présente les résultats de cette simulation. Comme dans la section 4.3.1, on se retrouve en présence de saisons dont on connaît les dates de changements. La section 4.3.4 présente une simulation dans le cas où les dates de changement de saison ne sont pas connues.

SIMULATION 20

CONDITIONS EXPERIMENTALES

Titre : Étude de puissance : Utilisation des variables muettes en présence de saisons.

Estimateurs étudiés : b_1 , l_1 , l_2 , l_3 , REQM, puissance.

Nombre de séries simulées : 500 X **Longueur des séries :** 300.

Modèle sous-jacent : $N(0,1)$, $\Delta_m = 0,002$.

Modèle de Saisons : $n_1 = 12$ et $\mu_1 = 0$, $n_2 = 13$ et $\mu_2 = 5$, $n_3 = 12$ et $\mu_3 = 3$, $n_4 = 13$ et $\mu_4 = 1$.

Modèle de régression utilisé : $Y_t = \beta_0 + \beta_1 X_t + \lambda_1 X_{s1} + \lambda_2 X_{s2} + \lambda_3 X_{s3} + e_t$, $X_t = t$

Tableau 4.2

Résultats numériques de la simulation 20; Puissance de la régression saisonnière en présence de saisons connues.

Estimateur	Moyenne	Ecart type	Précision
b_1	0,0020	0,0007	0,00006
REQM	1,002	0,041	0,003
l_1	4,998	0,159	0,014
l_2	3,002	0,164	0,014
l_3	0,999	0,168	0,015

Puissance empirique=0,846

Comme pour l'étude de seuil, cette simulation montre clairement que l'utilisation de variables muettes est tout à fait adéquate dans le cas de saisonnalités dont on connaît les dates de changement de saison: la puissance est équivalente à la puissance théorique et tous les estimateurs convergent vers les valeurs introduites dans les séries simulées.

4.3.3 Étude de seuil; présence de décalage.

Lorsque les dates de changements de saison sont inconnues, il est alors nécessaire d'estimer ces dates. Étant donné que le but principal des études de détection de tendance n'est pas d'obtenir une estimation précise des saisonnalités, une estimation visuelle des dates de changement de saison peut être suffisante pour obtenir une détection de tendance adéquate. L'utilisation des fonctions CUSUM (Hinkley, 1971; Pagé, 1954) peut également être un outil intéressant pour l'estimation des dates de changements de saison puisque des changements de niveau dans la série sont associés à des changements de pente dans la courbe CUSUM. Cette technique demande cependant plusieurs observations par saison. L'utilisation de ces deux méthodes d'estimation peut introduire un léger décalage entre le point de changement estimé et le point de changement réel. Afin de voir l'effet possible de ce biais, cette section étudie le seuil du test de détection de tendance en présence d'un décalage entre les dates de changements de saison estimées et réelles.

Dans un premier temps, le tableau 4.3 présente les résultats d'une simulation identique à la simulation 19 (seuil en présence de saisons; utilisation des variables muettes), sauf pour un décalage de deux unités de temps entre les saisons simulées (réelles) et les saisons estimées par le modèle de régression.

Il faut noter ici qu'on présente les résultats associés à l'ordonnée à l'origine (b_0), car la structure des décalages introduits fait que la première saison n'a pas une moyenne de 0. En effet, les décalages introduits amènent l'estimation d'une première saison avec 2 observations de moyenne 0 et 10 observations de moyenne 5, alors que les saisons estimées 2,3 et 4 contiennent respectivement 3 observations de moyenne 5 et 10 de moyenne 3, 2 observations de moyenne 3 et 10 de moyenne 1 et finalement 2 observations de moyenne 1 et 10 de moyenne 0.

Tableau 4.3
 Résultats numériques de la simulation 19; Seuil de la régression
 saisonnière en présence de saisons et d'un décalage.

Estimateur	Moyenne	Ecart type	Précision
b_1	0,00000	0,0007	0,00006
REQM	1,49	0,050	0,0044
b_0	4,169	0,140	0,012
l_1	-0,701	0,165	0,014
l_2	-2,837	0,170	0,015
l_3	-3,935	0,165	0,014

$$\hat{\alpha} = \text{Seuil empirique} = 0,004$$

Ce tableau montre que l'estimateur du paramètre d'échelle et les estimateurs des paramètres saisonniers sont biaisés en présence d'un décalage entre les saisons simulées et estimées, ce qui entraîne un seuil empirique beaucoup plus faible que le seuil théorique de 5%. Le biais associé au paramètre d'échelle consiste en une surestimation introduite par de grandes erreurs aux points de décalage. En ce qui concerne les paramètres des saisons, leur biais dépend de plusieurs facteurs dont les niveaux des saisons, l'emplacement des saisons et la structure du décalage.

On remarque donc qu'en présence de décalage dans l'estimation des saisons, le test de détection de tendance utilisant la régression linéaire est conservateur à cause d'une surestimation de la variance (le biais des estimateurs des paramètres de localisation jouant un rôle mineur pour des valeurs raisonnables). Notons cependant que le problème de décalage des saisons est inhérent à l'ensemble des méthodes de détections de tendance en sciences de l'eau, y compris le test de Kendall saisonnier qui sera discuté au chapitre 5.

4.3.4 Étude de puissance; présence de décalage.

Le tableau 4.4 présente les résultats d'une simulation identique à la simulation 20 (puissance en présence de saisons; utilisation des variables muettes), sauf pour un décalage

de deux unités de temps entre les saisons simulées (réelles) et les saisons estimées par le modèle de régression.

Tableau 4.4
Résultats numériques de la simulation 20; Puissance de la régression saisonnière en présence de saisons et d'un décalage.

Estimateur	Moyenne	Ecart type	Précision
b_1	0,00194	0,0007	0,00006
REQM	1,49	0,052	0,004
b_0	4,174	0,150	0,013
l_1	-0,710	0,171	0,015
l_2	-2,834	0,164	0,014
l_3	-3,928	0,157	0,014

Puissance empirique = 0,502

On remarque donc une perte de puissance significative en présence de décalage des saisons, cette perte de puissance étant causée par la surestimation du paramètre d'échelle comme dans le cas précédent de l'étude de seuil. En conclusion, la présence de décalage des saisons amène des estimateurs des paramètres d'échelle et des paramètres saisonniers biaisés. Ces biais amènent un test de détection de tendance conservateur et moins puissant. On verra au chapitre 5 la comparaison avec les tests non paramétriques utilisés dans le domaine de l'eau.

4.3.5 Étude complémentaire en présence de décalage.

Étant donnés les effets négatifs d'une mauvaise estimation des points de changements de saison, on pourrait se demander si une autre méthode d'estimation de ces changements permettrait de grands avantages sur l'estimation visuelle de ces changements. Un survol de la littérature permet de conclure que peu d'autres méthodes peuvent être utilisées efficacement pour solutionner le problème des dates de changement de saison. Des auteurs (Esterby et El-Shaarawi, 1981; Pettit, 1979; Smith, 1975) se sont attardés à l'estimation de ces dates de changement à l'aide d'approches utilisant la fonction de vraisemblance, les méthodes non paramétriques ou les méthodes bayésiennes. Hipel et

McCleod (1994) présentent une série plus complète de références sur ce sujet. Malheureusement ces méthodes sont souvent difficilement compatibles avec l'utilisation de la régression linéaire. De plus ces méthodes s'attardent à la détection d'un seul point de changement et sont donc peu adaptées à l'estimation de changements de saison qui implique plusieurs changements de niveau avec un petit nombre d'observations entre chaque changement.

En relation avec l'utilisation de la régression linéaire, une autre approche apparaît cependant intéressante, il s'agit de travailler avec des sous saisons, i.e. travailler avec les douze mois plutôt qu'avec quatre saisons de trois mois. Les effets saisonniers sont quand même pris en compte et l'effet de décalage des saisons doit être réduit. Afin de juger de l'intérêt de cette approche, la simulation 21 étudie l'effet du décalage lorsque douze saisons d'un mois sont modélisées plutôt que quatre saisons de trois mois. Les séries simulées demeurent cependant composées d'années de quatre saisons.

SIMULATION 21

CONDITIONS EXPERIMENTALES

Titre : Étude de seuil : Utilisation des variables muettes (11) en présence de saisons.

Estimateurs étudiés : $b_1, l_1, l_2, l_3, l_4, l_5, l_6, l_7, l_8, l_9, l_{10}, l_{11}, REQM_1, \text{seuil}$.

Nombre de séries simulées : 500 X **Longueur des séries :** 300.

Modèle sous-jacent : $N(0,1), \Delta_m = 0$.

Modèle de Saisons : $n_1 = 12$ et $\mu_1 = 0, n_2 = 13$ et $\mu_2 = 5, n_3 = 12$ et $\mu_3 = 3, n_4 = 13$ et $\mu_4 = 1$.

Modèle de régression utilisé : $Y_t = \beta_0 + \beta_1 X_t + \sum_{i=1}^{11} \lambda_i X_{si} + e_t, X_t = t$.

Comme pour la simulation 19, le tableau 4.5 montre que l'utilisation de 11 variables muettes (plutôt que 3 dans la simulation 19) amène un test de détection de tendance conservateur en présence de décalage dans l'estimation des saisons. Cependant,

l'utilisation de 11 variables muettes amène un biais plus faible de l'estimateur du paramètre d'échelle, ce qui permet un seuil empirique beaucoup plus près du seuil théorique comparativement à l'utilisation de 3 variables muettes. Cette simulation montre donc un intérêt certain à utiliser des sous saisons lorsqu'on ne connaît pas les dates des changements de saison.

Tableau 4.5

Résultats numériques de la simulation 21; Seuil de la régression saisonnière (12 saisons) en présence de 4 saisons et d'un décalage.

Estimateur	Moyenne	Ecart type	Précision
b_1	0,00012	0,0007	0.00006
REQM	1,31	0,054	0,005
l_1	-0,516	0,299	0,026
l_2	-0,529	0,297	0,026
l_3	1,991	0,285	0,025
l_4	4,476	0,287	0,025
l_5	4,482	0,279	0,024
l_6	3,481	0,281	0,025
l_7	2,486	0,293	0,026
l_8	2,476	0,296	0,026
l_9	1,489	0,291	0,026
l_{10}	0,499	0,309	0,027
l_{11}	0,482	0,289	0,025

$$\hat{\alpha} = \text{Seuil empirique} = 0,020$$

Afin de compléter l'étude sur l'intérêt d'utiliser des sous saisons pour réduire l'effet du décalage sur le test de détection de tendance, la simulation 22 étudie la puissance du test de détection de tendance en présence de décalage dans l'estimation des saisons mais avec la modélisation de 12 saisons de 1 mois. Le tableau 4.6 présente les résultats de cette simulation, on note alors que l'utilisation de 11 variables muettes en présence de décalage dans l'estimation des saisons, amène une perte de puissance significative pour le test de détection de tendance. Cependant, cette perte de puissance est plus petite que dans le cas où 3 variables muettes sont utilisées. Tout comme dans l'étude de seuil, l'utilisation de 11

variables muettes, plutôt que 3, amène un biais plus faible de l'estimateur du paramètre d'échelle qui se traduit par une perte de puissance plus faible.

SIMULATION 22

CONDITIONS EXPERIMENTALES

Titre: Étude de puissance : 11 variables muettes en présence de saisons décalées.

Estimateurs étudiés : $b_1, l_1, l_2, l_3, l_4, l_5, l_6, l_7, l_8, l_9, l_{10}, l_{11}$, REQM, puissance.

Nombre de séries simulées : 500 X **Longueur des séries :** 300.

Modèle sous-jacent : $N(0,1), \Delta_m = 0,002$.

Modèle de Saisons : $n_1 = 12$ et $\mu_1 = 0, n_2 = 13$ et $\mu_2 = 5, n_3 = 12$ et $\mu_3 = 3, n_4 = 13$ et $\mu_4 = 1$.

Modèle de régression utilisé : $Y_t = \beta_0 + \beta_1 X_t + \sum_{i=1}^{11} \lambda_i X_{st} + e_t, X_t = t$.

Tableau 4.6

Résultats numériques de la simulation 22; Puissance de la régression saisonnière (12 saisons) en présence de 4 saisons et d'un décalage.

Estimateur	Moyenne	Ecart type	Précision
b_1	0,00223	0,0007	0,00006
REQM	1,31	0,050	0,004
l_1	-0,502	0,294	0,026
l_2	-0,515	0,291	0,026
l_3	1,987	0,291	0,026
l_4	4,484	0,288	0,025
l_5	4,489	0,275	0,024
l_6	3,502	0,287	0,025
l_7	2,485	0,289	0,025
l_8	2,492	0,273	0,024
l_9	1,487	0,283	0,025
l_{10}	0,473	0,280	0,025
l_{11}	0,491	0,281	0,025

Puissance empirique = 0,752

Les résultats pour l'estimation des paramètres de localisation montrent un léger biais pour la pente de la tendance et les saisons. Le biais des estimateurs des saisons est facile à expliquer: Le décalage implique que les moyennes des premier et dernier mois des saisons contiennent deux observations de deux saisons plutôt que quatre d'une même saison: i.e. le premier mois estimé contient 2 observations de la saison 1 et deux de la saison 4 ce qui donne une moyenne théorique de 0,5. Tous les estimateurs des autres saisons (mois) montrent l'écart de la moyenne de la saison par rapport à 0,5 soit : -0,5, -0,5, 2, 4,5, 4,5, 3,5, 2,5, 2,5, 1,5, 0,5, 0,5.

Les simulations 21 et 22 montrent donc un intérêt certain à utiliser des sous saisons lorsqu'on ne connaît pas les dates des changements de saison. Cette technique permet un test plus puissant et moins conservateur. Rappelons cependant que le test de détection de tendance en présence de décalage dans les saisons demeure un test conservateur (seuil empirique significativement plus faible que 5%) et moins puissant qu'en absence de décalage. Les comparaisons du chapitre 5 permettront de voir la performance du test de Kendall saisonnier dans les mêmes conditions.

4.3.6 Étude complémentaire en absence de décalage

Devant les avantages d'utiliser 11 variables muettes plutôt que 3, on peut se demander s'il y a des inconvénients à une telle utilisation. Les tableaux 4.7 et 4.8 présentent les résultats de simulations identiques aux simulations 20 et 21 mais sans décalage. Le seuil et la puissance du test de détection de tendance sont donc étudiés dans la situation où l'on a 4 saisons de trois mois mais que l'on estime 12 saisons d'un mois.

Les tableaux 4.7 et 4.8 montrent que :1) les estimateurs des paramètres de localisation et d'échelle ne sont pas biaisés; 2) le seuil du test de détection de tendance n'est pas significativement différent de 5% et; 3) la puissance du test de détection de tendance n'est pas significativement différente de la puissance théorique (85%).

Tableau 4.7

Tableau 4.7

Résultats numériques de la simulation 21; Seuil de la régression saisonnière (12 saisons) en présence de 4 saisons mais sans décalage.

Estimateur	Moyenne	Ecart type	Précision
b_1	0,00003	0,0007	0,00006
REQM	0,999	0,042	0,004
l_1	-0,003	0,220	0,019
l_2	-0,005	0,302	0,026
l_3	4,991	0,280	0,025
l_4	5,012	0,289	0,025
l_5	4,997	0,282	0,025
l_6	2,995	0,275	0,024
l_7	2,994	0,286	0,025
l_8	3,001	0,299	0,026
l_9	1,010	0,285	0,025
l_{10}	0,998	0,285	0,025
l_{11}	0,992	0,297	0,026

$$\hat{\alpha} = \text{Seuil empirique} = 0,042$$

Tableau 4.8

Résultats numériques de la simulation 22; Puissance de la régression saisonnière (12 saisons) en présence de 4 saisons mais sans décalage.

Estimateur	Moyenne	Ecart type	Précision
b_1	0,00202	0,0007	0,00006
REQM	1,000	0,040	0,004
l_1	-0,019	0,296	0,026
l_2	-0,017	0,281	0,025
l_3	4,984	0,276	0,024
l_4	4,988	0,290	0,025
l_5	4,983	0,279	0,024
l_6	2,990	0,296	0,026
l_7	2,991	0,277	0,024
l_8	2,975	0,286	0,025
l_9	0,980	0,283	0,025
l_{10}	0,979	0,291	0,026
l_{11}	0,976	0,264	0,023

$$\text{Puissance empirique} = 0,846$$

En absence de décalage, l'utilisation de onze variables muettes plutôt que trois n'amène donc pas de désavantages apparents, alors qu'en présence de décalage on a vu des avantages à utiliser onze variables muettes (test moins conservateur et plus puissant). Dans l'incertitude quant à la structure des saisons, il apparaît donc préférable d'utiliser une variable muette par mois.

4.4 Détection de tendance en présence de saisonnalités et d'autocorrélation.

La section 2.2 a présenté une étude du test de détection de tendance, avec l'utilisation d'une variable avec retard, en présence d'autocorrélation, alors que la section 4.3 a présenté une étude similaire avec l'utilisation de variables muettes en présence de saisonnalités. La présente section combine les méthodes des sections 2.2 et 4.3 pour étudier la présence d'autocorrélation et de saisonnalités. On cherche donc seulement à s'assurer que l'utilisation conjointe de variable avec retard et de variables muettes n'amène pas de problème dans la détection de tendance avec régression multiple. On présente d'abord une étude de seuil et une étude de puissance sans décalage entre les saisons réelles et estimées et on complète avec des études en présence de décalage.

4.4.1 Étude de seuil et puissance; sans décalage.

La simulation 23 étudie donc l'utilisation de la régression autorégressive saisonnière en présence d'autocorrélation et de saisonnalités. On se place ici dans le cas de saisons de paramètres connus, c'est-à-dire qu'il n'y a pas de décalage entre les saisons réelles et estimées.

Le tableau 4.9, présente les résultats de cette simulation. Il montre que les estimateurs des paramètres associés aux saisons sont légèrement biaisés. Cependant, ce biais n'affecte pas l'estimation des paramètres de localisation et d'échelle de sorte que le seuil empirique n'est pas significativement différent de 5%.

SIMULATION 23

CONDITIONS EXPERIMENTALES

Titre: Étude de seuil : En présence d'autocorrélation et de saisonnalités.

Estimateurs étudiés : $b_1, d_1, l_1, l_2, l_3, \text{REQM}, \text{seuil}$.

Nombre de séries simulées : 500 X **Longueur des séries :** 300.

Modèle sous-jacent : $AR(1)$ avec $\phi_1 = 0,75$.

Modèle de Saisons : $n_1 = 12$ et $\mu_1 = 0, n_2 = 13$ et $\mu_2 = 5, n_3 = 12$ et $\mu_3 = 3, n_4 = 13$ et $\mu_4 = 1$.

Modèle de régression utilisé : $Y_t = \beta_0 + \beta_1 X_t + \phi Y_{t-1} + \sum_{s=1}^3 \lambda_s X_{st} + e_t, X_t = t$

Tableau 4.9

Résultats numériques de la simulation 23; Seuil de la régression autorégressive saisonnière en présence d'autocorrélation et de saisons.

Estimateur	Moyenne	Ecart type	Précision
b_1	0,00002	0,0007	0,00006
REQM	1,000	0,038	0,003
d_1	0,746	0,0156	0,0014
l_1	5,032	0,256	0,022
l_2	3,036	0,260	0,022
l_3	1,018	0,184	0,016

$$\hat{\alpha} = \text{Seuil empirique} = 0,044$$

La simulation 24 étudie la puissance du test de détection de tendance avec l'utilisation de la régression autorégressive saisonnière, en présence de saisonnalités et d'autocorrélation. Le tableau 4.10 montre que les estimateurs des paramètres associés aux saisons sont de nouveau légèrement biaisés, mais que le biais n'affecte pas l'estimation des paramètres de localisation et d'échelle. La puissance empirique n'est pas significativement différente de 85%.

SIMULATION 24

CONDITIONS EXPERIMENTALES

Titre: Étude de puissance : En présence d'autocorrélation et de saisonnalités.

Estimateurs étudiés : b_1 , d_1 , l_1 , l_2 , l_3 , REQM, puissance.

Nombre de séries simulées : 500 X **Longueur des séries :** 300.

Modèle sous-jacent : $AR(1)$ avec $\phi_1 = 0,75$ et $\Delta_m = 0,002$.

Modèle de Saisons : $n_1 = 12$ et $\mu_1 = 0$, $n_2 = 13$ et $\mu_2 = 5$, $n_3 = 12$ et $\mu_3 = 3$, $n_4 = 13$ et $\mu_4 = 1$.

Modèle de régression utilisé : $Y_t = \beta_0 + \beta_1 X_t + \phi Y_{t-1} + \sum_{s=1}^3 \lambda_s X_{st} + e_t, X_t = t$

Tableau 4.10

Résultats numériques de la simulation 24; Puissance de la régression autorégressive saisonnière en présence d'autocorrélation et de saisons.

Estimateur	Moyenne	Ecart type	Précision
b_1	0,00206	0,0007	0,00006
REQM	1,001	0,041	0,004
d_1	0,746	0,0158	0,0014
l_1	5,049	0,260	0,022
l_2	3,046	0,267	0,023
l_3	1,013	0,158	0,014

Puissance empirique = 0,832

Les simulations 23 et 24 montrent clairement que l'utilisation conjointe de variables avec retard et de variables muettes en présence d'autocorrélation et de saisonnalités n'amène pas de problèmes particuliers. Il est donc tout à fait adéquat d'utiliser la régression linéaire en présence de ces deux caractéristiques restrictives. Rappelons cependant qu'un

décalage entre les saisons estimées et les saisons réelles amène des seuils et des puissances empiriques différents des valeurs théoriques.

4.4.2 Étude de seuil et puissance; avec décalage.

Le tableau 4.11 présente les résultats d'une simulation identique à la simulation 23 (étude de seuil en présence d'autocorrélation et de saisonnalités) mais en présence de décalage (2 unités de temps) dans l'estimation des saisonnalités. Comme pour le tableau 4.3, les résultats associés à l'estimation de l'ordonnée à l'origine (b_0) sont présentés afin de tenir compte du fait que le décalage amène une ordonnée à l'origine significative. Il faut rappeler que les trois estimateurs associés aux saisons représentent la différence entre les moyennes saisonnières et l'ordonnée à l'origine.

Tableau 4.11
Résultats de la simulation 23; Seuil de la régression autorégressive saisonnière en présence d'autocorrélation, de saisons et de décalage.

Estimateur	Moyenne	Ecart type	Précision
b_1	-0,00006	0,0003	0,00003
REQM	1,334	0,050	0,004
d_1	0,906	0,0090	0,0008
b_0	2,546	0,119	0,010
l_1	-1,543	0,147	0,013
l_2	-2,450	0,116	0,010
l_3	-2,651	0,143	0,013

$$\hat{\alpha} = \text{Seuil empirique} = 0,000$$

On remarque facilement le biais important de tous les estimateurs ponctuels. Il n'est donc pas surprenant de voir que le test de détection de tendance ne conserve pas le seuil théorique de 5%: aucune tendance n'est détectée dans les 500 séries simulées. En comparant les résultats du tableau 4.11 à ceux du tableau 4.3 on remarque que les estimateurs des paramètres saisonniers ont un biais plus grand en présence d'autocorrélation (tableau 4.11), alors que le paramètre d'échelle présente un biais plus

grand lorsqu'aucune autocorrélation n'est présente (tableau 4.3). Cette situation particulière s'explique par une forme de lissage introduit par l'estimation de l'autocorrélation. En effet, les points de décalage ont la particularité d'être très fortement reliés aux points d'une saison adjacente; cette forte relation, qui devrait être prise en compte par les estimateurs saisonniers si aucun décalage n'était présent, est plutôt prise en compte dans l'estimation du coefficient d'autocorrélation. Il en découle une surestimation du coefficient d'autocorrélation qui amène un biais pour les paramètres saisonniers et de localisation beaucoup plus important que dans le cas sans autocorrélation. En ce qui concerne le paramètre d'échelle, le biais moins grand en présence d'autocorrélation s'explique par des résidus plus faibles aux points de décalage à cause de la forme de lissage introduit par l'estimation de l'autocorrélation.

Le tableau 4.12 présente les résultats d'une simulation identique à la simulation 24 (étude de puissance en présence d'autocorrélation et de saisonnalités) mais en présence de décalage (2 unités de temps) dans l'estimation des saisonnalités.

Tableau 4.12

Résultats de la simulation 24; Puissance de la régression autorégressive saisonnière en présence d'autocorrélation, de saisons et de décalage.

Estimateur	Moyenne	Ecart type	Précision
b_1	0,00073	0,0003	0,00003
REQM	1,327	0,051	0,004
d_1	0,906	0,0092	0,0008
b_0	2,560	0,112	0,010
l_1	-1,552	0,150	0,013
l_2	-2,450	0,109	0,010
l_3	-2,653	0,141	0,012

Puissance empirique = 0,000

Le tableau 4.12 amène les mêmes conclusions que le tableau 4.11 pour les estimateurs ponctuels des différents paramètres. Ces biais importants des estimateurs

ponctuels affectent significativement le test de détection de tendance utilisant la régression linéaire puisque encore une fois, aucune tendance n'a été détectée dans les 500 séries simulées.

Ces derniers résultats introduisent une importante limitation à l'utilisation de la régression linéaire pour la détection de tendance en environnement. Il faut cependant mettre cette limitation dans le contexte et rappeler que :

- les décalages peuvent être détectés à l'aide d'une étude des résidus de régression.
- on ne connaît pas la performance des tests non paramétriques dans la même situation; cette performance sera étudiée au chapitre 5.

4.5 Utilisation des variables muettes en absence de saisonnalités.

La section 3.6 a montré que l'utilisation des méthodes robustes en absence de valeurs aberrantes impliquait une perte de puissance en comparaison avec l'utilisation des méthodes classiques. Afin de s'assurer que l'utilisation de la régression saisonnière en absence de saisonnalités n'amène pas de problèmes similaires, les simulations 25 et 26 étudient le seuil et la puissance dans de telles circonstances.

Le tableau 4.13 présente les résultats de l'étude de seuil. On note alors que les estimateurs des paramètres de localisation et d'échelle ne sont pas biaisés et que le seuil empirique n'est pas significativement différent du seuil théorique.

SIMULATION 25CONDITIONS EXPERIMENTALES

Titre: Étude de seuil : en absence de saisonnalités.

Estimateurs étudiés : b_1, l_1, l_2, l_3 , REQM, seuil.

Nombre de séries simulées : 500 X **Longueur des séries :** 300.

Modèle sous-jacent : $N(0, 1)$.

Modèle de Saisons : Aucune saison.

Modèle de régression utilisé : $Y_t = \beta_0 + \beta_1 X_t + \phi Y_{t-1} + \sum_{s=1}^3 \lambda_s X_{st} + e_t, X_t = t$

SIMULATION 26CONDITIONS EXPERIMENTALES

Titre: Étude de puissance: En absence de saisonnalités.

Estimateurs étudiés : b_1, l_1, l_2, l_3 , REQM, puissance.

Nombre de séries simulées : 500 X **Longueur des séries :** 300.

Modèle sous-jacent : $N(0, 1)$, $\Delta_m = 0,002$.

Modèle de Saisons : Aucune saison.

Modèle de régression utilisé : $Y_t = \beta_0 + \beta_1 X_t + \phi Y_{t-1} + \sum_{s=1}^3 \lambda_s X_{st} + e_t, X_t = t$

Tableau 4.13

Résultats de la simulation 25; Seuil de la régression saisonnière en absence de saisons.

Estimateur	Moyenne	Ecart type	Précision
b_1	0,00001	0,0007	0,00006
REQM	1,001	0,041	0,004
b_0	-0,001	0,153	0,013
l_1	0,007	0,167	0,015
l_2	0,009	0,187	0,016
l_3	0,013	0,169	0,015

$$\hat{\alpha} = \text{Seuil empirique} = 0,056$$

Le tableau 4.14 présente les résultats de l'étude de puissance. On note de nouveau, que les estimateurs des paramètres de localisation et d'échelle ne sont pas biaisés et que la puissance empirique n'est pas significativement différente de la puissance théorique.

Tableau 4.14

Résultats de la simulation 26; Puissance de la régression saisonnière en absence de saisonnalités.

Estimateur	Moyenne	Ecart type	Précision
b_1	0,00201	0,0007	0,00006
REQM	1,002	0,043	0,004
b_0	-0,004	0,149	0,013
l_1	0,002	0,163	0,014
l_2	-0,005	0,162	0,014
l_3	0,003	0,162	0,014

$$\text{Puissance empirique} = 0,832$$

Les simulations 25 et 26 montrent donc que l'utilisation de la régression saisonnière en absence de saisonnalités n'empêche pas de conserver le seuil et n'amène pas de perte de puissance significative. En pratique, lorsque la signification des saisons est incertaine, on

peut donc utiliser les variables muettes sans crainte de réduire la validité des résultats du test de détection de tendance.

4.6 Discussion sur la présence de saisonnalités.

En présence de saisonnalités seulement, l'utilisation de la régression saisonnière permet d'obtenir un test de détection de tendance avec des propriétés semblables au cas théorique (régression linéaire simple sur des données $N(0,1)$ indépendantes à l'exception de la tendance). La principale difficulté de la méthode utilisée réside dans la modélisation des saisons. En effet, on a vu que la présence d'un décalage entre les saisons réelles et les saisons estimées peut amener le test de détection de tendance à présenter des propriétés indésirables.

En présence de saisonnalités et d'autocorrélation, l'utilisation de la régression autorégressive saisonnière permet également d'obtenir un test de détection de tendance avec des propriétés semblables au cas théorique. Cependant, la présence d'un décalage entre les saisons réelles et les saisons estimées a un effet encore plus important sur le test de détection de tendance en présence d'autocorrélation. En présence de saisonnalités et d'autocorrélation, il faut donc être certain de la modélisation des saisons pour se fier au résultat du test de détection de tendance utilisant la régression. Le chapitre 5 montre la comparaison avec les tests non paramétriques qui sont également affectés par la présence de décalage des saisons.

Si on désire porter un jugement sur la présence de décalage des saisons dans les séries étudiées, la régression a un avantage important sur les méthodes non paramétriques. En effet, l'étude des résidus permet généralement de détecter la présence d'une mauvaise modélisation des saisons. En absence d'autocorrélation, l'identification des décalages est directe, puisqu'une saison mal modélisée introduira des valeurs aberrantes aux extrémités des saisons et un ajustement de la longueur ou du début de la saison permettra d'éliminer la présence de valeurs aberrantes.

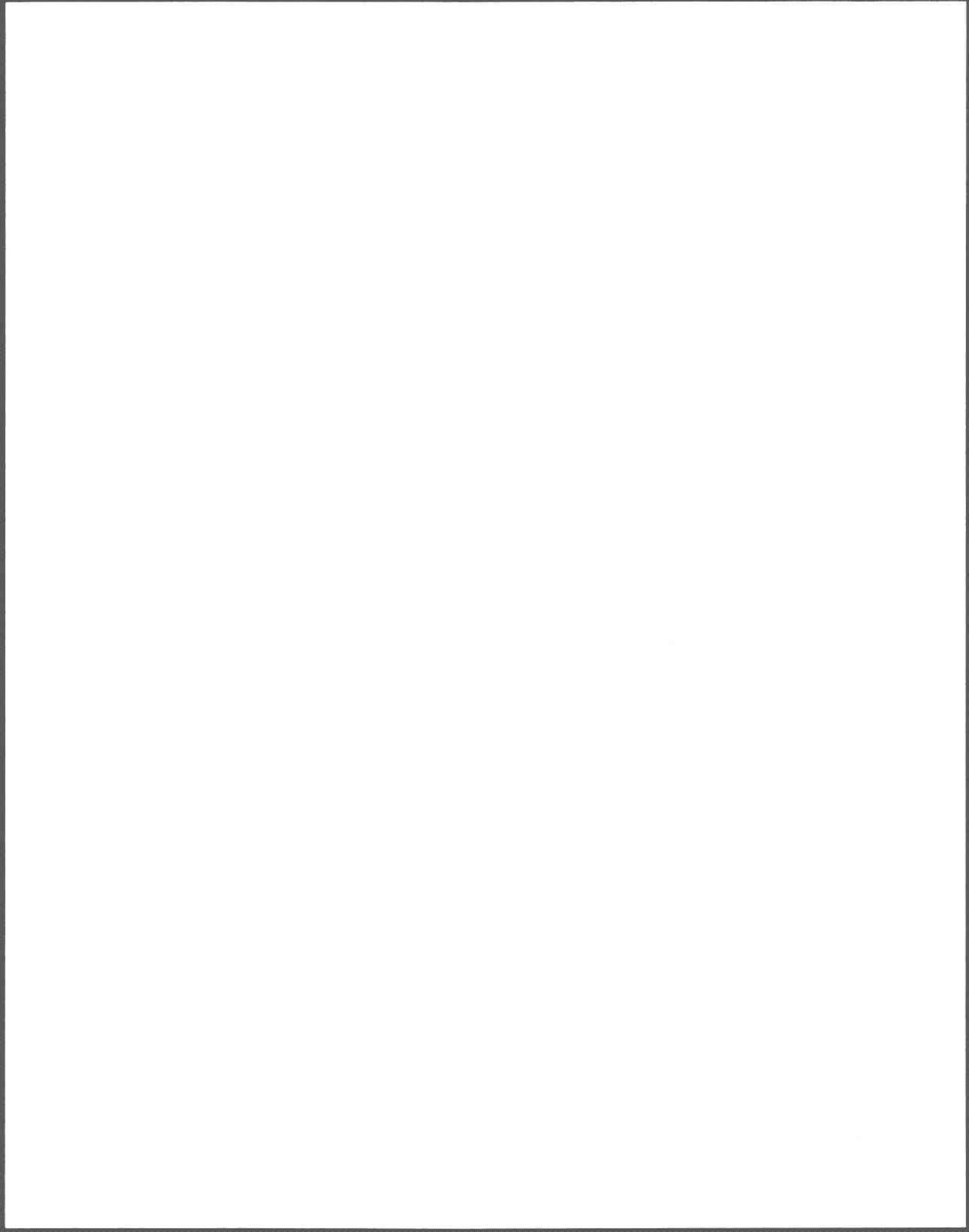
En présence d'autocorrélation, l'identification des décalages est moins directe. Afin de permettre une meilleure identification des décalages en présence d'autocorrélation, le tableau 4.15 présente les effets de la présence de décalage sur les séries de résidus de séries autocorrélées. On remarque que dans plusieurs situations les décalages impliquent l'introduction de valeurs aberrantes. Il est donc primordial de bien étudier les résidus aberrants pour savoir s'il est introduit par: 1) l'absence de ϕ dans le modèle; 2) un décalage dans les saisons ou ; 3) une raison externe à l'autocorrélation et les saisons.

Tableau 4.15

Effets sur les séries de résidus d'une mauvaise estimation des saisons en présence d'autocorrélation.

Présence de décalage?	ϕ dans le modèle?	Effet sur la série de résidus
Non	Non	Comme ϕ n'est pas dans le modèle, on a une mauvaise modélisation aux changements de saisons. En effet, en présence d'autocorrélation, les changements de saisons n'apparaissent pas instantanés, on a plutôt un changement progressif d'une saison à l'autre, plus ou moins long selon la force de l'autocorrélation. L'absence de ϕ dans le modèle ne permet pas de modéliser ces changements progressifs d'une saisons à l'autre.
Non	Oui	Aucune valeur aberrante, la modélisation est adéquate.
Oui	Non	Présence de valeurs aberrantes importantes.
Oui	Oui	Présence de valeurs aberrantes. ϕ est surestimé car le modèle tente d'expliquer les valeurs mal modélisées pour les saisons par une plus grande corrélation entre les données successives. Cette compensation vers le coefficient d'autocorrélation amène également des valeurs aberrantes plus petites que dans le cas de décalage avec ϕ absent du modèle.

En terminant, il est important de noter, qu'afin d'obtenir des détections de tendance valables dans des séries avec saisonnalités, ces séries doivent contenir au moins deux cycles (un cycle correspond généralement à une année) et que les cycles doivent être complets. Dans ces circonstances, on réduit au maximum les possibilités de confondre saisonnalités et tendance.



CHAPITRE 5

COMPARAISON AVEC LES TESTS NON PARAMÉTRIQUES

Ce chapitre présente une comparaison, au niveau du seuil et de la puissance, entre le test de détection de tendance développé dans les sections précédentes et les tests non paramétriques généralement utilisés dans le domaine de l'eau. Différentes situations contenant au moins une des caractéristiques restrictives (autocorrélation, saisonnalités et valeurs aberrantes) sont étudiées. Les comparaisons sont faites à l'aide de simulations de Monte Carlo.

Il est important de noter que les tests non paramétriques sont effectués sur les mêmes séries simulées que les tests utilisant la régression linéaire, il ne peut donc pas y avoir de différence significative introduite par des séries simulées différentes. On étudie de nouveau le seuil et la puissance, on cherche donc à savoir si le seuil empirique des tests non paramétriques est égal au seuil théorique et si la puissance de ces tests est meilleure que la puissance des tests utilisant la régression linéaire. En ce qui concerne le seuil, on s'attend à ce qu'il soit conservé pour les tests de Mann-Kendall saisonniers, puisque des développements théoriques appuient ce test, cependant, dans le cas du test de Spearman/Lettenmaier, l'absence de valeur critique théorique oblige une adaptation qui n'est

pas encore validée. L'étude de seuil permettra donc de s'assurer que cette adaptation permet de conserver le seuil à 5%.

5.1 Test de Spearman/Lettenmaier.

Pour permettre l'utilisation du test t en présence d'autocorrélation, Lettenmaier (1976) propose d'ajuster la variance de la moyenne à l'aide du nombre effectif d'observations, i.e.

$$Var(\bar{Y}) = \frac{\sigma_y^2}{n^*} \quad (5.1)$$

où n^* est le nombre effectif d'observations, il représente le nombre d'observations indépendantes équivalent au nombre d'observations autocorrélées de la série. Dans le cas d'un processus AR(1) on peut établir la relation suivante entre le nombre d'observations dans la série et le nombre effectif d'observations:

$$\frac{n}{n^*} = \frac{(1+\phi)}{(1-\phi)} + \frac{2\phi(\phi^n-1)}{n(\phi-1)} \quad (5.2)$$

on remarque que, pour n suffisamment grand, le deuxième terme de droite est négligeable. En remplaçant n^* dans l'équation (5.1) on obtient alors

$$Var(\bar{Y}) = \frac{\sigma_y^2 (1+\phi)}{n (1-\phi)} = \frac{\sigma^2}{n(1-\phi)^2}$$

qui représente la variance asymptotique de la moyenne en présence d'observations autocorrélées selon le modèle AR(1). Pour une démonstration, voir Laberge (1988).

Lettenmaier (1976) a ensuite montré qu'il est possible de connaître la puissance des tests de Spearman et de Mann-Whitney en présence d'autocorrélation. Il suffit d'utiliser le nombre effectif plutôt que le nombre d'observations autocorrélées. Ainsi, la puissance de ces tests avec n observations autocorrélées est égale à leur puissance avec n^* observations indépendantes.

Cependant, Lettenmaier (1976) ne discute que très peu des valeurs critiques à utiliser pour effectuer les tests de Spearman et de Mann-Whitney en présence d'autocorrélation. Il propose l'utilisation de valeurs critiques obtenues à l'aide de simulations de Monte Carlo. Lettenmaier présente des valeurs critiques pour quelques combinaisons de ϕ et de n , mais dans plusieurs cas l'utilisateur doit calculer ces propres valeurs critiques, à l'aide de simulation, avant de pouvoir effectuer le test. En pratique, l'utilisation de telles valeurs critiques devient donc rapidement un incitatif à laisser tomber ce test.

Afin de pouvoir effectuer le test de Spearman, en présence d'autocorrélation, sans l'obligation d'utiliser des valeurs critiques obtenues par simulation, une adaptation intuitive est maintenant présentée. Les développements proposés par Lettenmaier (1976) dans le cas du test t , sont alors généralisés au test de Spearman. Cette adaptation intuitive consiste donc à remplacer le nombre d'observations autocorrélées par le nombre effectif d'observation dans le calcul de la statistique utilisée dans le test de Spearman.

Ainsi, pour tester $H_0 : \rho = 0$ vs $H_1 : \rho \neq 0$, la statistique :

$$t = \frac{r_s \sqrt{(n-2)}}{\sqrt{(1-r_s^2)}}$$

est utilisée en présence d'observations indépendantes. L'adaptation intuitive est alors de diviser cette statistique par le ratio $\sqrt{n} / \sqrt{n^*}$ et de rejeter H_0 si :

$$t = \left| \frac{r_s \sqrt{n^*(n-2)}}{\sqrt{(1-r_s^2)} \sqrt{n}} \right| > t_{1-\alpha/2, n-2}$$

Il faut noter ici que plusieurs points doivent être validés avant d'utiliser cette adaptation avec de petits échantillons. En effet, il pourrait être préférable de diviser la statistique t par $\sqrt{n-2} / \sqrt{n^*-2}$ et il pourrait également être préférable de comparer la nouvelle statistique aux percentiles de la loi de Student à $(n^* - 2)$ degrés de liberté. Cependant, ces points ne seront pas discutés ici, puisque les séries traitées ont plus de 300 observations, ce qui

implique que l'utilisation de l'une ou l'autre des différentes possibilités n'affectera pas significativement les résultats du test.

Une simulation de Monte Carlo permet de valider l'utilisation du nombre effectif d'observations dans la construction du test. En effet, en simulant 500 séries autocorrélées avec $\phi = 0,75$ et $n=300$, on obtient un seuil empirique de 0,048 qui n'est pas significativement différent du seuil théorique de 5%. En introduisant en plus une tendance monotone de pente $\Delta_m=0,002$, on obtient une puissance empirique de 0,848 qui n'est pas significativement différente de la puissance théorique de 0,85. On peut donc conclure que l'utilisation du nombre effectif dans la construction de la statistique de Spearman/Lettenmaier permet un test avec des propriétés désirables: seuil conservé et puissance connue.

5.1.1 Comparaison des seuils en présence d'IO.

Les prochaines sections permettent maintenant de comparer les résultats du test de Spearman/Lettenmaier à ceux du test utilisant la régression autorégressive robuste (i.e. une régression utilisant les M-Estimeurs et une variable avec retard) si on introduit des valeurs aberrantes en plus de l'autocorrélation. Dans un premier temps, la simulation 27 compare le seuil des tests en présence d'IO.

SIMULATION 27

CONDITIONS EXPÉRIMENTALES

Titre : Comparaison avec Spearman/Lettenmaier en présence d'autocorrélation et de valeurs aberrantes.

Nombre de séries simulées: 500 X **Longueur des séries:** 300.

Modèle sous-jacent : AR(1) avec $\phi_1 = 0,75$ et $\sigma^2 = 1$.

Plusieurs simulations sont effectuées avec ces conditions expérimentales. Dans un premier temps, on cherche à comparer le seuil des tests avec différents emplacements et amplitudes des IO, tout en gardant le niveau d'autocorrélation à 0,75. Le tableau 5.1 présente les résultats de ces simulations pour: a) le test utilisant la régression autorégressive classique; b) le test utilisant la régression autorégressive robuste et; c) le test de Spearman/Lettenmaier.

Rappelons que, pour tenir compte de l'autocorrélation, une variable avec retard est utilisée dans les modèles de régressions autorégressives classiques et robustes. Dans le cas du test de Spearman/Lettenmaier, on utilise l'estimateur classique du coefficient d'autorégression pour effectuer le test.

Tableau 5.1
Comparaison des seuils des différents tests
de détection de tendance en présence d'IO et d'autocorrélation.

Type de IO	RÉG. AUTO. CLASSIQUE			RÉG. AUTO. ROBUSTE			SPEARMAN/LET.		
	b_1	d_1	$\hat{\alpha}$	b_{1M}	d_{1M}	$\hat{\alpha}$	b_{1NP}	d_{1NP}	$\hat{\alpha}$
$\mu=10, \sigma=3, t_0=45$	0,0004*	0,715*	0,016*	0,00004	0,740*	0,070	-	0,715*	0,070
$\mu=5, \sigma=2, t_0=45$	0,0002*	0,726*	0,046	0,00007	0,738*	0,068	-	0,726*	0,054
$\mu=0, \sigma=3, t_0=45$	0,00002	0,732*	0,066	0,00000	0,734*	0,062	-	0,732*	0,066
$\mu=10, \sigma=3, t_0=2$	-0,0008*	0,714*	0,040	-0,0001*	0,738*	0,054	-	0,714*	0,070
$\mu=5, \sigma=2, t_0=2$	-0,0004*	0,723*	0,066	-0,0002*	0,735*	0,082*	-	0,723*	0,086*
$\mu=0, \sigma=3, t_0=2$	-0,00002	0,732*	0,080*	-0,00001	0,734*	0,074	-	0,732*	0,088*

* Significativement différent de la valeur théorique au seuil 5%.

En ce qui concerne les résultats pour les tests utilisant les régressions autorégressives classiques et robustes, tout concorde avec les développements du chapitre 3. En comparant les seuils du test utilisant les M-Estimeurs et le test de Spearman/Lettenmaier on s'aperçoit que :

1. Pour les M-Estimeurs, les seuils empiriques ne sont pas significativement différents de 5% dans 5 cas sur 6. Dans le cas $\mu_{IO}=5, \sigma_{IO}=2$ et $t_{IO,\sigma}=2$, le seuil

empirique de 0,082 est significativement plus grand que 5%. Les quelques tendances détectées en trop peuvent être introduites par l'emplacement des valeurs aberrantes au début des séries simulées. On remarque cependant que l'amplitude et l'emplacement des valeurs aberrantes ont très peu d'influence sur le seuil de ce test.

2. Pour le test de Spearman/Lettenmaier, les seuils empiriques ne sont pas significativement différents de 5% dans 4 cas sur 6. Dans le cas $\mu_{IO} = 5, \sigma_{IO} = 2$ et $t_{IO,0} = 2$, tout comme pour le test utilisant les estimateurs robustes, le seuil empirique (0,086) est significativement plus grand que 5%. Cependant, le test de Spearman/Lettenmaier montre également un seuil empirique significativement plus grand que 5% dans le cas $\mu_{IO} = 0, \sigma_{IO} = 3$ et $t_{IO,0} = 2$. Les quelques tendances détectées en trop peuvent de nouveau être introduites par l'emplacement des valeurs aberrantes au début des séries simulées. On voit donc que l'emplacement des valeurs aberrantes semble avoir une plus grande influence sur le seuil de ce test comparativement au test utilisant les estimateurs robustes.

5.1.2 Comparaison des puissances en présence d'IO.

L'étude de comparaison des puissances qui suit confirme l'influence de l'emplacement des valeurs aberrantes sur le test de Spearman/Lettenmaier. Nous sommes encore dans les conditions de la simulation 27, avec l'introduction d'IO et d'une tendance monotone de pente $\Delta_m = 0,002$. Le tableau 5.2 présente les résultats de: a) le test utilisant la régression autorégressive classique; b) le test utilisant la régression autorégressive robuste et; c) le test de Spearman/Lettenmaier.

On remarque d'abord que les résultats pour les tests utilisant les estimateurs classiques et robustes concordent, de nouveau, avec les développements du chapitre 3. Le test utilisant la régression autorégressive robuste se présente de loin comme le meilleur des deux, à cause de ces estimateurs ponctuels qui estiment les paramètres du processus sous-jacent et la faible influence de l'emplacement des IO sur le seuil et la puissance du test.

Tableau 5.2
 Comparaison des puissances des différents tests
 de détection de tendance en présence d'IO et d'autocorrélation.

Type de IO	RÉG. AUTO. CLASSIQUE			RÉG. AUTO. ROBUSTE			SPEARMAN/LET.		
	b_1	d_1	$(1 - \hat{\beta})$	b_{1M}	d_{1M}	$(1 - \hat{\beta})$	b_{1NP}	d_{1NP}	$(1 - \hat{\beta})$
$\mu=10, \sigma=3, t_0 = 45$	0,0027*	0,717*	0,610*	0,0022*	0,741*	0,790*	-	0,717*	0,862
$\mu=5, \sigma=2, t_0 = 45$	0,0024*	0,723*	0,772*	0,0022*	0,734*	0,754*	-	0,723*	0,852
$\mu=0, \sigma=3, t_0 = 45$	0,0021*	0,733*	0,798*	0,0021*	0,735*	0,766*	-	0,733*	0,838
$\mu=10, \sigma=3, t_0 = 2$	0,0014*	0,716*	0,106*	0,0019*	0,743*	0,714*	-	0,716*	0,514*
$\mu=5, \sigma=2, t_0 = 2$	0,0018*	0,728*	0,514*	0,0020	0,736*	0,726*	-	0,728*	0,672*
$\mu=0, \sigma=3, t_0 = 2$	0,0022*	0,734*	0,808*	0,0022*	0,736*	0,786*	-	0,734*	0,844

* Significativement différent de la valeur théorique au seuil 5%.

En comparant les puissances du test utilisant les M-Estimeurs (régression autorégressive robuste) et le test de Spearman/Lettenmaier on note que :

1. Pour les M-Estimeurs, les puissances empiriques sont significativement plus faibles que la puissance théorique (0,85) dans tous les cas. En comparant avec les résultats obtenus à l'aide des estimateurs classiques (puissance empirique variant de 0,106 à 0,808), on remarque cependant que l'amplitude et l'emplacement des valeurs aberrantes ont très peu d'influence sur la puissance de ce test (puissance empirique variant de 0,714 à 0,790).
2. Pour le test de Spearman/Lettenmaier, les puissances empiriques ne sont pas significativement différentes de 85% dans 4 cas sur 6. Il faut cependant faire très attention à ce résultat puisque, dans un des cas où la puissance n'est pas significativement différente de la puissance théorique, le seuil était significativement plus grand que le seuil théorique, ce qui enlève toute valeur à la puissance. De plus, on voit que l'emplacement des valeurs aberrantes a une influence assez importante sur la puissance de ce test. Par exemple, dans le cas $\mu_{IO} = 10$ et $\sigma_{IO} = 3$, la puissance passe de 0,862 à 0,514 selon que les valeurs aberrantes se trouvent à la fin ou au début de la série.

3. En présence de valeurs aberrantes de moyenne nulle, les trois tests sont peu influencés par l'emplacement des valeurs aberrantes. On peut alors voir que le test de Spearman/Lettenmaier a une puissance significativement plus élevée que le test utilisant les M-Estimeurs.

Cette comparaison en présence d'IO et d'autocorrélation, entre le test utilisant la régression autorégressive robuste et le test de Spearman/Lettenmaier montre que :

1. Le test de Spearman/Lettenmaier est fortement influencé par l'emplacement des IO de moyenne non nulle mais pas par l'amplitude de ces dernières.
2. Le test utilisant les estimateurs robustes est beaucoup moins influencé par l'emplacement des valeurs aberrantes tout en demeurant peu influencé par l'amplitude des IO de moyenne non nulle.
3. Le test utilisant les estimateurs robustes est donc préférable si on ne connaît pas l'emplacement des valeurs aberrantes.
4. L'utilisation conjointe des deux tests (M-Estimeurs et Spearman/Lettenmaier), bien qu'encombrant, peut permettre des conclusions plus fiables.
5. En présence d'IO de moyenne nulle, le test de Spearman/Lettenmaier est préférable, car sa puissance est plus grande que celle du test utilisant les estimateurs robustes et qu'il est peu influencé par l'emplacement des valeurs aberrantes. Cependant, la majorité des séries dans le domaine de l'eau (concentrations et charges) ne contiennent pas de valeurs négatives; l'intérêt pratique d'IO de moyenne nulle est donc beaucoup plus faible que celui d'IO de moyenne non nulle. De plus, il faut faire attention dans l'application du test de Spearman/Lettenmaier en présence d'IO de moyenne nulle car, même si le seuil est conservé en moyenne pour les 500 séries simulées, on peut quand même avoir des séries où l'emplacement des IO a un effet sur le résultat du test.

Tous ces résultats montrent l'intérêt d'utiliser la régression, même si ce n'est que pour étudier les résidus afin de voir l'effet possible de l'emplacement des valeurs aberrantes.

5.1.3 Étude de seuil en présence d'AO.

Les conditions expérimentales sont identiques à celle de la simulation 27, sauf pour la présence de valeurs aberrantes statiques (AO) au lieu des IO. On cherche à s'assurer que l'adaptation du test de Spearman conserve un seuil α , tout en comparant cette adaptation avec le test utilisant la régression autorégressive robuste. Comme dans la section 5.1.1, 6 simulations sont effectuées: les combinaisons de deux emplacements et de trois amplitudes de valeurs aberrantes. Le tableau 5.3 présente les résultats de ces simulations pour les tests de régression autorégressive classique et robuste ainsi que le test de Spearman/Lettenmaier.

Tableau 5.3
Comparaison des seuils des différents tests
de détection de tendance en présence d'AO et d'autocorrélation.

Type de AO	RÉG. AUTO. CLASSIQUE			RÉG. AUTO. ROBUSTE			SPEARMAN/LET.		
	b_1	d_1	$\hat{\alpha}$	b_{1M}	d_{1M}	$\hat{\alpha}$	b_{1NP}	d_{1NP}	$\hat{\alpha}$
$\mu=10, \sigma=3, t_0=45$	0,0001*	0,402*	0,148*	-0,00002	0,612*	0,184*	-	0,402*	0,268*
$\mu=5, \sigma=2, t_0=45$	0,00009	0,606*	0,120*	0,00003	0,677*	0,106*	-	0,606*	0,146*
$\mu=0, \sigma=3, t_0=45$	0,00005	0,688*	0,064	0,00004	0,705*	0,070	-	0,688*	0,072
$\mu=10, \sigma=3, t_0=2$	-0,0006*	0,373*	0,154*	-0,00006	0,593*	0,168*	-	0,373*	0,270*
$\mu=5, \sigma=2, t_0=2$	-0,0002*	0,590*	0,094*	-0,00002	0,669*	0,104*	-	0,590*	0,118*
$\mu=0, \sigma=3, t_0=2$	-0,00003	0,678*	0,057	-0,00002	0,700*	0,069	-	0,678*	0,067

* Significativement différent de la valeur théorique au seuil 5%.

En ce qui concerne les résultats pour les tests utilisant les estimateurs classiques et robustes, tous concordent avec les développements du chapitre 3 (seuil non conservé en présence d'AO de moyenne non nulle). En comparant les seuils du test utilisant les M-Estimateurs et le test de Spearman/Lettenmaier on s'aperçoit que :

1. Pour les AO de moyenne nulle, les deux tests présentent des seuils empiriques qui ne sont pas significativement différents de 5%.

2. Lorsque les AO ont une moyenne non nulle, les deux tests présentent des seuils empiriques significativement plus grands que 5%. La sous-estimation du coefficient d'autorégression est à l'origine de ce trop grand nombre de détections de tendance en absence de tendance.
3. L'emplacement des valeurs aberrantes au début des séries simulées a une influence beaucoup plus faible qu'en présence d'IO, ce qui s'explique par la présence d'une seule observation élevée pour les AO contre plusieurs observations élevées (à cause de la propagation) pour les IO.

Devant ces comparaisons, il est clair qu'aucun test ne permet une détection de tendance adéquate en présence d'AO de moyenne non nulle, la sous-estimation du coefficient d'autorégression empêche la construction d'une statistique valable pour la détection de tendance.

L'utilisation des GM-Estimeurs (Denby et Martin 1979, Martin et Yohai 1985, Laberge, 1988) permet de diminuer le biais dans l'estimation du coefficient d'autorégression. Cependant, le GM-Estimeur pour le coefficient d'autorégression est également biaisé et son utilisation dans les modèles de régression linéaire est un sujet peu documenté qui dépasse les limites du présent travail.

Finalement, notons que l'utilisation du GM-Estimeur pour le coefficient d'autorégression permettrait sûrement un meilleur test de Spearman/Lettenmaier que dans le cas où l'on utilise l'estimateur classique, mais il n'est pas certain que cela serait suffisant pour rendre le test applicable en présence d'AO de moyenne non nulle.

En présence d'AO de moyenne nulle, l'étude de seuil montre que les deux tests conservent le seuil de 5%. La section suivante permet de voir comment se comporte la puissance.

5.1.4 Étude de puissance en présence d'AO.

Le tableau 5.4 présente les résultats des simulations pour la comparaison de puissance des différents tests en présence d'AO de moyenne nulle et d'autocorrélation. Les cas avec des AO de moyenne non nulle ne sont pas étudiés, car les différents tests ne conservent pas le seuil de 5%.

Tableau 5.4
Comparaison des puissances des différents tests
de détection de tendance en présence d'AO de moyenne nulle et d'autocorrélation.

Type de AO	RÉG. AUTO. CLASSIQUE			RÉG. AUTO. ROBUSTE			SPEARMAN/LET.		
	b_1	d_1	$(1 - \hat{\beta})$	b_{IM}	d_{IM}	$(1 - \hat{\beta})$	b_{INP}	d_{INP}	$(1 - \hat{\beta})$
$\mu=0, \sigma=3, t_0 = 45$	0,0029*	0,652*	0,890	0,0025*	0,694*	0,813	-	0,652*	0,894
$\mu=0, \sigma=3, t_0 = 2$	0,0029*	0,649*	0,868	0,0025*	0,692*	0,838	-	0,649*	0,932*

* Significativement différent de la valeur théorique au seuil 5%.

Les résultats pour les tests utilisant les estimateurs classiques et robustes concordent, de nouveau, avec les développements du chapitre 3. En comparant les puissances du test utilisant les M-Estimeurs et le test de Spearman/Lettenmaier on note que :

1. Pour les M-Estimeurs, les puissances empiriques ne sont pas significativement différentes de la puissance théorique (0.85) et ces puissances empiriques ne varient pas significativement malgré différents emplacements de valeurs aberrantes.
2. Pour le test de Spearman/Lettenmaier, la puissance empirique est significativement plus grande que la puissance théorique, à cause de la sous-estimation du coefficient d'autorégression.

Cette comparaison en présence d'AO de moyenne nulle et d'autocorrélation, entre le test utilisant les M-Estimeurs et le test de Spearman/Lettenmaier montre que :

1. Le test de Spearman/Lettenmaier est influencé par la sous-estimation du coefficient d'autorégression. L'utilisation de l'estimateur classique explique cette influence importante. Un autre estimateur est donc primordial pour que le test de Spearman/Lettenmaier puisse être utilisé en présence d'AO; ce sujet ne sera cependant pas approfondi ici.
2. Le test utilisant les estimateurs robustes apparaît donc préférable en présence d'AO de moyenne nulle et d'amplitude raisonnable.

Ces comparaisons permettent de conclure qu'en présence d'AO de moyenne nulle et d'autocorrélation, l'utilisation de la régression autorégressive robuste permet un test adéquat. Le test de Spearman/Lettenmaier, utilisé dans le domaine de l'eau, présente des faiblesses évidentes et ne devrait pas être utilisé à moins d'obtenir une bonne estimation du coefficient d'autorégression.

En présence d'AO de moyenne non nulle, aucun des tests étudiés ne présente des propriétés désirables, il est donc très hasardeux d'effectuer une détection de tendance dans cette situation. Il est évident que dans cette situation on a besoin de recherches complémentaires pour effectuer une détection de tendance adéquate.

5.2 Tests non paramétriques saisonniers.

Le test de Mann-Kendall (Mann, 1945; Kendall, 1975) est une application du test de Kendall pour les corrélations, cette application est généralement appelée Tau de Kendall. On teste alors l'hypothèse H_0 : les observations forment un échantillon de n variables aléatoires indépendantes et identiquement distribuées contre l'hypothèse alternative bilatérale H_1 : les distributions de x_k et x_j ne sont pas identiques pour tous $k, j \leq n$ avec $k \neq j$. La statistique du test est alors définie par :

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^n \operatorname{sgn}(x_j - x_k)$$

où:

$$\text{sgn}(\theta) = \begin{cases} 1 & \text{si } \theta > 0 \\ 0 & \text{si } \theta = 0 \\ -1 & \text{si } \theta < 0 \end{cases}$$

On peut montrer (Kendall, 1975) que la moyenne et la variance de S sous H_0 sont respectivement : $E[S] = 0$ et $\text{Var}[S] = n(n-1)(2n+5)/18$ et que la distribution de S est symétrique et tend vers une loi normale lorsque n tend vers l'infini. Notons finalement qu'il est facile d'ajuster la variance de S pour tenir compte de la présence de valeurs identiques (Hirsch, Slack et Smith, 1982; Kendall 1975). Pour $n < 11$ Mann (1945) et Kendall (1975) présentent la distribution exacte de S , alors que pour $n > 10$ les deux auteurs ont montré que l'approximation normale utilisant un facteur de correction de 1 est excellente. Comme les cas qui nous intéressent dans le présent travail supposent n grand, on s'attardera seulement à l'approximation normale: on calcule alors la statistique centrée réduite suivante:

$$Z = \begin{cases} \frac{S-1}{(\text{Var}(S))^{1/2}} & \text{si } S > 0 \\ 0 & \text{si } S = 0 \\ \frac{S+1}{(\text{Var}(S))^{1/2}} & \text{si } S < 0 \end{cases}$$

et on rejette H_0 si $|Z| > z_{1-\alpha/2}$, où $z_{1-\alpha/2}$ est le $(1-\alpha/2)$ e percentile de la distribution normale centrée réduite, pour obtenir un test au seuil de signification 5%.

On remarque facilement l'intérêt de ce test pour la détection de tendance car, en présence d'une tendance positive (augmentation du niveau de la série), on aura plusieurs cas où $x_k > x_j$ pour $k > j$ ce qui amènera une valeur de S élevée. Une valeur positive de S implique donc une tendance positive, alors qu'une valeur négative de S est associée à une tendance négative (diminution du niveau de la série dans le temps).

Afin de permettre la détection de tendance en présence de saisonnalités, Hirsch, Slack et Smith (1982) ont adapté ce test en un test saisonnier. On calcule alors une statistique S pour chacune des " k " saisons et on somme ces statistiques saisonnières pour

obtenir la statistique globale qui permet de tester la présence de tendance dans l'ensemble de la série. Comme la distribution asymptotique de chaque statistique saisonnière est une loi normale dont on connaît la moyenne et la variance, il est très simple de connaître la distribution asymptotique de la somme de ces statistiques. Soit S^* la somme des k statistiques saisonnières, on a alors:

$$\begin{aligned}
 S^* &= \sum_{i=1}^k S_i \\
 E[S^*] &= \sum_{i=1}^k E[S_i] = 0 \\
 \text{Var}[S^*] &= \sum_{i=1}^k \text{Var}[S_i] + 2 \sum_{g=1}^k \sum_{h=g+1}^k \text{cov}(S_g, S_h) \quad (5.1)
 \end{aligned}$$

En absence d'autocorrélation les termes de covariance sont nuls et la variance de S^* est très simple à obtenir en utilisant le calcul de variance de S . Le test de Mann Kendall saisonnier est alors effectué en utilisant l'approximation normale avec facteur de correction de 1:

$$Z = \begin{cases} \frac{S^*-1}{(\text{Var}(S^*))^{1/2}} & \text{si } S^* > 0 \\ 0 & \text{si } S^* = 0 \\ \frac{S^*+1}{(\text{Var}(S^*))^{1/2}} & \text{si } S^* < 0 \end{cases}$$

HIRSCH, SLACK et SMITH (1982) ont montré que le test de Kendall saisonnier conservait le seuil de 5% en présence de saisonnalités. La section 4.3.1 a également montré que le seuil empirique du test de détection de tendance utilisant la régression linéaire et les variables muettes n'était pas différent du seuil théorique de 5%. Devant ces résultats, aucune simulation n'est nécessaire pour comparer les seuils en présence de saisonnalités comme seule caractéristique restrictive.

Cependant, deux points doivent être étudiés dans la comparaison des tests de régression et de Kendall saisonnier. Dans un premier temps, on cherche à savoir si le test

de Kendall saisonnier est aussi affecté que le test utilisant la régression linéaire, en présence de décalages entre les saisons estimées et réelles. Dans un second temps, on s'attarde au problème de la présence de saisons contenant plusieurs observations à chaque année; on cherche alors à savoir s'il est préférable de prendre la moyenne mensuelle ou de garder toutes les observations de chaque mois pour effectuer le test de Kendall saisonnier.

HIRSCH, SLACK et SMITH (1982, p110) ont soulevé cette situation d'égalités dans la localisation d'observations de la série (i.e. plusieurs observations pour une même saison), mais les auteurs ne peuvent indiquer la méthode la plus appropriée pour tenir compte de ces égalités.

Les seuils et les puissances des tests de régression et de Kendall saisonnier seront donc comparés en présence de décalage des saisons et de plusieurs observations par saison. Dans ce dernier cas, deux formes de test de Kendall saisonnier seront utilisées: le test de Kendall saisonnier (KS) utilisera toutes les observations en traitant les égalités dans les deux séries (observations et localisation des observations), alors que le test de Kendall sur les moyennes saisonnières (KMS) utilisera les moyennes saisonnières, ce qui empêchera la présence d'égalité dans la série de localisation des observations.

5.2.1 Comparaison des seuils en présence de décalage.

En présence de décalage entre les saisons réelles et les saisons estimées, on a vu à la section 4.3.3 que les estimateurs ponctuels étaient significativement affectés et que le test de détection de tendance utilisant la régression linéaire devenait conservateur. La simulation 28 compare le seuil des tests de Kendall saisonniers (KS et KMS) au seuil conservateur du test de la régression saisonnière en présence de décalage des saisons. Le tableau 5.5 présente la comparaison des trois tests de détection de tendance.

SIMULATION 28

CONDITIONS EXPÉRIMENTALES

Titre : Comparaison avec les tests de Kendall saisonniers (seuil);
Présence de décalage des saisons.

Nombre de séries simulées: 500 X Longueur des séries: 300.

Modèle sous-jacent : $N(0,1)$ avec saisons;
 $n_1 = 12, \mu_1 = 5, n_2 = 13, \mu_2 = 3, n_3 = 12, \mu_3 = 1, n_4 = 13, \mu_4 = 0$

Tableau 5.5

Comparaison des seuils du test de régression saisonnière et des tests de Kendall saisonniers.

Décalage	Régression Saisonnière	Kendall Saisonnier	Kendall Moyennes Saisonnières
0	0,046	0,058	0,050
1	0,012*	0,064	0,044
2	0,004*	0,056	0,046

Le tableau 5.5 montre que les tests de Kendall saisonniers (KS et KMS) conservent le seuil de 5% malgré la présence de décalage entre les saisons estimées et les saisons réelles. On a vu à la section 4.3.3, que la raison principale amenant un seuil empirique plus petit que 5% pour le test de régression était une surestimation de la variance qui causait un plus petit nombre de rejets. L'étude des résidus de régression permet de constater que la surestimation de la variance est causée par la présence de résidus aberrants aux temps de décalage entre les saisons estimées et réelles.

Le seuil du test non paramétrique de Kendall saisonnier n'apparaît donc pas affecté par la présence de décalage des saisons. Une étude plus approfondie permet cependant de conclure que la présence de décalage a un effet sur le test de Kendall saisonnier. La structure des décalages introduits amène des valeurs plus grandes au début de trois saisons sur quatre (i.e. des valeurs de moyenne 5 dans la saison de moyenne 3, des valeurs de

moyenne 3 dans la saison de moyenne 1, des valeurs de moyenne 1 dans la saison de moyenne 0), alors que seulement une saison montre des valeurs plus faibles introduites par le décalage (des valeurs de moyenne 0 dans la saison de moyenne 5). Cette structure fait que le test de Kendall saisonnier devient asymétrique et que l'on détecte 31 tendances négatives sur 32 dans le cas du décalage de 1 unité, alors qu'on détecte 27 tendances négatives sur 28 dans le cas du décalage de 2 unités.

Notons finalement, que la présence de décalage systématique n'affecte pas le seuil du test non paramétrique de Kendall sur les moyennes saisonnières puisque les moyennes saisonnières sont toutes affectées de la même façon, d'une année à l'autre. Il serait intéressant d'évaluer l'effet de décalages non systématiques sur le seuil du test de Kendall sur les moyennes saisonnières. Ce sujet dépasse cependant les limites du présent travail.

5.2.2 Comparaison des puissances en présence de décalage.

En présence de décalage entre les saisons réelles et les saisons estimées, on a vu à la section 4.3.4 que les estimateurs ponctuels étaient significativement affectés et que le test de détection de tendance utilisant la régression linéaire perdait de la puissance. La simulation 29 compare la puissance des tests de Kendall saisonniers (KS et KMS) à la puissance du test de la régression saisonnière en présence de décalage des saisons.

Le tableau 5.6 compare les puissances des tests de régression et des tests de Kendall saisonniers (KS et KMS) en présence de saisonnalités. En absence de décalage (décalage égal à 0), on remarque que les tests de régression et de Kendall saisonnier présentent des puissances empiriques égales à la puissance théorique (85 %), alors que le test de Kendall sur les moyennes saisonnières présente une puissance significativement plus faible. Cette diminution de puissance, lorsqu'on utilise les moyennes saisonnières, n'est cependant pas discutée dans la littérature et il est hors des limites du présent travail de trouver les raisons de cette faible puissance; on peut cependant supposer que le petit nombre d'observations

utilisées (24) peut être la cause principale de ce résultat.

SIMULATION 29

CONDITIONS EXPÉRIMENTALES

Titre : Comparaison avec les tests de Kendall saisonniers (puissance);
Présence de décalage des saisons.

Nombre de séries simulées: 500 X Longueur des séries: 300.

Modèle sous-jacent : $N(0,1)$ avec saisons;

$$n_1 = 12, \mu_1 = 5, n_2 = 13, \mu_2 = 3, n_3 = 12, \mu_3 = 1, n_4 = 13, \mu_4 = 0$$

Tableau 5.6

Comparaison des puissances du test de régression saisonnière et des tests de Kendall saisonniers.

Décalage	Régression Saisonnière	Kendall Saisonnier	Kendall Moyennes Saisonnières
0	0,846	0,856	0,698*
1	0,660*	0,660*	0,708*
2	0,502*	0,488*	0,674*

Lorsqu'un décalage systématique est introduit entre les saisons estimées et réelles on s'aperçoit que la puissance des tests de régression et de Kendall saisonnier diminue significativement. La surestimation de la variance (présence de valeurs aberrantes) amène la plus faible puissance du test de régression alors que la structure des décalages amène la plus faible puissance du test de Kendall saisonnier (i.e. que la présence de valeurs plus grandes au début de trois saisons sur quatre amène un plus faible nombre de rejets de H_0).

Le tableau 5.6 montre finalement que la puissance du test de Kendall sur les moyennes saisonnières reste stable avec l'introduction d'un décalage. Ce dernier résultat s'explique facilement par le fait que les moyennes saisonnières sont toutes affectées de la même façon d'une année à l'autre. Comme dans le cas du seuil, une étude complémentaire

est nécessaire pour évaluer l'effet d'un décalage non systématique sur la puissance du test de Kendall sur les moyennes saisonnières.

La comparaison des tests de régression et des tests de Kendall saisonniers (KS et KMS) permet donc de conclure que:

1. Le test de régression utilisant les variables muettes est équivalent au test de Kendall saisonnier en présence de saisonnalités comme seule caractéristique restrictive.
2. Le test de Kendall sur les moyennes saisonnières est moins puissant que les deux autres en présence de saisonnalités comme seule caractéristique restrictive.
3. Les seuils des tests de régression et de Kendall saisonnier sont affectés par la présence de décalage systématique entre les saisons estimées et réelles.
4. Les puissances des tests de régression et de Kendall saisonnier diminuent significativement avec l'introduction de décalage systématique entre les saisons estimées et réelles.
5. Le test de Kendall sur les moyennes saisonnières n'est pas affecté par l'introduction de décalage systématique entre les saisons estimées et réelles.

Devant ces résultats, le choix du meilleur test en présence de saisonnalités comme seule caractéristique restrictive n'est pas évident. En pratique, la présence de décalage entre les saisons estimées et réelles est très probable, mais la présence de décalage "systématique" est très improbable. Devant ces considérations pratiques, il est impossible de choisir une forme ou l'autre du test de Kendall saisonnier pour une utilisation systématique quelle que soit la situation.

L'avantage du test de régression utilisant les variables muettes réside alors dans la possibilité d'évaluer la structure des décalages à l'aide de l'étude des résidus. La présence de valeurs aberrantes près des points de changement de saison permet d'identifier des décalages amenant des problèmes dans l'exécution du test de détection de tendance. Une telle étude est impossible avec les tests non paramétriques.

Il apparaît intéressant, pour fin de validation, d'effectuer les deux types de tests et de comparer les résultats. La connaissance des forces et des faiblesses des deux types de tests ainsi que l'étude des résidus de la régression permettent alors une conclusion beaucoup plus éclairée.

5.3 Test d'Hirsch et Slack en présence de saisonnalités et d'autocorrélation.

La section 5.2 a présenté le test de de Mann-Kendall (MANN ,1945; KENDALL, 1975) et l'adaptation qu'en ont fait HIRSCH, SLACK et SMITH (1982) pour la détection de tendance dans des séries chronologiques contenant des saisons. HIRSCH et SLACK (1984) ont complété l'étude du test de Mann-Kendall pour une utilisation en présence de saisonnalités et d'autocorrélation; le nouveau test qui en découle sera appelé test d'Hirsch et Slack dans les paragraphes suivants.

Le test d'Hirsch et Slack est donc également une application du test de Kendall pour les corrélations. On teste l'hypothèse H_0 : les observations forment un échantillon de n variables aléatoires indépendantes et identiquement distribuées contre l'hypothèse alternative bilatérale H_1 : les distributions de x_k et x_j ne sont pas identiques pour tous $k, j \leq n$ avec $k \neq j$. La différence entre le test d'Hirsch et Slack et le test de Kendall saisonnier réside dans l'estimation des covariances présentes dans l'équation (5.1), puisqu'on ne peut plus supposer que les statistiques saisonnières sont indépendantes.

Hirsch et Slack (1984, p. 728) présentent la façon d'estimer ces covariances:

$$\text{Cov}(S_g, S_h) = K_{gh}/3 + (n^3 - n)r_{gh}/9$$

où:

$$K_{gh} = \sum_{i < j} \text{sgn}[(x_{ig} - x_{ig})(x_{jh} - x_{ih})]$$

$$r_{gh} = \frac{3}{n^3 - n} \sum_{i,j,k} [(X_{ig} - X_{ig})(X_{jh} - X_{kh})]$$

Les auteurs notent qu'en absence d'égalités et de valeurs manquantes, r_{gh} est le coefficient de corrélation de Spearman pour les saisons g et h. En utilisant ces estimateurs de $\text{Cov}(S_g, S_h)$ dans le calcul de la variance de S^* (réf. équation 5.1), on obtient un test qui ne repose pas sur l'hypothèse d'indépendance des observations.

Une étude de simulation (Hirsch et Slack, 1984) montre que le seuil de ce test est conservé pour des séries contenant au moins 10 années de 12 saisons (120 valeurs) et pour un niveau d'autocorrélation inférieur ou égale à 0,60. Les auteurs ont également montré que ce test est moins puissant que le test de Kendall saisonnier en présence de données indépendantes.

Les prochains paragraphes présentent une comparaison (seuil et puissance) du test d'Hirsch et Slack et du test de détection de tendance utilisant la régression autorégressive saisonnière. Notons qu'aucune discussion dans l'article d'Hirsch et Slack n'explique le traitement possible d'égalités dans la série de localisation des observations (i.e. plusieurs observations dans la même saison). Comme le développement de tests non paramétriques n'est pas l'objet du présent travail, on ne s'attardera ici qu'au traitement utilisant une seule observation par saison: la moyenne.

Comme HIRSCH et SLACK (1984) ont montré que leur test ne conservait pas toujours le seuil de 5% en présence de saisonnalités et d'autocorrélation, il est nécessaire d'effectuer une simulation avec les caractéristiques des séries étudiées dans le chapitre sur les applications (6 années avec 4 moyennes saisonnières par année) pour savoir si le seuil est conservé dans cette situation. La section 4.4.1 a montré que le seuil du test de détection de tendance utilisant la régression autorégressive saisonnière n'était pas différent du seuil théorique de 5%.

La comparaison des seuils des tests de régression et d'Hirsch et Slack se fera également en présence de décalage systématique. En ce qui concerne les puissances des tests, les comparaisons seront faites dans plusieurs situations: a) présence de saisonnalités comme seule caractéristique restrictive; b) présence de saisonnalités et d'autocorrélation comme seules caractéristiques restrictives; c) présence de saisonnalités, d'autocorrélation et de décalage dans les saisons.

5.3.1 Comparaison des seuils.

En présence de décalage entre les saisons réelles et les saisons estimées, on a vu à la section 4.3.3 que les estimateurs ponctuels étaient significativement affectés et que le test de détection de tendance utilisant la régression linéaire devenait très conservateur. La simulation 30 compare le seuil du test d'Hirsch et Slack au seuil du test de la régression autorégressive saisonnière avec et sans décalage des saisons. Le tableau 5.7 présente les résultats de cette comparaison.

SIMULATION 30

CONDITIONS EXPÉRIMENTALES

Titre : Comparaison avec le test d'Hirsch et Slack (seuil);
Présence de décalage des saisons.

Nombre de séries simulées: 500 X Longueur des séries: 300.

Modèle sous-jacent : AR(1) avec $\phi = 0,75$ et saisons;

$$n_1 = 12, \mu_1 = 5, n_2 = 13, \mu_2 = 3, n_3 = 12, \mu_3 = 1, n_4 = 13, \mu_4 = 0$$

Tableau 5.7
 Comparaison des seuils du test de régression autorégressive
 saisonnière et du test d'Hirsch et Slack.

Décalage	Régression	Hirsch et Slack
0	0,056	0,036
2	0	0,038

Le tableau 5.7 permet de constater que le test de régression ne conserve pas le seuil de 5% en présence de décalage systématique entre les saisons estimées et réelles. Comme on l'a vu à la section 4.4.1, l'effet supplémentaire de la présence d'autocorrélation sur les estimateurs ponctuels rend le test de régression très conservateur. L'utilisation de la régression en présence de saisonnalités, d'autocorrélation et de décalage des saisons ne permet donc pas de détecter suffisamment de tendances significatives; on sous-évalue alors le nombre de tendances significatives dans les études utilisant la régression.

Le tableau 5.7 permet aussi de conclure que le seuil empirique du test d'Hirsch et Slack n'est pas significativement différent de 5%. La structure des séries (6 années avec 4 moyennes saisonnières par année) n'amène donc pas le test à être conservateur. Les seuils empiriques du tableau 5.7 sont comparables aux seuils obtenus par Hirsch et Slack (1984) dans les mêmes conditions. En effet, les auteurs ont obtenu des seuils empiriques de 0,014 pour 5 ans et $\phi=0,6$ et 0,056 pour 10 ans et $\phi=0,6$. De plus, comme le seuil empirique augmentait avec ϕ , le seuil empirique obtenu de 0,036 avec 6 ans et $\phi=0,75$ concorde avec les résultats d'Hirsch et Slack (1984). Dans les applications, on peut donc avoir une bonne idée du seuil du test d'Hirsch et Slack en utilisant les résultats de simulations de ces deux auteurs. Comme on travaillera toujours avec 6 années de données, on doit donc conclure que le test d'Hirsch et Slack aura tendance à être conservateur, en particulier pour des faibles valeurs de ϕ .

Devant les résultats du tableau 5.7, on pourrait être enclin à ne pas utiliser la régression linéaire en présence de décalage; rappelons cependant que la présence de décalage non systématique (le cas le plus probable en pratique) peut affecter le test d'Hirsch

et Slack. L'étude des résidus de la régression permet alors de détecter les valeurs aberrantes introduites par le décalage et d'ajuster notre estimation des saisons pour synchroniser les saisons réelles et estimées.

5.3.2 Comparaison des puissances.

Présence de saisonnalités comme seule caractéristique restrictive.

Hirsch et Slack (1984) expliquent que leur test est moins puissant que le test de Kendall saisonnier en présence de données indépendantes. À l'aide d'une simulation, ils ont illustré l'amplitude de la perte de puissance pour différents pourcentages de tendances. Cependant, la puissance de ce test est très peu documentée dans la littérature et il apparaît important de s'y attarder quelque peu dans les paragraphes suivants. La simulation 31 permet de comparer la puissance du test d'Hirsch et Slack avec celle du test de régression en présence de saisonnalités comme seule caractéristique restrictive.

SIMULATION 31

CONDITIONS EXPÉRIMENTALES

Titre : Comparaison avec le test d'Hirsch et Slack (puissance);
Présence de saisons comme seule caractéristique restrictive.

Nombre de séries simulées: 500 X Longueur des séries: 300.

Modèle sous-jacent : $N(0,1)$ avec saisons:

$$n_1 = 12, \mu_1 = 5, n_2 = 13, \mu_2 = 3, n_3 = 12, \mu_3 = 1, n_4 = 13, \mu_4 = 0$$

La puissance du test d'Hirsch et Slack dans les conditions expérimentales de la simulation 31 est de 0,39. Ce résultat montre une puissance très faible si l'on compare au 0,85 des tests de régression et de Kendall saisonnier dans les mêmes conditions. Il est important de noter que le fait d'utiliser les moyennes des observations de chaque saison

pourrait être en partie la cause de cette faible puissance, comme ce fût le cas pour le test de Kendall sur les moyennes saisonnières. Une étude de puissance serait donc nécessaire pour évaluer la puissance du test d'Hirsch et Slack dans différentes situations, cette étude dépasse cependant les limites de ce travail. Il est cependant important de noter que dans les conditions de la simulation 31 (situation que l'on retrouve en pratique, réf. chapitre 6), le test d'Hirsch et Slack est beaucoup moins puissant que le test de régression et le test de Kendall saisonnier.

Présence de saisons et d'autocorrélation.

Hirsch et Slack (1984) discutent très peu de la puissance de leur test, même en présence de saisonnalités et d'autocorrélation (situation idéale pour ce test). La simulation 32 permet de comparer la puissance du test d'Hirsch et Slack avec celle du test de régression en présence de saisonnalités et d'autocorrélation.

SIMULATION 32

CONDITIONS EXPÉRIMENTALES

Titre : Comparaison avec le test d'Hirsch et Slack (puissance);
Présence de saisons et d'autocorrélation.

Nombre de séries simulées: 500 X Longueur des séries: 300.

Modèle sous-jacent : AR(1) avec $\phi = 0,75$ et saisons:

$$n_1 = 12, \mu_1 = 5, n_2 = 13, \mu_2 = 3, n_3 = 12, \mu_3 = 1, n_4 = 13, \mu_4 = 0$$

La puissance du test d'Hirsch et Slack dans les conditions expérimentales de la simulation 32 est de 050. Ce résultat montre une puissance faible, si l'on compare au 083 du test de régression (réf. section 4.4.1) dans les mêmes conditions expérimentales. Le test d'Hirsch et Slack n'en demeure pas moins préférable au test de Kendall saisonnier qui détecte beaucoup trop de tendances dans les mêmes conditions.

Encore une fois, il est important de noter que le fait d'utiliser les moyennes des observations de chaque saison peut être en partie la cause de la faible puissance du test d'Hirsch et Slack comparativement au test de régression. Ce résultat appuie donc la nécessité d'une étude de puissance du test d'Hirsch et Slack. Il est important de noter que dans les conditions de la simulation 32 (une autre situation souvent retrouvée en pratique; réf. chapitre 6), le test d'Hirsch et Slack est de nouveau moins puissant que le test de régression.

Présence de saisons, d'autocorrélation et de décalage.

Hirsch et Slack (1984) ne discutent pas de l'effet d'un décalage sur leur test. On a cependant noté à la section 4.4.2 que le test de régression est fortement affecté par la présence de décalage et d'autocorrélation. La simulation 33 permet de juger de l'effet d'un décalage systématique sur la puissance du test d'Hirsch et Slack comparativement à l'effet très important sur la puissance du test de régression.

SIMULATION 33

CONDITIONS EXPÉRIMENTALES

Titre : Comparaison avec le test d'Hirsch et Slack (puissance); Présence de saisons, d'autocorrélation et de décalage des saisons (2 unités).

Nombre de séries simulées: 500 X Longueur des séries: 300.

Modèle sous-jacent : AR(1) avec $\phi = 0,75$ et saisons:

$$n_1 = 12, \mu_1 = 5, n_2 = 13, \mu_2 = 3, n_3 = 12, \mu_3 = 1, n_4 = 13, \mu_4 = 0$$

La puissance du test d'Hirsch et Slack dans les conditions expérimentales de la simulation 33 est de 0,48. Ce résultat montre une puissance non significativement différente de la puissance obtenue à la simulation 32 (absence de décalage). Il apparaît donc que,

comme dans le cas du test de Kendall sur les moyennes saisonnières, le test de Hirsch et Slack n'est pas affecté par la présence d'un décalage systématique. Rappelons cependant les doutes soulevés en présence probable de décalages non systématiques. La section 4.4.2 a montré que le test de régression n'avait réussi à détecter aucune tendance dans les mêmes conditions expérimentales. De plus, il faut rappeler que le fait d'utiliser les moyennes des observations de chaque saison peut amener une puissance plus faible du test d'Hirsch et Slack. Le test d'Hirsch et Slack apparaît donc préférable au test de régression en présence de saisonnalités, d'autocorrélation et de décalage systématique.

D'un autre côté il faut rappeler l'effet inconnu d'un décalage non systématique sur le test d'Hirsch et Slack et la possibilité de détecter les décalages à l'aide d'une étude de résidus dans l'analyse de régression. Ces deux points font qu'il peut quand même être utile de considérer la détection de tendance à l'aide de la régression puisqu'il est impossible d'étudier la structure des décalages avec le test d'Hirsch et Slack.

5.4 Synthèse des comparaisons.

Dans les chapitres précédents on a étudié et comparé plusieurs tests pouvant être utilisés dans différentes situations. Les tableaux suivants présentent les meilleurs tests paramétriques et non paramétriques pour les différentes combinaisons de caractéristiques restrictives.

Présence d'autocorrélation seulement.

Test	Restrictions	Commentaires	Seuil	Puissance
Spearman/ Lettenmaier	Aucune, sauf peut-être pour ϕ très grand.	Estimation de ϕ est nécessaire.	Conservé avec adaptation.	Égale puissance théorique.
Régression autorégressive	ϕ très grand.	Surparamétrisation peut être utilisée si p est inconnu.	Conservé pour $\phi < 0,75$.	Égale puissance théorique pour $\phi < 0,75$.

En présence d'autocorrélation seulement les deux tests les mieux adaptés sont le test de Spearman/Lettenmaier (non paramétrique) et la régression autorégressive (paramétrique). Dans cette situation, les deux tests donneront toujours des résultats très semblables. De façon générale, il faut faire attention à la structure de la série, car les développements supposent un processus AR(1). Pour d'autres structures, il faut ajuster le modèle de régression ou le calcul de la variance du test de Spearman/Lettenmaier.

En présence d'autocorrélation et d'IO.

Test	Restrictions	Commentaires	Seuil	Puissance
Spearman/Lettenmaier	Utilisation d'un estimateur robuste de ϕ est fortement recommandé.	Affecté par l'emplacement des valeurs aberrantes.	Généralement conservé.	Instable.
Régression autorégressive robuste	ϕ très grand.		Conservé.	Légère perte de puissance.

Dans cette situation, le test utilisant la régression autorégressive robuste présente de nombreux avantages: il est moins affecté par l'emplacement des IO et l'étude des résidus permet de juger du type et de l'emplacement des valeurs aberrantes. Le test de Spearman/Lettenmaier peut quand même être utilisé mais il est alors important de s'assurer que l'emplacement des IO est adéquat et que l'on utilise un estimateur robuste pour ϕ .

En présence d'autocorrélation et d'AO.

Test	Restrictions	Commentaires	Seuil	Puissance
Spearman/Lettenmaier	AO de moyenne non nulle.	Affecté par l'emplacement des valeurs aberrantes. Plus sûr si AO de moyenne nulle.	Pas toujours conservé.	Instable.
Régression autorégressive robuste.	AO de moyenne non nulle.	Plus sûr si AO de moyenne nulle.	Pas toujours conservé.	Instable.

En présence d'AO de moyenne non nulle, les deux tests présentent des propriétés très indésirables: un seuil pas toujours conservé ainsi qu'une puissance qui dépend de l'emplacement et de la moyenne des AO. L'utilisation de GM-Estimeurs pourrait permettre d'améliorer la situation mais des études connexes montrent que ce n'est pas une solution parfaite. Dans cette situation, il faut donc être très prudent lors d'étude de détection de tendance, car aucun test n'apparaît satisfaisant en présence d'AO de moyenne non nulle.

En présence d'AO de moyenne nulle, les deux tests présentent des propriétés beaucoup plus désirables. Le test utilisant la régression autorégressive robuste présente alors les mêmes avantages qu'en présence d'IO, 1) faible influence de l'emplacement des valeurs aberrantes et; 2) possibilité d'effectuer une étude des résidus.

En présence de saisonnalités.

Test	Restrictions	Commentaires	Seuil	Puissance
Kendall Saisonnier	Emplacement des valeurs aberrantes et à la présence de décalage systématique.		Conservé.	Égale puissance théorique.
Kendall sur les Moyennes Saisonniers	Effet inconnu des décalages non systématiques.	Pas affecté par la présence de décalage systématique.	Conservé.	Moins puissant que les deux autres en absence de décalage.
Régression Saisonnaire	Présence de décalage systématique.	Étude des résidus permet de juger de la présence de décalages.	Conservé.	Égale puissance théorique.

En présence de saisonnalités, l'utilisation des deux types de test (paramétrique et non paramétrique) apparaît très avantageuse. L'étude des résidus de la régression permet de porter un jugement sur la qualité de la modélisation. La connaissance des forces et des faiblesses des deux types de test permet ensuite une conclusion éclairée dans l'éventualité où les deux types de test n'amènent pas la même conclusion.

En présence de saisonnalités et d'autocorrélation.

Test	Restrictions	Commentaires	Seuil	Puissance
Hirsch et Slack sur les moyennes saisonnières	Test conservateur en présence de petits échantillons. Effet inconnu des décalages non systématiques.		Conservé.	Plus faible que puissance théorique.
Régression Saisonnière Autorégressive	Présence de décalage systématique.	Étude des résidus permet de juger de la présence de décalages.	Conservé.	Égale puissance théorique.

En absence de décalage systématique, le test utilisant la régression saisonnière autorégressive présente les propriétés désirées. Cependant, l'apparition de décalages amène très rapidement des propriétés indésirables pour ce test. Le test d'Hirsch et Slack sur les moyennes saisonnières présente une puissance plus faible en absence de décalage des saisons, mais il est beaucoup moins affecté par la présence de décalage systématique. Encore une fois, l'utilisation des deux types de test est appropriée, car l'étude des résidus de la régression permettra de valider les conclusions.

CHAPITRE 6

APPLICATION À DES SÉRIES ENVIRONNEMENTALES.

La préoccupation grandissante de la population pour l'environnement a poussé les gouvernements à travailler sur une solution des problèmes environnementaux. Les précipitations acides constituent un problème environnemental de premier plan dans l'est du Canada: Le transport à longue distance de l'anhydride sulfureux rejeté par les fonderies et les centrales électriques de l'est du Canada, ainsi que par les centrales électriques du Midwest et du nord des États-Unis, combiné à la faible capacité de neutralisation de la charge acide par les sols et la roche en place représentent les causes principales de ce problème. Les pluies acides induisent d'importants problèmes économiques et sociaux. Parmi ces problèmes, l'acidification des lacs et rivières qui amène une réduction progressive de la diversité biologique de la vie aquatique présente un intérêt certain.

Afin d'améliorer les connaissances sur l'effet des pluies acides sur les lacs canadiens, Environnement Canada a établi, au début des années 80, plusieurs bassins expérimentaux destinés à être étudiés pour évaluer les effets du transport à longue distance des polluants atmosphériques sur les écosystèmes aquatiques. Les cinq sites principaux qui ont servi au suivi intensif des écosystèmes sont: la région des lacs expérimentaux (Ont), les lacs Turkey (Ont), la région de Muskoka Haliburton (Ont), le lac Laflamme (Qué) et Kejimikujik (N.E.). Le

choix de ces bassins permettait d'avoir une bonne représentation des régions, des climats, de la sensibilité des territoires et des dépôts atmosphériques (Papineau, 1989).

Nous nous attardons ici au site du lac Laflamme. Le bassin versant du lac est situé à 80 km au nord de la ville de Québec dans la réserve des Laurentides (latitude 47°19', longitude 71°07'). Il se trouve sur le Bouclier Canadien et il est par le fait même caractérisé par une faible capacité de neutralisation de la charge acide. L'altitude du lac est de 884 m. Le bassin versant couvre une superficie de 68,4 ha alors que le lac occupe 6,1 ha soit environ 9% de la superficie du bassin versant. Le climat de la région est caractérisé par une température moyenne annuelle basse (0,2°C) et des précipitations annuelles abondantes (1,3 m). La prise des glaces a lieu vers la mi-novembre alors que leur départ se fait vers la mi-mai.

La banque de données disponible est constituée de séries temporelles longues et équidistantes pour :

1. les **concentrations** de 9 constituants dans les **précipitations** recueillies à la station Montmorency;
2. les **charges** des 9 mêmes constituants dans les **précipitations** recueillies à la station Montmorency;
3. les **concentrations** de 16 constituants dans le **lac Laflamme**;
4. les **concentrations** de 16 constituants à l'**exutoire** du lac Laflamme;
5. les **charges** de ces 13 constituants à l'**exutoire** du lac Laflamme.

Toutes les séries étudiées contiennent 312 observations hebdomadaires couvrant la période de six ans débutant le 17 décembre 1981 et se terminant le 12 décembre 1987. Toutes les variables représentant des concentrations sont exprimées en mg/l alors que les variables représentant des charges sont exprimées en kg/ha.sem.

Dans le cas des variables associées aux précipitations, la banque de données originale est épisodique journalière, c'est-à-dire qu'une observation est disponible si le volume des précipitations recueillies durant la journée est suffisant pour permettre une analyse physico-chimique. Afin de construire des séries hebdomadaires, les années de 365 jours ont été divisées en 51 semaines de 7 jours débutant le 1er janvier de chaque année et une semaine de 8 jours allant du 24 décembre au 31 décembre inclusivement; les volumes et les charges ont alors été sommés sur chaque semaine et la concentration moyenne d'une semaine a été obtenue en divisant la charge totale pour la semaine en question par le volume total des précipitations pour la même semaine.

La description complète des méthodes de cueillette des échantillons ainsi que les méthodes analytiques utilisées pour les différents paramètres physico-chimiques sont présentées dans Papineau (1989).

6.1 Choix des paramètres à étudier.

Étant donné le grand nombre de paramètres disponibles dans la banque de données complète (63 variables), une sélection a été faite afin de ne pas alourdir inutilement le présent chapitre. Dans un premier temps, les paramètres NO_3 , SO_4 et pH ont été sélectionnés à cause:

1. De leur importance dans le problème des précipitations acides, l'étude des tendances dans ces séries amène donc des conclusions d'un intérêt certain.
2. De leurs caractéristiques différentes: Fortes saisonnalités dans les séries de nitrates, présence de valeurs extrêmes dans les séries de sulfates et de pH. L'ensemble de ces paramètres permet donc de faire le tour des différentes méthodes présentées dans les chapitres précédents.
3. De leur présence dans les données du lac, de l'exutoire et des précipitations; leur étude permet donc un suivi avant, pendant et après le lac.

Dans un second temps, des analyses en composantes principales ont permis de mettre en évidence le comportement très homogène des paramètres Mg, Ca, Na et alcalinité. En ce qui concerne le sodium, une légère distinction apparaît cependant dans les séries de charges et de concentrations dans les précipitations. Malgré cette différence mineure, ce groupe d'ions a été regroupé en une seule variable qui sera appelée paramètre de minéralisation dans le reste du chapitre. Cette variable a été construite en sommant les concentrations (ou les charges) des trois paramètres Mg, Ca et Na. La grande similitude des comportements des trois ions nous porte à croire que la détection d'une tendance dans la série du paramètre de minéralisation doit se traduire par une tendance dans les trois séries indépendantes. Il est évident que, pour en être certain, une étude complémentaire doit être exécutée. Cependant, le regroupement effectué ici permet de réduire considérablement la taille de l'étude de détection de tendance tout en permettant une étude des quatre paramètres: Mg, Ca, Na et alcalinité.

Les autres séries disponibles dans la banque de données originale n'ont pas été étudiées. Ces séries, majoritairement des séries de métaux (Fe, Al, Zn, etc.), présentent souvent des valeurs tronquées dont le traitement n'a pas été discuté dans les chapitres précédents. D'autres paramètres présents dans les séries de précipitations seulement (NH_4) ou dans les séries du lac et de l'exutoire seulement (Couleur) présentaient un intérêt moins grand, à cause de l'impossibilité de faire un suivi dans l'étude des tendances.

À la suite de ces considérations, 18 séries sont étudiées dans la présente section, il s'agit des séries de **sulfates**, de **nitrate**s et du **paramètre de minéralisation** dans le lac (concentration), à l'exutoire (concentration et charge) et dans les précipitations (concentration et charge); ainsi que le **pH** dans le lac, à l'exutoire et dans les précipitations. La banque de données correspondante est présentée à l'appendice B. Afin de continuer la comparaison entre la régression linéaire et les méthodes non paramétriques, la détection de tendance dans les séries du lac Laflamme sera faite à l'aide des deux approches.

6.2 Étude des sulfates.

Les dépôts de sulfates dans les lacs de l'est du Canada sont généralement identifiés comme une cause majeure de l'acidification de ces lacs. L'étude de ce paramètre chimique est donc d'une importance primordiale dans l'étude de l'effet des pluies acides. L'étude suivante sur la détection de tendance temporelle permettra de juger de la signification des augmentations ou diminutions des concentrations (et des charges) de sulfates dans les précipitations sur le lac Laflamme, dans le lac lui-même et à son exutoire. L'étude des trois stades (avant, pendant et après) permet de constater la dynamique des tendances d'un stade à l'autre.

Étude des concentrations dans les précipitations.

L'étude de tendance de Papineau (1989) sur des données mensuelles de 1981 à 1987 n'a pas permis de détecter de tendance significative. L'auteur fait cependant remarquer la possibilité d'un changement de tendance après 1985 (i.e. augmentation des concentrations avant 1985 et diminution après 1985). L'étude des résidus de notre détection de tendance à l'aide de la régression permettra de juger de l'importance de ce changement de tendance possible.

La figure 6.1 présente la série de concentrations de sulfates dans les précipitations. Bien qu'aucune saisonnalité importante n'apparaisse, une analyse de variance permet de mettre en évidence une saisonnalité significative. Cette saisonnalité est caractérisée par de faibles concentrations pour les mois d'hiver (décembre à février, concentration moyenne de 0,70 mg/l) et par de fortes concentrations pour les mois d'été (juin à août, concentration moyenne de 2,01 mg/l).

Une première régression sans variable décalée amène un test de Durbin-Watson ($d=1,56$) significatif. On doit donc conclure que l'autocorrélation ($\phi_1 = 0,22$) est significative

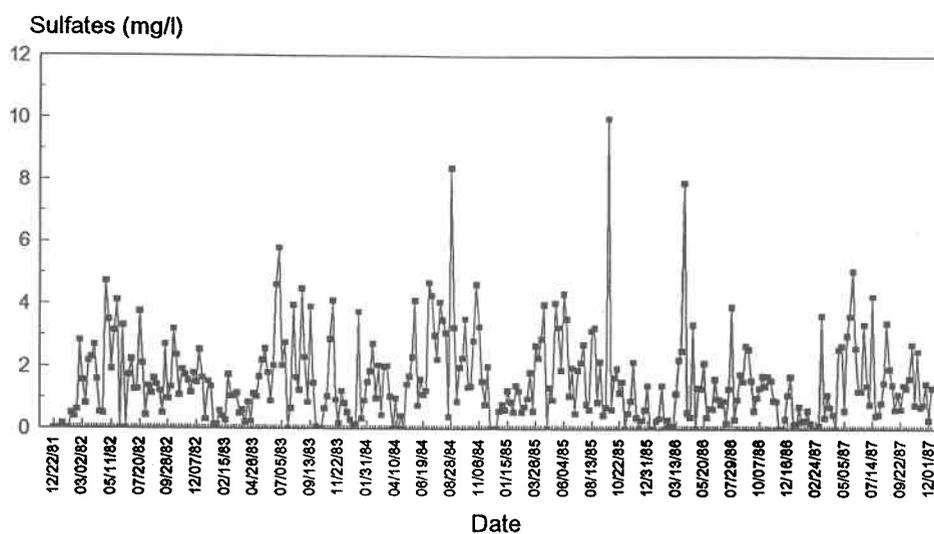


Figure 6.1. Série des concentrations de sulfates dans les précipitations à la station Montmorency.

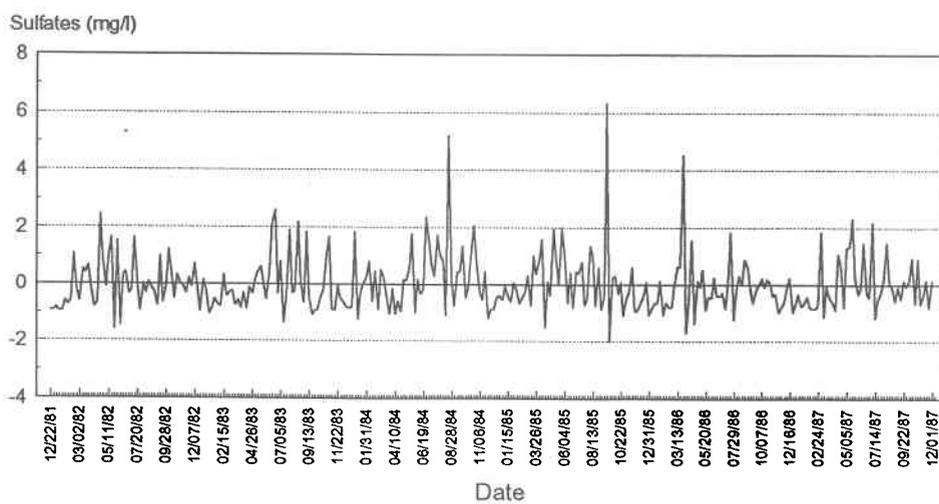


Figure 6.2. Résidus standardisés de la régression autorégressive classique; série des concentrations de sulfates dans les précipitations.

et que l'utilisation d'une régression autorégressive est nécessaire. Le tableau 6.1 présente les résultats des tests de détection de tendance pour les régressions autorégressives classique et robuste, ainsi que les résultats des tests de Kendall sur les Moyennes Saisonnières (4 moyennes saisonnières par année x 6 ans = 24 observations) et Spearman/Lettenmaier (312 observations). Le test d'Hirsch et Slack n'est pas utilisé puisque l'agrégation des données sous forme de moyennes saisonnières amène 24 observations non significativement autocorrélées.

Tableau 6.1
Résultats des tests de détection de tendance pour la série
de concentrations de sulfates dans les précipitations.

	Régression Autorégressive Classique	Régression Autorégressive Robuste	Kendall Moyennes Saisonnières	Spearman/ Lettenmaier
Estimateur b_1	-0,0006	-0,0005	n.a.	n.a.
Statistique du test	-0,66	-0,72	-0,88	-0,65
Significatif ?	non	non	non	non
Estimateur d_1	0,22	0,3	n.a.	0,22

Les résultats similaires des régressions classique et robuste montrent que les quelques valeurs extrêmes n'affectent que très peu la détection de tendance. Ces valeurs extrêmes, sans aucun doute de type statique, amènent cependant une légère sous-estimation de l'estimateur classique du coefficient d'autorégression. Comme le test de Spearman/Lettenmaier utilise l'estimateur classique du coefficient d'autorégression, les résultats similaires à ceux du test de régression classique ne sont pas surprenants. Finalement, bien que la saisonnalité soit apparue significative statistiquement, le résultat du test de Kendall Saisonnier (similaire aux résultats des tests non saisonniers) montre qu'elle n'a que très peu d'effet sur la présente détection de tendance.

La figure 6.2 présente les résidus standardisés de la régression autorégressive classique (N.B. pas de variable indicatrice pour les saisons). Les trois résidus standardisés extrêmes importants (> 4) correspondent à trois semaines de précipitations avec de fortes

concentrations en sulfates. Ces trois valeurs isolées ne peuvent être modélisées dans la régression mais les résultats de la régression robuste montrent que leur effet est négligeable sur le test de détection de tendance. Étant donné la faible autocorrélation (0,22), il est difficile d'identifier le type des valeurs aberrantes¹, puisque les "contrechocs" sur les deuxièmes résidus associés aux valeurs aberrantes statiques ne sont que de l'ordre de 1 à 1,25 ($0,22 \times 4.5$ à $0,22 \times 6$) et n'apparaissent donc pas comme extrêmes. Il faut cependant remarquer que les deux plus grands résidus négatifs suivent deux forts résidus positifs appuyant l'idée que les valeurs aberrantes sont de type statique. Le graphique des résidus standardisés permet également d'illustrer les faibles variations saisonnières et l'absence visible de changement de tendance. Ces deux points ne seront pas étudiés plus en profondeur puisqu'ils n'affectent que très légèrement la présente détection de tendance.

Étude des charges dans les précipitations.

La figure 6.3 présente la série de charges de sulfates dans les précipitations. Une comparaison avec la figure 6.1 (concentrations) montre que la saisonnalité apparaît plus importante dans la série des charges. Une analyse de variance permet de conclure à une saisonnalité significative. Cette saisonnalité est caractérisée par de faibles charges pour les mois d'hiver (décembre à février, charge moyenne de 0,14 kg/ha.se) et par de fortes charges pour les mois d'été (juin à août, charge moyenne de 0,49 kg/ha.se).

Une première régression sans variable avec retard amène un test de Durbin-Watson ($d=1,81$) non significatif. On doit donc conclure que les résidus ne sont pas significativement autocorrélés. Le tableau 6.2 présente les résultats des tests de détection de tendance pour les régressions classique et robuste ainsi que le résultat des tests de Kendall sur les Moyennes Saisonnières (4 moyennes saisonnières par année x 6 ans = 24 observations) et Kendall (312 observations).

¹ N.B. une valeur aberrante de type statique est caractérisée par deux résidus extrêmes successifs (r_t et r_{t+1}) de signes inverses avec une amplitude de l'ordre de ϕr_t pour le résidu r_{t+1} .

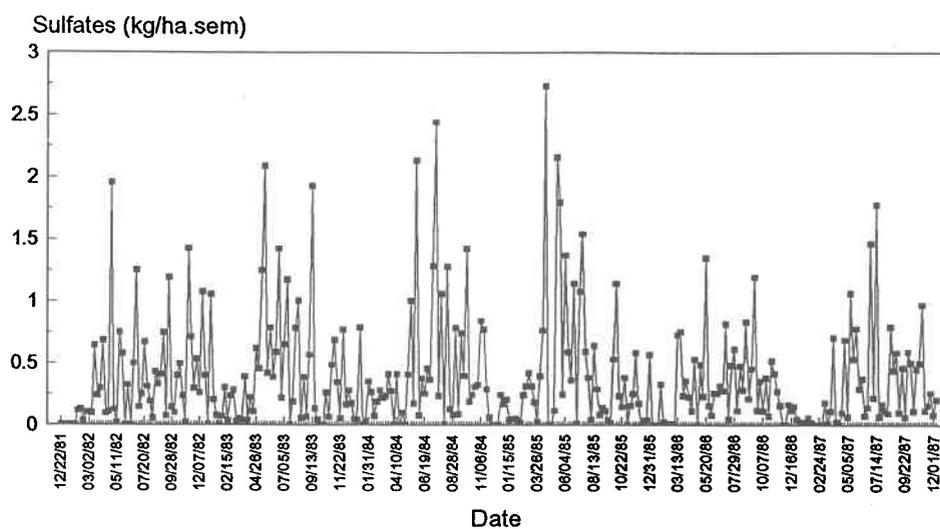


Figure 6.3. Série des charges de sulfates dans les précipitations à la station Montmorency.

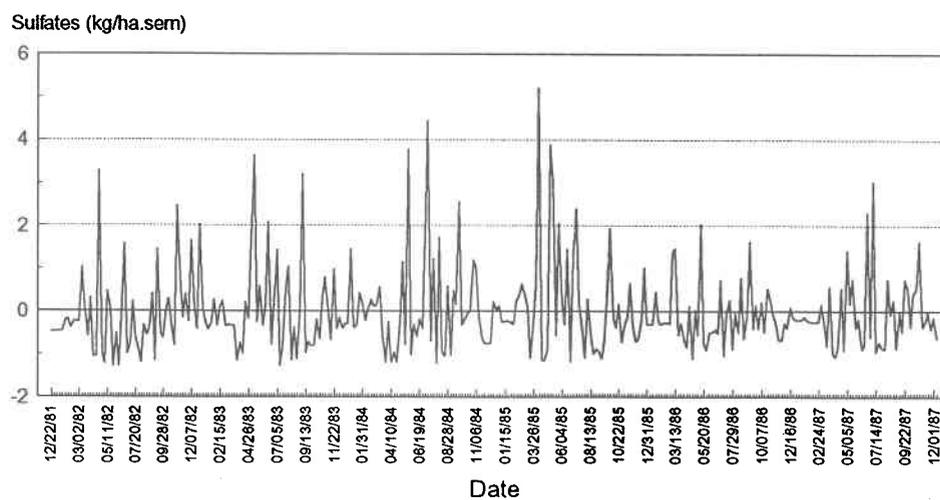


Figure 6.4. Résidus standardisés de la régression saisonnière classique; série des charges de sulfates dans les précipitations.

Tableau 6.2

Résultats des tests de détection de tendance pour la série
de charges de sulfates dans les précipitations.

	Régression classique	Régression robuste	Kendall Moyennes Saisonnieres	Kendall
Estimateur b_1	-0,0003	-0,0001	n.a.	n.a.
Statistique du test	-0,86	-0,57	-1,06	-0,36
Significatif ?	non	non	non	non

Les résultats sensiblement différents des régressions classique et robuste ainsi que les résultats encore plus différents du test de Kendall sur les Moyennes Saisonnieres montrent que les saisonnalités doivent être prises en compte pour fournir un résultat adéquat. Des régressions saisonnières classique et robuste ont donc été effectuées pour modéliser quatre saisons (hiver, mois de décembre à février; printemps, mois de mars à mai; été, mois de juin à août et; automne, mois de septembre à novembre). Ces régressions amènent une statistique de -1,29 pour la régression classique et -1,34 pour la régression robuste. Ces résultats concordent avec le résultat du test de Kendall sur les Moyennes Saisonnieres.

La figure 6.4 présente les résidus standardisés de la régression saisonnière classique (N.B. pas de variable décalée pour autocorrélation). Le graphique des résidus standardisés permet d'illustrer les faibles variations saisonnières et l'absence visible de changement de tendance dans les observations autres que les valeurs extrêmes. Ces deux points ne seront pas étudiés plus en profondeur puisqu'ils n'affectent que très légèrement la présente détection de tendance.

Étude des concentrations au centre du lac.

L'étude de tendance de Papineau (1989) sur des données mensuelles de 1981 à 1987 a permis de détecter une tendance significative (augmentation des concentrations de

sulfates dans le lac) avec un changement de l'amplitude de la tendance dans les deux dernières années. Ce changement de tendance sera discuté dans l'étude des résidus sur modèle de régression. La figure 6.5 présente la série de concentrations de sulfates dans le lac Laflamme. Aucune saisonnalité importante n'apparaît mais quelques valeurs extrêmes sont identifiables.

Une première régression sans variable avec retard amène un test de Durbin-Watson ($d=1,13$) fortement significatif. On doit donc conclure que l'autocorrélation ($\phi = 0,43$) est significative et que l'introduction d'une variable décalée est nécessaire afin de s'assurer que les tests sont adéquats. Le tableau 6.3 présente les résultats des tests de détection de tendance pour les régressions autorégressives classique et robuste ainsi que le résultat du test de Spearman/Lettenmaier.

Tableau 6.3

Résultats des tests de détection de tendance pour la série de concentrations de sulfates au centre du lac.

	Régression Autorégressive Classique	Régression Autorégressive Robuste	Spearman/ Lettenmaier
Estimateur b_1	0,0011	0,0010	n.a.
Statistique du test	4,2	4,8	4,2
Significatif ?	oui	oui	oui
Estimateur d_1	0,43	0,54	0,43

Les résultats quelque peu différents des régressions classique et robuste sont causés par la sous-estimation du coefficient d'autocorrélation par les méthodes classiques. Cette sous-estimation se répercute sur le résultat du test de Spearman/Lettenmaier, puisqu'on utilise ici l'estimateur des moindres carrés du paramètre d'autorégression pour effectuer ce test non paramétrique.

La figure 6.6 présente les résidus standardisés de la régression autorégressive classique. La présence de valeurs extrêmes de type statique à plusieurs endroits dans la série permet d'expliquer la présence de sous-estimation du paramètre d'autorégression par

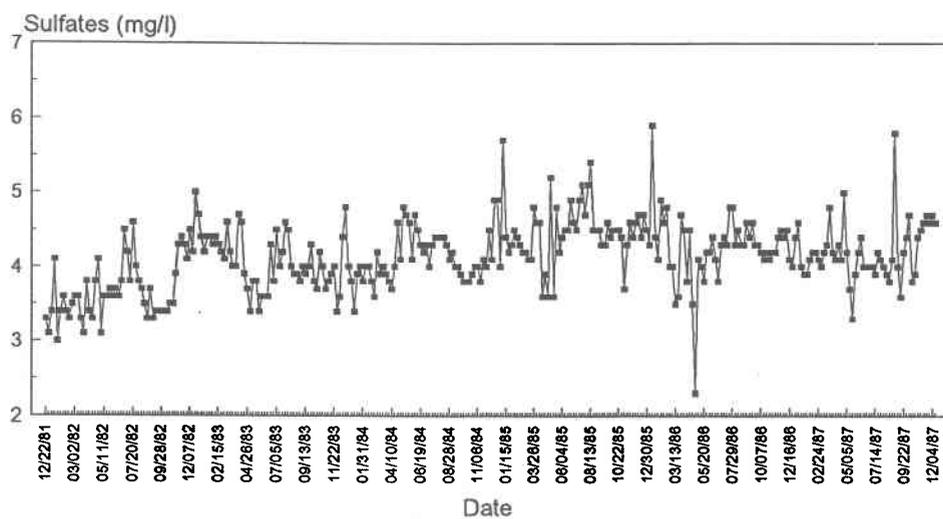


Figure 6.5. Série des concentrations de sulfates au centre du lac Laflamme.

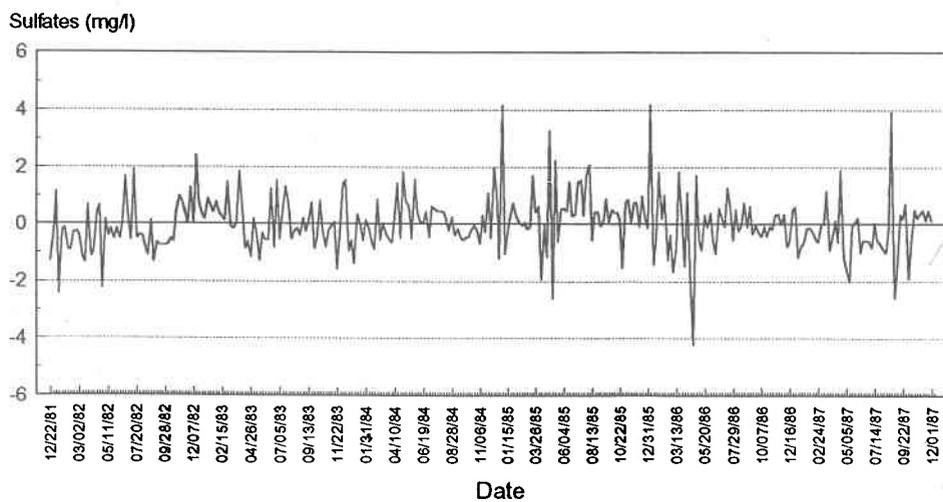


Figure 6.6. Résidus standardisés de la régression autorégressive classique; série des concentrations de sulfates au centre du lac.

l'estimateur des moindres carrés. Les résidus standardisés extrêmes (de l'ordre de 4) correspondent à trois semaines de fortes concentrations de sulfates (janvier 1985, janvier 1986 et septembre 1987) et une semaine de faible concentration de sulfates (avril 1986). Ces valeurs isolées ne peuvent être modélisées dans la régression et apparaissent assez difficiles à expliquer.

Devant ces observations on doit donner une plus grande valeur aux résultats de la régression autorégressive robuste qui ne sont pas affectés par la présence de ces valeurs douteuses. Dans l'éventualité où ces valeurs extrêmes puissent être expliquées et que leur effet sur la tendance générale doit être étudié, il est alors préférable de se fier aux résultats de la régression classique car l'estimateur du paramètre d'autorégression est alors beaucoup plus représentatif de l'autocorrélation réelle des concentrations de sulfates dans le lac.

Finalement, l'étude de la figure 6.6 permet de voir que le changement de tendance soulevé par Papineau (1989) se reflète par la présence de forts pourcentages de résidus négatifs au début (34 des 44 premiers résidus; 77%) et à la fin de la série (49 des 80 derniers résidus; 61%). Le traitement de la tendance en deux parties (avant et après 1985) permettra une meilleure connaissance de la structure de la tendance temporelle mais on peut quand même conclure que les concentrations de sulfates augmentent significativement entre 1981 et 1987.

Étude des concentrations à l'exutoire.

L'étude de tendance de Papineau (1989) sur les données mensuelles de 1981 à 1987 a permis de détecter une tendance significative (augmentation des concentrations de sulfates à l'exutoire du lac) avec un changement d'amplitude de la tendance dans les deux dernières années. La figure 6.7 présente la série de concentrations de sulfates à l'exutoire

du lac Laflamme. Aucune saisonnalité importante n'apparaît, mais quelques valeurs extrêmes sont identifiables; une valeur apparaît particulièrement extrême.

Une première régression sans variable avec retard amène un test de Durbin-Watson ($d=1,05$) fortement significatif. On doit donc conclure que l'autocorrélation ($\phi_1 = 0,52$) est significative et que l'utilisation d'une régression autorégressive est nécessaire. Le tableau 6.4 présente les résultats des tests de détection de tendance pour les régressions autorégressives classique et robuste ainsi que le résultat du test de Spearman/Lettenmaier.

Tableau 6.4

Résultats des tests de détection de tendance pour la série de concentrations de sulfates à l'exutoire.

	Régression classique	Régression robuste	Spearman/Lettenmaier
Estimateur b_1	0,001	0,0007	n.a.
Statistique du test	3,1	3	3,7
Significatif ?	oui	oui	oui
Estimateur d_1	0,52	0,68	0,52

Ces résultats permettent de conclure que les valeurs extrêmes amènent l'estimateur des moindres carrés à sous-estimer le paramètre d'autorégression. Cette sous-estimation explique la différence entre la statistique du test de Spearman/Lettenmaier et celle du test associé à la régression robuste. En ce qui concerne l'utilisation de la régression classique, la sous-estimation du paramètre d'autorégression est opposée à la surestimation du paramètre d'échelle, ce qui amène une statistique du test assez proche de celle associée à la régression robuste.

La figure 6.8 présente les résidus standardisés de la régression autorégressive classique. Un seul résidu apparaît aberrant, il est de type statique ce qui explique la sous-estimation, par l'estimateur du moindre carré, du paramètre d'autorégression. Cette valeur extrême isolée ne peut être modélisée dans la régression et aucune explication de cette observation aberrante n'est disponible.

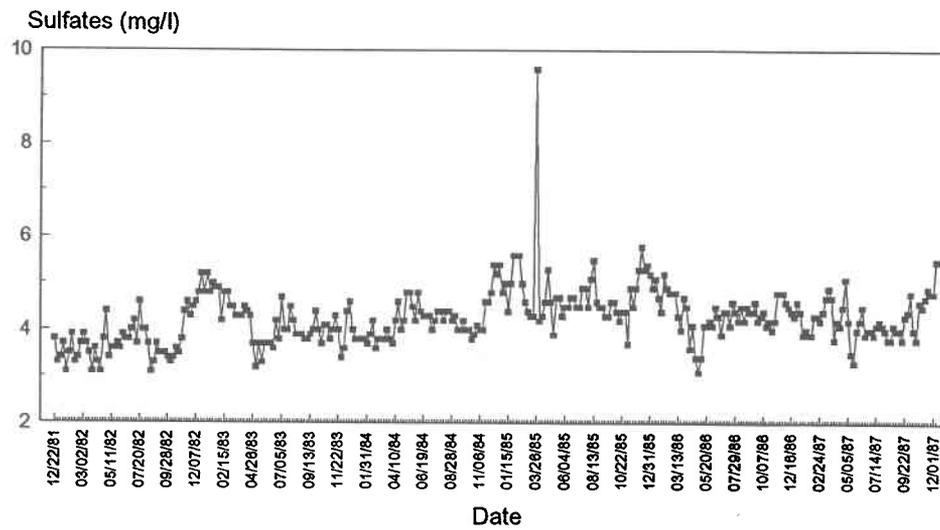


Figure 6.7. Série des concentrations de sulfates à l'exutoire du lac Laflamme.

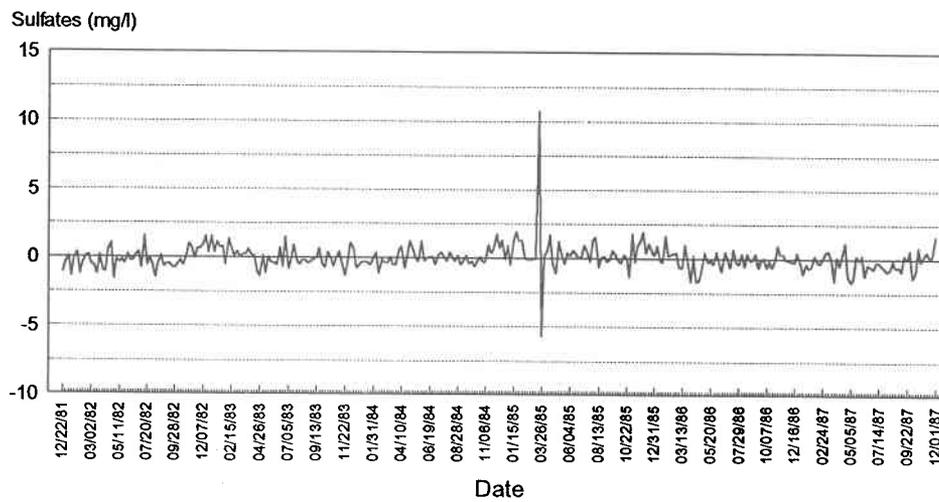


Figure 6.8. Résidus standardisés de la régression autorégressive classique; série des concentrations de sulfates à l'exutoire.

Devant ces observations on doit donner une plus grande valeur aux résultats de la régression autorégressive robuste qui ne sont pas affectés par la présence de cette valeur douteuse. Finalement, l'étude de la figure 6.8 permet de voir que le changement d'amplitude de tendance soulevé par Papineau (1989) se reflète par la présence de forts pourcentages de résidus négatifs au début (32 des 44 premiers résidus; 73%) et à la fin de la série (39 des 60 derniers résidus; 65%). Le traitement de la tendance en deux parties (avant et après 1985) permettra une meilleure connaissance de la structure de la tendance temporelle mais on peut quand même conclure que les concentrations de sulfates augmentent significativement entre 1981 et 1987.

Étude des charges à l'exutoire.

La figure 6.9 présente la série de charges de sulfates à l'exutoire du lac Laflamme. La série est caractérisée par la présence de pics importants à des intervalles d'environ 52 semaines. Ces pics correspondent aux crues de printemps, causées par la fonte des neiges, qui amènent une exportation importante de la charge de sulfates. On remarque également des creux importants associés aux charges exportées en hiver. Ces caractéristiques obligent l'utilisation de variables muettes pour tenir compte de ces saisonnalités.

Une première régression sans variable avec retard amène un test de Durbin-Watson ($d=0,87$) fortement significatif. On doit donc conclure que l'autocorrélation (en excluant l'effet des saisonnalités sur l'autocorrélation) est significative et que l'utilisation d'une régression autorégressive est nécessaire. Le tableau 6.5 présente les résultats des tests de détection de tendance pour les régressions saisonnières autorégressives classique et robuste ainsi que les résultats des tests de Kendall et de Kirsch et Slack sur les Moyennes Saisonnières.

Ce tableau montre une différence importante entre les méthodes de régression et les méthodes non paramétriques. En effet, les statistiques des tests de Kendall et d'Hirsch et

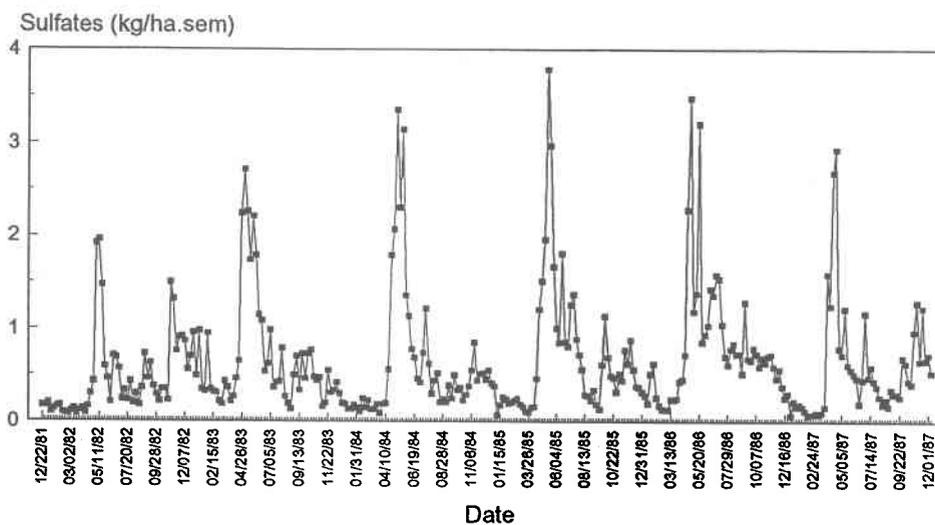


Figure 6.9. Série des charges de sulfates à l'exutoire du lac Laflamme.

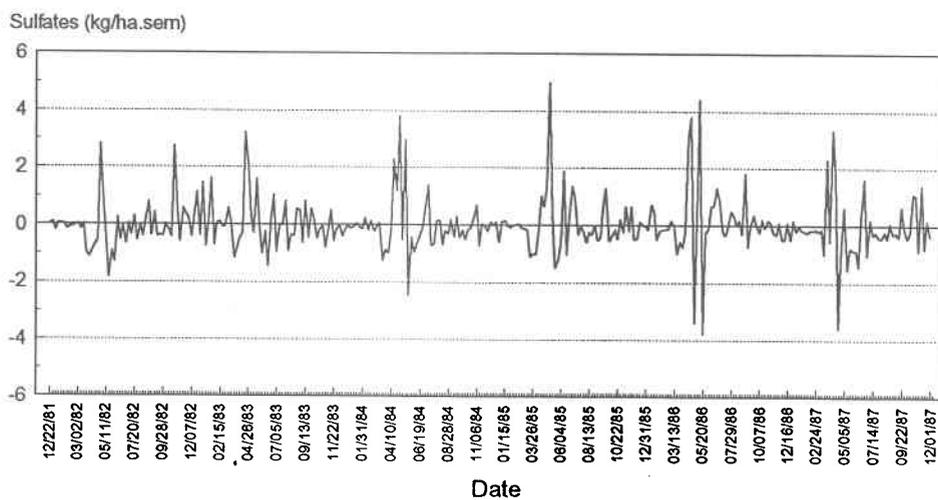


Figure 6.10. Résidus standardisés de la régression saisonnière autorégressive classique; série des charges de sulfates à l'exutoire.

Slack sur les moyennes saisonnières sont deux fois plus grandes que les statistiques des tests utilisant la régression. Malgré cette différence importante, aucun test n'est statistiquement significatif. Les résultats assez similaires des régressions classique et robuste montrent un effet mineur des valeurs aberrantes sur les tests de détection de tendance. La légère sous-estimation du paramètre d'autorégression par l'estimateur des moindres carrés laisse croire que la majorité des valeurs aberrantes sont de type dynamique.

Tableau 6.5

Résultats des tests de détection de tendance pour la série de charges de sulfates à l'exutoire.

	Régression classique	Régression robuste	Kendall Moyennes Saisonnières	Hirsch et Slack
Estimateur b_1	0,0002	0,0001	n.a.	n.a.
Statistique du test	0,52	0,66	1,32	1,17
Significatif ?	non	non	non	non
Estimateur d_1	0,6	0,67	n.a.	n.a.

La figure 6.10 présente les résidus standardisés de la régression saisonnière autorégressive classique. Plusieurs résidus apparaissent aberrants. Ces résidus aberrants surviennent presque exclusivement lors des crues de printemps. La figure 6.10 appuie l'idée que les valeurs douteuses sont majoritairement de type dynamique, car les résidus aberrants sont souvent suivis de résidus de même signe (i.e. pas de contrechoc caractéristique des valeurs aberrantes de type statique). Ces valeurs extrêmes peuvent difficilement être modélisées dans la régression, mais on sait que leur effet est grandement diminué par l'utilisation de la régression robuste. En particulier, la présence de valeurs aberrantes de type dynamique, amène une certaine confiance dans les résultats de la régression robuste.

Devant ces observations on doit maintenant expliquer la différence entre la régression robuste et les tests non paramétriques. L'utilisation des tests non paramétriques sur cette série montre que la différence peut venir du regroupement des observations sous forme de moyennes saisonnières. La statistique du test de Kendall Saisonnier (0,74) est en effet, très

près de la statistique du test de détection pour la régression robuste (0,66). Il semble donc que les moyennes saisonnières soient beaucoup plus près d'une tendance que les observations hebdomadaires. Étant donné la façon dont les statistiques de Kendall sont calculées, on peut conclure que les moyennes saisonnières ont plus tendance à être ordonnées (augmentation d'une année à l'autre) que les observations hebdomadaires. Autrement dit, à l'intérieur d'une même saison, la proportion d'observations concordantes (observation à l'année t plus petite qu'une observation à l'année $t+k$) est plus faible que la proportion de moyennes saisonnières concordantes (moyenne à l'année t plus petite qu'une moyenne à l'année $t+k$).

Il faut aussi se rappeler qu'en présence de saisonnalités et d'autocorrélation, la régression saisonnière autorégressive est très rapidement affectée par la présence de décalage systématique. Le graphique des résidus ne permet pas de conclure à la présence d'un décalage, mais les crues de printemps amènent une série de résidus assez semblable à celle que l'on obtiendrait en présence de décalage systématique. Devant ces observations, il apparaît plus sécuritaire de faire confiance aux tests non paramétriques sur les moyennes saisonnières.

6.3 Étude des nitrates.

L'étude suivante sur la détection de tendance temporelle permettra de juger de la signification des augmentations ou diminutions des concentrations (et des charges) de nitrates dans les précipitations sur le lac Laflamme, dans le lac lui-même et à son exutoire. L'étude des trois stades (avant, pendant et après le lac) permet de constater la dynamique des tendances d'un stade à l'autre.

Étude des concentrations dans les précipitations.

L'étude de tendance de Papineau (1989) sur des données mensuelles de 1981 à 1987 n'a pas permis de détecter de tendance significative. Dans la même étude, on conclut

cependant à une tendance croissante significative avant 1985. L'étude des résidus de notre détection de tendance à l'aide de la régression permettra de juger de l'importance de ce changement de tendance possible.

La figure 6.11 présente la série de concentrations de nitrates dans les précipitations. Bien qu'aucune saisonnalité importante n'apparaisse, une analyse de variance permet de mettre en évidence une saisonnalité significative. Cette saisonnalité est caractérisée par de fortes concentrations pour les mois d'hiver (janvier à mars, concentration moyenne de 0,38 mg/l) et par de faibles concentrations pour les mois d'été (juillet à septembre, concentration moyenne de 0,21 mg/l). Étant donné les faibles amplitudes des saisons, la régression saisonnière ne sera pas utilisée.

Une première régression sans variable avec retard amène un test de Durbin-Watson ($d=1,84$) non significatif. On doit donc conclure que l'autocorrélation n'est pas significative et que l'utilisation d'une régression autorégressive n'est pas nécessaire. Le tableau 6.6 présente les résultats des tests de détection de tendance pour les régressions classique et robuste, ainsi que les résultats des tests de Kendall, Kendall Saisonnier (13 valeurs par saison par année x 6 ans = 312 observations) et Kendall sur les Moyennes Saisonnières (4 moyennes saisonnières par année x 6 ans = 24 observations).

Tableau 6.6

Résultats des tests de détection de tendance pour la série de concentrations de nitrates dans les précipitations.

	Régression classique	Régression robuste	Kendall	Kendall Saisonnier	Kendall Moyennes Saisonnières
Estimateur b_1	0,00012	0,00007	n.a.	n.a.	n.a.
Statistique du test	0,67	0,54	1,21	0,94	0,38
Significatif ?	non	non	non	non	non

Les résultats des régressions classique et robuste montrent que les quelques valeurs extrêmes affectent l'estimation de la pente. En effet, la pente de la régression classique est

presque deux fois plus grande que la pente de la régression robuste. On note cependant que l'écart type des résidus est également plus grand pour la régression classique (0,29) comparativement à celui de la régression robuste (0,20). On remarque donc que les valeurs extrêmes amènent des biais sur les estimateurs ponctuels, mais que ces biais s'annulent dans la construction de la statistique utilisée pour la détection de tendance. Étant donné la structure de la série et le but qui est de détecter une tendance "générale", on désire donc une conclusion qui ne soit pas affectée par quelques valeurs aberrantes; on s'attardera donc au résultat de la régression robuste.

Les résultats des tests non paramétriques montrent que l'utilisation d'un test saisonnier n'est pas vraiment nécessaire, puisque les résultats des tests de Kendall et de Kendall Saisonnier donnent des résultats très similaires. Les distinctions sont beaucoup plus importantes lorsque l'on compare les résultats de ces deux tests avec les résultats du test de Kendall sur les Moyennes Saisonnières et le test utilisant la régression robuste. Ces distinctions s'expliquent par le regroupement sous forme de moyennes saisonnières dans un cas et par l'emplacement des valeurs aberrantes dans le deuxième cas.

Pour expliquer les valeurs plus élevées pour les statistiques des tests de Kendall, on sait que dans le cas de la régression robuste, les valeurs aberrantes n'ont que très peu d'effet sur l'estimation de la pente et de la variance, alors que, pour les tests de Kendall, l'emplacement des valeurs aberrantes (valeurs élevées dans la deuxième partie de la série) affectent la relation entre le temps et les concentrations sans affecter l'estimation de la variance. Les résultats des chapitres précédents ont montré cet effet de l'emplacement des valeurs aberrantes sur les tests non paramétriques. Devant ces observations, on devrait plutôt se fier aux résultats de la régression robuste et conclure que l'on est assez loin d'une tendance significative.

La figure 6.12 présente les résidus standardisés de la régression classique (N.B. pas de variable muette pour les saisons). Les résidus standardisés extrêmes (>3.5)

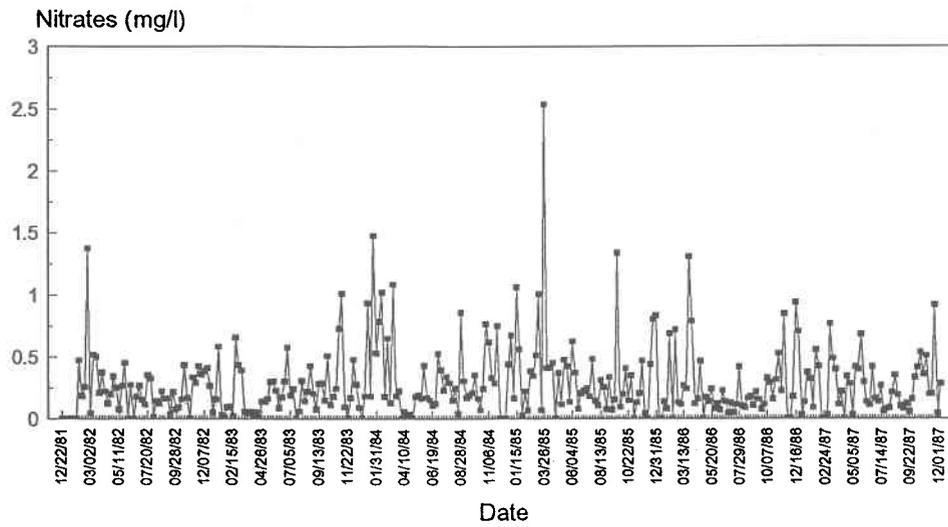


Figure 6.11. Série des concentrations de nitrates dans les précipitations à la station Montmorency.

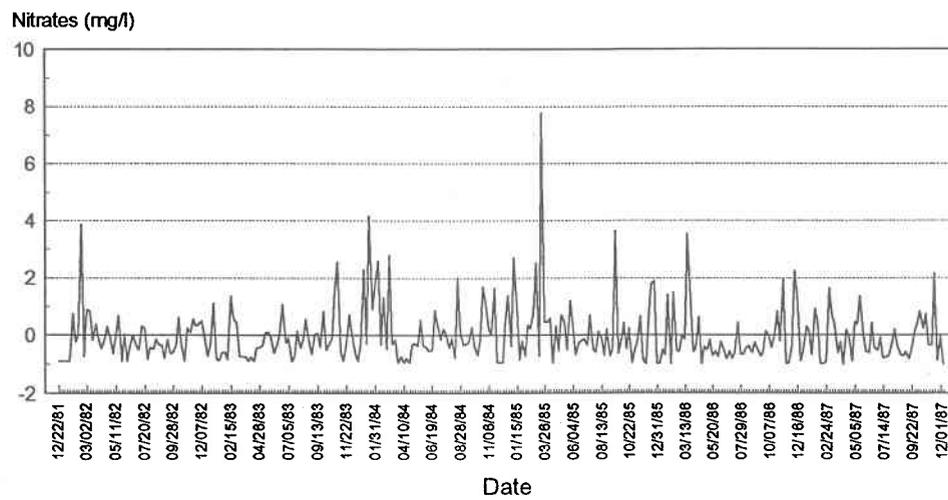


Figure 6.12. Résidus standardisés de la régression classique; série des concentrations de nitrates dans les précipitations.

correspondent à des semaines de précipitations avec de fortes concentrations en nitrates. Ces valeurs généralement isolées ne peuvent être modélisées dans la régression mais l'utilisation de la régression robuste permet de réduire leur effet sur le test de détection de tendance. Le graphique des résidus standardisés permet également d'illustrer les faibles fluctuations saisonnières et l'absence de preuve visuelle d'un changement de tendance après 1985 comme le conclut Papineau (1989). La tendance détectée par Papineau (1989) est sans doute introduite par l'emplacement des valeurs aberrantes en 1985 et il faut alors se poser la question à savoir si la présence d'une année exceptionnelle constitue une tendance. Si la réponse est non, on devrait plutôt s'attarder au résultat de la régression robuste qui ne permette pas de conclure que la tendance est significative au seuil 5%.

Étude des charges dans les précipitations.

La figure 6.13 présente la série de charges de nitrates dans les précipitations. Une comparaison avec la figure 6.11 (concentrations) montre que la saisonnalité apparaît un peu plus importante dans la série des charges. Une analyse de variance permet de conclure à une saisonnalité significative. La structure de cette saisonnalité est moins claire que pour les autres séries de nitrates et de sulfates dans les précipitations. Encore une fois, les faibles amplitudes des saisons nous amènent à ne pas utiliser la régression saisonnière.

Une première régression sans variable avec retard amène un test de Durbin-Watson ($d=1,77$) non significatif. On doit donc conclure que les résidus ne sont pas significativement autocorrélés. Le tableau 6.7 présente les résultats des tests de détection de tendance pour les régressions classique et robuste, ainsi que les résultats des tests de Kendall sur les moyennes saisonnières (4 moyennes saisonnières par année x 6 ans = 24 observations) et de Kendall Saisonnier (312 observations). Ce tableau montre que tous les tests sont loin d'être significatifs. Les résultats de signes inverses pour les régressions classique et robuste montrent un effet non négligeable des valeurs aberrantes sur la détection de tendance. Les charges élevées dans les précipitations en 1985 amènent une pente positive pour la

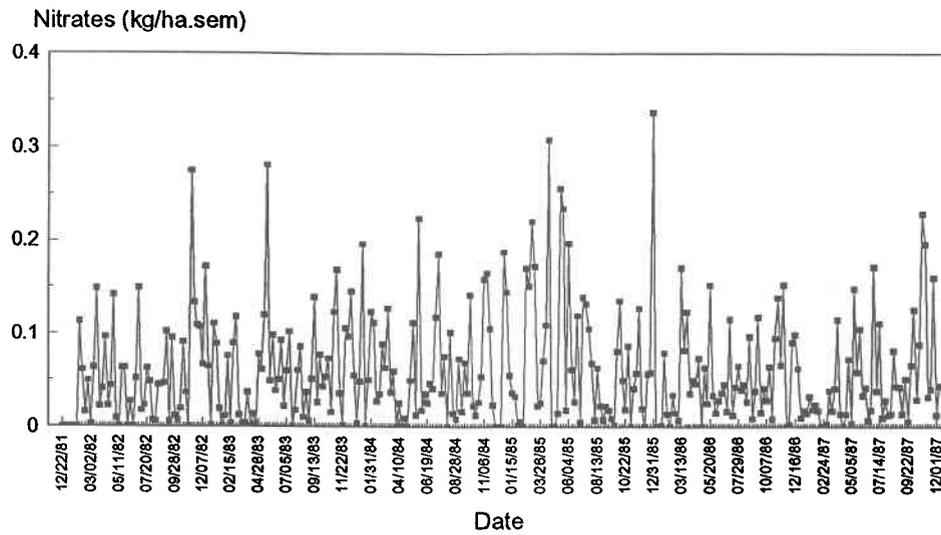


Figure 6.13. Série des charges de nitrates dans les précipitations à la station Montmorency.

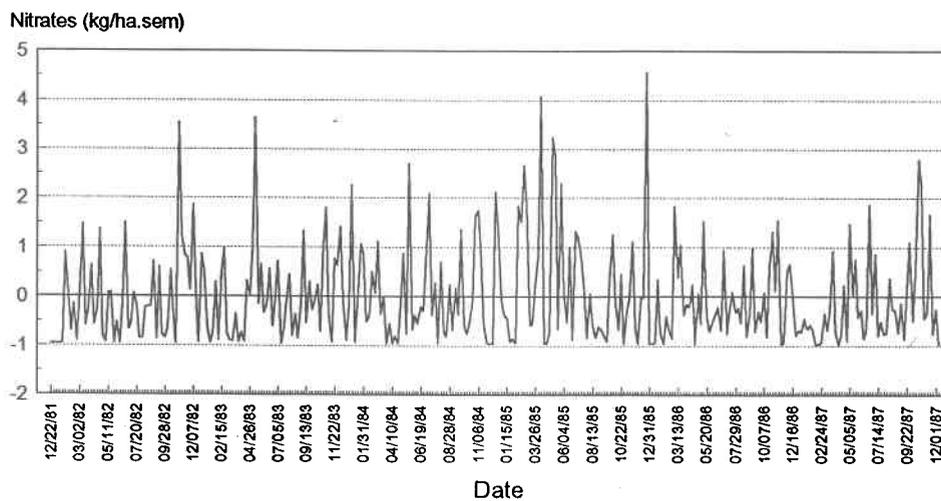


Figure 6.14. Résidus standardisés de la régression classique; série des concentrations de nitrates dans les précipitations.

régression classique mais lorsque l'effet de ces valeurs est réduit par l'utilisation des méthodes robustes, l'estimateur de la pente est alors négatif.

Tableau 6.7

Résultats des tests de détection de tendance pour la série de charges de nitrates dans les précipitations.

	Régression classique	Régression robuste	Kendall Saisonnier	Kendall Moyennes Saisonniers
Estimateur b_1	0,0000	-0,0001	n.a.	n.a.
Statistique du test	0,05	-0,38	0,14	-0,19
Significatif ?	non	non	non	non

Encore une fois, l'utilisation de l'une ou l'autre des méthodes dépend de l'importance que l'on désire donner à une conclusion qui repose sur quelques valeurs seulement. Dans le cas présent, le choix de l'une ou l'autre des méthodes est sans grande implication, puisque les deux méthodes montrent des statistiques beaucoup plus petites que les valeurs critiques.

La figure 6.14 présente les résidus standardisés de la régression classique. Le graphique des résidus standardisés montre la présence de quelques résidus douteux, tous positifs. Étant donné le modèle de régression utilisé (pas de variable avec retard, ni de variable muette), le graphique des résidus est très semblable au graphique de la série originale; il ne permet donc aucune information supplémentaire.

Étude des concentrations au centre du lac.

L'étude de tendance de Papineau (1989) sur des données mensuelles de 1981 à 1987 a permis de détecter une tendance significative (diminution des concentrations de nitrates dans le lac) avec un changement de pente de la tendance dans les deux dernières années. Ce changement de tendance sera discuté dans l'étude des résidus du modèle de régression. La figure 6.15 présente la série de concentrations de nitrates dans le lac Laflamme. Une très

forte saisonnalité apparaît, ce qui oblige l'utilisation d'une régression saisonnière. On note également la présence de quelques valeurs extrêmes.

Une première régression sans variable avec retard amène un test de Durbin-Watson ($d=0,16$) très fortement significatif. On doit donc conclure que les résidus sont autocorrélés et que l'utilisation d'une régression autorégressive est nécessaire. Le tableau 6.8 présente les résultats des tests de détection de tendance pour les régressions autorégressives saisonnières classique et robuste ainsi que les résultats des tests de Kendall Saisonnier, de Kendall sur les Moyennes Saisonnières et d'Hirsch et Slack.

Tableau 6.8

Résultats des tests de détection de tendance pour la série de concentrations de nitrates au centre du lac.

	Régression Saisonnière Autorégressive Classique	Régression Saisonnière Autorégressive Robuste	Kendall Saisonnier	Kendall Moyennes Saisonnières	Hirsch et Slack
Estimateur b_1	-0,00004	-0,00001	n.a.	n.a.	n.a.
Statistique du test	-1,5	-1,12	-7,67	-3,57	-2,36
Significatif ?	non	non	oui	oui	oui
Estimateur d_1	0,88	0,94	n.a.	n.a.	n.a.

Les résultats des régressions classique et robuste montrent que les quelques valeurs extrêmes affectent l'estimation de la pente. En effet, la pente de la régression classique est environ quatre fois plus grande que la pente de la régression robuste. On note cependant que l'écart type des résidus est également plus grand pour la régression classique (0,037) comparativement à celui de la régression robuste (0,015). On remarque donc que les valeurs extrêmes amènent un biais sur les estimateurs ponctuels mais que ces biais s'annulent partiellement dans la construction de la statistique utilisée pour la détection de tendance. Étant donné la structure de la série et le but qui est de détecter une tendance "générale", on désire donc une conclusion qui ne soit pas affectée par quelques valeurs aberrantes; on s'attardera donc au résultat de la régression robuste.

Les résultats des tests non paramétriques montrent des résultats très différents. La très grande valeur de la statistique du test de Kendall Saisonnier s'explique par la présence d'une forte autocorrélation qui biaise l'estimateur de la variance et qui n'est pas prise en compte par ce test. Le test de Kendall sur les Moyennes Saisonnières est moins affecté par la présence d'autocorrélation à cause du regroupement des observations sous forme de moyennes saisonnières; la présence d'une autocorrélation hebdomadaire de l'ordre de 0.90 permet cependant de croire que l'autocorrélation peut quand même avoir un effet sur ce test. Le test d'Hirsch et Slack est le test non paramétrique le plus fiable dans la présente situation.

En comparant maintenant les résultats des tests paramétrique (régression robuste) et non paramétrique (Hirsch et Slack) les plus fiables, on s'aperçoit de l'importante différence entre les deux. Le chapitre 4 a montré une diminution importante de la puissance de la régression en présence de décalage systématique entre les saisons estimées et réelles. La figure 6.16, qui présente la série des résidus standardisés pour la régression saisonnière autorégressive classique, ne montre pas la présence d'un décalage systématique dans l'estimation des saisons. Les valeurs douteuses sont plutôt attribuables aux données printanières qui peuvent être affectées par la fonte des neiges, reconnue pour avoir un effet important sur la concentration des nitrates dans les lacs. Devant ces caractéristiques, il est assez difficile de choisir le meilleur test. Cependant la plus grande stabilité du test d'Hirsch et Slack comparativement au test de régression milite en la faveur du test non paramétrique. On devrait alors conclure à une diminution significative des concentrations de nitrates dans le lac Laflamme. La figure 6.15 montre en effet une diminution des concentrations pour les années 1986 et 1987, alors qu'aucune tendance n'apparaît dans les quatre premières années.

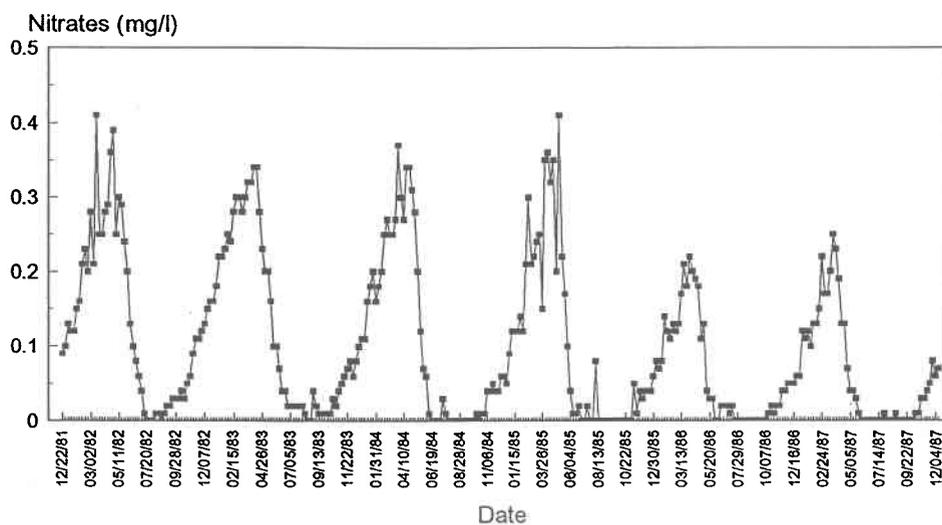


Figure 6.15. Série des concentrations de nitrates au centre du lac Laflamme.

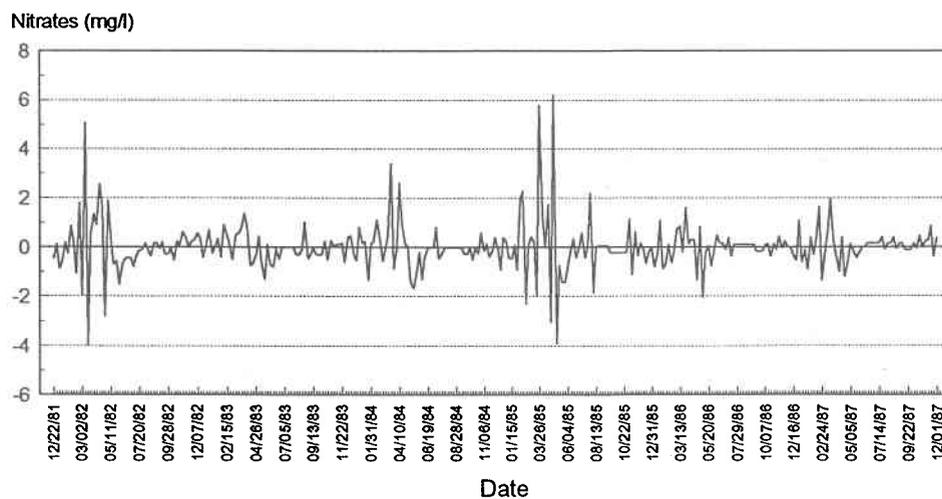


Figure 6.16. Résidus standardisés de la régression saisonnière autorégressive classique; série des concentrations de nitrates au centre du lac.

Étude des concentrations à l'exutoire.

L'étude de tendance de Papineau (1989) sur les données mensuelles de 1981 à 1987 a permis de détecter une tendance significative (diminution des concentrations de nitrates à l'exutoire du lac) avec un changement de pente de la tendance dans les deux dernières années. La figure 6.17 présente la série de concentrations de nitrates à l'exutoire du lac Laflamme. Une très forte saisonnalité apparaît en plus de quelques valeurs extrêmes beaucoup plus importantes que pour les concentrations au centre du lac (réf. figure 6.15).

Une première régression sans variable avec retard amène un test de Durbin-Watson ($d=0,20$) fortement significatif. On doit donc conclure que les résidus sont significativement autocorrélés et que l'introduction d'une variable avec retard est nécessaire. Le tableau 6.9 présente les résultats des tests de détection de tendance pour les régressions saisonnières autorégressives classique et robuste ainsi que les résultats des tests de Kendall Saisonnier, de Kendall sur les Moyennes Saisonnières et d'Hirsch et Slack.

Tableau 6.9

Résultats des tests de détection de tendance pour la série de concentrations de nitrates à l'exutoire.

	Régression Saisonnière Autorégressive Classique	Régression Saisonnière Autorégressive Robuste	Kendall Saisonnier	Kendall Moyennes Saisonnières	Hirsch et Slack
Estimateur b_1	-0,00004	-0,00001	n.a.	n.a.	n.a.
Statistique du test	-1,42	-1,12	-8,47	-3,51	-2,28
Significatif ?	non	non	oui	oui	oui
Estimateur d_1	0,87	0,94	n.a.	n.a.	n.a.

Le tableau 6.9 présente des résultats en tous points semblables à ceux du tableau 6.8 (série de concentrations au centre du lac). Comme la seule différence importante entre les deux séries est la présence de pics plus importants dans la série de concentrations à

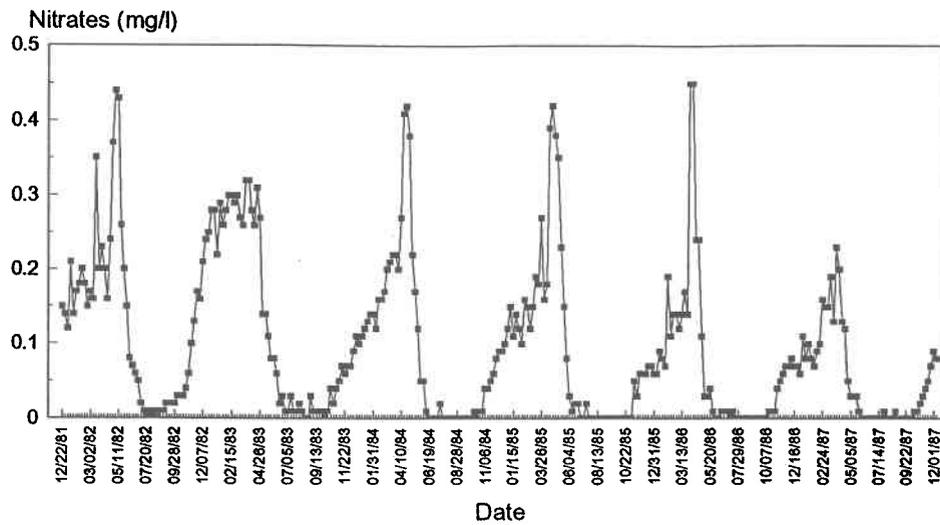


Figure 6.17. Série des concentrations de nitrates à l'exutoire du lac Laflamme.

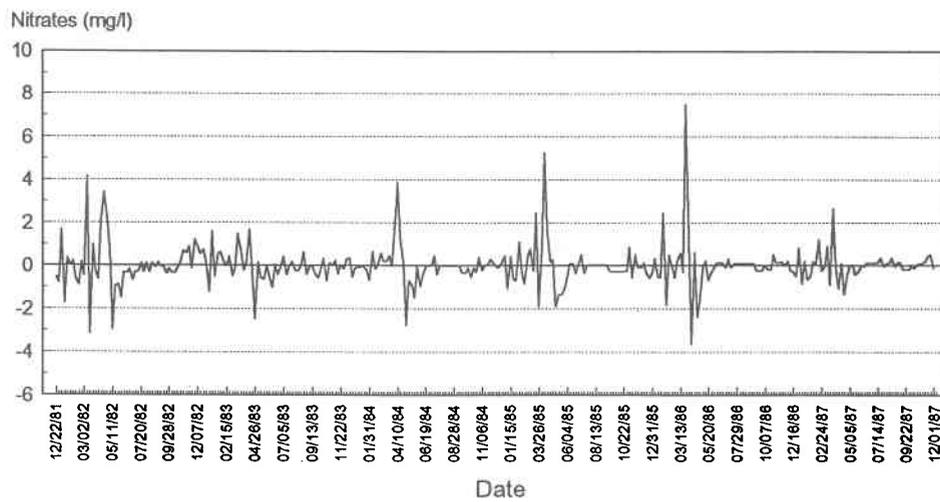


Figure 6.18. Résidus standardisés de la régression saisonnière autorégressive classique; série des concentrations de nitrates à l'exutoire.

l'exutoire, il est normal que le test de régression robuste ainsi que les tests non paramétriques ne soient pas affectés par une telle différence.

La figure 6.18 présente la série de résidus de la régression saisonnière autorégressive classique. Cette figure montre des résidus très semblables à ceux de la régression au centre du lac. La présence d'un très grand résidu dans la série de concentrations à l'exutoire au printemps 1986 représente la seule différence importante. Comme pour la série au centre du lac, on doit donc croire que le test d'Hirsch et Slack devrait présenter le résultat le plus fiable et conclure à une diminution significative des concentrations de nitrates à l'exutoire.

Étude des charges à l'exutoire.

La figure 6.19 présente la série de charges de nitrates à l'exutoire du lac Laflamme. La série est caractérisée par la présence de pics importants à des intervalles d'environ 52 semaines. Ces pics correspondent aux crues de printemps, causées par la fonte des neiges, qui amènent une exportation importante de la charge de nitrates. On remarque également des faibles charges hebdomadaires exportées le reste de l'année. L'année 1982 semble différente des autres années, à cause d'une plus forte charge exportée à l'été et l'automne et d'une plus faible charge exportée au printemps suivant (1983). La présence de saisonnalités importantes oblige l'utilisation d'une régression saisonnière.

Une première régression saisonnière sans variable avec retard amène un test de Durbin-Watson ($d=0,69$) fortement significatif. On doit donc conclure que l'autocorrélation (en excluant l'effet des saisonnalités sur l'autocorrélation) est significative et que l'utilisation d'une régression saisonnière autorégressive est nécessaire. Le tableau 6.10 présente les résultats des tests de détection de tendance pour les régressions saisonnières

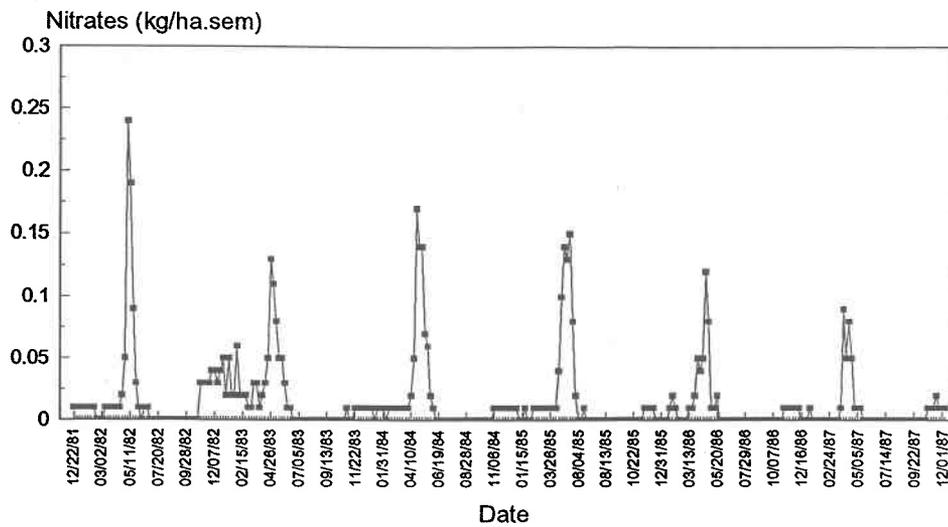


Figure 6.19. Série des charges de nitrates à l'exutoire du lac Laflamme.

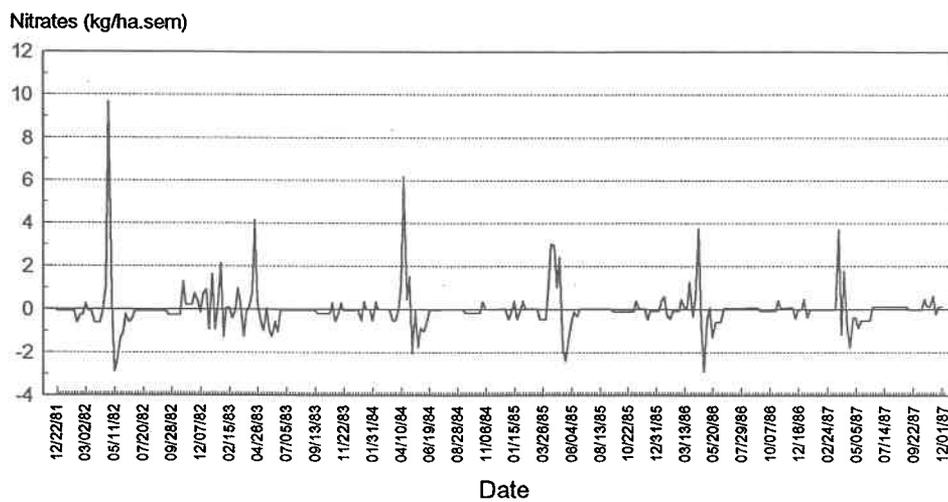


Figure 6.20. Résidus standardisés de la régression saisonnière autorégressive classique; série des charges de nitrates à l'exutoire.

autorégressives classique et robuste ainsi que les résultats des tests de Kendall Saisonnier, de Kendall sur les Moyennes Saisonnières et d'Hirsch et Slack.

Tableau 6.10

Résultats des tests de détection de tendance pour la série de charges de nitrates à l'exutoire.

	Régression Saisonnière Autorégressive classique	Régression Saisonnière Autorégressive robuste	Kendall Saisonnier	Kendall Moyennes Saisonnières	Hirsch et Slack
Estimateur b_1	-16×10^{-6}	-3×10^{-6}	n.a.	n.a.	n.a.
Statistique du test	-1,29	-3,03	-4,98	-2,21	-2,05
Significatif ?	non	oui	oui	oui	oui
Estimateur d_1	0,69	0,79	n.a.	n.a.	n.a.

Les résultats des régressions classique et robuste montrent que les valeurs extrêmes affectent l'estimation de la pente et la conclusion du test de détection de tendance. Étant donné la structure de la série, on s'attardera au résultat de la régression robuste.

Les résultats des tests non paramétriques montrent encore une fois l'effet de l'autocorrélation sur la statistique du test de Kendall Saisonnier (valeur trop grande à cause du biais sur la variance). Les résultats semblables des tests de Kendall sur les Moyennes Saisonnières et d'Hirsch et Slack montrent le peu d'effet de cette même autocorrélation une fois que les observations sont regroupées sous forme de moyennes saisonnières.

Les résultats des tests paramétrique (régression robuste) et non paramétrique (Hirsch et Slack) les plus fiables, permettent d'arriver à la même conclusion. On peut donc conclure en toute confiance à la présence d'une diminution significative des charges de nitrates exportées à l'exutoire du lac.

La figure 6.20, qui présente la série des résidus standardisés pour la régression saisonnière autorégressive classique, ne montre pas la présence d'un décalage

systematique dans l'estimation des saisons. Les valeurs douteuses sont attribuables à la fonte printanière qui amène de fortes charges de nitrates exportées à l'exutoire du lac Laflamme.

6.4 Étude du pH.

L'étude du pH des lacs est d'une grande importance puisque plusieurs études ont montré une diminution de la diversité biologique à mesure que le pH diminue dans un lac. Devant de telles causes, une étude de détection de tendance permet de porter un jugement sur l'évolution du pH d'un lac et d'élaborer un plan d'action dans l'éventualité d'une tendance pouvant affecter la diversité biologique. L'étude suivante sur la détection de tendance temporelle permettra de juger de la signification des augmentations ou diminutions du pH dans les précipitations sur le lac Laflamme, dans le lac lui-même et à son exutoire.

Étude des précipitations.

L'étude de tendance de Papineau (1989) sur des données mensuelles de 1981 à 1987 n'a pas permis la détection d'une tendance significative. La figure 6.21 présente la série de pH dans les précipitations, on remarque la présence de quelques valeurs aberrantes et l'absence de saisonnalité.

Il faut noter que certaines semaines ne présentaient pas de volumes de précipitations suffisants pour permettre une analyse physico-chimique. Pour ces semaines le pH moyen a été utilisé comme estimateur des données manquantes. Une première régression sans variable avec retard amène un test de Durbin-Watson ($d=1.53$) significatif. On doit donc conclure que les résidus sont autocorrélés et que l'introduction d'une variable avec retard est nécessaire. Le tableau 6.11 présente les résultats des tests de détection de tendance pour les régressions autorégressives classique et robuste ainsi que le résultat du test de Kendall.

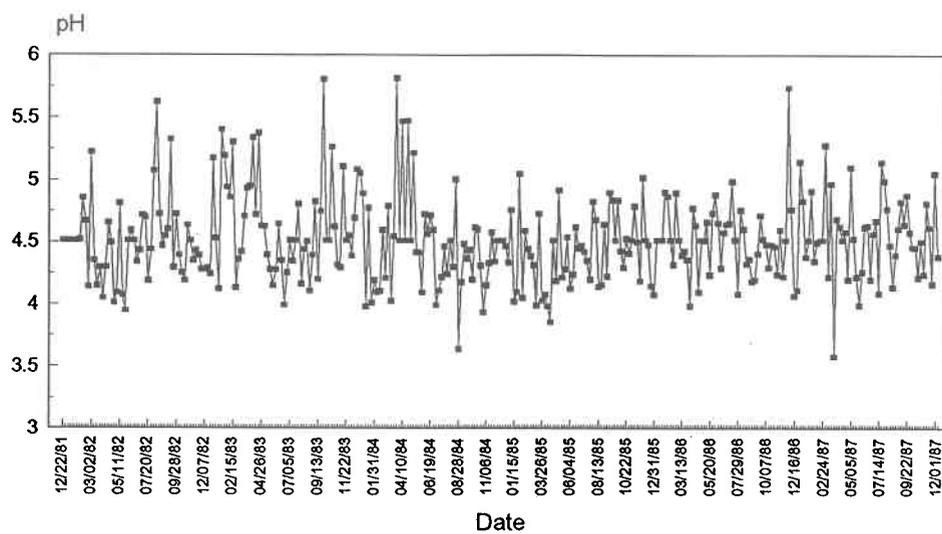


Figure 6.21. Série des pH dans les précipitations à la station Montmorency.

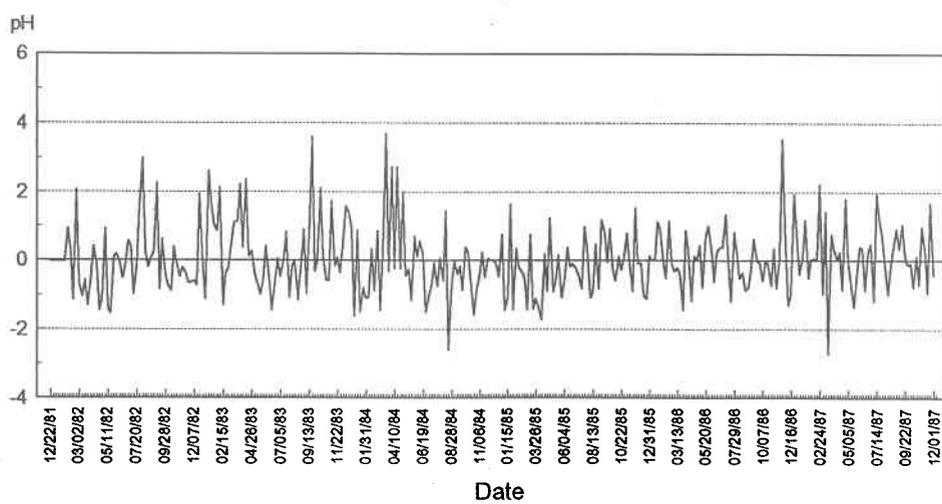


Figure 6.22. Résidus standardisés de la régression autorégressive classique; série des pH dans les précipitations.

Tableau 6.11

Résultats des tests de détection de tendance pour la série de pH dans les précipitations.

	Régression classique	Régression robuste	Kendall
Estimateur b_1	-0,00011	0,00009	n.a.
Statistique du test	-0,5	0,44	0,23
Significatif ?	non	non	non
Estimateur d_1	0,1	0,09	n.a.

Les résultats des régressions classique et robuste montrent l'effet des valeurs aberrantes au début de la série. La concordance des tests de la régression robuste et de Kendall permet de conclure que: 1) la tendance générale est à une augmentation des pH; 2) cette tendance générale n'est pas statistiquement significative et: 3) l'effet de l'autocorrélation est négligeable. La figure 6.22 présente la série de résidus du pH dans les précipitations pour la régression autorégressive classique. On remarque les quelques valeurs aberrantes qui affectent l'estimation de la pente. Aucune autre irrégularité ne permet de douter des conclusions des tests.

Étude au centre du lac.

L'étude de tendance de Papineau (1989) sur des données mensuelles de 1981 à 1987 a permis de détecter une augmentation significative des pH au centre du lac. La figure 6.23 présente la série de pH au centre du lac, on remarque la présence de valeurs aberrantes à la fonte printanière et l'absence de saisonnalité.

Une première régression sans variable avec retard amène un test de Durbin-Watson ($d=0,98$) significatif impliquant l'utilisation d'une régression autorégressive. Le tableau 6.12 présente les résultats des tests de détection de tendance pour les régressions autorégressives classique et robuste ainsi que le résultat du test de Spearman/Lettenmaier.

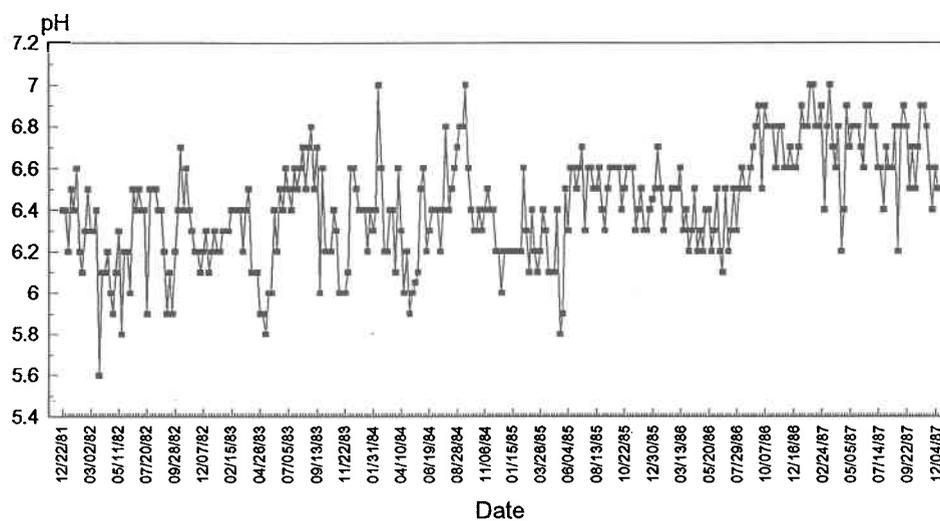


Figure 6.23. Série des pH au centre du lac Laflamme.

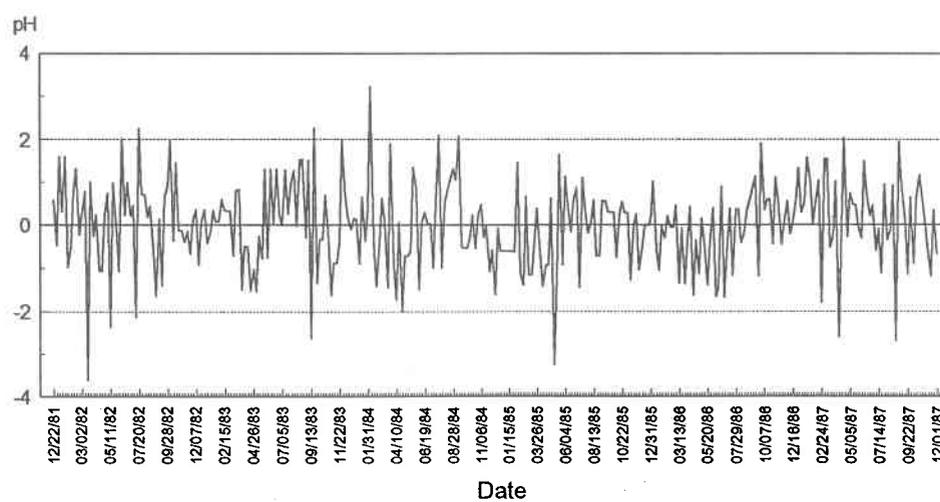


Figure 6.24. Résidus standardisés de la régression autorégressive classique; série des pH au centre du lac.

Tableau 6.12

Résultats des tests de détection de tendance pour la série de pH dans le lac.

	Régression classique	Régression robuste	Spearman/ Lettenmaier
Estimateur b_1	0,0008	0,0007	n.a.
Statistique du test	5,46	5,31	5,04
Significatif ?	oui	oui	oui
Estimateur d_1	0,49	0,55	0,49

Les résultats du tableau 6.12 montrent que les valeurs aberrantes ont peu d'effet sur la détection de tendance dans la série de pH au centre du lac. La forte signification de tous les tests de détection de tendance ne laisse aucun doute à la présence d'une augmentation significative du pH au lac Laflamme entre 1981 et 1987. La figure 6.24 présente les résidus de la régression autorégressive classique pour le pH au centre du lac. Les résidus aberrants associés aux fontes printanières sont facilement identifiables et aucune autre irrégularité ne permet de douter des conclusions des tests.

Étude à l'exutoire.

L'étude de tendance de Papineau (1989) sur des données mensuelles de 1981 à 1987 a permis de détecter une augmentation significative des pH à l'exutoire du lac. La figure 6.25 présente la série de pH à l'exutoire. On remarque la présence de valeurs aberrantes (plus importantes que pour la série au centre du lac) à la fonte printanière et l'absence de saisonnalité.

Une première régression sans variable avec retard amène un test de Durbin-Watson ($d=0,89$) significatif ce qui explique l'utilisation d'une régression autorégressive. Le tableau 6.13 présente les résultats des tests de détection de tendance pour les régressions autorégressives classique et robuste ainsi que le résultat du test de Spearman/Lettenmaier.

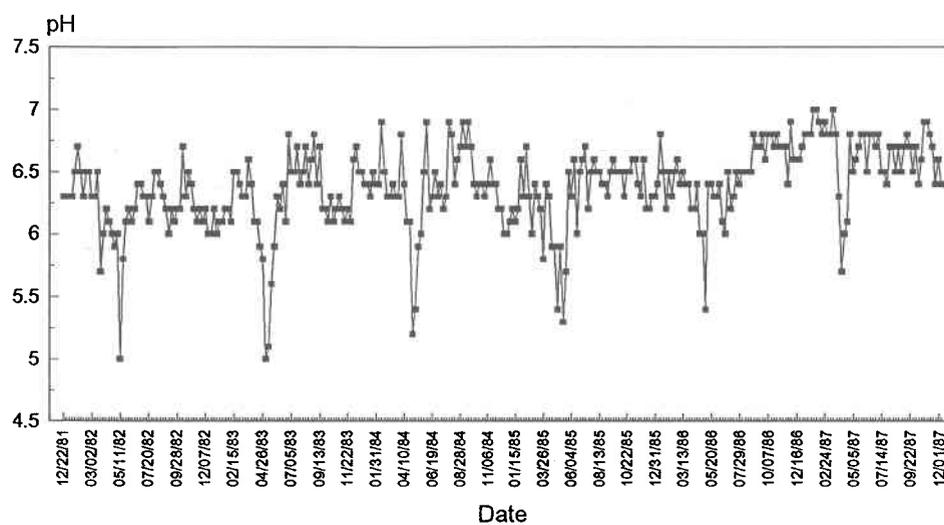


Figure 6.25. Série des pH à l'exutoire du lac Laflamme.

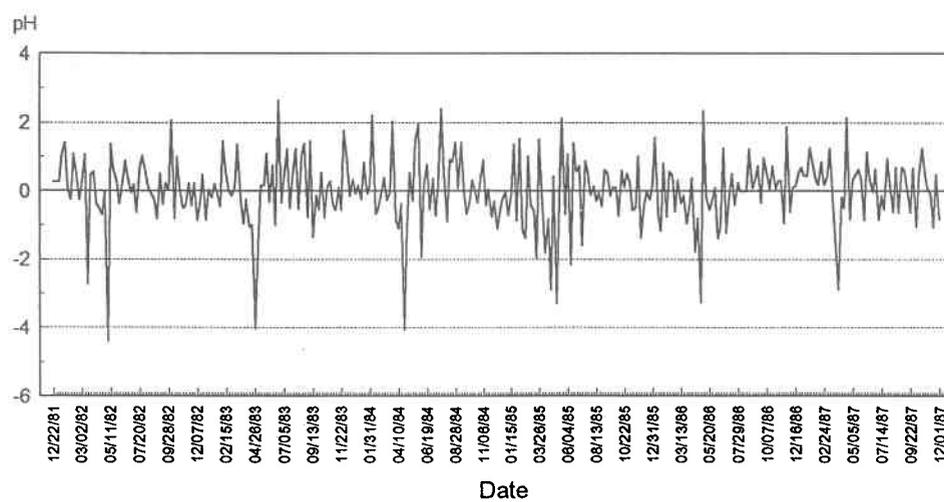


Figure 6.26. Résidus standardisés de la régression autorégressive classique; série des pH à l'exutoire.

Tableau 6.13

Résultats des tests de détection de tendance pour la série de pH à l'exutoire.

	Régression classique	Régression robuste	Spearman/ Lettenmaier
Estimateur b_1	0,0006	0,0007	n.a.
Statistique du test	3,6	4,88	4,13
Significatif ?	oui	oui	oui
Estimateur d_1	0,6	0,57	0,6

La différence entre les statistiques des tests des régressions classique et robuste est introduite par l'estimation de la variance qui est gonflée par la présence de valeurs aberrantes dans le cas de la régression classique. La forte signification de tous les tests de détection de tendance ne laisse cependant aucun doute à la présence d'une augmentation significative du pH à l'exutoire du lac Laflamme entre 1981 et 1987. La figure 6.26 présente les résidus de la régression autorégressive classique pour le pH à l'exutoire du lac. Les résidus aberrants associés aux fontes printanières sont facilement identifiables et le fait qu'ils sont plus importants que dans le centre du lac explique la plus grande différence entre les régressions classique et robuste. Aucune autre irrégularité ne permet de douter des conclusions des tests.

6.5 Étude du paramètre de minéralisation.

Le paramètre de minéralisation a une certaine importance dans l'étude de l'effet des pluies acides. En effet, les ions composant la construction de ce paramètre (Ca, Mg et Na) sont généralement associés à la capacité de neutralisation de la charge acide. Une tendance détectée dans les séries du paramètre de minéralisation pourra donc être associée à un changement dans la capacité de neutralisation de la charge acide; on peut alors mettre sur pied un plan d'action pour contrer un effet négatif. L'étude suivante sur la détection de tendance temporelle permettra de juger de la signification des augmentations ou diminutions du paramètre de minéralisation dans les précipitations (concentrations et charges) sur le lac Laflamme, dans le lac lui-même et à son exutoire (concentrations et charges).

du paramètre de minéralisation dans les précipitations (concentrations et charges) sur le lac Laflamme, dans le lac lui-même et à son exutoire (concentrations et charges).

Étude des concentrations dans les précipitations.

L'étude de tendance de Papineau (1989) sur des données mensuelles de 1981 à 1987 n'étudiait pas les concentrations des paramètres de minéralisation dans les précipitations. La figure 6.27 présente la série de concentrations du paramètre de minéralisation dans les précipitations, on remarque la présence de quelques valeurs aberrantes importantes et l'absence de saisonnalité.

Une première régression sans variable avec retard amène un test de Durbin-Watson ($d=1,79$) non significatif. On doit donc conclure que les résidus ne sont pas significativement autocorrélés et que l'introduction d'une variable avec retard est inutile. Le tableau 6.14 présente les résultats des tests de détection de tendance pour les régressions classique et robuste ainsi que les résultats des tests de Kendall et de Spearman/Lettenmaier.

Tableau 6.14

Résultats des tests de détection de tendance pour la série de concentrations du paramètre de minéralisation dans les précipitations.

	Régression classique	Régression robuste	Kendall	Spearman/Lettenmaier
Estimateur b_t	-0,0002	-0,0002	n.a.	n.a.
Statistique du test	-1,06	-3,91	-2,62	-2,41
Significatif ?	non	oui	oui	oui
Estimateur d_t	0,1	0,06	n.a.	0,1

Les résultats des tests de détection utilisant la régression montrent l'effet important des valeurs aberrantes sur la détection de tendance dans la série de concentrations dans les précipitations du paramètre de minéralisation. Les estimateurs de la pente de la régression

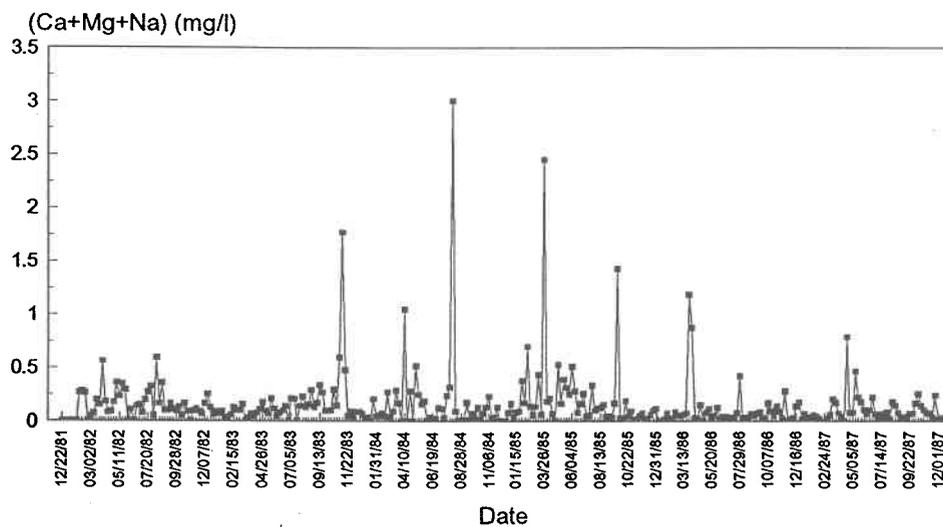


Figure 6.27. Série des concentrations du paramètre de minéralisation dans les précipitations à la station Montmorency.

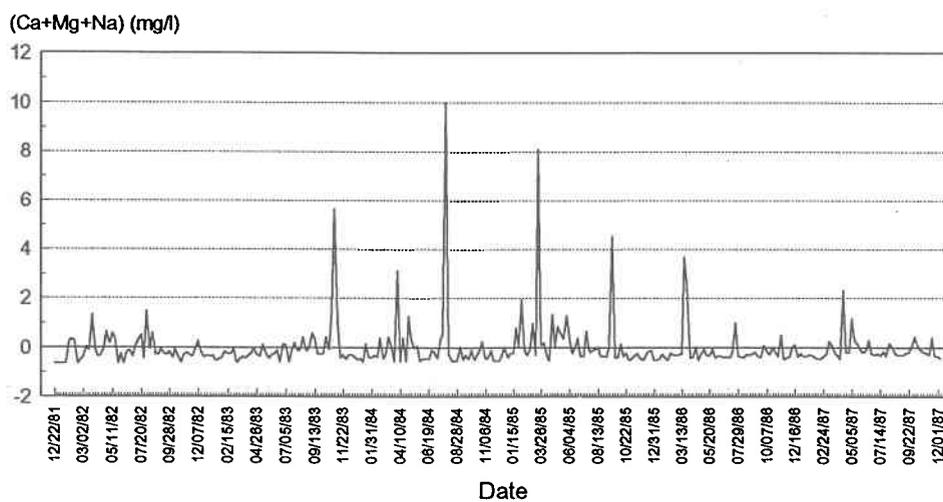


Figure 6.28. Résidus standardisés de la régression classique; série des concentrations du paramètre de minéralisation dans les précipitations.

montrent que les valeurs aberrantes n'affectent pas ce paramètre de localisation. Les statistiques des tests reflètent cependant l'effet majeur des valeurs aberrantes sur l'estimation de la variance. La surestimation importante de la variance par l'estimation classique amène les conclusions différentes des deux tests. Devant ces résultats, on doit donner une plus grande importance aux résultats de la régression robuste qui n'est pas affectée par la présence de quelques valeurs aberrantes et, par conséquent, amène une conclusion beaucoup plus représentative de l'ensemble de la série.

À titre d'illustration, les résultats de régressions autorégressives sont maintenant présentés. On note alors une statistique de $-0,92$ pour la régression autorégressive classique et de $-3,55$ pour la régression autorégressive robuste. Ces résultats, très près des résultats du tableau 6.14, permettent de voir que l'effet de l'autocorrélation est mineur lorsque le test de Durbin-Watson rejette la présence de résidus autorégressifs. L'effet mineur de l'autocorrélation est également confirmé par les résultats des tests non paramétriques, puisque les tests de Spearman/Lettenmaier et de Kendall donne des statistiques très semblables.

Les résultats du tableau 6.14 permettent donc de conclure que la série de concentrations du paramètre de minéralisation diminue significativement de 1981 à 1987. La construction de ce paramètre demande cependant une étude plus approfondie sur chaque paramètre si l'on veut porter une conclusion sur un ion en particulier (Ca, Na ou Mg). On peut cependant conclure qu'au moins un ion important utilisé dans le paramètre de minéralisation diminue significativement entre 1981 et 1987.

La figure 6.28 présente les résidus de la régression classique pour les concentrations dans les précipitations du paramètre de minéralisation. Les résidus aberrants associés aux précipitations de fortes concentrations sont facilement identifiables, mais leur modélisation est impossible, on doit simplement se fier aux méthodes peu affectées par leur présence: i.e.

régression robuste et tests non paramétriques. Comme ces méthodes donnent des conclusions identiques on n'a donc pas de raison de rejeter cette conclusion.

Étude des charges dans les précipitations

L'étude de tendance de Papineau (1989) sur des données mensuelles de 1981 à 1987 n'étudiait pas les charges des paramètres de minéralisation dans les précipitations. La figure 6.29 présente la série des charges du paramètre de minéralisation dans les précipitations, on remarque la présence de quelques valeurs aberrantes importantes, à des emplacements différents de la série de concentrations, ainsi que l'absence de saisonnalité.

Une première régression sans variable avec retard amène un test de Durbin-Watson ($d=1,72$) non significatif. On doit donc conclure que les résidus ne sont pas significativement autocorrélés et que l'introduction d'une variable avec retard est inutile. Le tableau 6.15 présente les résultats des tests de détection de tendance pour les régressions classique et robuste ainsi que les résultats des tests de Kendall et de Spearman/Lettenmaier.

Tableau 6.15

Résultats des tests de détection de tendance pour la série de charges du paramètre de minéralisation dans les précipitations.

	Régression classique	Régression robuste	Kendall	Spearman/Lettenmaier
Estimateur b_1	-0,00007	-0,00004	n.a.	n.a.
Statistique du test	-2,67	-3,05	-2,36	-1,99
Significatif ?	oui	oui	oui	oui
Estimateur d_1	0,14	0,07	n.a.	0,14

Les résultats des tests de détection utilisant la régression montrent l'effet important des valeurs aberrantes sur la détection de tendance dans la série de charges dans les précipitations du paramètre de minéralisation. Les estimateurs des pentes de la régression

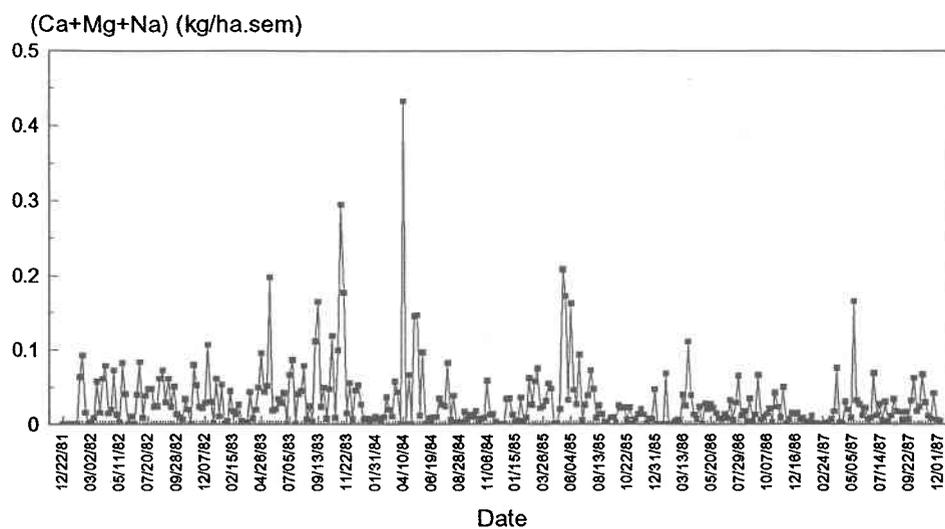


Figure 6.29. Série des charges du paramètre de minéralisation dans les précipitations à la station Montmorency.

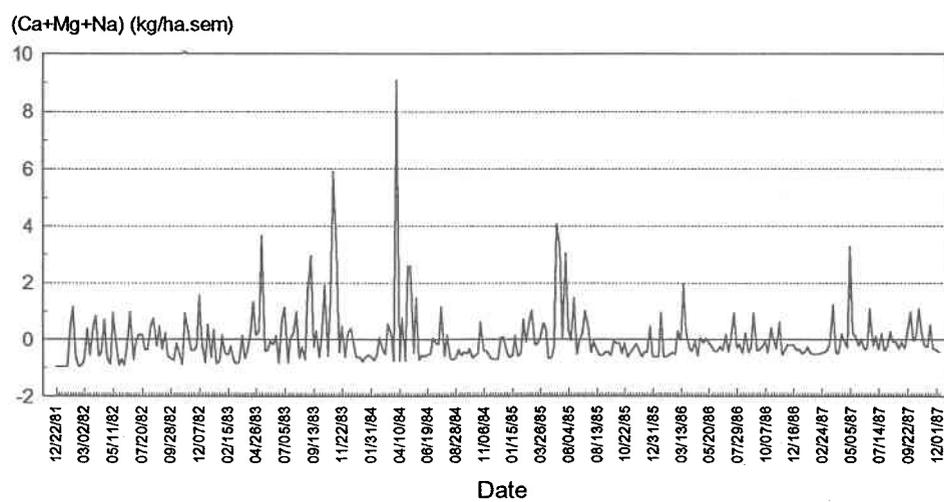


Figure 6.30. Résidus standardisés de la régression classique; série des charges du paramètre de minéralisation dans les précipitations.

montrent que les valeurs aberrantes affectent ce paramètre de localisation; les valeurs aberrantes positives au début de la série accentuent artificiellement l'amplitude de la pente négative. Les statistiques des tests ne reflètent cependant pas l'effet majeur des valeurs aberrantes à cause de l'effet des valeurs aberrantes sur l'estimation de la variance. La surestimation importante de la variance par l'estimation classique amène des statistiques assez proches pour les deux tests paramétriques. On doit quand même donner une plus grande fiabilité aux résultats de la régression robuste qui n'est pas affectée par la présence de quelques valeurs aberrantes.

À titre d'illustration, l'utilisation d'une variable avec retard amène des statistiques de tests de -2,25 pour la régression autorégressive classique et de -2,52 pour la régression autorégressive robuste. Ces résultats permettent à nouveau de conclure que l'effet de l'autocorrélation est mineur lorsque le test de Durbin Watson rejette l'autocorrélation dans les résidus. L'effet mineur de l'autocorrélation est confirmé par les résultats des tests non paramétriques, puisque les tests de Spearman/Lettenmaier et de Kendall donne des statistiques assez semblables. Les tests non paramétriques concluent à la présence d'une tendance négative significative (le résultat du test de Spearman/Lettenmaier est cependant à la limite de la signification).

Ces différents résultats permettent de conclure que la série de charges du paramètre de minéralisation diminue significativement de 1981 à 1987. Rappelons que la construction de ce paramètre demande cependant une étude plus approfondie si l'on veut conclure sur un ion en particulier (Ca, Na ou Mg).

La figure 6.30 présente les résidus de la régression classique pour les charges dans les précipitations du paramètre de minéralisation. Les résidus aberrants associés aux précipitations de fortes charges sont facilement identifiables, mais sont très difficile à modéliser, on doit donc se fier aux méthodes peu affectées par leur présence: i.e. régression robuste et tests non paramétriques.

Étude du centre du lac.

L'étude de tendance par Papineau (1989) sur des données mensuelles de 1981 à 1987 a montré que la concentration de calcium diminuait significativement au centre du lac, mais n'a détecté aucune tendance pour les concentrations de sodium et de magnésium. La figure 6.31 présente la série de concentrations du paramètre de minéralisation au centre du lac, on remarque la présence d'une forte saisonnalité et de valeurs aberrantes (diminutions importantes) aux printemps. Ces valeurs aberrantes sont associées à la dilution des paramètres de minéralisation lors de la fonte printanière. La saisonnalité s'explique, selon Papineau (1989), par une augmentation des concentrations en hiver à cause de la recharge du lac par un écoulement souterrain.

Une première régression sans variable avec retard amène un test de Durbin-Watson ($d=0,39$) fortement significatif. On doit donc conclure que les résidus sont significativement autocorrélés et que l'utilisation d'une régression autorégressive est nécessaire. Le tableau 6.16 présente les résultats des tests de détection de tendance pour les régressions autorégressives saisonnières classique et robuste, ainsi que les résultats des tests de Kendall Saisonnier, Kendall sur les Moyennes Saisonnières et d'Hirsch et Slack.

Tableau 6.16

Résultats des tests de détection de tendance pour la série de concentrations du paramètre de minéralisation au centre du lac.

	Régression Saisonnière Autorégressive Classique	Régression Saisonnière Autorégressive Robuste	Kendall Saisonnier	Kendall Moyennes Saisonnières	Hirsch et Slack
Estimateur b_1	-0,0002	-0,00007	n.a.	n.a.	n.a.
Statistique du test	-0,92	-0,53	-1,24	-0,94	-1,04
Significatif ?	non	non	non	non	non
Estimateur d_1	0,71	0,83	n.a.	n.a.	n.a.

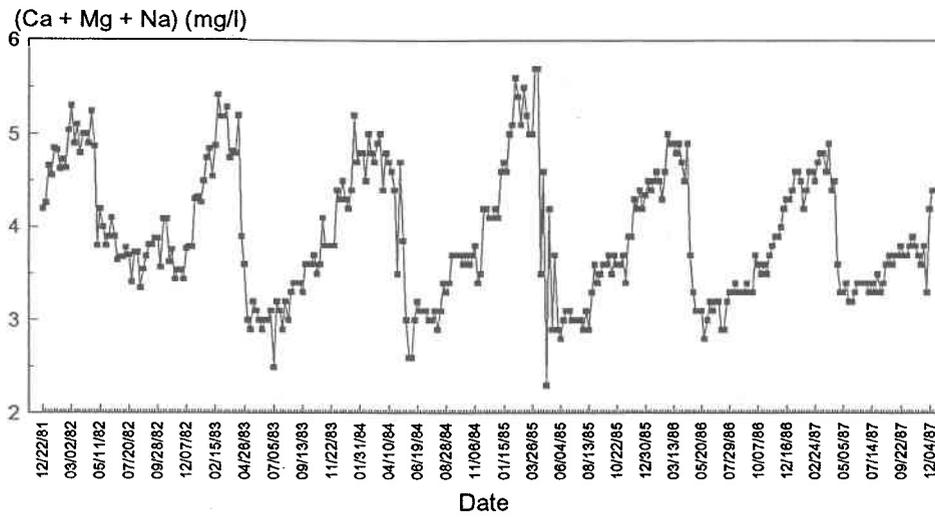


Figure 6.31. Série des concentrations du paramètre de minéralisation au centre du lac Laflamme.

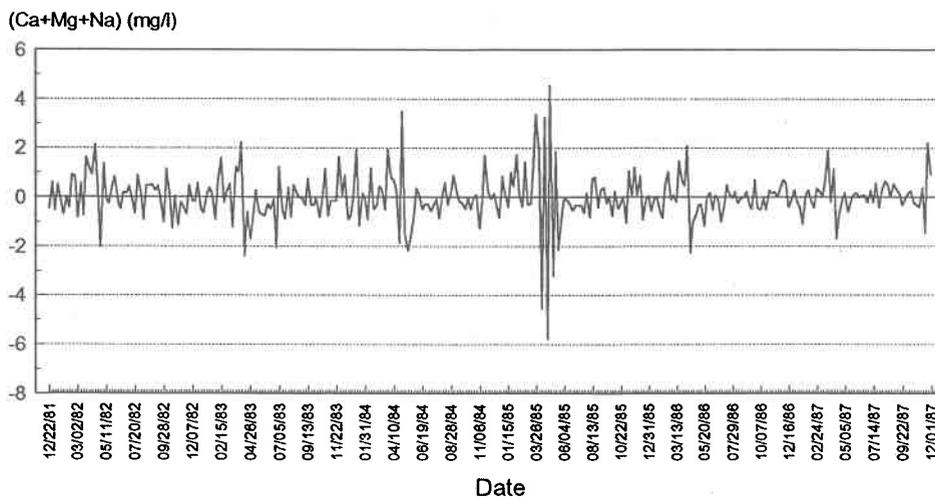


Figure 6.32. Résidus standardisés de la régression saisonnière autorégressive classique; série des concentrations du paramètre de minéralisation au centre du lac.

Les résultats des tests de régression montrent l'effet important des valeurs aberrantes: 1) pente trois fois plus grande pour l'estimation classique; 2) estimation de la variance plus grande pour l'estimation classique et; 3) statistique plus petite pour l'estimation robuste. Les résultats des trois tests non paramétriques sont très près les uns des autres; il n'y a donc pas d'effet important de l'autocorrélation une fois que les saisons sont prises en compte et le regroupement en moyennes saisonnières n'affecte pas la détection de tendance pour cette série.

La figure 6.32 présente les résidus de la régression saisonnière autorégressive classique sur les concentrations au centre du lac. Cette figure montre la présence de résidus fortement aberrants ce qui permet généralement de conclure qu'il est préférable de se fier à la régression robuste. Cependant, les résultats des tests non paramétriques sont plus près de la régression classique que de la régression robuste. Pour expliquer cette situation, il faut se rappeler que l'étude du chapitre 5 a montré que les tests non paramétriques sont beaucoup plus affectés par l'emplacement des valeurs aberrantes que la régression robuste. Comme les résidus aberrants très importants sont négatifs et situés dans la seconde partie de la série, ces résidus ont tendance à accentuer artificiellement l'amplitude de la pente. Il est fort probable que cet emplacement des résidus aberrants ait un effet sur les tests non paramétriques et il est certain qu'elle ait un effet sur la régression classique. Devant de tels résultats, la régression robuste apparaît la méthode la plus fiable puisqu'elle est moins influencée par l'emplacement des valeurs aberrantes.

Étude des concentrations à l'exutoire.

L'étude de tendance de Papineau (1989) ne présente pas de détection de tendance pour les concentrations des paramètres de minéralisation à l'exutoire. La figure 6.33 présente la série de concentrations du paramètre de minéralisation à l'exutoire, on remarque la présence d'une forte saisonnalité et de valeurs aberrantes (diminutions importantes) aux

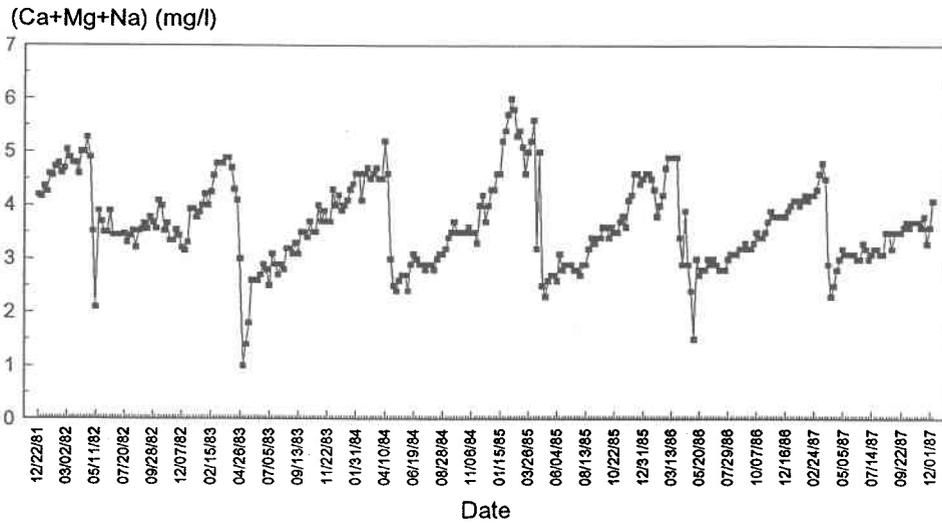


Figure 6.33. Série des concentrations du paramètre de minéralisation à l'exutoire du lac Laflamme.

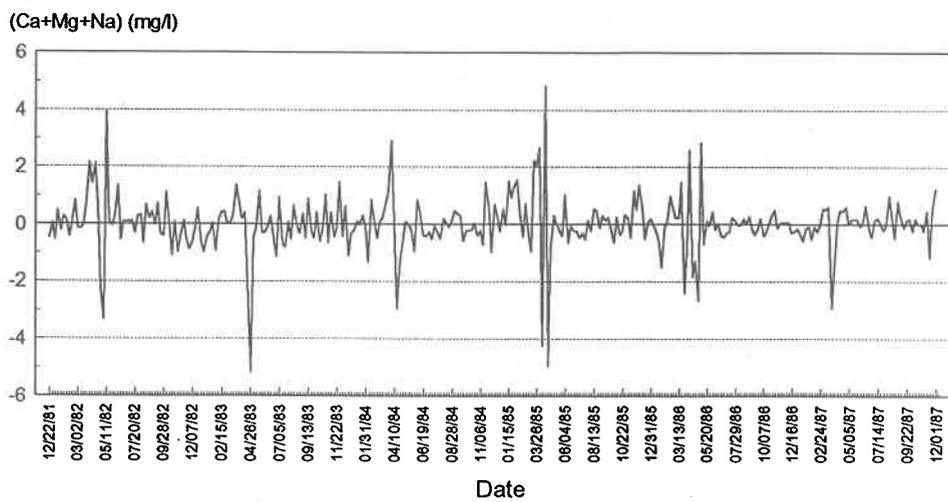


Figure 6.34. Résidus standardisés de la régression saisonnière autorégressive classique; série des concentrations du paramètre de minéralisation à l'exutoire.

printemps. Ces valeurs aberrantes sont associées à la dilution des paramètres de minéralisation lors de la fonte printanière.

Une première régression sans variable avec retard amène un test de Durbin-Watson ($d=0,73$) fortement significatif. On doit donc conclure que les résidus sont significativement autocorrélés et que l'utilisation d'une régression autorégressive est nécessaire. Le tableau 6.17 présente les résultats des tests de détection de tendance pour les régressions autorégressives saisonnières classique et robuste, ainsi que les résultats des tests de Kendall Saisonnier, Kendall sur les Moyennes Saisonnières et d'Hirsch et Slack.

Tableau 6.17

Résultats des tests de détection de tendance pour la série de concentrations du paramètre de minéralisation à l'exutoire.

	Régression Saisonnière Autorégressive Classique	Régression Saisonnière Autorégressive Robuste	Kendall Saisonnier	Kendall Moyennes Saisonnières	Hirsch et Slack
Estimateur b_1	-0,00021	-0,00004	n.a.	n.a.	n.a.
Statistique du test	-0,86	-0,27	-0,86	-0,19	-0,17
Significatif ?	non	non	non	non	non
Estimateur d_1	0,72	0,82	n.a.	n.a.	n.a.

Les résultats du tableau 6.17 montrent un effet important des valeurs aberrantes sur le résultat du test utilisant la régression classique. De son côté, la statistique du test de Kendall Saisonnier est gonflée par la présence d'autocorrélation qui n'est pas prise en compte par ce test. Finalement, les trois autres tests présentent des statistiques très semblables en plus d'être très appropriés à la situation. On peut donc conclure que la tendance de diminution des concentrations à l'exutoire est loin d'être acceptée.

La figure 6.34 présente la série de résidus de la régression saisonnière autorégressive. Les valeurs aberrantes sont très visibles, mais aucune caractéristique des résidus ne permet de douter de la conclusion du paragraphe précédent.

Étude des charges à l'exutoire.

L'étude de tendance de Papineau (1989) sur des données mensuelles à l'exutoire de 1981 à 1987 n'a montré aucune tendance significative dans les séries de charges des différents ions associés à la minéralisation (Ca, Na et Mg). La figure 6.35 présente la série de charges du paramètre de minéralisation à l'exutoire, on remarque la présence d'une forte saisonnalité et de fortes valeurs aberrantes aux printemps. Ces valeurs aberrantes sont associées aux fortes crues à l'exutoire lors de la fonte printanière.

Une première régression sans variable avec retard amène un test de Durbin-Watson ($d=1,04$) significatif. On utilise donc la régression autorégressive. Le tableau 6.18 présente les résultats des tests de détection de tendance pour les régressions autorégressives saisonnières classique et robuste, ainsi que les résultats des tests de Kendall Saisonnier, Kendall sur les Moyennes Saisonnières et d'Hirsch et Slack.

Tableau 6.18

Résultats des tests de détection de tendance pour la série de charges du paramètre de minéralisation à l'exutoire.

	Régression Saisonnière Autorégressive Classique	Régression Saisonnière Autorégressive Robuste	Kendall Saisonnier	Kendall Moyennes Saisonnières	Hirsch et Slack
Estimateur b_1	0,00006	0,00002	n.a.	n.a.	n.a.
Statistique du test	0,33	0,21	0,21	0,19	0,16
Significatif ?	non	non	non	non	non
Estimateur d_1	0,53	0,6	n.a.	n.a.	n.a.

Les résultats des tests des régressions classique et robuste, assez semblables, montrent que les valeurs aberrantes ont une influence sur l'estimateur classique de la pente, mais que l'effet sur l'estimateur classique de la variance annule la surestimation de la pente dans la construction de la statistique du test.

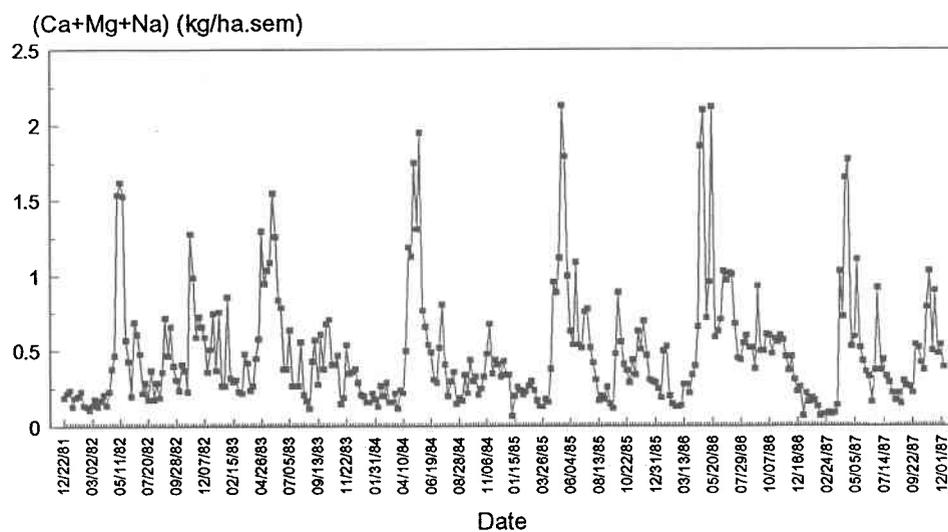


Figure 6.35. Série des charges du paramètre de minéralisation à l'exutoire du lac Laflamme.

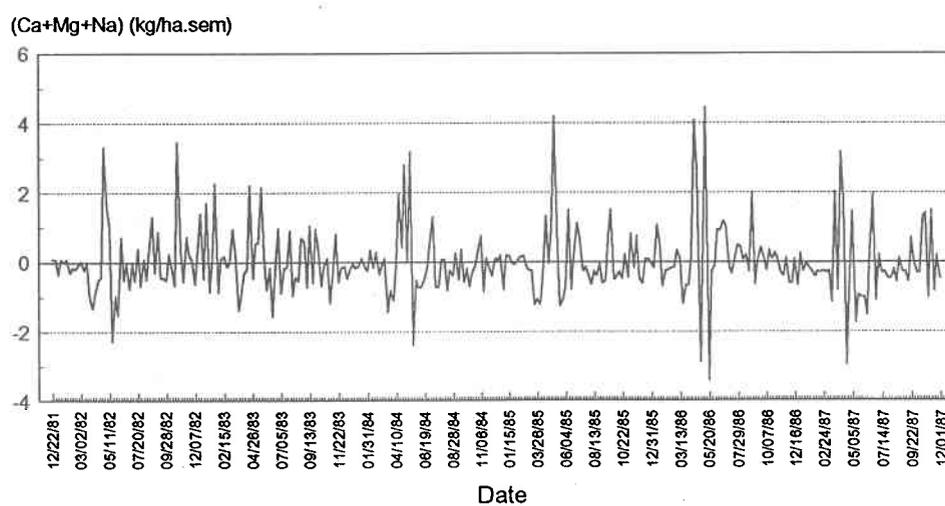


Figure 6.36. Résidus standardisés de la régression saisonnière autorégressive classique; série des charges du paramètre de minéralisation à l'exutoire.

Les résultats assez proches des tests non paramétriques montrent la faible influence de l'autocorrélation une fois que les saisonnalités sont prises en compte, puisque le test de Kendall Saisonnier ne corrige pas la présence d'autocorrélation mais la statistique obtenue est quand même presque identique aux statistiques des tests prenant en compte la présence de saisonnalité et d'autocorrélation. Les résultats des tests non paramétriques montrent également la faible influence d'un regroupement des observations sous forme de moyennes saisonnières.

La figure 6.36 présente les résidus de la régression saisonnière autorégressive classique pour la série des charges à l'exutoire du paramètre de minéralisation. Les fortes charges exportées lors des fontes printanières sont associées à de forts résidus qu'il est très difficile de modéliser dans les modèles de régression. Il est donc préférable de se fier aux résultats des méthodes qui sont peu influencées par la présence de telles valeurs: i.e. la régression robuste et les tests non paramétriques. Comme ces méthodes permettent la même conclusion, il est possible de conclure à l'absence de tendance dans la série des charges à l'exutoire du paramètre de minéralisation. La présence d'une forte corrélation entre les différents ions utilisés pour construire ce paramètre intégrateur permet également de conclure en l'absence de tendance pour chacune des séries de ces ions. En effet, les fortes corrélations entre les ions calcium, sodium et magnésium annulent la possibilité de tendances inverses qui pourraient s'annuler et amener un test non significatif pour la détection de tendance dans la série du paramètre de minéralisation.

6.6 Discussion sur les applications.

Le tableau 6.19 présente le résumé de l'étude de détection de tendance utilisant les tests paramétriques. Les pentes des tendances détectées sont alors présentées ainsi que les niveaux initial (en décembre 1981) et final (décembre 1987) de la série. Étant donné la présence fréquente de valeurs aberrantes, les estimateurs des régressions robustes sont

présentés pour ces différents paramètres de localisation. En absence de tendance significative, la moyenne (non robuste) est présentée.

Tableau 6.19
Résumé de l'étude de détection de tendance utilisant la régression linéaire.

	Sulfates	pH	Nitrates	Paramètre de minéralisation
Concentration dans les précipitations	<u>pas de tendance</u> moy.=1,52 mg/l	<u>pas de tendance</u> moyenne=4,5	<u>pas de tendance</u> moy.=0,28 mg/l	<u>diminution</u> -0,010mg/l.an de 0,12 mg/l à 0,06 mg/l
Charge dans les précipitations	<u>pas de tendance</u> moy.=0,37 kg/ha.sem		<u>pas de tendance</u> moy.=0,058 kg/ha.sem	<u>diminution</u> -0,002kg/ha.an de 0,027 kg/ha.sem à 0,015kg/ha.sem
Concentration au centre du lac	<u>augmentation</u> 0,115 mg/l/an de 3,74 mg/l à 4,43 mg/l	<u>augmentation</u> 0,086/an de 6,16 à 6,67	<u>pas de tendance</u> moy.=0,099mg/l	<u>pas de tendance</u> moy.=3,92 mg/l
Concentration à l'exutoire du lac	<u>augmentation</u> 0,109 mg/l/an 3,88 mg/l à 4,53 mg/l	<u>augmentation</u> 0,083/an 6,14 à 6,64	<u>pas de tendance</u> moy.=0,096mg/l	<u>pas de tendance</u> moy.=3,66 mg/l
Charge à l'exutoire du lac	<u>pas de tendance</u> moy.=0,64 kg/ha.sem		<u>diminution</u> multicollinéarité brouille l'estimation	<u>pas de tendance</u> moy.=0,48 kg/ha.sem

Le tableau 6.20 présente, de son côté, les résultats de l'étude de détection de tendance en utilisant les tests non paramétriques les plus appropriés. On remarque rapidement que les conclusions sont les mêmes dans les deux tableaux à l'exception des séries de concentrations de nitrates au centre du lac et à l'exutoire. Dans ces deux derniers cas, on a conclu que les résultats des tests non paramétriques étaient les plus fiables. Dans tous les autres cas, les informations du tableau 6.19 présentent un grand avantage sur les

résultats du tableau 6.20, puisqu'il est alors possible de connaître les niveaux des séries et l'amplitude des tendances.

Tableau 6.20

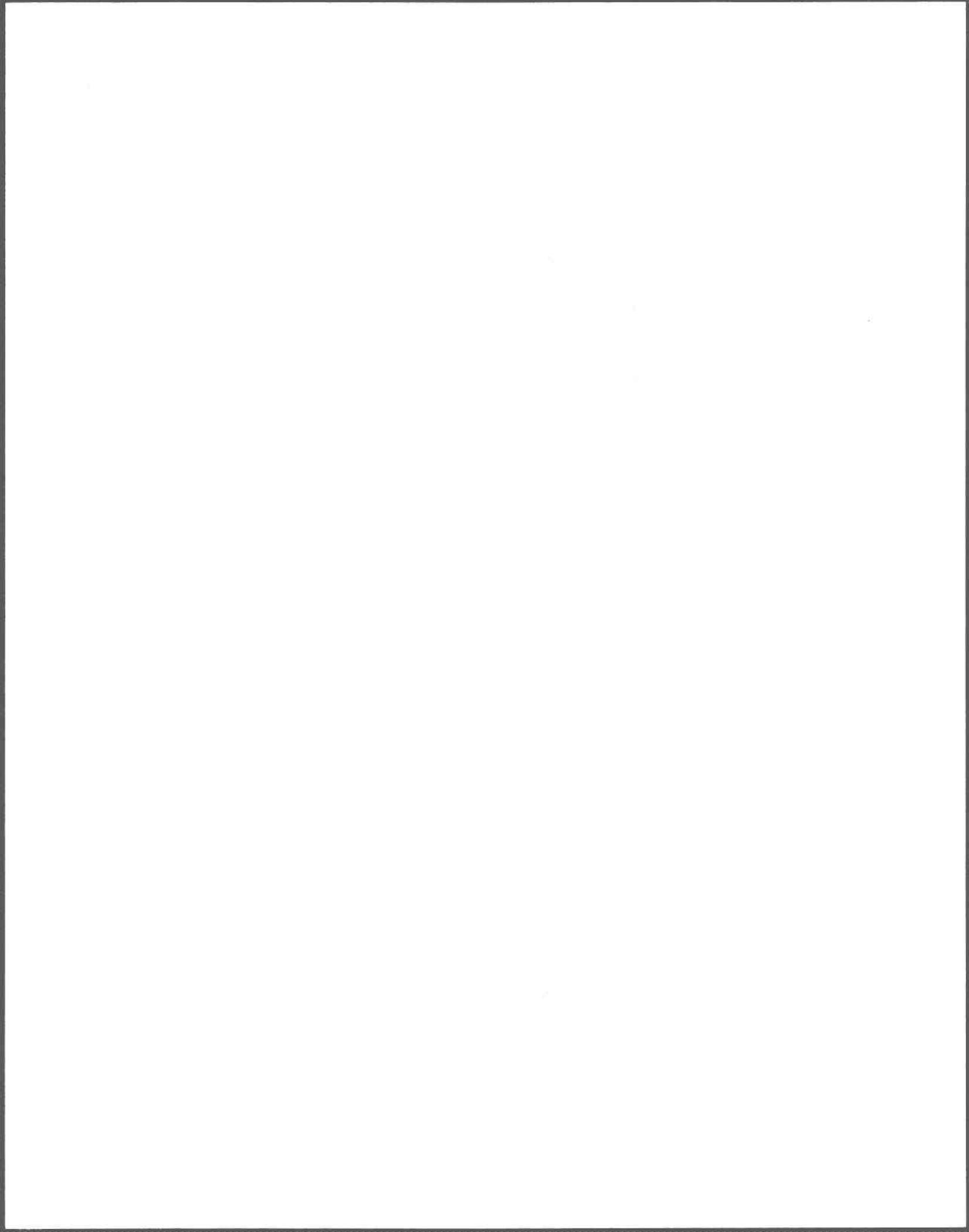
Résumé de l'étude de détection de tendance utilisant les tests non paramétriques.

	Sulfates	pH	Nitrates	Paramètre de minéralisation
Concentration dans les précipitations	<u>pas de tendance</u>	<u>pas de tendance</u>	<u>pas de tendance</u>	<u>diminution</u>
Charge dans les précipitations	<u>pas de tendance</u>		<u>pas de tendance</u>	<u>diminution</u>
Concentration au centre du lac	<u>augmentation</u>	<u>augmentation</u>	<u>diminution</u>	<u>pas de tendance</u>
Concentration à l'exutoire du lac	<u>augmentation</u>	<u>augmentation</u>	<u>diminution</u>	<u>pas de tendance</u>
Charge à l'exutoire du lac	<u>pas de tendance</u>		<u>diminution</u>	<u>pas de tendance</u>

De façon générale, on peut conclure :

1. à une diminution du paramètre de minéralisation dans les précipitations. Cette diminution se retrouve dans les concentrations (-0,01 mg/l par année de 1981 à 1987) et dans les charges (-0,002 kg/ha par année de 1981 à 1987).
2. à une augmentation des concentrations de sulfates au centre du lac et à l'exutoire. Dans les deux cas, l'augmentation est de l'ordre de 0,11 mg/l par année de 1981 à 1987.
3. à une augmentation du pH au centre du lac et à l'exutoire de l'ordre de 0,08 unité de pH par année de 1981 à 1987.
4. à une diminution des concentrations de nitrates au centre du lac et à l'exutoire en plus d'une diminution des charges de nitrates à l'exutoire. Pour ce paramètre, il est cependant difficile d'obtenir une estimation directe de l'amplitude de la tendance.

L'étude de tendance pour différentes séries des mêmes paramètres permet de mettre en évidence la grande correspondance : 1) des tendances des charges et des concentrations dans les précipitations et; 2) des tendances des concentrations au centre du lac et à l'exutoire.



CONCLUSION

Dans ce travail on cherchait à établir une procédure simple et efficace pour la détection de tendance à l'aide de la régression linéaire. La procédure développée suppose que les séries chronologiques sont bien échantillonnées et propose des adaptations pour tenir compte de la présence d'autocorrélation, de valeurs aberrantes et de saisonnalité.

Bien qu'individuellement chacune des méthodes statistiques utilisées soit bien connue, l'étude de leur utilisation conjointe sera d'une grande utilité pour quiconque désire utiliser la régression linéaire pour détecter des tendances dans des séries chronologiques. En effet, comme il est assez rare qu'une seule caractéristique restrictive soit présente dans une série, il est important de savoir ce qui se passe lorsqu'on utilise plus d'une adaptation pour tenir compte de la présence de plus d'une caractéristique restrictive.

L'étude des propriétés (seuil et puissance) du test de détection de tendance utilisant la régression linéaire dans des séries chronologiques contenant différentes combinaisons de caractéristiques restrictives constitue également de l'information nouvelle pour les utilisateurs potentiels de la régression linéaire. Plusieurs de ces utilisateurs ont souvent renoncé à cette méthode statistique simple à cause du manque de documentation sur les adaptations possibles pour tenir compte de la présence d'autocorrélation, de valeurs aberrantes et de saisonnalités. Ces utilisateurs pourront maintenant avoir, dans un seul document, les informations nécessaires à une bonne utilisation de la régression linéaire pour détecter des tendances.

Le côté pratique de ce travail amène également des informations qui ne sont pas discutées dans d'autres documents: effet de l'emplacement des valeurs aberrantes sur les

tests de détection de tendance, effet d'un mauvais choix du type de tendance, effet d'un décalage entre les saisons réelles et les saisons estimées, détection de tendance lorsque l'ordre du processus autorégressif est inconnu. Avec ces informations, on sait maintenant que certains tests ne sont pas aussi fiables qu'on le croyait et on peut donc effectuer des études de détection de tendance plus éclairées.

De façon générale, les adaptations utilisées pour la régression linéaire permettent de détecter des tendances dans la majorité des séries chronologiques bien échantillonnées. En présence d'autocorrélation seulement, l'utilisation de variable avec retard permet un test de détection de tendance avec des propriétés équivalentes à celles du test de Spearman/Lettenmaier. En présence de saisonnalités seulement, l'utilisation de variable muette permet également un test avec des propriétés équivalentes à celles du test non paramétrique approprié: Kendall Saisonnier. Finalement, en présence de valeurs aberrantes seulement, l'utilisation des M-Estimeurs de Huber permet un test qui conserve le seuil de 5% mais qui présente une puissance un peu plus faible que la puissance théorique.

En présence d'autocorrélation et de valeurs aberrantes de type dynamique (IO), l'utilisation conjointe de variable avec retard et de M-Estimeurs permet un test avec des propriétés intéressantes en plus d'être moins affecté que le test de Spearman/Lettenmaier par l'emplacement des valeurs aberrantes.

En présence d'autocorrélation et de valeurs aberrantes de type statique (AO), l'utilisation conjointe de variable avec retard et de M-Estimeurs ne permet pas toujours un test approprié. En effet, lorsque les AO n'ont pas une moyenne nulle, le seuil n'est pas conservé et la puissance est instable. Il faut cependant noter que le test de Spearman/Lettenmaier présente les mêmes restrictions puisque l'estimation du coefficient d'autorégression est à l'origine de ces problèmes et ce test "non paramétrique" nécessite un estimateur pour le "paramètre" d'autorégression.

En présence de saisonnalité et d'autocorrélation, l'utilisation conjointe de variables muettes et avec retard permet un test qui conserve le seuil et une puissance plus élevée que

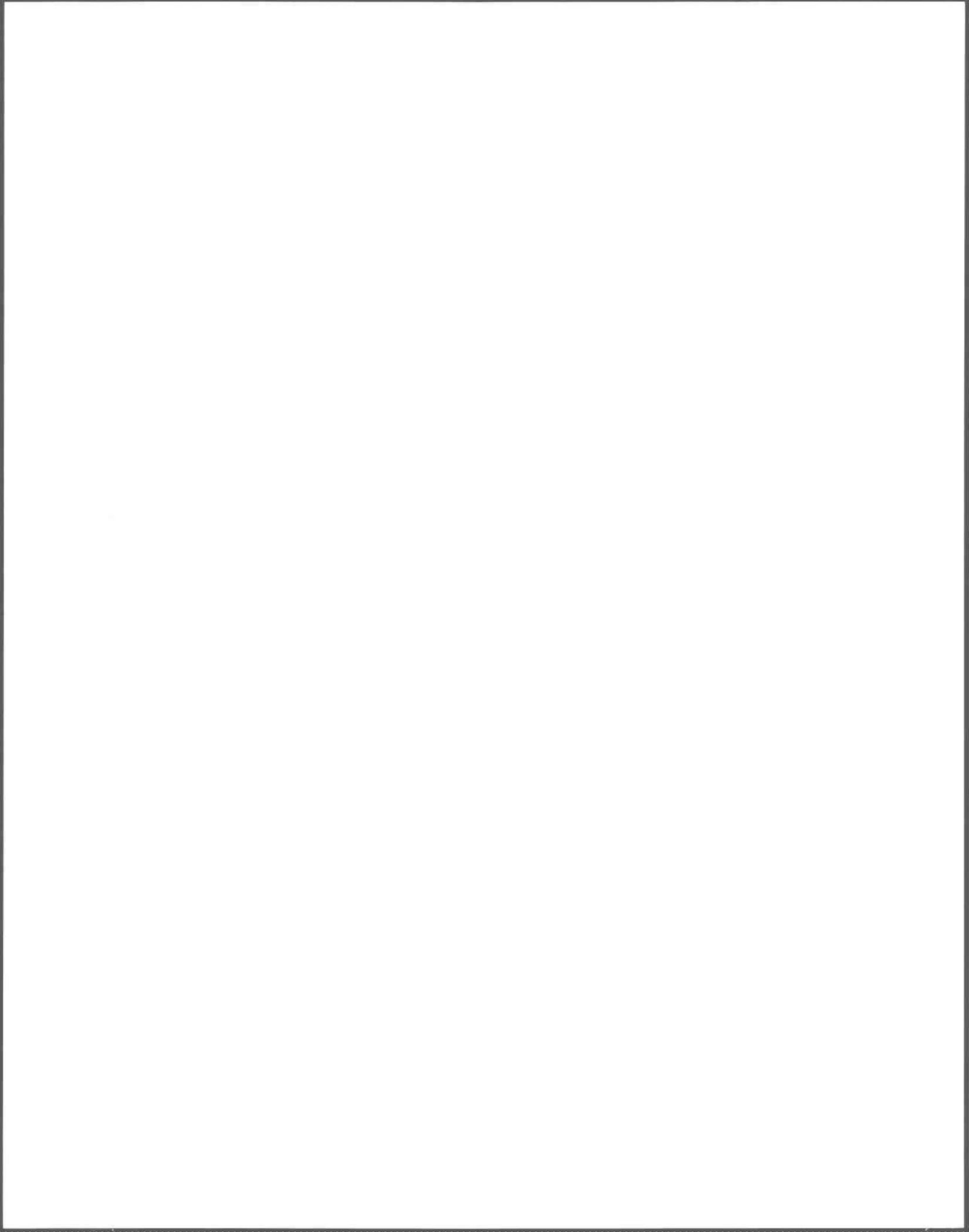
le test de Hirsch et Slack. La présence de décalage entre les saisons réelles et estimées affecte énormément le test utilisant la régression saisonnière autorégressive.

Il reste donc certaines situations qui demandent des études complémentaires avant de pouvoir utiliser adéquatement la régression : 1) estimation du coefficient d'autorégression en présence de AO; 2) détection de tendance en présence de décalage entre les saisons réelles et estimées et; 3) détection de tendance en présence de forte autocorrélation ($\phi > 0,75$).

La comparaison avec les tests non paramétriques a également permis de valider une construction nouvelle de la statistique du test de Spearman/Lettenmaier. Auparavant, il était nécessaire de trouver les valeurs critiques à l'aide de simulations de Monte Carlo. Avec la nouvelle statistique on peut effectuer le test automatiquement en utilisant les valeurs critiques de la loi de Student.

Finalement, les applications ont montré la simplicité d'utilisation de la régression linéaire pour une étude de détection de tendances et l'intérêt d'effectuer une étude de résidus afin de valider les conclusions. L'étude des résidus permet également de mettre en évidence des situations pouvant affecter la détection de tendance : emplacement des valeurs aberrantes, présence de AO, présence d'une mauvaise modélisation des saisons et présence de deux tendances de pentes différentes. Le chapitre des applications a également montré l'intérêt d'utiliser conjointement les méthodes paramétriques et non paramétriques et de comparer les résultats afin de valider les conclusions.

Le but premier de ce travail était de donner une vue d'ensemble de l'utilisation de la régression linéaire pour la détection de tendance dans des séries bien échantillonnées. En atteignant ce but on a cependant été obligé de laisser tomber l'étude complémentaire de certains points précis associés à l'une ou l'autre des caractéristiques restrictives. On n'a qu'à penser à la présence de AO de moyenne non nulle, de forte autocorrélation ou encore de décalage non systématique. Ces points devront cependant être considérés dans des études ultérieures afin de permettre une utilisation plus répandue de la régression linéaire.

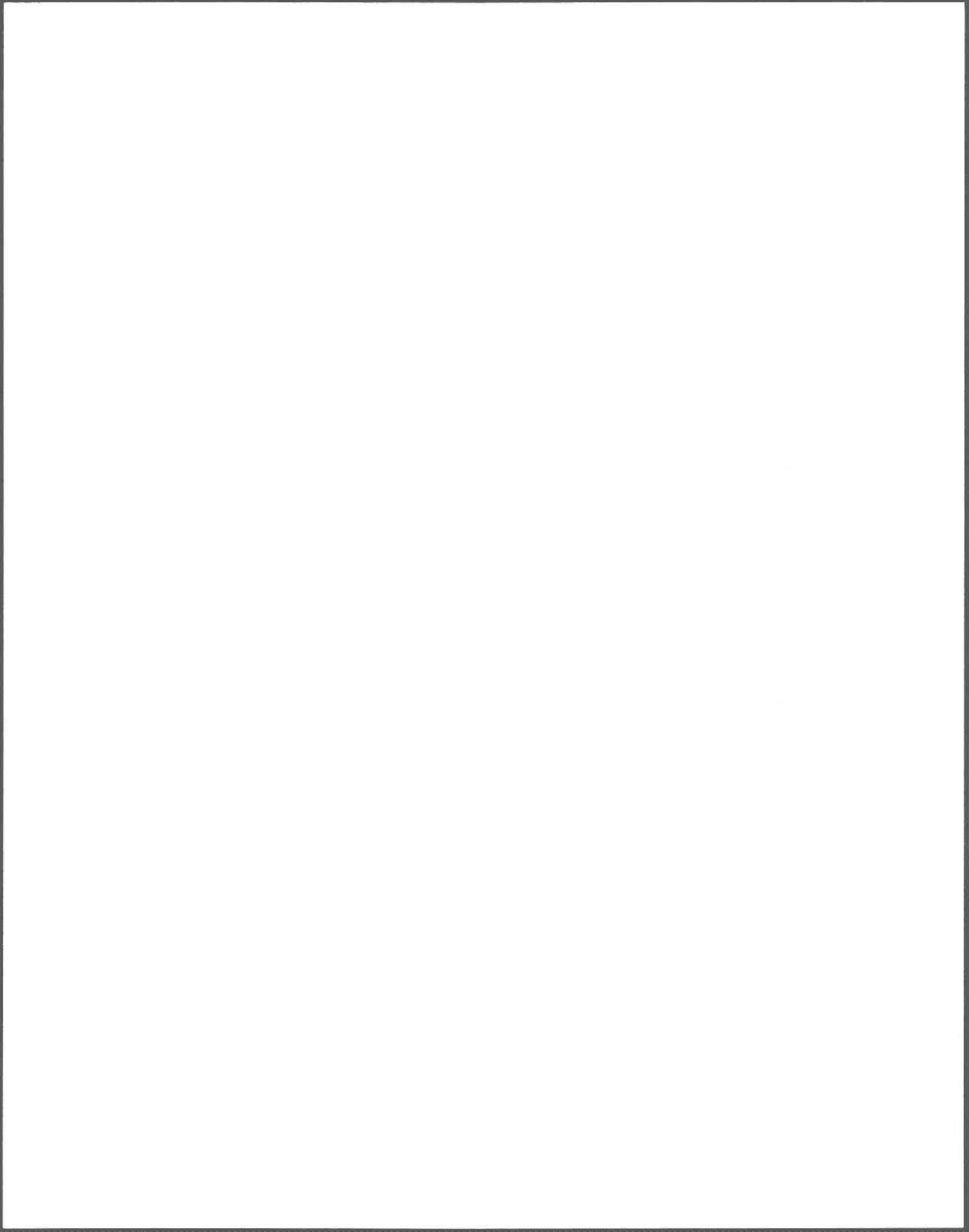


BIBLIOGRAPHIE

- ABRAHAM, B., CHUANG, A., "Outlier detection and time series modeling." , Technometrics, 1989, 31(2), 241-248.
- BARNETT, V., LEWIS, T., "Outliers in statistical data", Wiley, Chichester, 1978.
- BICKEL, P.J., DOKSUM, K.A., "Mathematical statistics", Holden-Day, Oakland, 1977, 492 pages.
- BOX, G.E.P, JENKINS, G.M., "Time series analysis: Forecasting and control", Holden-Day, Oakland, 1976, 575 pages.
- BOX, G.E.P., MULLER, "A note on the generation of random deviates.", Annals of Mathematical Statistics, 1958, 29, 610-611.
- BUSTOS, O.H., YOHAI, V.J., "Robust estimates of ARMA models", Journal of the American Statistical Association, 1986, 81, 155-168.
- CHANG, I., TIAO, G.C., CHEN, C., "Estimation of time series parameters in the presence of outliers.", Technometrics, 1988, 30(2), 193-204.
- CLUIS, D.A., LABERGE, C., HOULE, C., "Détection des tendances en qualité de l'eau". I.Méthodologie. INRS-eau, rapport scientifique #229, 122 pages; II. Applications. INRS-eau, rapport scientifique #230, 430 pages; III. Organigramme et listes. INRS-eau, rapport scientifique #238, 230 pages, 1987.
- DENBY, L., MARTIN, R.D., "Robust estimation of the first autoregressive parameter", Journal of the American Statistical Association, 1979, 74, 140-146.
- ESTERBY, S.R., EL-SHAARAWI, A.H., "Inference about the point of change in a regression model", J. R. Statist. Soc. C, 1981, 30(3), 277-285.
- FOX, A.J., "Outliers in time series", Jour. Royal Stat. Society, ser B, 1972, 34, 350-363.
- HAWKINS, D.M., "Identification of outliers", Chapman and Hall, London, 1980.
- HINKLEY, D.V., "Inference about the change-point from cumulative sum tests", Biometrika, 1971, 58(3), 509-523.
- HIPEL, K.W., McCLEOD, A.I., "Time series modelling of water resources and environmental systems." Elsevier, Amsterdam, 1994.
- HIRSCH, R.M., SLACK, J.R., SMITH, R.A., "Techniques of trend analysis for monthly water quality data", Water Resources Research, 1982, 18(1), 107-121.

- HIRSCH, R.M., SLACK, J.R., "A nonparametric trend test for seasonal data with serial dependence", Water Resources Research, 1984, 20(6), 727-732.
- HUBER, P.J., "Robust Statistics", John Wiley and Sons, New York, 1981, 308 pages.
- KENDALL, M.G., "Rank correlation methods", Charles Griffin, Londres, 1975.
- KMENTA, J., "Elements of econometrics", Macmillan, New York, 1971, 655 pages.
- LABERGE, C., "Estimation robuste des paramètres du modèle AR(1): Application à des données environnementales." Mémoire de maîtrise, Université Laval, 1988, 129 pages.
- L'ÉCUYER, P., "Efficient and portable combined pseudo-random number generators." Rapport de recherche #DIUL-RR-8612, Département d'informatique de l'Université Laval, Université Laval, Québec, Canada, G1K 7P4. 1986, 24 pages.
- LETTENMAIER, D.P., "Detection of trends in water quality data from records with dependent observations", Water Resources Research, 1976, 12(5), 1037-1046.
- MANN, H.B., "Non-parametric tests against trend", Econometrica, 1945, 13, 245-259.
- MARTIN, R.D., "Robust estimation in autoregressive models", Dans D.R. Brillinger et G.C. Tiao, Eds, Directions in time series, Institute of mathematical statistics publication, Hayward, CA, 1980, 228-254.
- MARTIN, R.D., YOHAI, V.J., "Robustness in time series and estimating ARMA models", Dans Handbook of statistics 5, Eds, D.R. Brillinger et P.R. Krishnaiah, Elsevier, New York, 1985, 119-155.
- McKIE, C., "Seasonality in Canadian homicides, 1961-1980", The Can. Journ. Statistics, 1985, 13(4) 266-268.
- NETER, J., WASSERMAN, W., "Applied linear statistical models", Richard Irwin, Homewood, 1974, 842 pages.
- PAGÉ, E.S., "Continuous inspection schemes", Biometrika, 1954, 41, 100-114.
- PAPINEAU, M., "Bilans ioniques et changements dans la qualité de l'eau du bassin calibré du lac Laflamme de 1981 à 1988." Environnement Canada, 1989, 156 pages.
- PETTIT, A.N., "A non-parametric approach to the change-point problem", Appl. Statist., 1979, 28(2), 126-135.
- PIERCE, D.A., "A survey of recent developments in seasonal adjustment", The American Statistician, 1980, 34(3) 125-134.

- SHISKIN, J., YOUNG, A.H., MUSGRAVE, J.C., "The X-11 variant of the census Method-11 seasonal adjustment program", Technical report #15, U.S. Bureau of Census, 1967.
- SMITH, A.F.M., "A Bayesian approach to inference about a change-point in a sequence of random variables, Biometrika, 1975, 62(2), 407-416.
- TSAY, R.S., TIAO, G.C., "Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and nonstationary ARMA models." Journal of the American Statistical Association, 1984, 79, 84-96.
- WONNACOTT, T.H., WONNACOTT, R.J., "Regression: A second course in statistics.", Wiley, New York, 1980, 556 pages.



APPENDICE A

Sous-routine FORTRAN
pour le calcul des M-estimateurs de Huber.

```

SUBROUTINE REGROB (NOBS, P, X, Y, B0, SIGMA0, VARB)
C   * * * * *
C   *
C   * Cette sous-routine FORTRAN 77 permet de calculer les
C   * M-Estimeurs de HUBER dans une régression multiple.
C   *
C   * NOBS : Nombre d'observations.
C   * P : Nombre de variables indépendantes.
C   * X : Matrice NOBS x (P+1), contenant des 1 sur la première colonne
C   *     et les variables indépendantes sur les autres colonnes
C   * Y : Vecteurs de longueur NOBS, contenant les valeurs de la
C   *     variable dépendante.
C   * B0 : Valeurs initiales pour les p+1 estimateurs associés aux para-
C   *     mètres des variables indépendantes. On utilise généralement
C   *     les estimateurs des moindres carrés.
C   * SIGMA0 : Valeur initiale pour le paramètre d'échelle. On utilise
C   *     généralement l'estimateur des moindres carrés.
C   * VARB : Vecteur de longueur (p+1)x(p+2)/2 contenant les entrées
C   *     de l'inverse de la matrice d'information (Inf=(X'X)-1): les
C   *     valeurs sont V(1)=Inf(1,1), V(2)=Inf(2,1), V(3)=Inf(2,2),
C   *     V(4)=Inf(3,1), V(5)=Inf(3,2), V(6)=Inf(3,3), etc.
C   *
C   * Certains paramètres sont lus dans un fichier nommé regrob.dat afin
C   * de ne pas avoir à recompiler à chaque changement de paramètre.
C   * KK : Valeur de K à partir de laquelle on limite l'influence des
C   *     résidus standardisés.
C   * BK : Valeur de B(K) associée à KK, voir thèse pour définition
C   * EPS: Critère de convergence des estimateurs (.01, .001, etc)
C   * * * * *
C
REAL X(350,11), Y(350), B0(11), B1(11), XB(350), Z(350), SOM(11)
REAL TAU(11), IX(11,11), SB(11), B1ST(11), DIFF(11), R(350)
REAL VARB(66), KK, KS, BK, EPS, CC, Q, S, SS, S0
INTEGER TEST(11), P
OPEN (7, FILE='REGROB.DAT', STATUS='OLD')
READ (7, *) KK, BK, EPS
Q=BK*(NOBS-P)
ITER=0
TS=0
T=0
10 IF (ITER.GE.100) GOTO 900
IF (TS.EQ.0) GOTO 15
IF (T.EQ.1) GOTO 500
C   *****
C   CALCUL DU PARAMETRE D'ECHELLE
C   *****
15 CARRE=0
CMPT=0
KS=KK*SIGMA0
DO 50 I=1,NOBS
TEMP=0.0
DO 40 J=1,P
TEMP=TEMP+X(I,J)*B0(J)

```

```

40    CONTINUE
      XB(I)=TEMP
      Z(I)=Y(I)-XB(I)
      IF (ABS(Z(I)).GT.KS) THEN
          CMPT=CMPT+1
      ELSE
          CARRE=CARRE+Z(I)*Z(I)
      ENDIF
50    CONTINUE
      S0=0.5*Q
      S=Q-(KK*KK*CMPT)
      SS=KK*KK*SIGMA0*SIGMA0*CMPT
      SIGMA1H=SQRT((CARRE+SS)/Q)
      SIGMA1=SIGMA1H
C     *****
C     CALCUL DU VECTEUR BETA
C     *****
      CMPT2=0
      S=KK*SIGMA1
      DO 70 I=1,NOBS
          IF (ABS(Z(I)).GT.S) THEN
              R(I)=S*ABS(Z(I))/Z(I)
          ELSE
              R(I)=Z(I)
              CMPT2=CMPT2+1
          ENDIF
70    CONTINUE
      DO 80 J=1,P
          TEMP2=0.0
          DO 75 I=1,NOBS
              TEMP2=TEMP2+X(I,J)*R(I)
75    CONTINUE
          TAU(J)=TEMP2
80    CONTINUE
          IT=0
          DO 90 I=1,P
              DO 90 J=1,I
                  IT=IT+1
                  IX(I,J)=VARB(IT)
                  IX(J,I)=VARB(IT)
90    CONTINUE
          ZP=P
          ZNOB=NOBS
          CTE=ZNOB**2*SIGMA1**2*BK*((1+(ZP/ZNOB))*((ZNOB-CMPT2)/CMPT2))**2
          RMSE=CTE/CMPT2**2
          DO 100 I=1,P
              TEMP3=0.0
              DO 95 J=1,P
                  TEMP3=TEMP3+IX(I,J)*TAU(J)
95    CONTINUE
          B1(I)=B0(I)+TEMP3
          SB(I)=RMSE*IX(I,I)
100   CONTINUE

```

```

C *****
C TEST SUR LA CONVERGENCE DES ESTIMATEURS
C *****
  W=EPS*SIGMA1
  IF (ABS(SIGMA1-SIGMA0).LT.W) THEN
    DO 110 I=1,P
      TEST(I)=0
      DIFF(I)=ABS(B1(I)-B0(I))
      VAR=SB(I)*EPS
      IF (DIFF(I).LE.VAR) TEST(I)=1
      B1ST(I)=B1(I)/SB(I)**0.5
      TS=1
      T=1
110 CONTINUE
    ELSE
      TS=0
    ENDIF
120 DO 130 I=1,P
      T=T*TEST(I)
      B0(I)=B1(I)
130 CONTINUE
    SIGMA0=SIGMA1
    ITER=ITER+1
    GOTO 10
C *****
C IMPRESSION DES RÉSULTATS
C *****
500 WRITE (8,'(20F6.2)')(B0(I),I=1,P),(SB(I),I=1,P),(B1ST(I),I=1,P)
    $,SIGMA0
    WRITE (11,'(4F9.5)')B0(2),B1ST(2),SIGMA0,RMSE
    GOTO 999
900 WRITE (8,*) 'L' 'ALGORITHME N' 'A PAS CONVERGE'
999 RETURN
    END

```

APPENDICE B

Banque des données utilisées du lac Laflamme.

Liste des variables associées aux colonnes A à R.

A : Paramètre de minéralisation : Charges à l'exutoire en kg/ha.sem.

B : Paramètre de minéralisation : Concentrations à l'exutoire en mg/l.

C : Paramètre de minéralisation : Concentrations au centre du lac en mg/l.

D : Paramètre de minéralisation : Concentrations dans les précipitations en kg/ha.sem.

E : Paramètre de minéralisation : Charges dans les précipitations en mg/l.

F : Nitrates : Charges à l'exutoire en kg/ha.sem.

G : Nitrates : Concentrations à l'exutoire en mg/l.

H : Nitrates : Concentrations au centre du lac en mg/l.

I : Nitrates : Concentrations dans les précipitations en kg/ha.sem.

J : Nitrates : Charges dans les précipitations en mg/l.

K : pH à l'exutoire.

L : pH au centre du lac.

M : pH dans les précipitations.

N : Sulfates : Charges à l'exutoire en kg/ha.sem.

O : Sulfates : Concentrations à l'exutoire en mg/l.

P : Sulfates : Concentrations au centre du lac en mg/l.

Q : Sulfates : Concentrations dans les précipitations en kg/ha.sem.

R : Sulfates : Charges dans les précipitations en mg/l.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
0.19	4.20	4.20	0.00	0.000	0.01	0.054	0.09	0.00	0.000	6.3	6.4	4.5	0.17	3.8	3.3	0.00	0.00
0.22	4.17	4.26	0.00	0.000	0.01	0.052	0.10	0.00	0.000	6.3	6.4	4.5	0.17	3.3	3.1	0.00	0.00
0.24	4.36	4.66	0.00	0.000	0.01	0.052	0.13	0.00	0.000	6.3	6.2	4.5	0.20	3.4	3.4	0.00	0.00
0.13	4.26	4.56	0.00	0.000	0.01	0.051	0.12	0.00	0.000	6.3	6.5	4.5	0.10	3.7	4.1	0.14	0.00
0.19	4.59	4.85	0.00	0.000	0.01	0.052	0.12	0.00	0.000	6.5	6.4	4.5	0.14	3.1	3.0	0.00	0.00
0.20	4.56	4.83	0.00	0.000	0.01	0.046	0.15	0.00	0.000	6.7	6.6	4.5	0.16	3.5	3.4	0.00	0.00
0.23	4.73	4.63	0.27	0.064	0.01	0.040	0.16	0.47	0.113	6.5	6.2	4.5	0.17	3.9	3.6	0.49	0.12
0.14	4.79	4.73	0.28	0.093	0.01	0.041	0.21	0.19	0.061	6.3	6.1	4.9	0.10	3.3	3.4	0.39	0.13
0.13	4.61	4.64	0.26	0.016	0.00	0.047	0.23	0.26	0.016	6.5	6.3	4.7	0.09	3.4	3.3	0.61	0.04
0.11	4.69	5.04	0.00	0.000	0.00	0.030	0.20	1.38	0.050	6.5	6.5	4.1	0.08	3.7	3.5	2.83	0.10
0.14	5.04	5.30	0.04	0.003	0.00	0.040	0.28	0.05	0.003	6.3	6.3	5.2	0.11	3.9	3.6	1.55	0.10
0.18	4.90	4.90	0.07	0.009	0.01	0.043	0.21	0.52	0.063	6.3	6.3	4.4	0.14	3.7	3.6	0.80	0.10
0.13	4.80	5.10	0.20	0.058	0.01	0.044	0.41	0.50	0.149	6.5	6.4	4.1	0.08	3.5	3.3	2.18	0.64
0.17	4.80	4.80	0.15	0.016	0.01	0.045	0.25	0.21	0.022	5.7	5.6	4.3	0.12	3.1	3.1	2.30	0.24
0.21	4.60	5.00	0.56	0.062	0.01	0.046	0.25	0.37	0.041	6.0	6.1	4.1	0.14	3.6	3.8	2.69	0.30
0.14	5.00	5.00	0.18	0.079	0.01	0.048	0.28	0.22	0.097	6.2	6.1	4.3	0.09	3.3	3.4	1.59	0.68
0.23	5.00	4.90	0.08	0.015	0.01	0.046	0.29	0.13	0.023	6.1	6.2	4.7	0.16	3.1	3.3	0.52	0.10
0.38	5.27	5.24	0.09	0.020	0.02	0.043	0.36	0.20	0.044	6.0	6.0	4.5	0.30	3.8	3.8	0.50	0.11
0.47	4.90	4.87	0.18	0.073	0.05	0.059	0.39	0.34	0.142	5.9	5.9	4.0	0.43	4.4	4.1	4.73	1.96
1.54	3.52	3.80	0.36	0.013	0.24	0.044	0.25	0.25	0.009	6.0	6.1	4.1	1.92	3.4	3.1	3.50	0.13
1.62	2.10	4.20	0.23	0.003	0.19	0.038	0.30	0.08	0.001	5.0	6.3	4.8	1.96	3.6	3.6	1.92	0.03
1.53	3.90	4.00	0.35	0.083	0.09	0.056	0.29	0.27	0.063	5.8	5.8	4.1	1.47	3.6	3.6	3.15	0.75
0.57	3.70	3.80	0.29	0.041	0.03	0.036	0.24	0.45	0.063	6.1	6.2	3.9	0.59	3.7	3.7	4.12	0.58
0.43	3.50	3.90	0.00	0.000	0.01	0.028	0.20	0.00	0.000	6.2	6.2	4.5	0.46	3.6	3.6	0.00	0.00
0.20	3.50	4.10	0.11	0.011	0.00	0.026	0.13	0.28	0.027	6.1	6.0	4.6	0.21	3.9	3.7	3.32	0.33
0.69	3.90	3.90	0.00	0.000	0.01	0.031	0.10	0.00	0.000	6.2	6.5	4.5	0.71	3.8	3.6	0.00	0.00
0.61	3.45	3.65	0.14	0.040	0.01	0.021	0.08	0.18	0.052	6.4	6.4	4.3	0.69	3.8	3.8	1.74	0.50
0.48	3.45	3.68	0.15	0.084	0.00	0.046	0.06	0.27	0.149	6.4	6.5	4.4	0.57	4.0	4.5	2.23	1.25
0.22	3.45	3.68	0.08	0.009	0.00	0.041	0.04	0.15	0.018	6.3	6.4	4.7	0.24	4.2	4.2	1.28	0.15
0.29	3.45	3.78	0.20	0.039	0.00	0.027	0.01	0.12	0.023	6.3	6.4	4.7	0.34	3.7	3.8	1.28	0.25
0.18	3.47	3.70	0.27	0.048	0.00	0.025	0.00	0.35	0.063	6.1	5.9	4.2	0.23	4.6	4.6	3.76	0.67
0.37	3.31	3.41	0.33	0.048	0.00	0.023	0.00	0.33	0.048	6.3	6.5	4.4	0.43	4.0	4.0	2.11	0.31
0.18	3.43	3.73	0.05	0.024	0.00	0.023	0.00	0.02	0.007	6.5	6.5	5.1	0.20	4.0	3.8	0.44	0.19
0.29	3.53	3.73	0.60	0.025	0.00	0.024	0.01	0.14	0.006	6.5	6.5	5.6	0.30	3.7	3.7	1.38	0.06
0.19	3.22	3.35	0.16	0.062	0.00	0.023	0.01	0.12	0.045	6.4	6.4	4.7	0.18	3.1	3.5	1.14	0.43
0.36	3.52	3.55	0.36	0.073	0.00	0.024	0.00	0.22	0.045	6.3	6.4	4.5	0.36	3.3	3.3	1.62	0.33
0.72	3.55	3.69	0.10	0.030	0.00	0.038	0.01	0.16	0.047	6.2	6.2	4.6	0.73	3.7	3.7	1.42	0.41
0.47	3.66	3.81	0.10	0.062	0.00	0.050	0.02	0.17	0.102	6.0	5.9	4.6	0.46	3.5	3.3	1.22	0.75
0.66	3.56	3.81	0.16	0.024	0.00	0.032	0.02	0.04	0.006	6.2	6.1	5.3	0.63	3.5	3.4	0.51	0.08
0.40	3.78	3.88	0.12	0.051	0.00	0.025	0.03	0.22	0.095	6.1	5.9	4.3	0.38	3.5	3.4	2.71	1.19
0.31	3.68	3.88	0.09	0.014	0.00	0.025	0.03	0.08	0.012	6.2	6.2	4.7	0.29	3.4	3.4	0.97	0.14
0.24	3.57	3.57	0.13	0.010	0.00	0.028	0.03	0.09	0.007	6.2	6.4	4.4	0.22	3.3	3.4	1.35	0.10
0.41	4.09	4.09	0.06	0.007	0.00	0.042	0.04	0.16	0.020	6.7	6.7	4.3	0.35	3.4	3.4	3.20	0.40
0.37	3.99	4.09	0.16	0.034	0.00	0.039	0.03	0.44	0.091	6.3	6.4	4.2	0.35	3.6	3.5	2.36	0.49
0.23	3.53	3.63	0.09	0.020	0.00	0.039	0.05	0.17	0.037	6.5	6.6	4.6	0.23	3.5	3.5	1.09	0.24
1.28	3.66	3.76	0.00	0.000	0.03	0.035	0.06	0.00	0.000	6.4	6.4	4.5	1.50	3.8	3.9	1.92	0.02
0.99	3.34	3.44	0.10	0.080	0.03	0.036	0.09	0.33	0.275	6.2	6.3	4.4	1.32	4.4	4.3	1.74	1.42
0.59	3.34	3.54	0.12	0.053	0.03	0.040	0.11	0.29	0.134	6.1	6.2	4.4	0.76	4.6	4.4	1.56	0.71
0.73	3.54	3.54	0.09	0.024	0.03	0.036	0.11	0.43	0.109	6.2	6.2	4.4	0.91	4.3	4.3	1.17	0.30
0.66	3.43	3.44	0.07	0.022	0.04	0.036	0.12	0.36	0.107	6.1	6.1	4.3	0.92	4.5	4.1	1.79	0.53
0.59	3.21	3.77	0.17	0.029	0.04	0.025	0.13	0.38	0.067	6.2	6.2	4.3	0.87	4.6	4.5	1.50	0.26
0.36	3.16	3.79	0.25	0.107	0.03	0.025	0.15	0.41	0.172	6.0	6.3	4.3	0.56	4.8	4.2	2.54	1.07

0.51	3.31	3.79	0.13	0.031	0.04	0.028	0.16	0.27	0.064	6.0	6.1	4.2	0.70	5.2	5.0	1.65	0.40
0.75	3.93	4.31	0.07	0.002	0.05	0.030	0.16	0.05	0.002	6.2	6.2	5.2	0.96	4.8	4.7	0.30	0.01
0.37	3.93	4.33	0.09	0.062	0.02	0.034	0.18	0.16	0.111	6.0	6.3	4.5	0.49	5.2	4.4	1.53	1.05
0.76	3.77	4.27	0.07	0.011	0.05	0.029	0.22	0.58	0.089	6.1	6.2	4.1	0.98	4.8	4.2	1.36	0.21
0.27	3.87	4.50	0.09	0.054	0.02	0.034	0.22	0.03	0.019	6.1	6.2	5.4	0.35	5.0	4.4	0.13	0.08
0.27	4.00	4.75	0.02	0.001	0.02	0.033	0.23	0.01	0.001	6.2	6.3	5.2	0.33	4.9	4.4	0.13	0.01
0.86	4.22	4.85	0.04	0.005	0.06	0.035	0.25	0.09	0.011	6.2	6.3	4.9	0.95	4.9	4.3	0.57	0.07
0.32	4.00	4.55	0.06	0.045	0.02	0.023	0.24	0.10	0.076	6.1	6.3	4.9	0.35	4.2	4.4	0.40	0.30
0.30	4.25	4.88	0.13	0.018	0.02	0.024	0.28	0.02	0.003	6.5	6.4	5.3	0.32	4.8	4.3	0.27	0.04
0.31	4.56	5.42	0.10	0.014	0.02	0.025	0.30	0.66	0.090	6.5	6.4	4.1	0.31	4.8	4.2	1.74	0.24
0.23	4.79	5.19	0.10	0.027	0.01	0.031	0.30	0.44	0.118	6.4	6.4	4.4	0.22	4.5	4.1	1.04	0.28
0.22	4.79	5.19	0.16	0.005	0.01	0.026	0.28	0.39	0.013	6.3	6.4	4.4	0.19	4.5	4.6	1.06	0.03
0.48	4.79	5.29	0.00	0.000	0.03	0.026	0.30	0.06	0.003	6.3	6.2	4.7	0.43	4.3	4.2	1.15	0.05
0.42	4.89	4.75	0.03	0.003	0.03	0.026	0.32	0.05	0.004	6.6	6.4	4.9	0.37	4.3	4.0	0.50	0.04
0.24	4.89	4.82	0.07	0.044	0.01	0.034	0.32	0.06	0.037	6.4	6.5	5.0	0.22	4.3	4.0	0.59	0.39
0.27	4.70	4.80	0.05	0.008	0.02	0.047	0.34	0.01	0.001	6.1	6.1	5.3	0.27	4.5	4.7	0.21	0.04
0.45	4.30	5.20	0.08	0.020	0.03	0.045	0.34	0.05	0.014	6.1	6.1	4.7	0.46	4.4	4.6	0.85	0.22
0.58	4.10	3.90	0.11	0.050	0.05	0.044	0.28	0.01	0.003	5.9	6.1	5.4	0.65	4.3	3.9	0.24	0.11
1.30	3.00	3.60	0.17	0.096	0.13	0.037	0.23	0.14	0.078	5.8	5.9	4.6	2.24	3.7	3.7	1.12	0.62
0.95	1.00	3.00	0.10	0.044	0.11	0.029	0.20	0.14	0.061	5.0	5.9	4.6	2.71	3.2	3.4	1.03	0.46
1.04	1.40	2.90	0.07	0.052	0.08	0.034	0.20	0.16	0.120	5.1	5.8	4.4	2.26	3.7	3.8	1.68	1.25
1.09	1.80	3.20	0.21	0.198	0.05	0.027	0.16	0.30	0.281	5.6	6.0	4.3	1.74	3.3	3.8	2.21	2.09
1.55	2.60	3.10	0.12	0.019	0.05	0.031	0.10	0.30	0.049	5.9	6.0	4.2	2.21	3.7	3.4	2.57	0.42
1.26	2.60	3.00	0.05	0.021	0.03	0.025	0.10	0.23	0.098	6.3	6.4	4.3	1.79	3.7	3.6	1.81	0.78
0.84	2.60	2.90	0.08	0.034	0.01	0.021	0.07	0.09	0.039	6.2	6.2	4.6	1.15	3.7	3.6	0.90	0.39
0.79	2.70	3.00	0.10	0.029	0.01	0.019	0.04	0.18	0.050	6.4	6.5	4.4	1.09	3.6	3.6	2.04	0.59
0.38	2.90	3.00	0.14	0.043	0.00	0.019	0.04	0.30	0.093	6.1	6.4	4.0	0.54	4.2	4.3	4.63	1.42
0.38	2.80	3.10	0.00	0.000	0.00	0.021	0.02	0.58	0.022	6.8	6.6	4.3	0.62	3.8	3.8	5.82	0.22
0.64	2.50	2.50	0.21	0.067	0.00	0.024	0.02	0.19	0.060	6.5	6.5	4.5	0.99	4.7	4.5	2.04	0.65
0.27	3.10	3.20	0.21	0.087	0.00	0.029	0.02	0.24	0.102	6.5	6.4	4.3	0.37	4.0	4.0	2.77	1.17
0.27	2.90	3.10	0.00	0.000	0.00	0.025	0.02	0.00	0.000	6.7	6.6	4.5	0.42	4.0	4.2	0.00	0.00
0.27	2.70	2.90	0.14	0.041	0.00	0.015	0.02	0.06	0.017	6.4	6.5	4.8	0.43	4.5	4.6	0.66	0.19
0.56	2.90	3.20	0.23	0.045	0.00	0.014	0.02	0.31	0.060	6.5	6.6	4.2	0.79	4.2	4.5	3.97	0.78
0.21	2.80	3.00	0.13	0.079	0.00	0.021	0.01	0.14	0.086	6.7	6.7	4.4	0.27	3.9	4.0	1.66	1.00
0.17	3.20	3.30	0.16	0.007	0.00	0.022	0.00	0.22	0.010	6.4	6.5	4.5	0.20	3.9	3.9	1.24	0.06
0.12	3.20	3.40	0.29	0.025	0.00	0.026	0.00	0.43	0.037	6.6	6.7	4.1	0.14	3.9	3.9	4.50	0.39
0.43	3.10	3.40	0.13	0.004	0.00	0.028	0.04	0.20	0.006	6.8	6.8	4.4	0.50	3.8	3.8	2.30	0.07
0.57	3.30	3.40	0.17	0.112	0.00	0.026	0.02	0.08	0.051	6.4	6.5	4.8	0.70	3.8	4.0	0.85	0.57
0.28	3.10	3.30	0.34	0.165	0.00	0.020	0.01	0.28	0.139	6.7	6.7	4.2	0.34	3.9	3.9	3.92	1.92
0.61	3.50	3.60	0.27	0.024	0.00	0.020	0.01	0.29	0.026	6.2	6.0	4.8	0.73	4.0	4.0	1.48	0.13
0.38	3.50	3.60	0.10	0.050	0.00	0.020	0.01	0.15	0.077	6.2	6.6	5.8	0.46	4.4	4.3	0.08	0.04
0.68	3.40	3.60	0.10	0.008	0.00	0.021	0.01	0.51	0.043	6.1	6.2	4.5	0.73	4.0	3.8	0.00	0.00
0.71	3.70	3.70	0.10	0.048	0.00	0.022	0.01	0.11	0.053	6.3	6.2	4.5	0.77	3.7	3.7	0.00	0.00
0.41	3.50	3.50	0.29	0.119	0.00	0.023	0.03	0.18	0.073	6.1	6.2	5.3	0.48	4.1	4.2	0.65	0.26
0.41	3.50	3.60	0.14	0.009	0.00	0.026	0.02	0.24	0.015	6.2	6.4	4.6	0.44	4.1	4.0	1.05	0.07
0.47	4.00	4.10	0.60	0.100	0.01	0.028	0.04	0.73	0.123	6.3	6.3	4.3	0.47	3.8	3.7	2.88	0.48
0.15	3.70	3.80	1.77	0.295	0.00	0.028	0.05	1.01	0.169	6.2	6.0	4.3	0.16	4.0	3.8	4.11	0.69
0.19	3.90	3.80	0.48	0.177	0.00	0.031	0.06	0.10	0.035	6.1	6.0	5.1	0.20	4.3	3.9	0.93	0.35
0.54	3.70	3.80	0.05	0.015	0.01	0.035	0.07	0.01	0.002	6.2	6.0	4.5	0.55	4.0	4.0	0.18	0.06
0.35	3.70	3.80	0.09	0.056	0.01	0.017	0.08	0.17	0.106	6.1	6.1	4.6	0.31	3.4	3.4	1.22	0.77
0.36	4.30	4.40	0.03	0.007	0.01	0.023	0.06	0.48	0.097	6.6	6.6	4.4	0.34	3.6	3.6	0.82	0.17
0.38	4.00	4.30	0.09	0.046	0.01	0.021	0.08	0.28	0.145	6.7	6.6	4.7	0.42	4.4	4.4	0.54	0.28
0.29	4.20	4.50	0.09	0.053	0.01	0.021	0.10	0.09	0.055	6.5	6.5	5.1	0.30	4.6	4.8	0.29	0.17
0.21	3.90	4.30	0.07	0.027	0.01	0.021	0.11	0.01	0.004	6.5	6.4	5.1	0.20	4.0	4.0	0.12	0.05

0.20	4.00	4.20	0.03	0.007	0.01	0.024	0.11	0.18	0.048	6.4	6.4	4.9	0.19	3.8	3.8	0.16	0.04
0.16	4.10	4.40	0.04	0.008	0.00	0.022	0.16	0.94	0.197	6.4	6.4	4.0	0.14	3.8	3.4	3.75	0.79
0.16	4.30	5.20	0.00	0.000	0.01	0.026	0.18	0.18	0.001	6.3	6.2	4.8	0.14	3.8	3.9	0.33	0.00
0.22	4.40	4.70	0.21	0.007	0.01	0.026	0.20	1.48	0.050	6.5	6.4	4.0	0.18	3.8	4.0	0.91	0.03
0.18	4.60	4.80	0.05	0.011	0.01	0.024	0.16	0.53	0.123	6.4	6.3	4.2	0.15	3.7	3.8	1.51	0.35
0.13	4.60	4.80	0.06	0.008	0.00	0.025	0.18	0.79	0.112	6.4	6.4	4.1	0.11	3.9	4.0	1.87	0.27
0.27	4.10	4.50	0.07	0.002	0.01	0.025	0.20	1.02	0.028	6.9	7.0	4.1	0.25	4.2	4.0	2.74	0.07
0.20	4.60	5.00	0.05	0.010	0.01	0.037	0.25	0.18	0.035	6.5	6.6	4.6	0.15	3.6	3.8	0.98	0.19
0.29	4.70	4.80	0.27	0.037	0.01	0.037	0.27	0.65	0.089	6.3	6.2	4.2	0.23	3.8	3.6	2.06	0.28
0.16	4.50	4.70	0.04	0.020	0.01	0.038	0.25	0.13	0.063	6.3	6.2	4.8	0.13	3.8	4.2	0.44	0.22
0.16	4.60	4.90	0.09	0.011	0.01	0.037	0.25	1.09	0.127	6.4	6.4	4.0	0.14	3.8	3.9	2.01	0.24
0.22	4.70	5.00	0.29	0.058	0.01	0.038	0.27	0.18	0.037	6.3	6.4	4.5	0.19	4.0	4.0	2.03	0.41
0.12	4.50	4.40	0.17	0.044	0.01	0.037	0.37	0.23	0.059	6.3	6.1	5.8	0.09	3.8	3.9	1.05	0.28
0.24	4.50	4.80	0.00	0.000	0.01	0.030	0.30	0.00	0.000	6.8	6.6	4.5	0.19	3.7	3.8	0.00	0.00
0.22	5.20	4.70	1.05	0.433	0.02	0.030	0.27	0.06	0.025	6.4	6.3	5.5	0.20	4.2	3.7	0.99	0.41
0.50	4.60	4.60	0.00	0.000	0.05	0.031	0.34	0.00	0.000	6.1	6.0	4.5	0.56	4.6	4.0	0.00	0.00
1.19	3.00	4.40	0.28	0.067	0.17	0.019	0.34	0.04	0.009	6.1	6.2	5.5	1.79	4.0	4.6	0.42	0.10
1.13	2.50	3.50	0.00	0.000	0.14	0.021	0.31	0.00	0.000	5.2	5.9	4.5	2.07	4.2	4.1	0.00	0.00
1.75	2.40	4.70	0.52	0.146	0.14	0.014	0.28	0.18	0.050	5.4	6.0	5.2	3.35	4.8	4.8	1.44	0.41
1.31	2.60	3.85	0.25	0.147	0.07	0.019	0.20	0.19	0.112	5.9	6.1	4.4	2.30	4.8	4.7	1.69	1.00
1.95	2.70	3.00	0.16	0.012	0.06	0.025	0.12	0.16	0.013	6.0	6.1	4.4	3.14	4.5	4.6	2.32	0.18
0.77	2.70	2.60	0.19	0.097	0.02	0.025	0.07	0.43	0.224	6.5	6.5	4.1	1.36	4.2	4.1	4.11	2.14
0.66	2.40	2.60	0.01	0.001	0.01	0.019	0.06	0.17	0.017	6.9	6.6	4.7	1.14	4.8	4.7	0.75	0.08
0.54	2.90	3.00	0.04	0.009	0.00	0.020	0.01	0.15	0.036	6.2	6.2	4.6	0.78	4.4	4.5	1.60	0.38
0.49	3.10	3.20	0.03	0.006	0.00	0.021	0.00	0.11	0.025	6.3	6.3	4.7	0.69	4.3	4.3	1.11	0.25
0.31	3.00	3.10	0.03	0.010	0.00	0.021	0.00	0.13	0.046	6.5	6.4	4.6	0.46	4.3	4.2	1.24	0.46
0.29	2.90	3.10	0.13	0.010	0.00	0.019	0.00	0.53	0.041	6.3	6.4	4.0	0.42	4.3	4.3	4.69	0.37
0.52	2.90	3.10	0.12	0.035	0.00	0.032	0.00	0.39	0.118	6.4	6.4	4.1	0.74	4.0	4.0	4.27	1.28
0.81	2.80	3.00	0.03	0.027	0.00	0.022	0.03	0.23	0.186	6.2	6.2	4.2	1.22	4.2	4.3	3.01	2.44
0.41	2.90	3.00	0.24	0.025	0.00	0.024	0.01	0.34	0.036	6.3	6.4	4.5	0.62	4.4	4.4	2.24	0.24
0.20	2.90	3.10	0.32	0.083	0.00	0.023	0.00	0.29	0.076	6.9	6.8	4.2	0.30	4.4	4.4	4.05	1.06
0.30	2.80	2.90	3.00	0.006	0.00	0.020	0.00	0.15	0.000	6.8	6.4	4.5	0.44	4.2	4.4	3.50	0.01
0.36	3.00	3.10	0.09	0.039	0.00	0.021	0.00	0.25	0.102	6.4	6.5	4.3	0.52	4.4	4.4	3.09	1.28
0.15	3.10	3.40	0.01	0.003	0.00	0.040	0.00	0.04	0.014	6.6	6.6	5.0	0.21	4.4	4.3	0.40	0.13
0.19	3.10	3.30	0.00	0.000	0.00	0.040	0.00	0.86	0.009	6.7	6.7	3.6	0.24	4.2	4.1	8.40	0.08
0.17	3.20	3.40	0.01	0.003	0.00	0.021	0.00	0.30	0.073	6.9	6.8	4.2	0.21	4.3	4.2	3.25	0.78
0.34	3.40	3.70	0.17	0.017	0.00	0.021	0.00	0.17	0.017	6.7	6.8	4.5	0.39	4.0	4.0	0.89	0.09
0.22	3.50	3.70	0.02	0.007	0.00	0.026	0.00	0.18	0.069	6.9	7.0	4.4	0.25	4.0	4.0	1.99	0.74
0.44	3.70	3.70	0.07	0.013	0.00	0.020	0.00	0.21	0.036	6.7	6.6	4.4	0.50	4.2	3.9	2.28	0.40
0.30	3.50	3.70	0.02	0.010	0.00	0.020	0.00	0.35	0.142	6.4	6.4	4.2	0.34	4.0	3.8	3.53	1.43
0.33	3.50	3.60	0.13	0.018	0.00	0.013	0.01	0.16	0.023	6.3	6.3	4.6	0.36	4.0	3.8	1.34	0.19
0.21	3.50	3.70	0.02	0.003	0.00	0.027	0.00	0.07	0.012	6.4	6.3	4.6	0.23	3.8	3.8	1.38	0.24
0.25	3.50	3.60	0.07	0.008	0.00	0.031	0.01	0.24	0.026	6.4	6.4	4.3	0.29	3.9	3.9	2.83	0.31
0.33	3.60	3.70	0.13	0.009	0.00	0.024	0.01	0.77	0.054	6.3	6.3	3.9	0.38	4.1	4.0	4.64	0.33
0.48	3.50	3.80	0.23	0.059	0.01	0.026	0.04	0.62	0.159	6.4	6.4	4.2	0.55	4.0	4.0	3.28	0.84
0.68	3.50	3.40	0.03	0.013	0.01	0.023	0.04	0.33	0.166	6.6	6.5	4.3	0.85	4.0	3.8	1.54	0.77
0.35	3.30	3.50	0.04	0.014	0.01	0.031	0.05	0.29	0.106	6.4	6.4	4.6	0.44	4.6	4.1	0.78	0.29
0.44	4.00	4.20	0.13	0.004	0.01	0.031	0.04	0.75	0.023	6.4	6.4	4.3	0.51	4.6	4.0	2.00	0.06
0.41	4.20	4.20	0.00	0.000	0.01	0.030	0.04	0.00	0.000	6.2	6.2	4.5	0.52	4.8	4.5	0.00	0.00
0.33	3.70	4.10	0.00	0.000	0.01	0.031	0.06	0.00	0.000	6.2	6.2	4.5	0.45	5.4	4.1	0.00	0.00
0.43	4.00	4.10	0.00	0.000	0.01	0.034	0.06	0.00	0.000	6.0	6.0	4.5	0.55	5.2	4.9	0.00	0.00
0.34	4.30	4.20	0.08	0.034	0.01	0.040	0.05	0.44	0.188	6.0	6.2	4.5	0.41	5.4	4.9	0.57	0.24
0.34	4.30	4.10	0.16	0.035	0.01	0.042	0.09	0.68	0.145	6.1	6.2	4.3	0.38	4.8	4.0	0.79	0.17
0.07	4.60	4.60	0.04	0.013	0.00	0.042	0.12	0.16	0.056	6.2	6.2	4.8	0.07	5.0	5.7	0.61	0.21

0.20	4.60	4.70	0.09	0.003	0.00	0.042	0.12	1.06	0.036	6.1	6.2	4.0	0.18	4.4	4.4	1.23	0.04
0.26	5.20	4.60	0.09	0.005	0.01	0.061	0.12	0.56	0.033	6.2	6.2	4.1	0.26	5.0	4.2	0.86	0.05
0.24	5.40	5.00	0.38	0.036	0.00	0.054	0.14	0.03	0.003	6.6	6.2	5.1	0.24	5.6	4.3	0.55	0.05
0.21	5.70	5.10	0.17	0.005	0.00	0.065	0.12	0.22	0.006	6.3	6.2	4.1	0.20	5.6	4.5	1.39	0.04
0.23	6.00	5.60	0.70	0.009	0.01	0.052	0.21	0.07	0.001	6.7	6.6	4.6	0.21	5.6	4.4	1.22	0.02
0.27	5.80	5.40	0.14	0.062	0.01	0.048	0.30	0.38	0.171	6.3	6.3	4.4	0.23	5.0	4.3	0.54	0.24
0.30	5.30	5.10	0.06	0.027	0.01	0.048	0.21	0.35	0.151	6.0	6.1	4.4	0.25	4.6	4.2	0.71	0.31
0.24	5.40	5.50	0.13	0.058	0.01	0.049	0.22	0.51	0.221	6.4	6.4	4.3	0.19	4.4	4.2	0.97	0.42
0.17	5.10	5.20	0.43	0.075	0.01	0.052	0.24	1.01	0.173	6.3	6.2	4.0	0.16	4.3	4.1	1.82	0.31
0.13	4.60	5.00	0.07	0.022	0.01	0.045	0.25	0.07	0.023	6.2	6.1	4.7	0.11	4.3	4.1	0.57	0.19
0.13	5.00	5.00	2.45	0.025	0.01	0.043	0.15	2.54	0.025	5.8	6.2	4.0	0.10	9.6	4.8	2.68	0.03
0.18	5.20	5.70	0.18	0.032	0.01	0.047	0.35	0.41	0.071	6.4	6.4	4.1	0.15	4.2	4.6	2.28	0.40
0.16	5.60	5.70	0.21	0.055	0.01	0.057	0.36	0.41	0.109	6.3	6.3	4.0	0.16	4.3	4.6	2.90	0.77
0.38	3.20	3.50	0.07	0.048	0.04	0.028	0.32	0.45	0.308	5.9	6.1	3.9	0.46	4.6	3.6	3.98	2.73
0.96	5.00	4.60	0.00	0.000	0.10	0.042	0.35	0.00	0.000	5.9	6.1	4.5	1.21	5.3	3.9	0.00	0.00
0.89	2.50	2.30	0.53	0.002	0.14	0.037	0.20	0.37	0.001	5.4	6.1	4.2	1.51	4.6	3.6	1.33	0.00
1.12	2.30	4.20	0.17	0.020	0.13	0.025	0.41	0.12	0.014	5.9	6.4	4.9	1.96	3.9	5.2	0.94	0.12
2.13	2.60	2.90	0.39	0.209	0.15	0.032	0.22	0.48	0.256	5.3	5.8	4.2	3.79	4.7	3.6	4.04	2.16
1.79	2.70	3.70	0.31	0.173	0.08	0.033	0.17	0.43	0.234	5.7	5.9	4.3	2.97	4.7	4.8	3.26	1.80
1.00	2.70	2.90	0.25	0.033	0.02	0.028	0.10	0.14	0.018	6.5	6.5	4.5	1.67	4.3	4.2	1.89	0.25
0.63	2.60	2.80	0.51	0.163	0.00	0.024	0.04	0.63	0.198	6.3	6.3	4.1	1.01	4.5	4.4	4.34	1.37
0.54	3.10	3.00	0.28	0.047	0.00	0.024	0.01	0.37	0.062	6.6	6.6	4.2	0.85	4.5	4.5	3.54	0.59
1.09	2.80	3.10	0.08	0.027	0.01	0.021	0.01	0.08	0.027	6.0	6.6	4.6	1.81	4.7	4.5	1.07	0.36
0.54	2.90	3.10	0.16	0.094	0.00	0.024	0.02	0.21	0.120	6.5	6.5	4.5	0.86	4.7	4.9	1.97	1.14
0.52	2.90	3.00	0.26	0.006	0.00	0.047	0.00	0.23	0.005	6.6	6.6	4.5	0.81	4.5	4.6	0.50	0.01
0.76	2.90	3.00	0.05	0.027	0.00	0.022	0.00	0.25	0.139	6.7	6.7	4.4	1.26	4.5	4.5	1.90	1.08
0.78	2.80	3.00	0.05	0.040	0.00	0.026	0.02	0.18	0.132	6.2	6.3	4.4	1.37	4.9	4.9	2.14	1.54
0.52	2.80	3.00	0.34	0.073	0.00	0.025	0.00	0.48	0.105	6.5	6.6	4.2	0.89	4.9	5.1	2.72	0.59
0.42	2.70	2.90	0.10	0.048	0.00	0.025	0.00	0.14	0.068	6.6	6.6	4.8	0.72	4.5	4.7	0.80	0.38
0.31	2.90	3.10	0.12	0.009	0.00	0.022	0.08	0.11	0.008	6.5	6.5	4.7	0.56	5.1	5.1	0.63	0.05
0.17	2.90	2.90	0.13	0.026	0.00	0.032	0.00	0.31	0.063	6.5	6.5	4.1	0.29	5.5	5.4	3.16	0.64
0.20	3.20	3.30	0.15	0.014	0.00	0.029	0.00	0.26	0.023	6.4	6.6	4.2	0.28	4.6	4.5	3.26	0.29
0.18	3.40	3.60	0.04	0.004	0.00	0.030	0.00	0.08	0.007	6.4	6.4	4.6	0.23	4.5	4.5	0.87	0.08
0.26	3.30	3.40	0.05	0.003	0.00	0.032	0.00	0.34	0.022	6.3	6.3	4.2	0.34	4.5	4.5	2.17	0.14
0.15	3.40	3.50	0.04	0.009	0.00	0.031	0.00	0.07	0.018	6.5	6.5	4.9	0.19	4.3	4.3	0.45	0.11
0.12	3.40	3.60	0.17	0.010	0.00	0.031	0.00	0.15	0.009	6.6	6.6	4.8	0.14	4.3	4.3	0.70	0.04
0.48	3.60	3.60	1.43	0.003	0.00	0.031	0.00	1.34	0.003	6.5	6.6	4.5	0.61	4.6	4.6	10.00	0.02
0.89	3.60	3.70	0.03	0.026	0.00	0.027	0.00	0.10	0.081	6.5	6.6	4.8	1.14	4.6	4.4	0.62	0.53
0.56	3.40	3.50	0.03	0.023	0.00	0.031	0.00	0.20	0.135	6.5	6.6	4.4	0.69	4.4	4.5	1.67	1.14
0.41	3.60	3.70	0.19	0.023	0.00	0.032	0.00	0.41	0.050	6.3	6.4	4.3	0.49	4.2	4.5	1.94	0.24
0.37	3.50	3.60	0.05	0.006	0.00	0.038	0.00	0.15	0.019	6.5	6.5	4.5	0.47	4.4	4.5	1.17	0.15
0.29	3.50	3.60	0.09	0.023	0.00	0.038	0.00	0.35	0.087	6.5	6.6	4.4	0.32	4.4	4.4	1.52	0.38
0.44	3.70	3.70	0.00	0.000	0.00	0.042	0.00	0.00	0.000	6.6	6.6	4.5	0.51	3.7	3.7	0.00	0.00
0.34	3.80	3.40	0.02	0.007	0.00	0.041	0.05	0.13	0.041	6.6	6.6	4.8	0.44	4.9	4.3	0.51	0.16
0.63	3.60	3.90	0.05	0.014	0.01	0.043	0.01	0.20	0.057	6.4	6.3	4.5	0.77	4.5	4.6	0.90	0.25
0.51	4.10	3.90	0.08	0.021	0.01	0.046	0.04	0.47	0.127	6.3	6.4	4.2	0.62	4.9	4.4	2.16	0.58
0.70	4.20	4.30	0.03	0.013	0.01	0.056	0.03	0.04	0.020	6.6	6.5	5.0	0.88	5.3	4.6	0.37	0.17
0.47	4.60	4.20	0.00	0.000	0.01	0.065	0.04	0.00	0.000	6.2	6.3	4.5	0.56	5.8	4.7	0.00	0.00
0.31	4.60	4.40	0.06	0.008	0.00	0.063	0.04	0.44	0.057	6.2	6.3	4.5	0.38	5.3	4.4	0.28	0.04
0.30	4.40	4.20	0.10	0.007	0.00	0.062	0.04	0.81	0.058	6.3	6.4	4.1	0.37	5.4	4.7	0.61	0.04
0.29	4.50	4.35	0.12	0.047	0.00	0.050	0.06	0.83	0.337	6.3	6.5	4.1	0.32	5.2	4.5	1.40	0.57
0.25	4.60	4.50	0.00	0.000	0.00	0.040	0.08	0.00	0.000	6.4	6.5	4.5	0.28	4.9	4.3	0.00	0.00
0.19	4.60	4.40	0.00	0.000	0.00	0.042	0.07	0.00	0.000	6.8	6.7	4.5	0.20	5.1	5.9	0.00	0.00
0.50	4.50	4.50	0.02	0.000	0.01	0.043	0.08	0.14	0.001	6.5	6.5	4.5	0.51	4.7	4.4	0.26	0.00

0.53	4.30	4.60	0.07	0.069	0.02	0.042	0.14	0.08	0.079	6.2	6.3	4.9	0.62	4.4	4.1	0.34	0.33
0.20	3.80	4.50	0.03	0.001	0.01	0.040	0.12	0.69	0.014	6.5	6.4	4.9	0.26	5.2	4.9	1.41	0.03
0.15	4.00	4.30	0.00	0.000	0.00	0.055	0.11	0.00	0.000	6.3	6.4	4.5	0.18	4.9	4.6	0.00	0.00
0.13	4.20	4.60	0.08	0.004	0.00	0.063	0.13	0.72	0.034	6.5	6.5	4.3	0.14	4.8	4.8	0.30	0.01
0.13	4.70	5.00	0.05	0.006	0.00	0.060	0.12	0.13	0.015	6.6	6.5	4.9	0.13	4.8	4.0	0.11	0.01
0.14	4.90	4.90	0.05	0.003	0.00	0.063	0.13	0.12	0.007	6.4	6.5	4.5	0.13	4.8	4.0	0.08	0.00
0.28	4.90	4.90	0.06	0.041	0.01	0.063	0.17	0.27	0.171	6.5	6.6	4.4	0.24	4.3	3.5	1.15	0.73
0.28	4.90	4.80	0.08	0.026	0.01	0.067	0.21	0.24	0.082	6.4	6.3	4.4	0.24	4.0	3.6	2.23	0.75
0.22	4.90	4.90	1.19	0.111	0.02	0.066	0.18	1.31	0.123	6.4	6.4	4.4	0.25	4.7	4.7	2.52	0.24
0.34	3.40	4.70	0.88	0.039	0.05	0.036	0.22	0.79	0.036	6.2	6.2	4.0	0.43	4.5	4.5	7.91	0.36
0.40	2.90	4.50	0.03	0.013	0.04	0.034	0.20	0.12	0.050	6.2	6.3	4.8	0.45	3.6	3.8	0.54	0.22
0.66	3.90	4.90	0.03	0.008	0.05	0.042	0.19	0.16	0.046	6.4	6.5	4.6	0.72	4.1	4.5	0.39	0.11
1.86	2.90	3.70	0.15	0.024	0.12	0.043	0.18	0.47	0.074	6.0	6.2	4.1	2.28	3.4	3.5	3.35	0.53
2.10	2.40	3.30	0.00	0.000	0.08	0.032	0.11	0.00	0.000	6.0	6.3	4.5	3.48	3.1	2.3	0.00	0.00
0.72	1.50	3.10	0.08	0.029	0.01	0.032	0.13	0.17	0.063	5.4	6.2	4.5	1.19	3.4	4.1	1.33	0.49
0.96	3.00	3.10	0.12	0.021	0.01	0.036	0.04	0.14	0.025	6.4	6.4	4.7	1.38	4.1	4.0	1.30	0.23
2.12	2.70	3.10	0.04	0.028	0.02	0.034	0.03	0.24	0.152	6.4	6.4	4.2	3.20	4.1	3.8	2.13	1.35
0.59	2.80	2.80	0.05	0.022	0.00	0.031	0.03	0.08	0.033	6.3	6.2	4.7	0.86	4.2	4.2	0.38	0.16
0.63	2.80	3.00	0.13	0.015	0.00	0.029	0.00	0.12	0.015	6.3	6.3	4.9	0.94	4.1	4.2	0.66	0.08
0.71	3.00	3.20	0.02	0.008	0.00	0.025	0.00	0.07	0.028	6.4	6.5	4.7	1.04	4.5	4.4	0.65	0.26
1.03	2.90	3.10	0.05	0.007	0.00	0.029	0.02	0.23	0.036	6.1	6.2	4.3	1.43	4.3	4.1	1.62	0.26
0.97	3.00	3.20	0.04	0.014	0.00	0.024	0.02	0.14	0.045	6.0	6.1	4.6	1.36	3.9	3.8	0.97	0.32
1.02	2.90	3.20	0.03	0.009	0.00	0.025	0.02	0.05	0.017	6.5	6.5	4.6	1.58	4.4	4.3	0.79	0.28
1.01	2.80	2.90	0.04	0.033	0.00	0.032	0.01	0.13	0.116	6.2	6.2	4.6	1.54	4.4	4.4	0.91	0.82
0.68	2.80	2.90	0.03	0.006	0.00	0.030	0.02	0.05	0.013	6.3	6.3	5.0	1.05	4.1	4.3	0.20	0.05
0.45	2.80	3.20	0.08	0.029	0.00	0.026	0.00	0.12	0.043	6.5	6.5	4.5	0.70	4.6	4.8	1.30	0.48
0.44	3.00	3.30	0.42	0.066	0.00	0.020	0.00	0.42	0.065	6.4	6.3	4.1	0.61	4.4	4.8	3.93	0.61
0.55	3.10	3.30	0.03	0.012	0.00	0.025	0.00	0.10	0.039	6.5	6.5	4.8	0.78	4.2	4.3	0.30	0.12
0.60	3.10	3.40	0.04	0.018	0.00	0.022	0.00	0.09	0.045	6.5	6.6	4.6	0.84	4.5	4.5	0.95	0.48
0.52	3.10	3.30	0.02	0.003	0.00	0.023	0.00	0.17	0.027	6.5	6.5	4.3	0.73	4.2	4.3	1.77	0.28
0.52	3.20	3.30	0.07	0.035	0.00	0.019	0.00	0.18	0.097	6.5	6.5	4.4	0.73	4.5	4.3	1.55	0.83
0.38	3.20	3.30	0.05	0.004	0.00	0.018	0.00	0.11	0.009	6.8	6.6	4.2	0.52	4.4	4.6	2.68	0.21
0.93	3.30	3.40	0.07	0.013	0.00	0.019	0.00	0.22	0.038	6.7	6.7	4.2	1.29	4.4	4.4	2.57	0.45
0.50	3.20	3.30	0.09	0.066	0.00	0.022	0.00	0.16	0.118	6.7	6.8	4.4	0.68	4.6	4.6	1.57	1.19
0.50	3.20	3.30	0.03	0.007	0.00	0.023	0.00	0.08	0.015	6.8	6.9	4.7	0.66	4.2	4.3	0.59	0.12
0.61	3.30	3.70	0.03	0.009	0.00	0.022	0.00	0.12	0.041	6.6	6.5	4.5	0.79	4.3	4.3	1.02	0.35
0.60	3.50	3.60	0.17	0.014	0.00	0.026	0.00	0.33	0.028	6.8	6.9	4.5	0.73	4.4	4.2	1.32	0.11
0.48	3.40	3.50	0.10	0.022	0.00	0.027	0.01	0.29	0.064	6.8	6.8	4.3	0.59	4.1	4.1	1.71	0.38
0.58	3.40	3.60	0.05	0.003	0.00	0.031	0.02	0.16	0.009	6.7	6.8	4.5	0.69	4.2	4.2	1.39	0.07
0.56	3.50	3.50	0.14	0.043	0.00	0.044	0.01	0.31	0.095	6.8	6.8	4.5	0.63	4.0	4.1	1.70	0.52
0.60	3.70	3.70	0.09	0.024	0.01	0.047	0.02	0.53	0.139	6.7	6.6	4.2	0.70	4.2	4.2	1.58	0.41
0.57	3.90	3.80	0.03	0.010	0.01	0.051	0.02	0.23	0.066	6.7	6.8	4.6	0.72	4.8	4.2	0.93	0.27
0.46	3.80	3.90	0.28	0.050	0.01	0.052	0.04	0.85	0.153	6.7	6.8	4.2	0.59	4.8	4.4	0.90	0.16
0.37	3.80	3.90	0.00	0.000	0.01	0.055	0.04	0.00	0.000	6.4	6.6	4.5	0.46	4.8	4.5	0.00	0.00
0.46	3.80	4.00	0.02	0.007	0.01	0.054	0.05	0.01	0.002	6.9	6.6	5.7	0.56	4.6	4.4	0.01	0.00
0.31	3.80	4.20	0.03	0.016	0.01	0.052	0.05	0.18	0.091	6.6	6.7	4.8	0.38	4.5	4.5	0.33	0.17
0.23	3.80	4.30	0.14	0.015	0.00	0.055	0.05	0.94	0.099	6.6	6.6	4.1	0.26	4.4	4.1	1.10	0.12
0.26	3.90	4.30	0.17	0.015	0.00	0.054	0.06	0.71	0.062	6.6	6.6	4.1	0.30	4.3	4.0	1.69	0.15
0.07	4.00	4.40	0.02	0.007	0.00	0.052	0.06	0.03	0.009	6.7	6.7	5.1	0.07	4.6	4.4	0.16	0.05
0.22	4.10	4.60	0.07	0.008	0.01	0.051	0.12	0.14	0.017	6.8	6.9	4.8	0.22	4.4	4.6	0.18	0.02
0.16	4.10	4.60	0.03	0.001	0.00	0.038	0.11	0.38	0.014	6.8	6.8	4.4	0.16	3.9	4.0	0.72	0.03
0.19	4.00	4.50	0.04	0.004	0.00	0.038	0.12	0.32	0.032	6.8	6.8	4.5	0.19	4.0	3.9	0.25	0.02
0.17	4.10	4.20	0.06	0.012	0.00	0.043	0.10	0.09	0.018	7.0	7.0	4.9	0.16	3.9	3.9	0.28	0.06
0.13	4.20	4.40	0.04	0.002	0.00	0.041	0.13	0.56	0.023	7.0	7.0	4.3	0.12	3.9	4.1	0.60	0.03

0.07	4.10	4.60	0.02	0.001	0.00	0.056	0.13	0.43	0.017	6.9	6.8	4.5	0.07	4.3	4.2	0.20	0.01
0.08	4.20	4.60	0.00	0.000	0.00	0.053	0.15	0.00	0.000	6.8	6.8	4.5	0.08	4.3	4.2	0.00	0.00
0.08	4.20	4.50	0.00	0.000	0.00	0.052	0.22	0.00	0.000	6.9	6.9	4.5	0.08	4.2	4.1	0.00	0.00
0.09	4.30	4.70	0.03	0.002	0.00	0.062	0.17	0.03	0.003	6.8	6.4	5.3	0.10	4.4	4.0	0.11	0.01
0.08	4.60	4.80	0.06	0.003	0.00	0.053	0.17	0.77	0.039	6.8	6.8	4.2	0.08	4.7	4.2	3.65	0.18
0.09	4.80	4.80	0.21	0.007	0.00	0.060	0.20	0.49	0.017	7.0	7.0	5.0	0.10	4.9	4.3	0.38	0.01
0.14	4.50	4.60	0.17	0.018	0.01	0.062	0.25	0.40	0.041	6.8	6.7	3.6	0.16	4.7	4.8	1.10	0.11
1.03	2.90	4.90	0.08	0.076	0.09	0.033	0.23	0.12	0.115	6.3	6.6	4.7	1.59	3.8	4.2	0.71	0.70
0.73	2.30	4.40	0.04	0.003	0.05	0.031	0.19	0.22	0.014	5.7	6.8	4.6	1.25	4.2	4.1	0.48	0.03
1.65	2.50	4.50	0.00	0.000	0.08	0.033	0.13	0.00	0.000	6.0	6.2	4.5	2.68	4.1	4.3	0.00	0.00
1.77	2.80	3.60	0.79	0.031	0.05	0.039	0.13	0.34	0.013	6.1	6.4	4.6	2.93	4.5	4.1	2.57	0.10
0.53	3.00	3.30	0.08	0.020	0.01	0.039	0.07	0.28	0.073	6.8	6.9	4.2	0.79	5.1	5.0	2.69	0.69
0.59	3.20	3.30	0.08	0.009	0.01	0.033	0.04	0.03	0.004	6.5	6.7	5.1	0.72	4.2	4.2	0.60	0.07
1.11	3.10	3.40	0.47	0.166	0.01	0.031	0.04	0.42	0.149	6.6	6.8	4.5	1.22	3.5	3.7	3.00	1.06
0.52	3.10	3.20	0.22	0.032	0.00	0.028	0.03	0.40	0.059	6.7	6.8	4.2	0.61	3.3	3.3	3.63	0.53
0.43	3.10	3.20	0.18	0.028	0.00	0.028	0.01	0.68	0.105	6.8	6.8	4.0	0.56	4.0	3.9	5.09	0.78
0.36	3.10	3.30	0.11	0.012	0.00	0.032	0.00	0.30	0.033	6.8	6.7	4.3	0.51	4.2	4.2	2.61	0.29
0.33	3.00	3.40	0.07	0.022	0.00	0.027	0.00	0.14	0.042	6.5	6.6	4.6	0.47	4.5	4.4	1.22	0.37
0.16	3.00	3.40	0.10	0.006	0.00	0.026	0.00	0.11	0.007	6.8	6.9	4.6	0.20	3.9	4.0	1.22	0.08
0.37	3.30	3.40	0.22	0.009	0.00	0.027	0.00	0.42	0.018	6.8	6.9	4.2	0.45	4.0	4.0	3.36	0.14
0.92	3.20	3.40	0.07	0.069	0.00	0.022	0.00	0.17	0.172	6.7	6.8	4.6	1.17	4.0	4.0	1.41	1.47
0.37	3.00	3.30	0.04	0.012	0.00	0.019	0.00	0.14	0.038	6.8	6.8	4.7	0.48	3.9	4.0	0.80	0.22
0.44	3.10	3.40	0.07	0.028	0.00	0.026	0.00	0.27	0.111	6.5	6.6	4.1	0.59	4.1	3.9	4.26	1.78
0.33	3.20	3.30	0.03	0.005	0.00	0.018	0.00	0.06	0.010	6.5	6.6	5.1	0.44	4.2	4.2	0.43	0.07
0.29	3.20	3.50	0.09	0.031	0.00	0.023	0.01	0.08	0.028	6.4	6.4	5.0	0.38	4.1	4.1	0.48	0.17
0.22	3.10	3.30	0.03	0.004	0.00	0.017	0.00	0.09	0.012	6.7	6.7	4.8	0.27	4.0	4.0	0.84	0.11
0.17	3.10	3.40	0.18	0.012	0.00	0.023	0.00	0.21	0.014	6.7	6.6	4.5	0.19	3.8	3.9	1.50	0.10
0.22	3.50	3.60	0.15	0.034	0.00	0.022	0.00	0.35	0.082	6.5	6.6	4.1	0.24	3.8	3.8	3.42	0.79
0.15	3.50	3.70	0.08	0.018	0.00	0.019	0.01	0.19	0.043	6.7	6.8	4.4	0.17	4.1	4.1	1.94	0.44
0.30	3.20	3.60	0.04	0.017	0.00	0.023	0.00	0.10	0.042	6.5	6.2	4.6	0.35	4.0	5.8	1.44	0.58
0.27	3.50	3.70	0.04	0.006	0.00	0.020	0.00	0.08	0.014	6.7	6.8	4.8	0.30	4.0	4.0	0.62	0.10
0.26	3.50	3.70	0.04	0.017	0.00	0.021	0.00	0.13	0.051	6.8	6.9	4.6	0.30	3.8	3.6	1.13	0.46
0.22	3.50	3.80	0.07	0.007	0.00	0.023	0.00	0.06	0.006	6.7	6.8	4.9	0.27	4.3	4.2	0.65	0.07
0.54	3.60	3.70	0.08	0.032	0.00	0.000	0.00	0.16	0.066	6.5	6.5	4.6	0.69	4.4	4.4	1.42	0.59
0.52	3.70	3.70	0.17	0.063	0.00	0.000	0.00	0.33	0.126	6.7	6.7	4.5	0.63	4.8	4.7	1.33	0.50
0.42	3.60	3.80	0.26	0.018	0.00	0.000	0.01	0.41	0.029	6.4	6.5	4.5	0.44	4.0	3.8	1.63	0.11
0.37	3.70	3.90	0.15	0.024	0.00	0.000	0.01	0.54	0.089	6.6	6.7	4.2	0.41	3.8	3.9	2.73	0.45
0.79	3.70	3.80	0.11	0.068	0.01	0.000	0.03	0.36	0.229	6.9	6.9	4.5	0.97	4.6	4.4	0.78	0.50
1.03	3.70	3.70	0.08	0.030	0.01	0.000	0.03	0.51	0.197	6.9	6.9	4.2	1.29	4.5	4.5	2.51	0.97
0.50	3.60	3.60	0.07	0.012	0.01	0.000	0.04	0.20	0.032	6.8	6.8	4.8	0.65	4.7	4.6	0.71	0.12
0.90	3.80	3.80	0.04	0.008	0.02	0.000	0.05	0.20	0.040	6.7	6.6	4.6	1.23	4.9	4.7	0.78	0.16
0.48	3.30	3.30	0.24	0.042	0.01	0.000	0.08	0.92	0.161	6.4	6.4	4.2	0.66	4.8	4.6	1.48	0.26
0.54	3.60	4.20	0.02	0.006	0.01	0.000	0.06	0.04	0.012	6.6	6.6	5.1	0.72	4.8	4.7	0.29	0.08
0.39	4.10	4.40	0.03	0.005	0.01	0.000	0.07	0.28	0.044	6.4	6.5	4.4	0.53	5.5	4.6	1.33	0.20