

Université de Montréal

Modélisation des apports naturels de réservoirs

par

Lyes Ouhib

Département de mathématiques et de statistique

Faculté des arts et des sciences

Mémoire présenté à la Faculté des études supérieures

en vue de l'obtention du grade de

Maître ès sciences (M.Sc.)
en Statistique

octobre 2005



QA

3

USF

2005

V.017



Direction des bibliothèques

AVIS

L'auteur a autorisé l'Université de Montréal à reproduire et diffuser, en totalité ou en partie, par quelque moyen que ce soit et sur quelque support que ce soit, et exclusivement à des fins non lucratives d'enseignement et de recherche, des copies de ce mémoire ou de cette thèse.

L'auteur et les coauteurs le cas échéant conservent la propriété du droit d'auteur et des droits moraux qui protègent ce document. Ni la thèse ou le mémoire, ni des extraits substantiels de ce document, ne doivent être imprimés ou autrement reproduits sans l'autorisation de l'auteur.

Afin de se conformer à la Loi canadienne sur la protection des renseignements personnels, quelques formulaires secondaires, coordonnées ou signatures intégrées au texte ont pu être enlevés de ce document. Bien que cela ait pu affecter la pagination, il n'y a aucun contenu manquant.

NOTICE

The author of this thesis or dissertation has granted a nonexclusive license allowing Université de Montréal to reproduce and publish the document, in part or in whole, and in any format, solely for noncommercial educational and research purposes.

The author and co-authors if applicable retain copyright ownership and moral rights in this document. Neither the whole thesis or dissertation, nor substantial extracts from it, may be printed or otherwise reproduced without the author's permission.

In compliance with the Canadian Privacy Act some supporting forms, contact information or signatures may have been removed from the document. While this may affect the document page count, it does not represent any loss of content from the document.

Université de Montréal

Faculté des études supérieures

Ce mémoire intitulé

Modélisation des apports naturels de réservoirs

présenté par

Lyes Ouhib

a été évalué par un jury composé des personnes suivantes :

Jean-François Angers

(président-rapporteur)

Roch Roy

(directeur de recherche)

Luc Perreault

(co-directeur)

Pierre Duchesne

(membre du jury)

Mémoire accepté le:

20 octobre 2005

REMERCIEMENTS

Je tiens avant tout à remercier mon directeur de recherche, M. Roch Roy, pour sa grande patience, sa disponibilité et son soutien tant scientifique que financier et moral. Il a été pour moi un directeur exceptionnel et je l'en remercie de tout mon coeur.

Ce mémoire n'aurait pas été possible sans l'étroite collaboration de l'institut de recherche d'Hydro-Québec (IREQ) et surtout de mon co-directeur M. Luc Perreault, qui a su m'engager dans un projet stimulant et utile. Je tiens à lui exprimer toute ma gratitude et mon profond respect.

J'aimerais également exprimer ma reconnaissance envers les personnes qui m'ont aidé de près ou de loin à la réalisation de mon projet de recherche surtout le personnel du Département de mathématiques et de statistique de l'Université de Montréal. Mes remerciements les plus vifs vont aussi à mes chers camarades de classe en particulier à James Merleau qui est pour moi un frère et un grand ami. Je le remercie énormément pour son aide et ses conseils.

Je dois témoigner un hommage à ma famille, surtout les êtres les plus chers de ma vie, ma mère qui m'a encouragé à poursuivre mes études et m'a donné la meilleure éducation qui soit, ma femme Nawel pour sa patience et ses encouragements dans les moments difficiles et surtout mon petit Yanis pour son courage et sa soif de vivre.

Finalement, je dédie ce mémoire à mon défunt père qui aurait été si fier de moi.

SOMMAIRE

Dans ce mémoire, nous introduisons les modèles autorégressifs moyennes mobiles périodiques (*PARMA*) comme une alternative aux modèles *ARMA*, du point de vue de la modélisation des séries hydrologiques. En effet, les modèles autorégressifs périodiques (*PAR*) sont plus appropriés pour la génération de séries saisonnières puisqu'ils permettent de représenter une structure d'autocorrélation qui varie selon les saisons. Dans ce mémoire, nous décrivons les principales caractéristiques des séries d'apports saisonniers ainsi que les principales propriétés des modèles *PAR*. Par la suite, nous construisons des modèles périodiques capables de reproduire les principales caractéristiques statistiques des apports saisonniers dont la structure d'autocorrélation mensuelle, la corrélation interannuelle ainsi que la corrélation spatiale. Afin de générer des séries d'apports, deux méthodes sont utilisées pour générer les erreurs. La première de type paramétrique est basée sur la loi lognormale et la deuxième de type non paramétrique est basée sur le rééchantillonnage des résidus.

Mots clés : Modélisation des apports, processus périodiques, modèles *PAR*, test d'ajustement, autocorrélation périodique, loi lognormale, rééchantillonnage.

SUMMARY

We introduce the periodic autoregressive moving average models (*PARMA*) as an alternative to the *ARMA* models, from the point of view of the modeling of hydrological time series. Indeed, the periodic autoregressive models (*PAR*) are more suitable for the generation of seasonal series since they make it possible to represent an autocorrelation structure that varies according to seasons. In this master thesis, we describe the principal characteristics of the seasonal water flow series as well as the main properties of the *PAR* models. Thereafter, we build periodic models able to reproduce the principal statistical characteristics of seasonal water flows namely the monthly autocorrelation structure as well as the interannual and spatial correlations. In order to generate series of water flows, two methods are used to generate the errors. The first one which is of the parametric type, is based on the lognormal distribution whilst the second one which is based on the resampling of the residuals is purely nonparametric.

Keywords : Water flow modeling, periodic process, *PAR* models, goodness-of-fit test, periodic autocorrelation, lognormal distribution, resampling.

TABLE DES MATIÈRES

Remerciements	iii
Sommaire.....	iv
Summary	v
Liste des figures	ix
Liste des tableaux	xiv
Introduction	1
Contexte général.....	1
Problématique et objectifs de l'étude.....	2
Plan du mémoire	3
Chapitre 1. Principales caractéristiques des séries d'apports historiques	6
1.1. Les données.....	6
1.2. Périodicité.....	10
1.3. Dépendance à court terme	15
1.4. Dépendance à long terme	19
1.5. Cohérence spatiale	23
1.6. Non-normalité	26
1.7. Conclusion.....	29

Chapitre 2. Modèles de séries chronologiques en hydrologie	30
2.1. Introduction	30
2.2. Généralités	30
2.2.1. Stationnarité	30
2.2.2. Processus <i>ARMA</i>	31
2.2.3. Processus <i>ARIMA</i>	36
2.2.4. Processus <i>SARIMA</i>	36
2.3. Processus périodiques	38
2.4. Processus <i>ARMA</i> périodiques	41
2.4.1. Condition pour la causalité d'un <i>PAR(1)</i>	48
2.5. Processus autoregressifs périodiques	50
2.5.1. Motivation	50
2.5.2. Équations de Yule-Walker pour un <i>PAR(p)</i>	50
2.5.3. Estimation par la méthode des moments	51
2.5.4. Validation d'un modèle <i>PAR</i>	53
2.6. Modèles <i>PAR</i> avec une variable explicative	53
2.7. Modèle <i>PAR</i> multisite et corrélation spatiale	54
2.7.1. Modèle <i>PAR(p)</i> multisite	54
2.7.2. Estimation de la matrice de covariance spatiale	57
2.8. Génération de réalisations du modèle <i>CPARX</i>	57
2.8.1. Transformation des erreurs	58
2.8.2. Rééchantillonnage des vecteurs de résidus	59
Chapitre 3. Application des modèles <i>PAR</i> et <i>PARX</i> aux apports	61
3.1. Plan des simulations	62
3.1.1. Algorithme d'estimation pour le modèle <i>PAR</i>	63

3.1.2.	Algorithme d'estimation pour le modèle $PARX$	63
3.1.3.	Algorithme de génération d'apports pour le modèle PAR et $PARX$	64
3.2.	Mise en oeuvre du modèle $PAR(1)$	65
3.2.1.	Modèle estimé $PAR(1)$ et test d'adéquation de Ljung et Box ..	65
3.2.2.	Analyse des hydrogrammes simulés.....	68
3.2.3.	Discussion des résultats de simulation dans le cas d'un modèle $PAR(1)$	73
3.2.4.	Discussion des résultats des corrélations interannuelles simulées avec un $PAR(1)$	87
3.2.5.	Discussion des corrélations spatiales simulées avec un $PAR(1)$..	89
3.2.6.	Conclusion.....	95
3.3.	Mise en oeuvre du modèle $PAR(1)$ avec variable explicative	95
3.3.1.	Modèle estimé $PARX(1)$ et test d'adéquation de Ljung et Box ..	95
3.3.2.	Analyse des hydrogrammes simulés.....	98
3.3.3.	Discussion des résultats de simulation dans le cas d'un modèle $PARX(1)$	102
3.3.4.	Discussion des résultats des corrélations interannuelles simulées avec un $PARX(1)$	115
3.3.5.	Discussion des corrélations spatiales simulées avec un $PARX(1)$	117
	Conclusion	122
	Annexe : code matlab	125
3.4.	Programmes des simulations avec un modèle $PAR(1)$ pour la méthode transformée.....	125
3.4.1.	Programme principal de modélisation des apports.....	125
3.4.2.	Les fonctions.....	128
	Bibliographie	134

LISTE DES FIGURES

1.1	Localisation géographique des différents réservoirs d'Hydro-Québec. . .	7
1.2	Représentation graphique des apports mensuels aux réservoirs Caniapiscau, Churchills-Falls et Gouin durant la période janvier 1961 à décembre 2002.	9
1.3	Hydrogrammes du réservoir Caniapiscau pour les années 1972, 1982, 1986 et 1987.	11
1.4	Hydrogrammes du réservoir Churchill Falls pour les années 1972, 1982, 1986 et 1987.	11
1.5	Hydrogrammes du réservoir Gouin pour les années 1972, 1982, 1986 et 1987.	12
1.6	Les moyennes et écarts types hebdomadaires pour le réservoir Caniapiscau. 13	
1.7	Les moyennes et écarts types hebdomadaires pour le réservoir Churchill Falls.	14
1.8	Les moyennes et écarts types hebdomadaires pour le réservoir Gouin.	14
1.9	Autocorrélations périodiques mensuelles pour le réservoir Caniapiscau.	16
1.10	Autocorrélations périodiques mensuelles pour le réservoir Churchill Falls.	17
1.11	Autocorrélations périodiques mensuelles pour le réservoir Gouin.	18
1.12	Autocorrélations des apports annuels pour le réservoir Caniapiscau. . .	20
1.13	Autocorrélations des apports annuels pour le réservoir Churchill Falls.	21
1.14	Autocorrélations des apports annuels pour le réservoir Gouin.	22

1.15	Graphes comparatifs des corrélations historiques entre les 3 réservoirs de référence pour les 12 mois.	26
1.16	Représentation graphique quantiles des apports-quantiles de la loi normale au réservoir Caniapiscau pour les 11 ^e , 13 ^e , 17 ^e et 22 ^e semaines.	27
1.17	Représentation graphique quantiles des apports-quantiles de la loi normale au réservoir Churchill Falls pour les 11 ^e , 17 ^e , 19 ^e et 28 ^e semaines.	27
1.18	Représentation graphique quantiles des apports-quantiles de la loi normale au réservoir Gouin pour les 11 ^e , 14 ^e , 16 ^e et 19 ^e semaines.	28
3.1	Hydrogrammes comparatifs pour les séries historiques, transformées et rééchantillonnées au réservoir Caniapiscau.	70
3.2	Hydrogrammes comparatifs pour les séries historiques, transformées et rééchantillonnées au réservoir Churchill Falls.	71
3.3	Hydrogrammes comparatifs pour les séries historiques, transformées et rééchantillonnées au réservoir Gouin.	72
3.4	Moyenne mensuelle historique (★) au réservoir Caniapiscau et diagrammes en boîte des moyennes mensuelles de 100 séries générées avec un modèle <i>PAR</i> (1) suite à la transformation logarithmique ou suite au rééchantillonnage des résidus.	75
3.5	Moyenne mensuelle historique (★) au réservoir Churchill Falls et diagrammes en boîte des moyennes mensuelles de 100 séries générées avec un modèle <i>PAR</i> (1) suite à la transformation logarithmique ou suite au rééchantillonnage des résidus.	76
3.6	Moyenne mensuelle historique (★) au réservoir Gouin et diagrammes en boîte des moyennes mensuelles de 100 séries générées avec un modèle <i>PAR</i> (1) suite à la transformation logarithmique ou suite au rééchantillonnage des résidus.	77

- 3.7 Écart type mensuel historique (★) au réservoir Caniapiscou et diagrammes en boîte des écarts types mensuels de 100 séries générées avec un modèle $PAR(1)$ suite à la transformation logarithmique ou suite au rééchantillonnage des résidus. 79
- 3.8 Écart type mensuel historique (★) au réservoir Churchill Falls et diagrammes en boîte des écarts types mensuels de 100 séries générées avec un modèle $PAR(1)$ suite à la transformation logarithmique ou suite au rééchantillonnage des résidus. 80
- 3.9 Écart type mensuel historique (★) au réservoir Gouin et diagrammes en boîte des écarts types mensuels de 100 séries générées avec un modèle $PAR(1)$ suite à la transformation logarithmique ou suite au rééchantillonnage des résidus. 81
- 3.10 Autocorrélations périodiques de délai 1, 2 et 3 au réservoir Caniapiscou et diagrammes en boîte des autocorrélations mensuelles de 100 séries générées avec un modèle $PAR(1)$ suite à une transformation logarithmique ou suite au rééchantillonnage des résidus. 84
- 3.11 Autocorrélations périodiques de délai 1, 2 et 3 au réservoir Churchill Falls et diagrammes en boîte des autocorrélations mensuelles de 100 séries générées avec un modèle $PAR(1)$ suite à une transformation logarithmique ou suite au rééchantillonnage des résidus. 85
- 3.12 Autocorrélations périodiques de délai 1, 2 et 3 au réservoir Gouin et diagrammes en boîte des autocorrélations mensuelles de 100 séries générées avec un modèle $PAR(1)$ suite à une transformation logarithmique ou suite au rééchantillonnage des résidus. 86
- 3.13 Hydrogrammes comparatifs pour les séries historiques, transformées et rééchantillonnées au réservoir Caniapiscou. 99
- 3.14 Hydrogrammes comparatifs pour les séries historiques, transformées et rééchantillonnées au réservoir Churchill Falls. 100

- 3.15 Hydrogrammes comparatifs pour les séries historiques, transformées et rééchantillonnées au réservoir Gouin..... 101
- 3.16 Moyenne mensuelle historique (★) au réservoir Caniapiscou et diagrammes en boîte des moyennes mensuelles de 100 séries générées avec un modèle *PARX*(1) suite à la transformation logarithmique ou suite au rééchantillonnage des résidus..... 105
- 3.17 Moyenne mensuelle historique (★) au réservoir Churchill Falls et diagrammes en boîte des moyennes mensuelles de 100 séries générées avec un modèle *PARX*(1) suite à la transformation logarithmique ou suite au rééchantillonnage des résidus..... 106
- 3.18 Moyenne mensuelle historique (★) au réservoir Gouin et diagrammes en boîte des moyennes mensuelles de 100 séries générées avec un modèle *PARX*(1) suite à la transformation logarithmique ou suite au rééchantillonnage des résidus..... 107
- 3.19 Écart type mensuel historique (★) au réservoir Caniapiscou et diagrammes en boîte des écarts types mensuels de 100 séries générées avec un modèle *PARX*(1) suite à la transformation logarithmique ou suite au rééchantillonnage des résidus..... 108
- 3.20 Écart type mensuel historique (★) au réservoir Churchill Falls et diagrammes en boîte des écarts types mensuels de 100 séries générées avec un modèle *PARX*(1) suite à la transformation logarithmique ou suite au rééchantillonnage des résidus..... 109
- 3.21 Écart type mensuel historique (★) au réservoir Gouin et diagrammes en boîte des écarts types mensuels de 100 séries générées avec un modèle *PARX*(1) suite à la transformation logarithmique ou suite au rééchantillonnage des résidus..... 110
- 3.22 Autocorrélations périodiques au réservoir Caniapiscou et diagrammes en boîte des autocorrélations mensuelles de 100 séries générées avec un

	modèle $PARX(1)$ suite à la une transformation logarithmique ou suite au rééchantillonnage des résidus.	112
3.23	Autocorrélations périodiques au réservoir Churchill Falls et diagrammes en boîte des autocorrélations mensuelles de 100 séries générées avec un modèle $PARX(1)$ suite à la transformation logarithmique ou suite au rééchantillonnage des résidus.	113
3.24	Autocorrélations périodiques au réservoir Gouin et diagrammes en boîte des autocorrélations mensuelles de 100 séries générées avec un modèle $PARX(1)$ suite à la transformation logarithmique ou suite au rééchantillonnage des résidus.	114

LISTE DES TABLEAUX

1.1	Résultats des statistiques annuelles pour les 17 réservoirs de référence.	8
1.2	Autocorrélations de délai 1 des apports annuels aux 17 réservoirs. . . .	23
1.3	Corrélations spatiales entre les 17 réservoirs (données historiques). . . .	25
3.1	Estimation et validation du modèle $PAR(1)$ pour les réservoirs Caniapiscou et Churchill Falls	66
3.2	Estimation et validation du modèle $PAR(1)$ pour le réservoir Gouin..	67
3.3	Résultats comparatifs des statistiques annuelles observées et simulées d'un modèle $PAR(1)$ pour les différents réservoirs d'Hydro-Québec. . . .	74
3.4	Résultats comparatifs des corrélations interannuelles observées et simulées d'un modèle $PAR(1)$ pour les différents réservoirs d'Hydro-Québec.	88
3.5	Corrélations spatiales entre les 17 réservoirs obtenus avec un modèle $PAR(1)$ et une transformation logarithmique des résidus.	91
3.6	Écarts types des corrélations spatiales pour les 17 réservoirs obtenus avec un modèle $PAR(1)$ et une transformation logarithmique des résidus.	92
3.7	Corrélations spatiales entre les 17 réservoirs obtenus avec un modèle $PAR(1)$ et en rééchantillonnant les résidus.	93
3.8	Écarts types des corrélations spatiales pour les 17 réservoirs obtenus avec un modèle $PAR(1)$ et en rééchantillonnant les résidus.	94
3.9	Estimation et validation du modèle $PARX(1)$ pour les réservoirs Caniapiscou et Churchill Falls.	96

3.10	Estimation et validation du modèle $PARX(1)$ pour le réservoir Gouin.	97
3.11	Résultats comparatifs des statistiques annuelles observées et simulées d'un modèle $PARX(1)$ pour les différents réservoirs d'Hydro-Québec.	104
3.12	Résultats comparatifs des corrélations interannuelles observées et simulées d'un modèle $PARX(1)$ pour les différents réservoirs d'Hydro-Québec.	116
3.13	Corrélations spatiales entre les 17 réservoirs obtenues avec un modèle $PARX(1)$ et une transformation logarithmique des résidus.	118
3.14	Écarts types des corrélations spatiales pour les 17 réservoirs obtenue avec un modèle $PARX(1)$ et une transformation logarithmique des résidus.	119
3.15	Corrélations spatiales entre les 17 réservoirs obtenues avec un modèle $PARX(1)$ et en rééchantillonnant les résidus.	120
3.16	Écarts types des corrélations spatiales pour les 17 sites obtenus avec un modèle $PARX(1)$ et en rééchantillonnant les résidus.	121

INTRODUCTION

CONTEXTE GÉNÉRAL

Depuis plusieurs années, les modèles périodiques suscitent un vif intérêt dans la littérature des séries chronologiques. Présentés comme une alternative aux modèles *ARIMA* saisonniers, les modèles *ARMA* à coefficients périodiques (*PARMA*) sont importants particulièrement en hydrologie. Plusieurs classes de modèles ont été développées pour les séries périodiques, voir par exemple Vecchia (1985) qui a développé et appliqué l'estimation par maximum de vraisemblance conditionnelle aux modèles *ARMA* périodiques. Hipel et McLeod (1994) ont discuté de l'identification empirique des modèles *AR* périodiques. Basawa et Lund (2001) ont étudié les propriétés asymptotiques des estimateurs pour les modèles autorégressifs moyennes mobiles périodiques (*PARMA*), causals et inversibles. L'ouvrage récent de Franses et Paap (2005) discute l'application des modèles périodiques en économétrie.

Dans ce mémoire, nous nous intéressons principalement à la modélisation des séries hydrologiques à pas de temps mensuel et hebdomadaire. Notre étude s'inscrit dans le cadre du projet innovation intitulé "Amélioration au processus global de planification de la production du parc Hydro-Québec" piloté par Louis Lafond, chargé de projet à l'IREQ (Institut de recherche d'Hydro-Québec). Elle est également basée sur le rapport technique de Luc Perreault (Perreault et Latraverse, 2001) concernant la modélisation des apports naturels. Notre étude consiste à construire des modèles périodiques complets. Des modèles de séries chronologiques capables de reproduire en moyenne les caractéristiques statistiques des

apports mensuels et hebdomadaires ainsi que la dépendance spatiale et temporelle.

Nous allons décrire dans cette étude les aspects pratiques et théoriques ainsi que la mise en oeuvre opérationnelle des modèles statistiques employés pour générer des séries d'apports mensuels et hebdomadaires pour le logiciel d'optimisation de la production SDDP (Stochastic Dual Dynamic Programming). Il s'agit de modèles d'apports de type périodiques autorégressifs (*PAR*).

PROBLÉMATIQUE ET OBJECTIFS DE L'ÉTUDE

Hydro-Québec gère un immense parc hydroélectrique duquel elle tire la majeure partie de sa production d'électricité (96%). La gestion efficace de ces infrastructures vise surtout à optimiser la production d'électricité tout en respectant certaines contraintes telles que :

- satisfaire la demande d'électricité,
- garantir un débit minimal des rivières,
- éviter les inondations sur les différents sites.

Pour ce faire, l'IREQ a développé le logiciel d'optimisation stochastique SDDP pour répartir la production aux différents sites. Cet outil permet de déterminer les plans de production qui minimiseront les coûts, tout en tenant compte notamment de l'aléa des apports naturels et des contraintes d'exploitation.

La prévision d'apports en eau dans les réservoirs est la problématique principale de cette étude puisque les apports en eau dans les différents sites constituent un phénomène aléatoire qui change de saison en saison, ce qui le rend plus difficile à modéliser. Ainsi, sachant que les apports en eau dans les réservoirs varient selon les saisons, nous pensons que la connaissance des caractéristiques statistiques et stochastiques de ces apports peut contribuer à prévoir l'apport en eau dans chaque réservoir et permettre ainsi une amélioration de la planification à moyen et à long termes de la production d'électricité.

Pour réaliser cette étude, nous nous sommes fixé comme objectifs :

- De construire des modèles d'apports capables d'une part de reproduire les caractéristiques statistiques des apports historiques tels que la moyenne, l'écart type et les autocorrélations et d'autre part de préserver la dépendance temporelle et la corrélation spatiale existante entre les différents réservoirs.
- De fournir les paramètres nécessaires à la mise en oeuvre de modèles statistiques de simulation dans SDDP (fichiers de paramètres, algorithmes de simulation, documentation technique, etc.) et ceci pour l'ensemble des réservoirs du parc d'Hydro-Québec.

Il est à noter aussi qu'il est primordial que les modèles choisis soient compatibles avec les techniques de programmation dynamique implantées dans SDDP, c'est à-dire que les modèles statistiques qui relient les apports d'une période à l'autre soient linéaires. Par ailleurs pour la réalisation de ce travail, nous allons utiliser lors de la mise en oeuvre des modèles des séries d'apports fournies par l'unité PQDH (Prévisions et Qualité des Données Hydroélectriques). Il s'agit de mesures d'apport en eau enregistrées entre 1950 et 2002 à des pas de temps mensuels et hebdomadaires pour 17 réservoirs. Elles serviront à la construction de modèles périodiques permettant de générer des séries d'apports synthétiques dont les caractéristiques statistiques seront similaires à celles des apports historiques.

PLAN DU MÉMOIRE

Ce mémoire décrit les différentes étapes de la modélisation des apports. Il est divisé en trois chapitres. Dans le premier chapitre, nous présentons les principales caractéristiques des séries d'apports utilisées. Cette présentation est utile puisqu'elle nous permet d'analyser le comportement des séries d'apports et d'orienter le choix de modèle.

Au deuxième chapitre, nous introduisons d'abord la notion de processus stochastique et spécialement la classe des processus *ARMA*. Par la suite, nous mettons l'accent sur l'étude des modèles autorégressifs moyenne mobile périodiques

(*PARMA*). Cette présentation nous amène à définir un certain nombre de notions essentielles à l'analyse des séries chronologiques. Ensuite, nous discutons des modèles *PAR* et de l'avantage d'inclure une variable exogène aux modèles périodiques que nous notons *PARX*. Afin de décrire la corrélation spatiale, nous avons considéré la classe de modèles *PAR* multivariés dont les coefficients sont des matrices diagonales. La corrélation entre les composantes d'un tel modèle multivarié est entièrement décrite par la matrice de covariance du bruit. Il est question aussi de l'estimation de la matrice de covariance. La méthode d'estimation est différente de l'approche traditionnelle puisqu'au lieu d'utiliser les données observées pour estimer la matrice de covariance, nous allons utiliser les résidus calculés. Cette modification de l'approche nous permet d'éviter de transformer les données et d'employer une relation non linéaire. Nous allons proposer à la fin de ce chapitre, deux méthodes pour l'estimation des corrélations spatiales. Une méthode paramétrique en utilisant la transformation lognormale et une méthode non paramétrique en utilisant les techniques simples du bootstrap.

Un aspect important de l'hydrologie statistique est la génération de séries (hydrogrammes) se rapprochant le plus possible des séries historiques. Avec les modèles *PAR*, il faut d'abord générer des réalisations de bruits blancs. Pour tenir compte du fait que les apports d'eau sont en général plus près d'une distribution lognormale que d'une distribution normale, il est coutume en hydrologie de générer des bruits blancs gaussiens et d'appliquer la transformation exponentielle afin d'obtenir des bruits lognormaux. Une approche qui semble nouvelle en hydrologie consiste à rééchantillonner les résidus des modèles *PAR* estimés. Il s'agit d'une approche purement non paramétrique et en plus, elle est très simple à implanter.

Finalement, au dernier chapitre nous présentons les résultats de simulations afin d'évaluer la performance des deux méthodes appliquées aux modèles *PAR* et *PARX*. Nous faisons une comparaison détaillée des résultats obtenus avec la transformation exponentielle et la méthode bootstrap, toutes deux appliquées aux

résidus de modèles $PAR(1)$ et $PARX(1)$. Nous terminerons ce mémoire avec la conclusion de nos travaux.

Chapitre 1

PRINCIPALES CARACTÉRISTIQUES DES SÉRIES D'APPORTS HISTORIQUES

Dans ce chapitre, nous présentons une analyse descriptive des séries d'apports historiques. Nous allons analyser les moyennes, les écarts types et les corrélations existantes entre les apports (hebdomadaires et mensuels) d'une part et entre les apports de différents réservoirs d'Hydro-Québec (corrélation spatiale) d'autre part.

1.1. LES DONNÉES

Les séries d'apports hebdomadaires et mensuels sont disponibles pour la période 1950 à 2002. Ces séries représentent les apports naturels observés par bilan hydrique des réservoirs. Le débit d'eau est mesuré en $m^3/seconde$. Notons qu'à l'issue de l'exercice de validation et de correction effectué par PQDH, nous considérons pour l'estimation des modèles que l'historique débute en 1961 et se termine en 2002. Nous avons donc des observations sur 42 années.

– Sites considérés dans l'étude

La figure 1.1 montre l'emplacement des principaux réservoirs d'Hydro-Québec. La connaissance de la situation géographique de ces réservoirs nous permettra de mettre ainsi en évidence les dépendances spatiales existantes entre les différents réservoirs.

– *Statistiques descriptives*

Nous présentons dans le tableau 1.1 les 17 réservoirs considérés dans notre étude ainsi que les caractéristiques statistiques annuelles de chacun d’eux. Notons que ce sont ces réservoirs que nous allons considérer dans ce travail. Dans la suite de notre étude, nous allons présenter seulement les résultats de 3 réservoirs (Churchill Falls, Caniapiscou et Gouin) où Caniapiscou (1) et Churchill Falls (6) sont situés au nord et géographiquement voisins et Gouin (11) est situé plus au sud. En effet, ces réservoirs par leurs situations géographiques et les différentes quantités d’apports qu’ils reçoivent périodiquement, semblent être représentatifs des 14 autres réservoirs.

TABLEAU 1.1. Résultats des statistiques annuelles pour les 17 réservoirs de référence.

Code	Sites	Moyenne en m^3/s	Écart type
1	Caniapiscou	768,03	131,93
2	Laforge	163,73	27,11
3	LaGrande4	556,24	91,64
4	EOL Total	850,32	118,54
5	Bersimis	318,10	40,30
6	Churchill Falls	1422,40	223,41
7	Manic 5	646,89	96,40
8	Lac Ste-Anne	191,63	25,12
9	Outardes 4	364,46	46,62
10	Ste-Marguerite	108,93	19,34
11	Gouin	175,26	29,15
12	Manouane	45,67	9,23
13	Rapide Blanc	167,82	32,32
14	Mattawin	73,88	13,74
15	TR-BE-LT2	173,88	28,26
16	GM-SH-Lagab	106,92	29,92
17	LG2-SAK	977,58	172,58

La figure 1.2 nous donne pour les réservoirs Caniapiscou, Churchill Falls et Gouin un aperçu du comportement des apports mensuels durant ces 42 années. La périodicité de chacune des 3 séries ressort clairement.

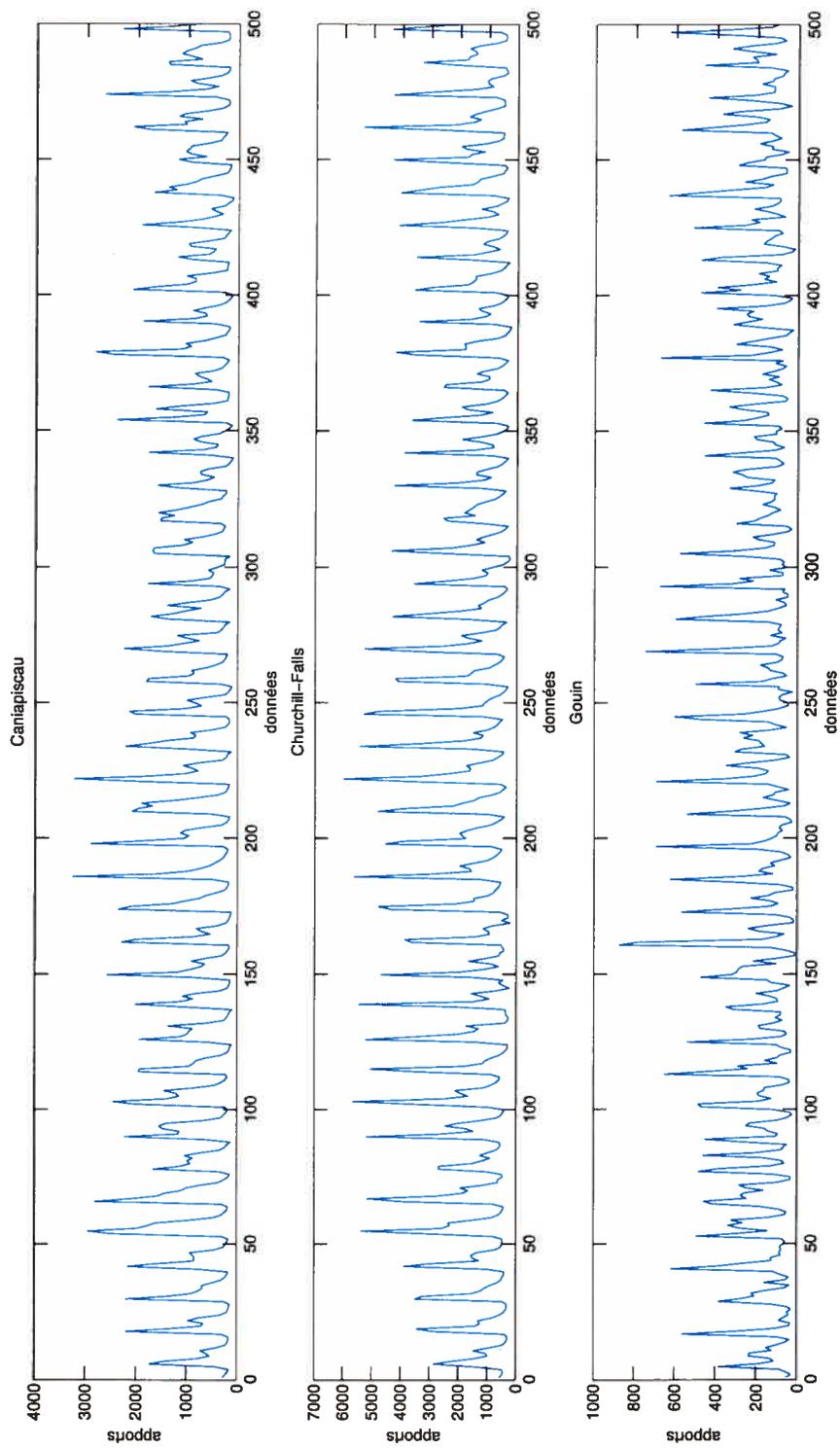


FIGURE 1.2. Représentation graphique des apports mensuels aux réservoirs Caniapiscau, Churchill-Falls et Gouin durant la période janvier 1961 à décembre 2002.

1.2. PÉRIODICITÉ

Dans notre travail, nous nous sommes intéressés aux séries périodiques dont le pas de temps de base est inférieur à l'année. Ainsi, l'apport en eau au réservoir i ($i = 1, \dots, I$) à la date t sera noté $X_t^{(i)} = X_{rs+m}^{(i)} = X_{r,m}^{(i)}$, où $t = rs + m$, $r = 0, \dots, n - 1$, représente l'année, avec $n = 42$ et $m = 1, \dots, s$ représente la saison, avec $s = 12$, dans le cas de données mensuelles et $s = 52$ dans le cas de données hebdomadaires et $I = 17$ pour l'ensemble des réservoirs. L'ensemble X des observations d'apports au réservoir (i), est représenté par la matrice suivante :

$$X^{(i)} = \begin{bmatrix} X_{0,1}^{(i)} & X_{0,2}^{(i)} & \dots & X_{0,m}^{(i)} & \dots & X_{0,s}^{(i)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ X_{r,1}^{(i)} & X_{r,2}^{(i)} & \dots & X_{r,m}^{(i)} & \dots & X_{r,s}^{(i)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ X_{n-1,1}^{(i)} & X_{n-1,2}^{(i)} & \dots & X_{n-1,m}^{(i)} & \dots & X_{n-1,s}^{(i)} \end{bmatrix}$$

où les lignes représentent les années et les colonnes correspondent aux saisons. Les figures ci-dessous représentent le comportement des apports hebdomadaires tout au long d'une année pour les réservoirs Caniapiscau, Churchill Falls et Gouin. En hydrologie, ce type de représentation est appelé hydrogramme. On définit ce dernier comme une expression ou une représentation graphique de la variation des débits d'eau au cours d'une année.

L'analyse graphique des hydrogrammes sur 52 semaines pour le réservoir Caniapiscau présentée dans la figure 1.3, montre une certaine périodicité variable entre les apports pour les 4 années : 1972, 1982, 1986 et 1987. Chaque hydrogramme admet aussi des débits maximums, plus précisément deux pointes. La première est située entre la 15^e et la 25^e semaine correspondant à la période mai-juillet. La seconde de moindre importance, est située entre la 35^e et la 40^e semaine correspondant à la période octobre-novembre. L'augmentation subite du volume des apports dans la première période est due à la crue printanière où la fonte des neiges conjuguée aux précipitations liquides ont engendré une augmentation importante des débits d'eau dans les réservoirs. L'augmentation perçue dans la seconde période correspond à la crue automnale causée par la forte pluie d'automne.

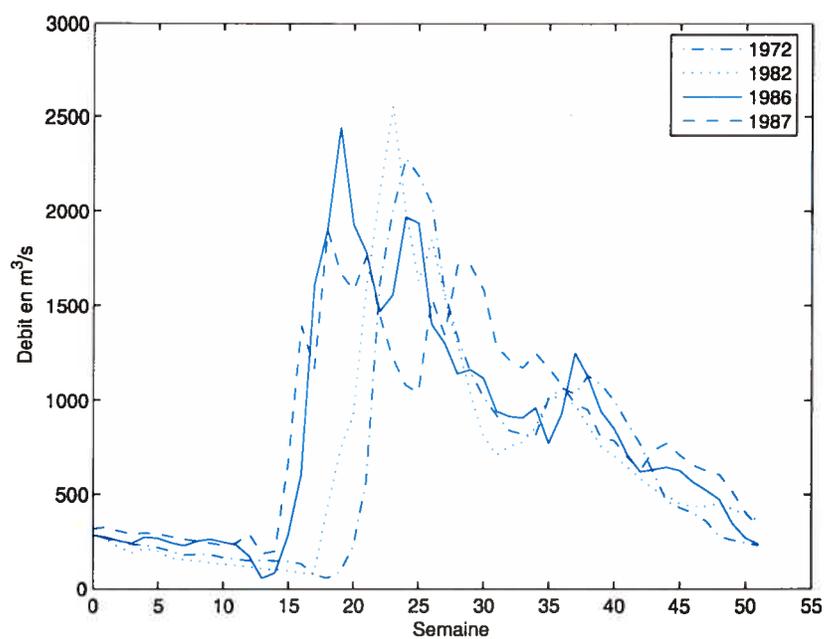


FIGURE 1.3. Hydrogrammes du réservoir Caniapiscau pour les années 1972, 1982, 1986 et 1987.

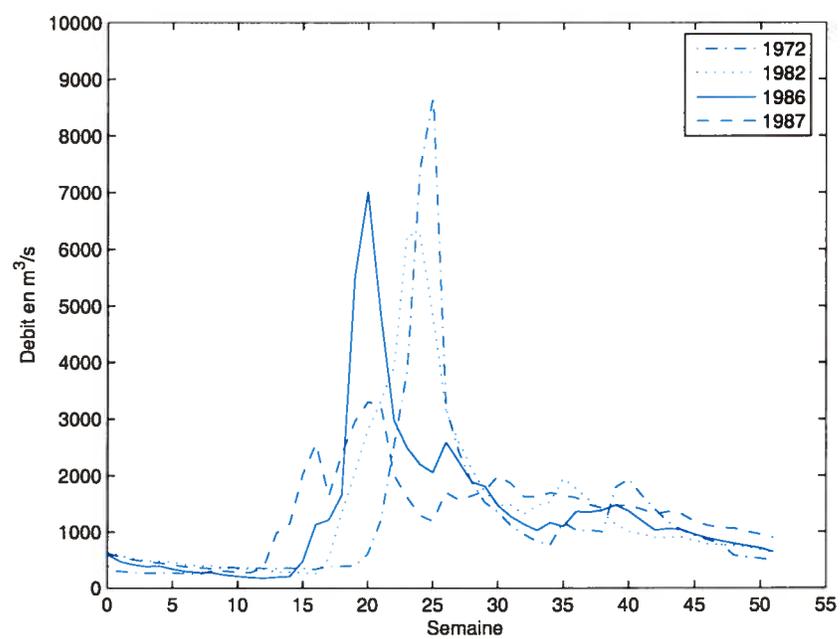


FIGURE 1.4. Hydrogrammes du réservoir Churchill Falls pour les années 1972, 1982, 1986 et 1987.

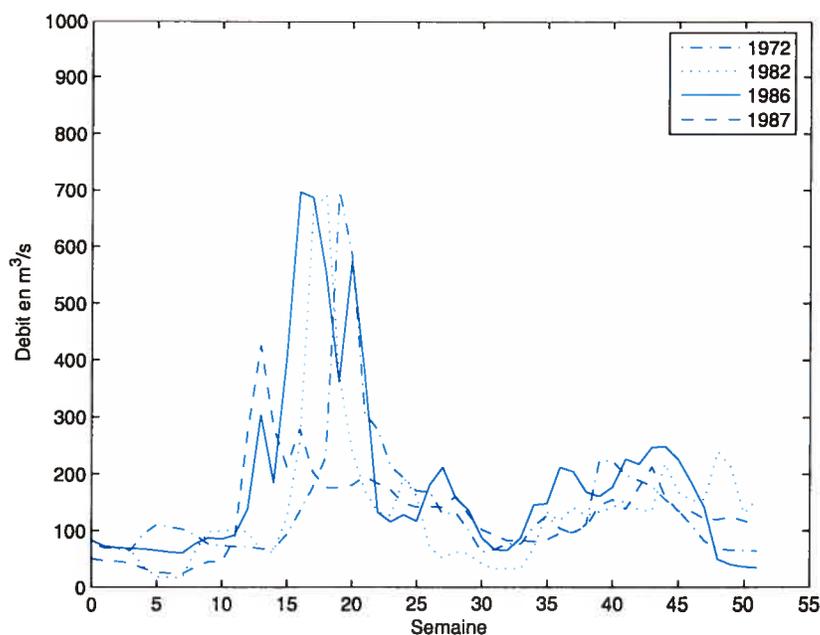


FIGURE 1.5. Hydrogrammes du réservoir Gouin pour les années 1972, 1982, 1986 et 1987.

Par son débit d'eau dans les réservoirs, la crue automnale n'est pas vraiment importante en comparaison à celle du printemps et la figure 1.3 le montre bien. Un phénomène similaire est aussi observé dans les figures 1.4 et 1.5 pour les réservoirs Churchill Falls et Gouin.

Nous avons constaté que l'augmentation subite des débits durant des périodes précises de l'année n'est pas fixe dans le temps. Certaines années, les crues printanières sont précoces comme celles survenues à la 15^e et la 17^e semaine alors que d'autres sont tardives comme celles survenues à la 23^e et la 25^e semaine. C'est pourquoi nous pouvons supposer que globalement les crues printanières surviennent dans un intervalle compris entre la 15^e et 25^e semaine et les crues automnales entre la 40^e et 50^e semaine.

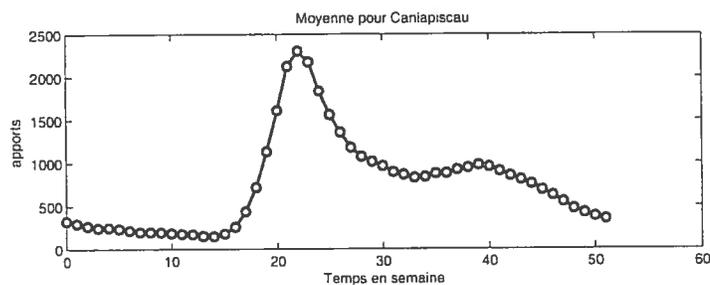
Dans le cadre de cette étude, nous allons surtout essayer de reproduire ces différentes tendances lors des simulations. Il est intéressant aussi d'analyser la moyenne et l'écart type des 3 réservoirs à chaque saison, définis par

$$\hat{\mu}_{x,m} = \frac{1}{n} \sum_{r=0}^{n-1} X_{rs+m}, \quad (1.2.1)$$

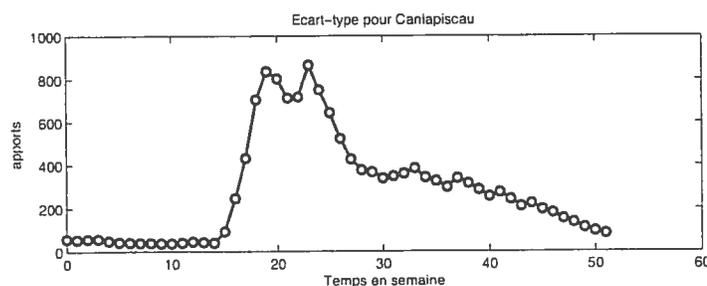
$$\hat{\sigma}_{x,m} = \left[\frac{1}{n} \sum_{r=0}^{n-1} (X_{rs+m} - \hat{\mu}_m)^2 \right]^{\frac{1}{2}} \quad (1.2.2)$$

où $m = 1, \dots, s$.

Pour les tendances statistiques (moyenne et écart type), les figures 1.6, 1.7, 1.8 pour les réservoirs Caniapiscou, Churchill-Falls et Gouin confirment l'existence d'une périodicité annuelle dont nous devons tenir compte lors de la génération des apports. En effet, nous constatons que durant la période hivernale, les débits semblent homogènes mais au moment de la crue printanière les observations deviennent hétérogènes à cause de la périodicité variable déjà observée dans les figures 1.3, 1.4 et 1.5. Notons que le même phénomène a été observé pour les 14 autres réservoirs.

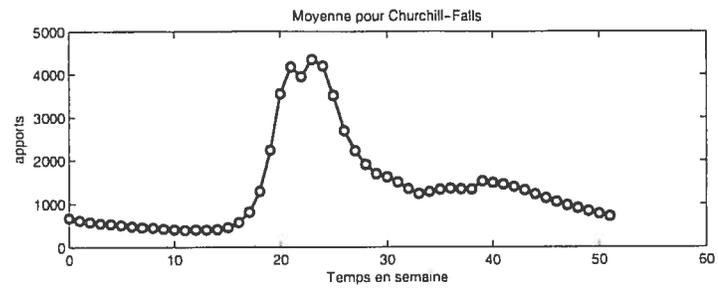


(a)

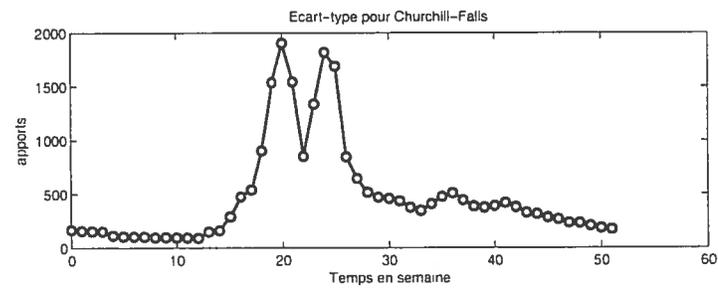


(b)

FIGURE 1.6. Les moyennes et écarts types hebdomadaires pour le réservoir Caniapiscou.

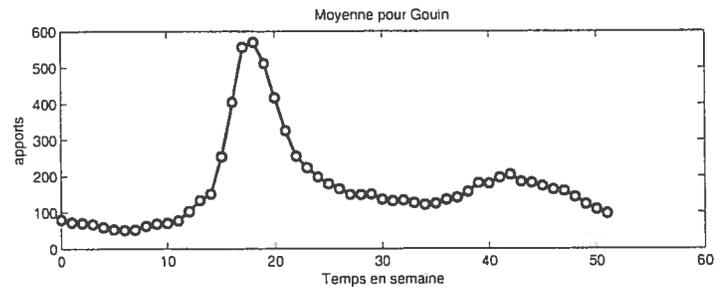


(a)

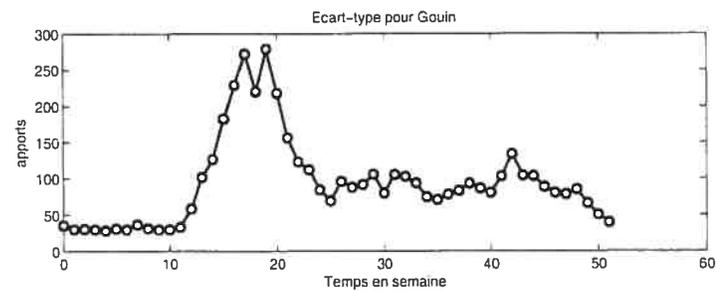


(b)

FIGURE 1.7. Les moyennes et écarts types hebdomadaires pour le réservoir Churchill Falls.



(a)



(b)

FIGURE 1.8. Les moyennes et écarts types hebdomadaires pour le réservoir Gouin.

Pour ne pas allonger le texte, nous allons présenter dans la section 1.3 que les résultats mensuels.

1.3. DÉPENDANCE À COURT TERME

Pour mesurer la dépendance à court terme entre les apports, nous allons calculer pour chaque série les autocorrélations périodiques, c'est à-dire pour chaque saison. Pour une réalisation donnée, on estime l'autocorrélation périodique de délai h pour la saison m par

$$\hat{\rho}_m(h) = \frac{\hat{\gamma}_m(h)}{\sqrt{\hat{\gamma}_m(0)\hat{\gamma}_{m-h}(0)}}, \quad (1.3.1)$$

où $\hat{\gamma}_m(h)$ représente l'autocovariance périodique estimée de délai h pour la saison m définie par

$$\hat{\gamma}_m(h) = \frac{1}{n} \sum_{r=0}^{n-1} (X_{rs+m} - \bar{X}_m)(X_{rs+m-h} - \bar{X}_m) \quad (1.3.2)$$

où $m = 1, \dots, s$ et $h = 0, 1, \dots, ns - 1$. Dans la définition de $\hat{\gamma}_m(h)$, nous posons $X_{rs+m-h} - \bar{X}_m = 0$ lorsque $X_{rs+m-h} = X_{t-h}$ n'est pas disponible, c'est-à-dire si $t - h \leq 0$. Les figures 1.9, 1.10 et 1.11 montrent les autocorrélations périodiques au délai h , $h = 1, 2, 3$ pour les 12 mois. Les droites horizontales représentent plus et moins 2 écarts types sous l'hypothèse de bruit blanc et cet écart type est $\frac{1}{\sqrt{n}}$.

L'analyse des figures 1.9, 1.10 et 1.11 montre au début une autocorrélation périodique mensuelle d'ordre 1 très forte pour les 3 réservoirs mais à partir du 4e mois, on observe une nette tendance à la baisse de l'autocorrélation. Cela est due surtout à la crue printanière qui survient entre le 4e et le 6e mois, dépendamment des réservoirs. En effet, pour le réservoir Caniapiscou la crue printanière débute à partir du 5^e mois, alors que pour le réservoir Churchill Falls, il y a des moments où la crue débute à partir du 4^e mois et d'autres où elle ne débute qu'à partir du 6^e mois. Pour le réservoir Gouin, la crue printanière débute un peu plus tôt. C'est-à-dire à partir du 4^e mois, ce qui est normal vu que ce réservoir est situé plus au sud par rapport à Caniapiscou et Churchill Falls. Ainsi, afin de simuler des hydrogrammes de crue réalistes, les modèles que nous allons construire doivent donc impérativement préserver les caractéristiques des autocorrélations.

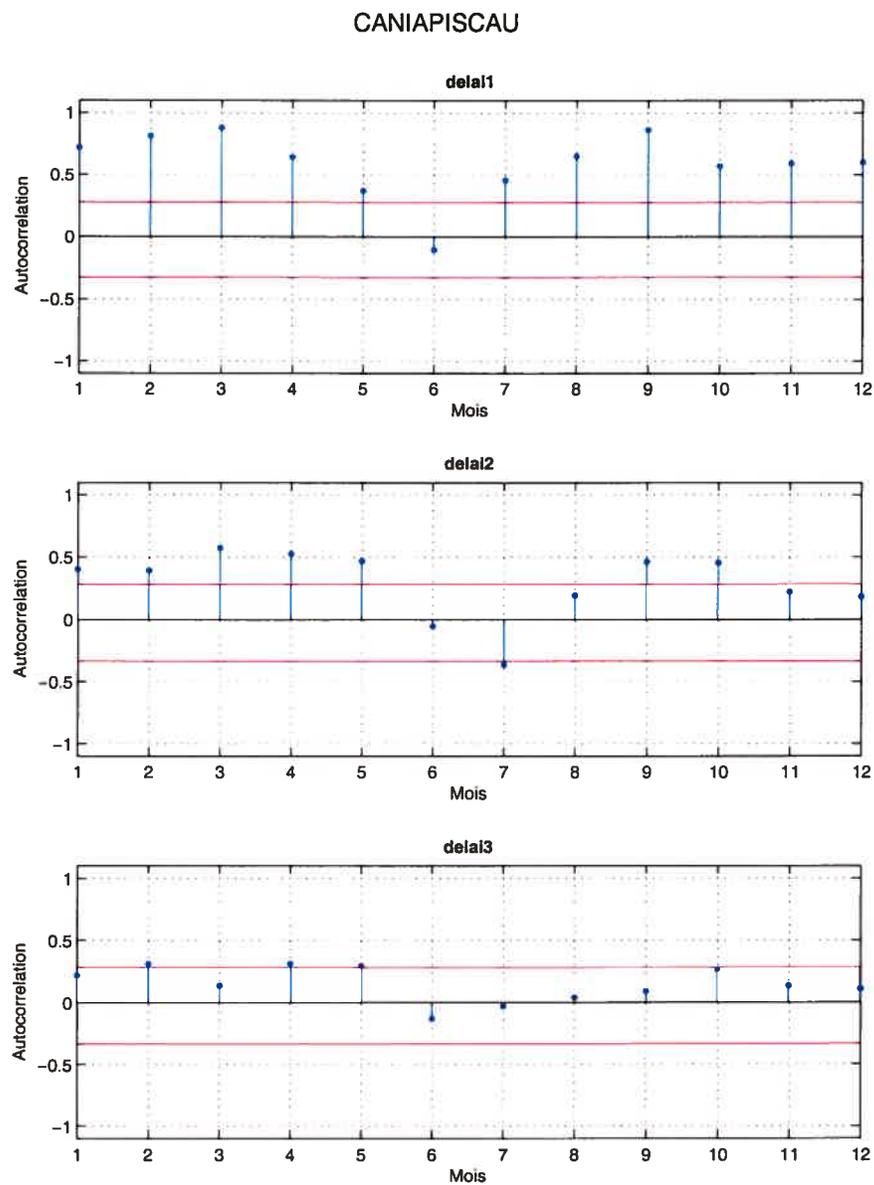


FIGURE 1.9. Autocorrélations périodiques mensuelles pour le réservoir Caniapiscau.

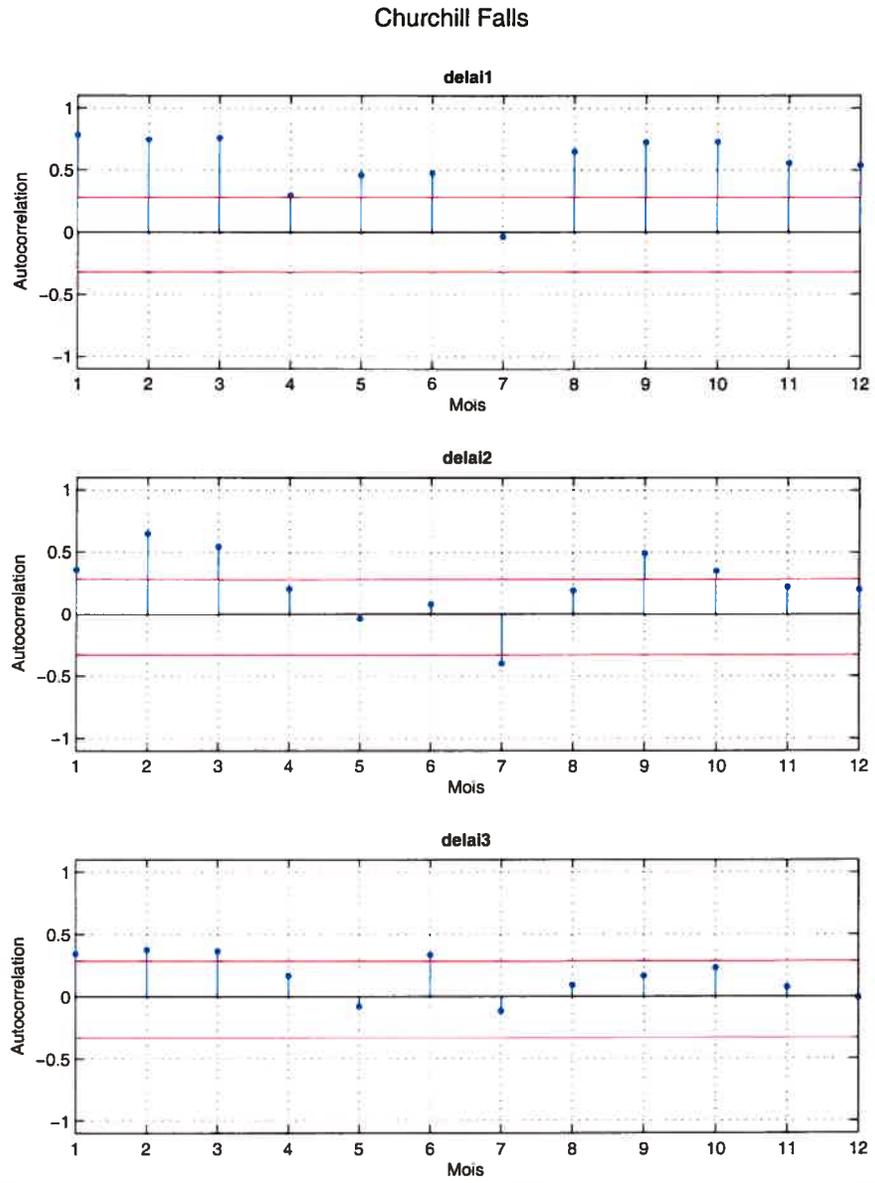


FIGURE 1.10. Autocorrélations périodiques mensuelles pour le réservoir Churchill Falls.

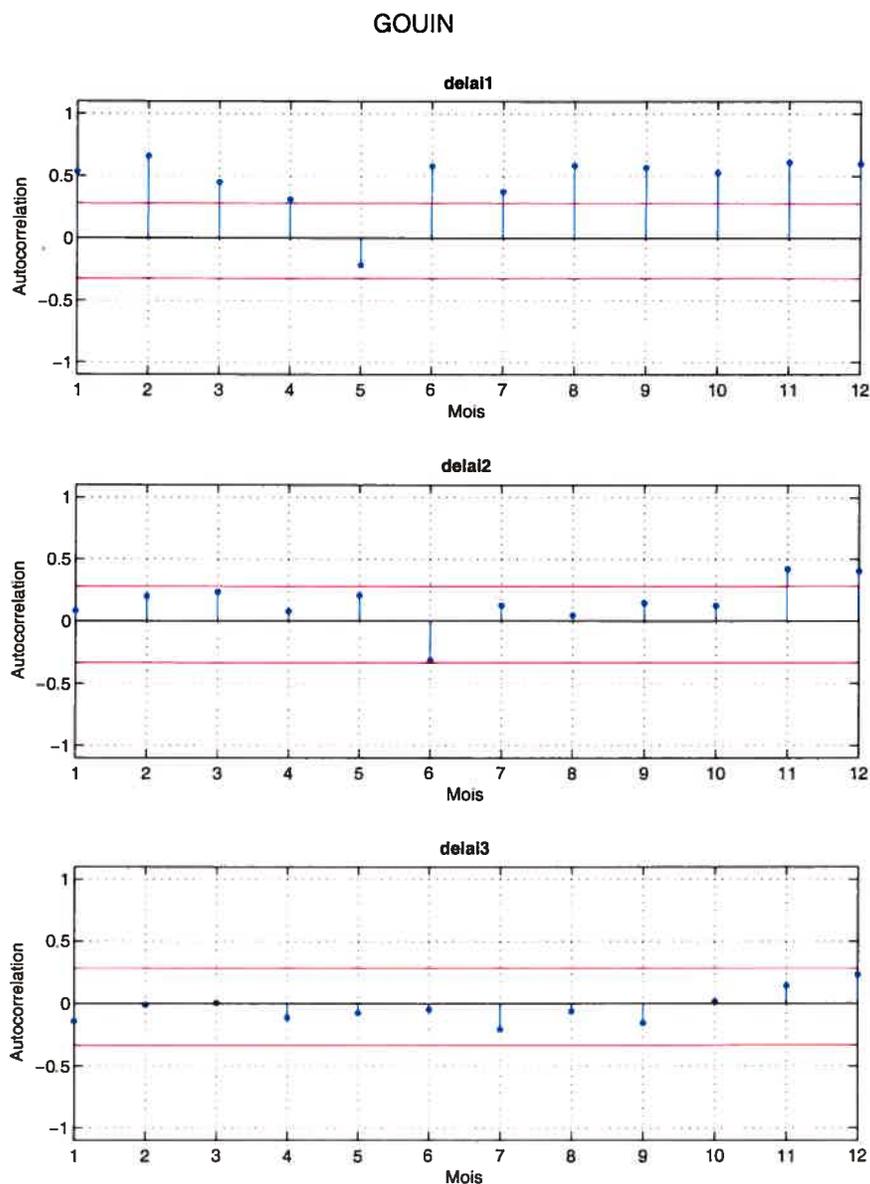


FIGURE 1.11. Autocorrélations périodiques mensuelles pour le réservoir Gouin.

Ainsi, en plus de la périodicité déjà constatée, nous observons l'existence d'une corrélation assez forte entre les apports naturels d'un mois à un autre. Cette corrélation est observée aussi dans le cas des séries d'apports historiques hebdomadaires.

1.4. DÉPENDANCE À LONG TERME

Il est important lors de la modélisation de reproduire la corrélation interannuelle. En effet, cette corrélation a une influence sur les caractéristiques statistiques ainsi que sur le volume des apports qu'on ne peut ignorer. Tenir compte de la corrélation interannuelle est indispensable pour la gestion à plus long terme. Les modèles utilisés dans le logiciel SDDP doivent reproduire cette persistance. Ainsi, en plus de la présence de corrélation entre les apports à court terme (mensuels et hebdomadaires), il y a l'autocorrélation interannuelle d'une année à l'autre qui est présente pour certains réservoirs. Cette corrélation est calculée à partir du volume annuel des apports, c'est-à-dire, $R^{(i)}(1) = \text{corr}(V_r^{(i)}, V_{r-1}^{(i)})$ où le volume

$$V_r^{(i)} = \sum_{m=1}^s X_{rs+m}^{(i)} \quad (1.4.1)$$

avec l'année $r = 0, \dots, n - 1$ et le réservoir $i = 1, \dots, I$.

Les figures 1.12, 1.13 et 1.14 montrent les autocorrélations des apports annuels pour les délais, $h = 1, \dots, 10$ années. Les droites représentent plus et moins 2 écarts types ou l'écart type de la corrélation estimée $\hat{\rho}(h)$ est :

$$\sqrt{\frac{1}{n}(1 + 2(\hat{\rho}(1)^2 + \hat{\rho}(2)^2 + \dots + \hat{\rho}(h-1)^2))},$$

l'écart type sous l'hypothèse que le vrai modèle est $MA(h)$, $h = 1, 2, \dots$ (Hipel et McLeod, 1994, p.72-73). Au délai 1, l'écart type est $\frac{1}{\sqrt{n}}$, l'écart type sous l'hypothèse de bruit blanc. Les limites de signification au niveau 5% sont approximativement $\pm 0,30$. Ainsi, dans la figure 1.13 pour le réservoir Churchill Falls, nous observons une corrélation interannuelle significative (0,42). Dans la figure 1.12, pour le réservoir Caniapiscou (0,26), la corrélation n'est pas significative alors que pour Gouin (-0,10), on n'observe aucune corrélation significative, c'est un bruit blanc. Le tableau 1.2 donne les corrélations annuelles de délai 1 des 14 autres réservoirs.

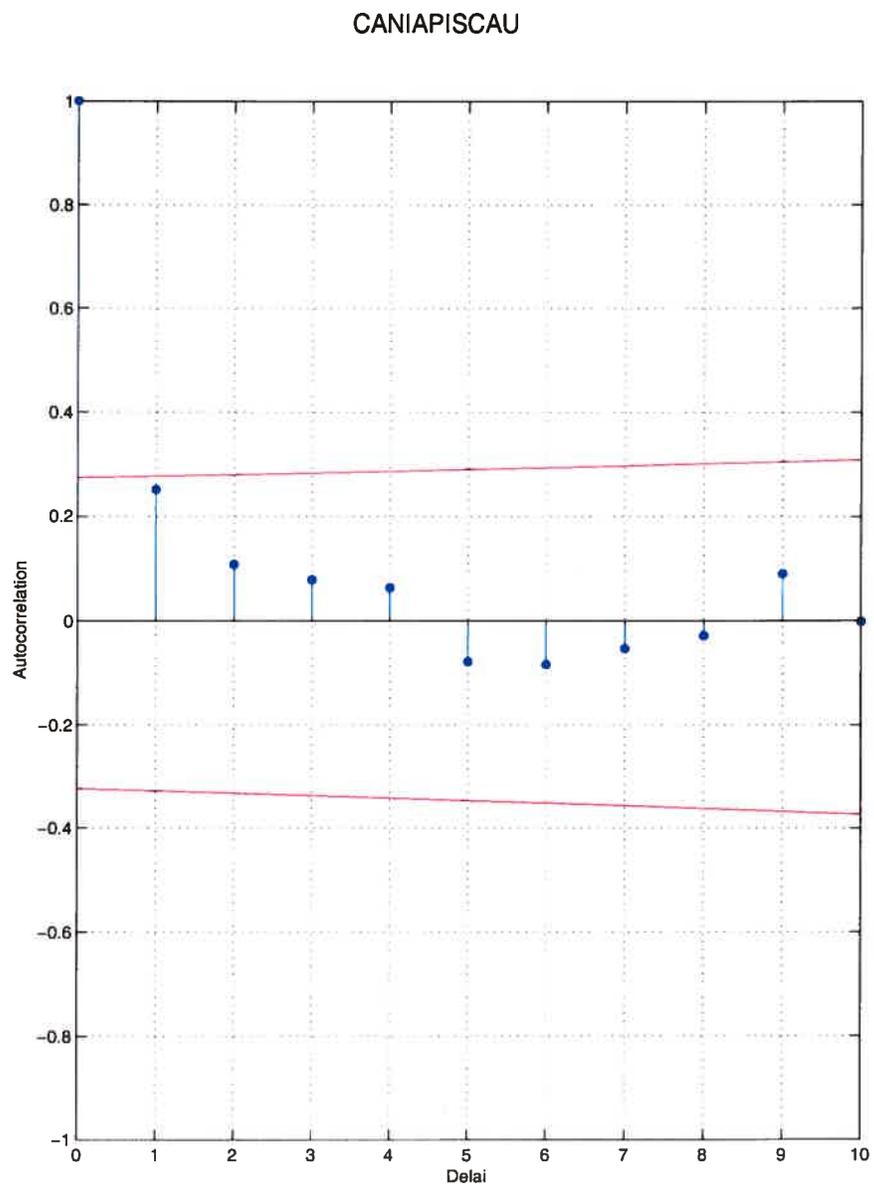


FIGURE 1.12. Autocorrélations des apports annuels pour le réservoir Caniapiscau.

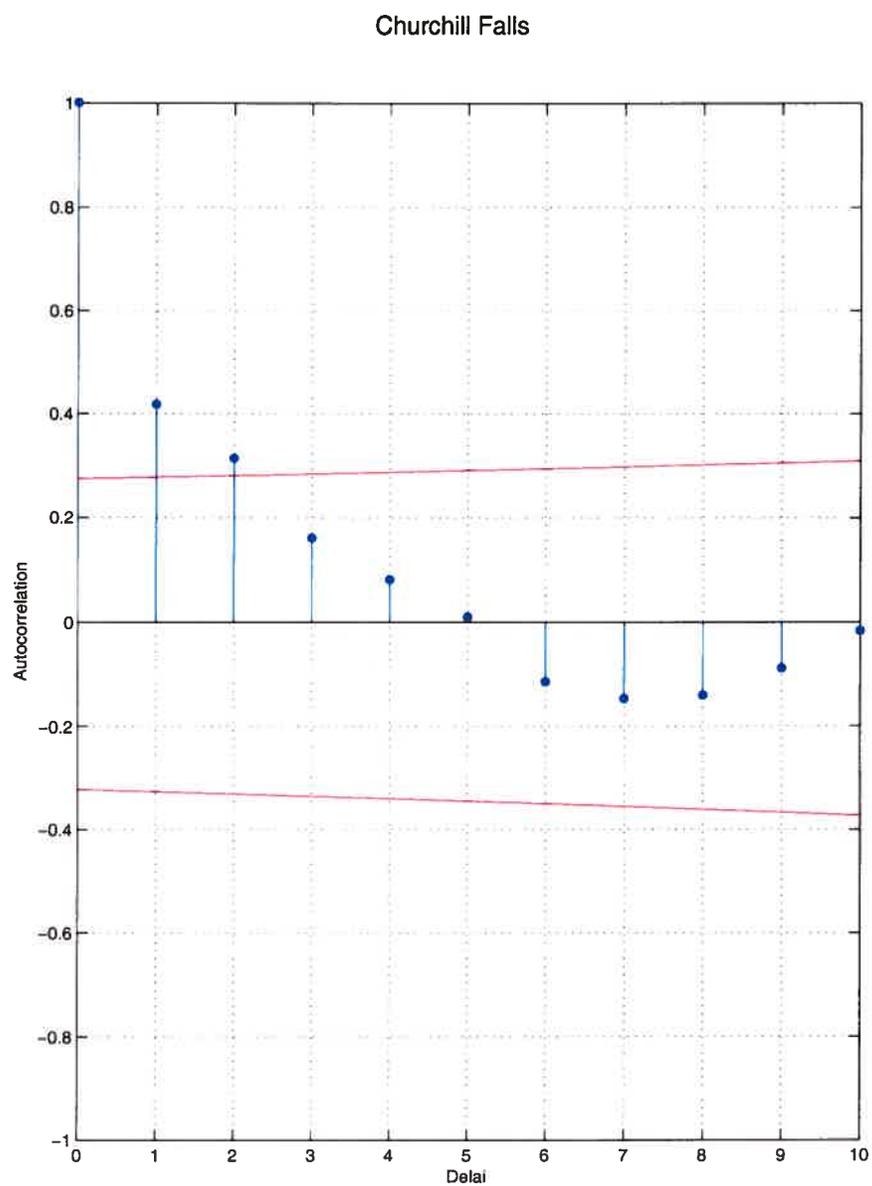


FIGURE 1.13. Autocorrélations des apports annuels pour le réservoir Churchill Falls.

GOUIN

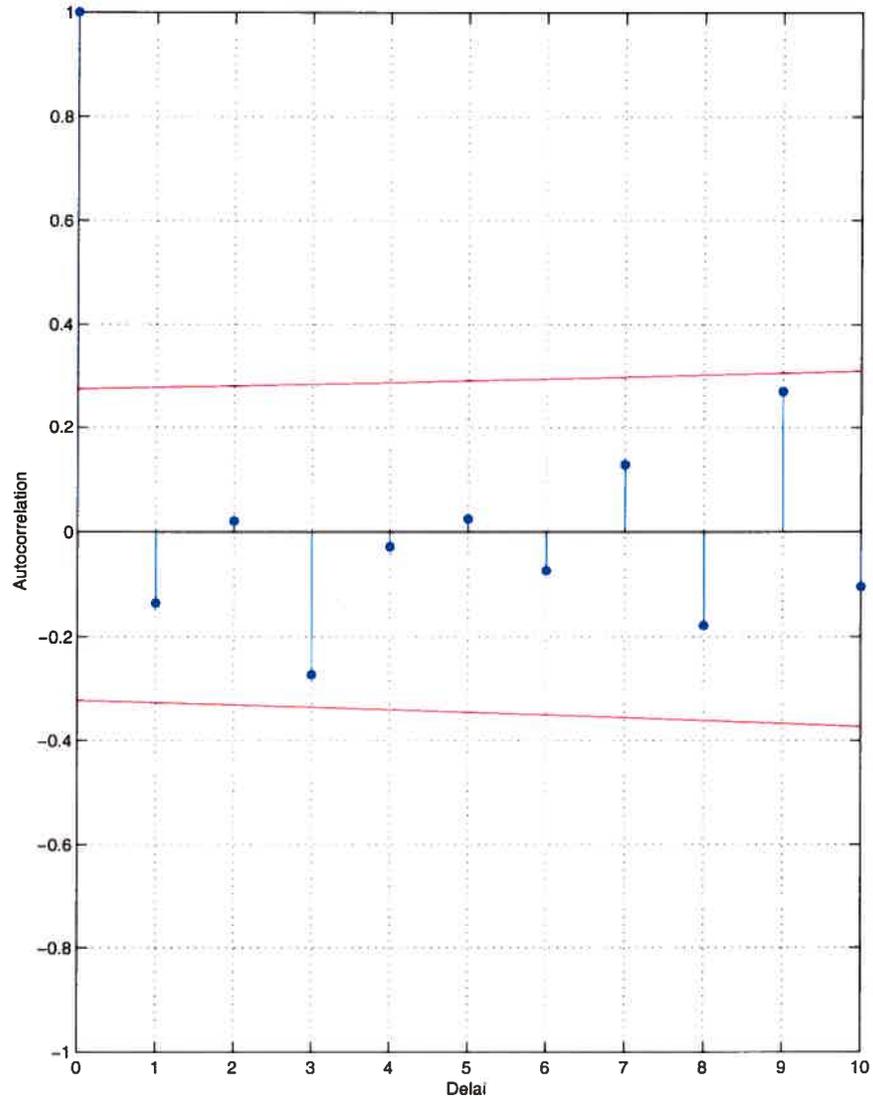


FIGURE 1.14. Autocorrélation des apports annuels pour le réservoir Gouin.

TABLEAU 1.2. Autocorrélations de délai 1 des apports annuels aux 17 réservoirs.

Code	Réservoirs	Corrélation
1	Caniapiscau	0,26
2	Laforge	0,34
3	LaGrande 4	0,38
4	EOL-Total	0,24
5	Bersimis	0,23
6	Churchill Falls	0,42
7	Manic 5	0,27
8	Lac Ste-Anne	0,15
9	Outardes 4	0,29
10	Ste-Marguerite	0,52
11	Gouin	-0,10
12	Manouane	-0,06
13	Rapide Blanc	-0,09
14	Mattawin	0,08
15	TR-BE-LT2	0,17
16	GM-SH-Lagab	0,10
17	LG2-SAK	0,33

1.5. COHÉRENCE SPATIALE

L'apport naturel est soumis aux phénomènes climatiques qui se manifestent à une échelle régionale. On peut alors penser qu'il existe une certaine corrélation régionale ou spatiale entre les caractéristiques statistiques des apports périodiques de chaque série. Cette corrélation spatiale, qui est calculée à partir des volumes annuels décrits par (1.4.1), est définie par

$$\hat{\rho}_{i,j} = \frac{\hat{\gamma}_{i,j}}{\sqrt{\hat{\gamma}_{i,i} \hat{\gamma}_{j,j}}}, \quad i, j = 1, 2, \dots, 17 \quad (1.5.1)$$

où

$$\hat{\gamma}_{i,j} = \frac{1}{n} \sum_{r=0}^{n-1} (V_r^{(i)} - \bar{V}^{(i)})(V_r^{(j)} - \bar{V}^{(j)}). \quad (1.5.2)$$

Si on applique l'équation (1.5.2) aux volumes annuels, nous observons dans le tableau 1.3 que les réservoirs qui sont situés dans la même région que Caniapiscau sont fortement corrélés avec ce dernier. C'est le cas avec Laforge (0,87), La Grande 4 (0,89) et Churchill Falls (0,83). Par contre, pour des réservoirs situés

plus au sud (voir la figure 1.1) la corrélation est très faible, comme c'est le cas avec Mattawin $(-0,01)$ et GM-SH-Lagab $(0,01)$. On observe aussi que les réservoirs Gouin et Rapide Blanc sont assez fortement corrélés $(0,73)$. Dans la figure 1.15, on voit bien que Caniapiscau et Churchill Falls sont fortement corrélés alors que Caniapiscau et Gouin et Churchill Falls et Gouin le sont moins. Ceci est normal puisque les réservoirs voisins subissent les mêmes phénomènes climatiques, ce qui se répercute généralement dans les corrélations régionales. Par conséquent, lors de la validation, les modèles choisis doivent être capables de reproduire cette corrélation spatiale lorsqu'elle existe.

TABLEAU 1.3. Corrélations spatiales entre les 17 réservoirs (données historiques).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	1.00																
2	0.87	1.00															
3	0.89	0.99	1.00														
4	0.68	0.78	0.76	1.00													
5	0.22	0.36	0.36	0.31	1.00												
6	0.83	0.71	0.73	0.56	0.36	1.00											
7	0.67	0.69	0.69	0.60	0.50	0.67	1.00										
8	0.37	0.42	0.42	0.42	0.72	0.69	1.00										
9	0.50	0.61	0.60	0.52	0.83	0.79	0.76	1.00									
10	0.45	0.44	0.48	0.51	0.56	0.53	0.63	0.60	1.00								
11	0.35	0.41	0.39	0.40	0.46	0.27	0.35	0.43	0.25	1.00							
12	0.09	0.09	0.08	0.10	0.37	0.29	0.39	0.37	0.32	0.42	1.00						
13	0.16	0.24	0.21	0.21	0.45	0.29	0.40	0.43	0.16	0.73	0.69	1.00					
14	-0.01	-0.05	-0.07	-0.03	0.28	0.23	0.40	0.28	0.19	0.31	0.74	0.59	1.00				
15	0.12	0.17	0.15	0.09	0.59	0.39	0.58	0.52	0.33	0.45	0.71	0.72	0.71	1.00			
16	0.01	-0.06	-0.05	-0.04	0.34	0.23	0.35	0.26	0.30	0.28	0.82	0.61	0.74	0.59	1.00		
17	0.77	0.85	0.85	0.83	0.33	-0.05	0.35	0.52	0.56	0.34	0.17	0.28	0.74	0.74	0.02	0.19	1.00

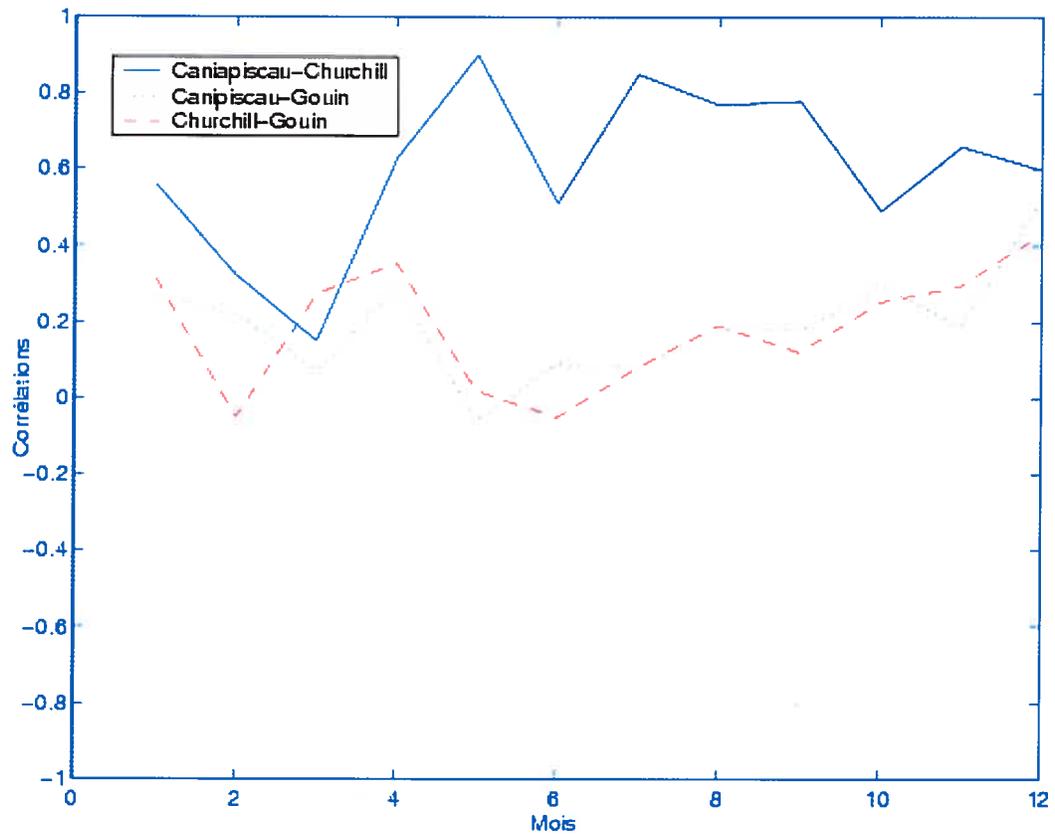


FIGURE 1.15. Graphe comparatif des corrélations historiques entre les 3 réservoirs de référence pour les 12 mois.

1.6. NON-NORMALITÉ

La majorité des modèles stochastiques de simulation repose sur l'hypothèse que les données sont distribuées selon une loi normale, alors qu'en réalité ce n'est pas toujours le cas, (voir, par exemple, la figure 1.16 pour le réservoir Caniapiscou). Les séries sont normales au début, jusqu'à la 11^e semaine mais à partir de la 13^e semaine, nous observons un changement significatif dans la distribution des données où les apports observés ne sont plus normaux. Ceci est dû au fait que durant les périodes allant de la 13^e à la 17^e semaine (dépendant des années), il y a la crue printanière. Cette crue qui commence pour certaines années plus tôt ou plus tard, augmente le débit d'eau dans les réservoirs.

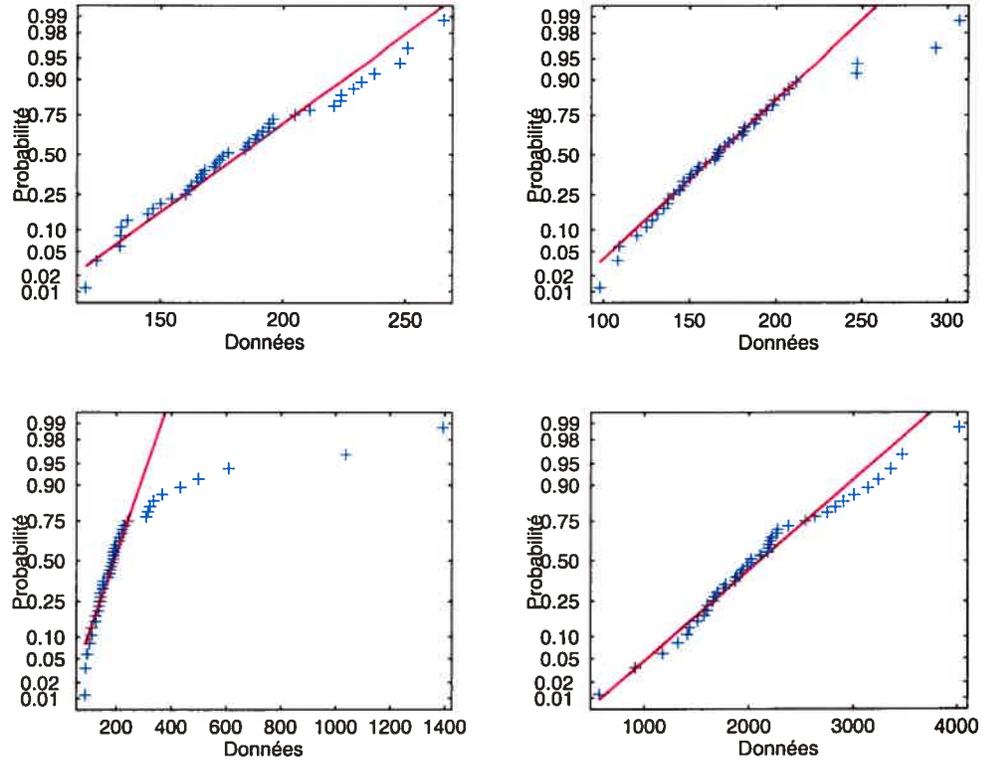


FIGURE 1.16. Représentation graphique quantiles des apports-quantiles de la loi normale au réservoir Caniapiscau pour les 11^e, 13^e, 17^e et 22^e semaines.

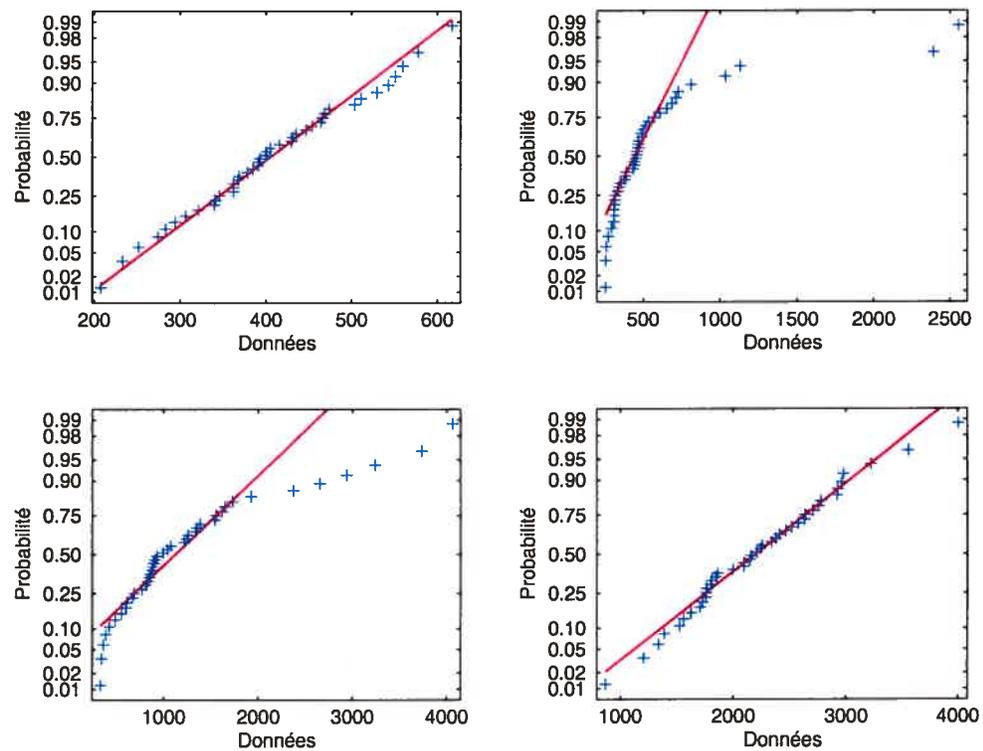


FIGURE 1.17. Représentation graphique quantiles des apports-quantiles de la loi normale au réservoir Churchill Falls pour les 11^e, 17^e, 19^e et 28^e semaines.

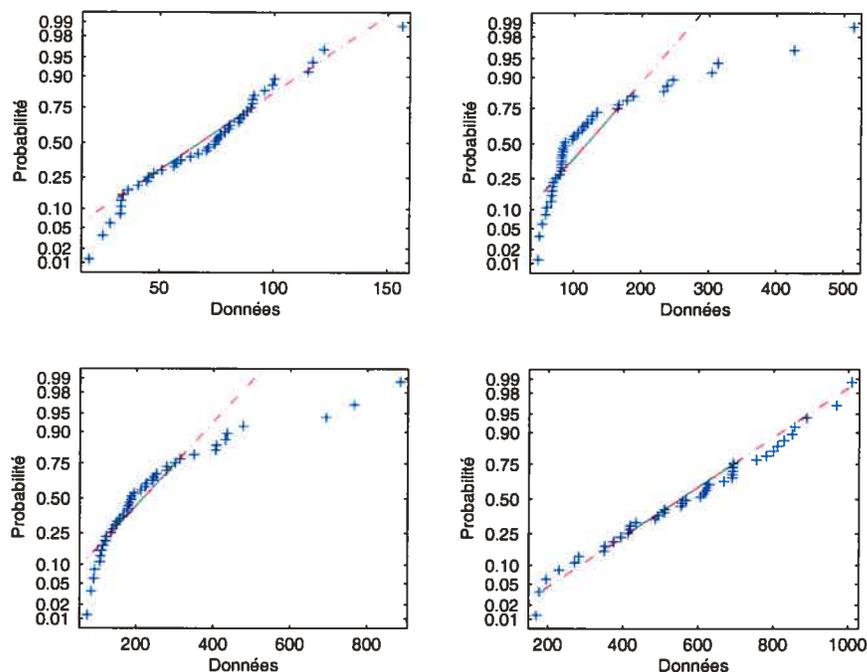


FIGURE 1.18. Représentation graphique quantiles des apports-quantiles de la loi normale au réservoir Gouin pour les 11^e, 14^e, 16^e et 19^e semaines.

Concernant la figure 1.17 pour le réservoir Churchill Falls, nous observons que les apports ne sont plus normaux à partir de la 17^e semaine. Ceci est dû, comme pour le réservoir Caniapiscau, à la crue printanière mais à partir de la 28^e semaine les graphiques quantiles-quantiles se rapprochent d'une droite. Pour la figure 1.18 pour le réservoir Gouin, nous remarquons que les apports ne sont plus normaux à partir de la 14^e semaine. C'est-à-dire que la crue printanière survient avant Caniapiscau et Churchill Falls. Ceci est normal, puisque le réservoir Gouin est situé plus au sud. Mais à partir de la 19^e semaine les graphiques quantiles-quantiles se rapprochent d'une droite.

Nous avons présenté les graphiques quantiles-quantiles des données hebdomadaires car les changements attribuables à la crue printanière sont plus facilement perceptibles qu'avec les données mensuelles.

Une approche classique pour se rapprocher de la normalité est d'effectuer des transformations des données. En général, cette procédure facilite la modélisation et permet de considérer une gamme plus large de modèles. Dans notre cas, nous ne pouvons pas transformer directement les données à cause de la contrainte de

linéarité du logiciel SDDP. Pour les simulations, la transformation sera plutôt appliquée aux résidus.

1.7. CONCLUSION

L'analyse de ces différentes séries pour les réservoirs Caniapiscou, Churchill Falls et Gouin, choisis comme référence parmi les 17 réservoirs, a montré d'une part que les apports historiques mensuels et hebdomadaires sont périodiques, non-normaux, très autocorrélés et d'autre part qu'il existe une dépendance spatiale entre les apports en eau. Ainsi, les modèles qui vont être choisis dans le chapitre suivant doivent tenir compte de ces caractéristiques, en plus de la contrainte de linéarité des modèles d'apports imposée par le logiciel d'optimisation SDDP.

Chapitre 2

MODÈLES DE SÉRIES CHRONOLOGIQUES EN HYDROLOGIE

2.1. INTRODUCTION

Avant d'introduire les modèles les plus adéquats pour ce type de séries, il est important de présenter certaines notions de base sur les processus stochastiques. Cela va nous permettre de faire la distinction entre les modèles stationnaires applicables aux apports annuels et les modèles périodiques permettant la modélisation des apports périodiques et guider notre choix de modèle.

2.2. GÉNÉRALITÉS

2.2.1. Stationnarité

La stationnarité est une propriété fondamentale dans l'analyse des séries chronologiques.

Définition 2.2.1. (*Stationnarité au sens large*) Le processus $\{X_t : t \in \mathbb{Z}\}$ tel que $E[X_t]^2 < \infty$ est dit stationnaire au sens large (SSL) si :

- i) $E[X_t] = \mu_X$ est indépendante de t , $\forall t \in \mathbb{Z}$,
- ii) $Cov(X_t, X_{t-h}) = \gamma(h)$ est indépendante de t , $\forall t, h \in \mathbb{Z}$.

La stationnarité au sens large ne fait appel qu'aux deux premiers moments des variables X_t et revient à supposer que les deux premiers moments sont invariants

par rapport aux translations des indices de temps. On l'appelle aussi stationnarité du second ordre. Si un processus possède des moments d'ordre $m \geq 1$ et que ceux-ci sont invariants par translations des indices de temps, alors on dit que le processus est stationnaire à l'ordre m . Des exemples de processus stationnaires sont fournis par les bruits blancs.

Définition 2.2.2. (*Bruit blanc faible*) *Le processus $\{X_t\}$ est un bruit blanc faible $(0, \sigma^2)$ si :*

- i) $E[X_t] = \mu_X = 0, \quad \forall t,$
- ii) $Var(X_t) = \gamma(0) = \sigma_x^2, \quad \forall t,$
- iii) $\gamma(h) = Cov(X_t, X_{t-h}) = 0, \quad \forall t, \forall h \neq 0.$

Définition 2.2.3. (*Bruit blanc fort*) *Le processus $\{X_t\}$ est un bruit blanc fort de moyenne zéro si les X_t sont des variables aléatoires indépendantes et identiquement distribuées (iid) selon la fonction de répartition $F(x) = P(X_t \leq x)$, $\forall x \in \mathbb{R}, \forall t \in \mathbb{Z}$*

Si la variance existe, un bruit blanc fort est aussi un bruit blanc faible mais l'inverse n'est pas vrai en général.

2.2.2. Processus ARMA

Une classe particulièrement populaire de processus stationnaires est la classe des processus $ARMA(p, q)$ stationnaires non saisonniers tels que popularisés par la publication du livre de Box and Jenkins en 1970 (dernière édition : Box et al., 1994). Ils sont caractérisés par un nombre fini de paramètres, ce qui constitue une propriété fort importante lors de la modélisation d'une série finie.

Définition 2.2.4. *Un processus $\{X_t : t \in \mathbb{Z}\}$, tel que $E[X_t^2] < \infty$ et $E[X_t] = 0$, est un processus autoregressif-moyenne mobile d'ordre (p, q) , noté $ARMA(p, q)$, s'il est stationnaire et s'il satisfait la représentation suivante :*

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = \epsilon_t - \theta_1 \epsilon_{t-1} - \dots - \theta_q \epsilon_{t-q}, \quad \forall t \in \mathbb{Z}, \quad (2.2.1)$$

où $\{\epsilon_t\}$ est un bruit blanc fort $(0, \sigma_\epsilon^2)$ et les paramètres $\phi_i, i = 1, \dots, p$, sont appelés les paramètres autorégressifs et les $\theta_j, j = 1, \dots, q$, les paramètres moyenne mobile.

L'équation (2.2.1) peut s'écrire également sous la forme

$$\phi(B)X_t = \theta(B)\epsilon_t \quad (2.2.2)$$

où B est l'opérateur retard vérifiant : $B^m X_t = X_{t-m}$ pour $m \geq 1$ avec $B^0 = 1$, $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$, $\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$ sont respectivement les opérateurs autorégressif et moyenne mobile. Si $q = 0$, $\{X_t\}$ est un processus purement autorégressif, noté $AR(p)$. Si $p = 0$, $\{X_t\}$ est un processus moyenne mobile, noté $MA(q)$.

Notons qu'un processus $ARMA(p, q)$ est *stationnaire* si et seulement si les racines du polynôme $\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$ sont de module strictement supérieur à 1. Un processus définie par (2.2.2) qui est stationnaire est aussi causal, c'est-à-dire qu'il admet une représentation de la forme

$$X_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j} = \psi(B)\epsilon_t$$

où $\psi(B) = \phi(B)^{-1}\theta(B) = \sum_{j=0}^{\infty} \psi_j B^j$ et $\{\epsilon_t\}$ est un bruit blanc faible.

Un processus $ARMA$ est *inversible*, c'est-à-dire que X_t peut s'écrire comme une combinaison linéaire des valeurs passées ;

$$X_t - \sum_{j=1}^{\infty} \pi_j X_{t-j} = \pi(B)X_t = \epsilon_t \text{ et } \pi(B) = 1 - \sum_{j=0}^{\infty} \pi_j B^j.$$

Un processus $ARMA$ satisfaisant (2.2.1) est inversible si et seulement si les racines de $\theta(z)$ sont toutes de module strictement supérieur à 1 et $\pi(B) = \theta(B)^{-1}\phi(B)$. Enfin, si $E[X_t] = \mu_X \neq 0$, il suffit de remplacer dans (2.2.2) X_t par $X_t - \mu_X$.

Nous pouvons résumer la modélisation d'une série chronologique selon la méthode de Box et Jenkins en trois étapes principales (Roy, 1980) :

- *Identification*, étape qui consiste à déterminer, à partir de l'analyse des comportements des autocorrélations et des autocorrélations partielles échantillonnelles, les ordres p et q du modèle *ARMA*.
- *Estimation*, que ce soit par la méthode de vraisemblance maximale ou par la méthode des moments, des paramètres autorégressifs et moyennes mobiles, de la moyenne et de la variance du processus. Concernant la méthode des moments, puisque c'est celle qu'on utilise dans cette étude, le principe est simple : Il suffit d'exprimer les paramètres en fonction des autocovariances. Par exemple, pour un *AR*(p) on a (Brockwell et Davis, 2002, section 3.2)

$$\gamma(h) - \phi_1\gamma(h-1) - \dots - \phi_p\gamma(h-p) = 0, \quad h > 0.$$

Les p premières équations nous donnent

$$\begin{bmatrix} \gamma(0) & \gamma(1) & \dots & \gamma(p-1) \\ \gamma(1) & \gamma(0) & \dots & \gamma(p-2) \\ \vdots & \vdots & \vdots & \vdots \\ \gamma(p-1) & \gamma(p-2) & \dots & \gamma(0) \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix} = \begin{bmatrix} \gamma(1) \\ \gamma(2) \\ \vdots \\ \gamma(p) \end{bmatrix}$$

que nous écrivons de façon abrégée

$$\Gamma_p \phi = \gamma_p(1) \tag{2.2.3}$$

où Γ_p représente la matrice des coefficients du système d'équations linéaires, $\phi = (\phi_1, \dots, \phi_p)'$ est le vecteur des paramètres *AR* et $\gamma_p(1) = (\gamma(1), \dots, \gamma(p))'$. Si $\sigma_x^2 = \gamma(0) \neq 0$, Γ_p est inversible et nous avons

$$\phi = \Gamma_p^{-1} \gamma_p(1). \tag{2.2.4}$$

L'estimateur par la méthode des moments de ϕ est obtenu en remplaçant dans (2.2.4), les autocovariances théoriques par les autocovariances échantillonnelles :

$$\hat{\Phi}_{mm} = \hat{\Gamma}_p^{-1} \hat{\gamma}_p(1). \quad (2.2.5)$$

L'estimateur (2.2.5) s'exprime aussi en fonction des autocorrélations échantillonnelles

$$\hat{\Phi}_{mm} = \hat{R}_p^{-1} \hat{\rho}_p(1) \quad (2.2.6)$$

où

$$\hat{R}_p = \begin{bmatrix} \hat{\rho}(0) & \hat{\rho}(1) & \dots & \hat{\rho}(p-1) \\ \hat{\rho}(1) & \hat{\rho}(0) & \dots & \hat{\rho}(p-2) \\ \vdots & \vdots & \vdots & \vdots \\ \hat{\rho}(p-1) & \hat{\rho}(p-2) & \dots & \hat{\rho}(0) \end{bmatrix}$$

et $\hat{\rho}_p(1) = (\hat{\rho}(1), \dots, \hat{\rho}(1))'$.

- *Validation* du modèle estimé par une analyse des résidus. Lorsque le modèle estimé est satisfaisant, les résidus correspondant à ce modèle se comportent essentiellement comme un bruit blanc. Les deux outils suivants sont particulièrement importants. Notons $\hat{\rho}_\varepsilon(h)$ l'autocorrélation résiduelle de délai h et soit le vecteur

$$\rho_{\varepsilon, H} = (\hat{\rho}_\varepsilon(1), \dots, \hat{\rho}_\varepsilon(H))'$$

des H premières autocorrélations résiduelles.

Sous l'hypothèse

H_0 : le modèle estimé est le vrai modèle,

si $\{\varepsilon_t\}$ est un bruit blanc fort et si les coefficients ψ_j du vrai modèle sont négligeables pour $j > H - (p + q)$, alors

$$\hat{\rho}_{\varepsilon, H} \text{ est } AN_H \left(\mathbf{0}, \frac{1}{n}(I_H - D) \right) \quad (2.2.7)$$

où I_H représente la matrice identité de dimension H et D est une matrice idempotente de rang $H - l$, l étant le nombre de paramètres *AR* et *MA*

estimés. Le symbole "AN" veut dire "asymptotiquement normal" lorsque n tends vers l'infini. De (2.2.7), il s'ensuit que les $\hat{\rho}_\varepsilon(h)$ sont asymptotiquement *iid* $N(0, \frac{1}{n})$ pour les grandes valeurs de h , d'où les limites de signification $\pm \frac{2}{\sqrt{n}}$ au niveau 5%. De (2.2.7), il découle aussi que la statistique portemanteau de Box-Pierce

$$Q_{BP}(H) = n \sum_{h=1}^H \hat{\rho}_\varepsilon(h)^2$$

est approximativement χ_{H-1}^2 et au niveau α , H_0 est rejetée si

$$Q_{BP}(H) > \chi_{H-1, 1-\alpha}^2$$

où $\chi_{m,p}^2$ désigne le quantile d'ordre p de la loi χ_m^2 .

Dans des études subséquentes à l'article de Box et Pierce, il a été remarqué que ce test conduit souvent au non rejet de l'hypothèse nulle alors que les autocorrélations résiduelles à certains délais sont importantes. Des études menées par certains auteurs notamment Davis, Triggs et Newbold (1977) et Prothero et Wallis (1976), font ressortir que la distribution asymptotique de $Q_{BP}(m)$ peut dévier considérablement de celle d'une χ_{m-1}^2 . C'est pourquoi dans le but d'améliorer l'approximation de la statistique $Q_{BP}(m)$ par une khi-deux, Ljung et Box (1978) ont proposé une modification de cette statistique qui est basée sur la variance exacte des autocorrélations d'un bruit blanc gaussien. Leur statistique notée $Q_{LB}(H)$ est définie par

$$Q_{LB}(H) = n(n+2) \sum_{h=1}^H (n-h)^{-1} \hat{\rho}_\varepsilon(h)^2.$$

Ces auteurs ont montré par des simulations que le test modifié fournit une approximation sensiblement améliorée qui pourrait être adéquate dans la plupart des cas pratiques. C'est ce test d'adéquation que nous utilisons dans notre étude.

Notons que le processus de modélisation à l'aide des modèles *ARMA* que nous venons de décrire peut être adapté à des classes plus générales de processus dont les processus *ARIMA* et *SARIMA*.

2.2.3. Processus *ARIMA*

En hydrologie, il semble raisonnable de supposer que plusieurs types de séries annuelles (sur plusieurs dizaines d'années) sont stationnaires et nous pouvons utiliser des modèles *ARMA* pour les décrire. Mais certains types de séries hydrologiques (même annuelles, Hipel et McLeod, 1994) sont non stationnaires et les modèles *ARIMA* (Autoregressive Integrated Moving Average) sont davantage appropriés pour les décrire.

Définition 2.2.5. *Un processus $ARIMA(p,d,q)$ est un processus de la forme $\{X_t : t \in N\}$ qui admet une représentation de la forme*

$$\phi(B)\nabla^d X_t = \theta_0 + \theta(B)\epsilon_t \quad (2.2.8)$$

où $\nabla = 1 - B$ est l'opérateur différence régulière, les polynômes $\phi(z)$ et $\theta(z)$ ont leurs racines à l'extérieur du cercle unité, θ_0 est une constante et $\{\epsilon_t\}$ est un bruit blanc fort $(0, \sigma_\epsilon^2)$.

Un processus *ARIMA* satisfait l'équation (2.2.8) qu'on peut aussi écrire

$$\phi^*(B)X_t = \theta(B)\epsilon_t$$

où $\phi^*(z) = \phi(z)(1 - z)^d$ est le polynôme autorégressif généralisé de degré $p + d$. Le polynôme $\phi^*(z)$ admet la racine 1 de multiplicité d et donc, $\{X_t\}$ est non stationnaire si $d \geq 1$. Notons que les modèles *ARIMA* sont utiles pour décrire des séries avec des tendances polynomiales. Cependant, la fonction moyenne d'un processus *ARIMA* n'est pas déterminée de façon unique par les équations (2.2.2) à (2.2.8) et peut être 0, une constante ou un polynôme de degré inférieur ou égal à d .

2.2.4. Processus *SARIMA*

Les modèles *SARIMA* constituent une généralisation des modèles *ARMA* et *ARIMA* permettant de décrire des séries saisonnières.

Définition 2.2.6. *On dit que le processus $\{X_t\}$ est un processus SARIMA (Seasonal AutoRegressive Integrated Moving Average) d'ordre $(p, d, q) * (P, D, Q)_s$ si ce dernier a une saisonnalité de période s et qu'on peut écrire :*

$$\phi(B)\Phi(B^s)\nabla^d\nabla_s^D X_t = \theta_0 + \theta(B)\Theta(B^s)\epsilon_t \quad (2.2.9)$$

où

$$\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p, \quad \Phi(B) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} \dots - \Phi_p B^{Ps},$$

$$\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q, \quad \Theta(B) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} \dots - \Theta_q B^{Qs}$$

θ_0 est une constante et $\{\epsilon_t\}$ est un bruit blanc fort. De plus, on suppose que les 4 polynômes $\phi(z)$, $\theta(z)$, $\Phi(z)$ et $\Theta(z)$ ont leurs racines à l'extérieur du cercle unité.

Les opérateurs $\Phi(B^s)$ et $\Theta(B^s)$ sont appelés respectivement opérateur autoregressif saisonnier et opérateur moyenne mobile saisonnier alors que $\phi(B)$ et $\theta(B)$ sont les opérateurs non saisonniers. Notons qu'un processus SARIMA $\{X_t\}$ n'est pas stationnaire, mais $\{\nabla^d\nabla_s^D X_t\}$ est stationnaire. Dans un processus SARIMA défini par (2.2.9), le modèle pour chaque saison de la période est le même et est décrit par l'équation,

$$\Phi(B^s)\nabla_s^D X_t = \Theta_0 + \Theta(B^s)\epsilon_t,$$

voir par exemple Box et Jenkins (1970, chap.9), Abraham et Ledolter (1983, chap.6).

Nous avons vu au chapitre précédent que le comportement des séries d'apports mensuels varie considérablement durant l'année, en particulier à cause de la crue printanière. Pour cette raison, on ne peut s'attendre à ce que les modèles SARIMA puissent décrire de telles séries de façon satisfaisante. Pour cette raison, plusieurs chercheurs se sont intéressés à la classe des processus périodiques. Cette classe de modèles est plus générale que la classe des modèles SARIMA dans la mesure où les modèles décrivant les données de chaque saison peuvent varier d'une saison à l'autre. Les principales propriétés des processus périodiques sont présentées dans la section qui suit.

2.3. PROCESSUS PÉRIODIQUES

Considérons le processus du second ordre $\{X_t : t \in \mathbb{Z}\}$ et posons

- $E[X_t] = \mu(t)$,
- $Cov(X_t, X_{t'}) = g(t, t'), \quad t, t' \in \mathbb{Z}$.

Définition 2.3.1. *Le processus du second ordre $\{X_t\}$ est périodique de période s si les fonctions $\mu(t)$ et $g(t, t')$ sont périodiques de période s , c'est à-dire*

$$\mu(t + s) = \mu(t), \quad (2.3.1)$$

$$g(t + s, t' + s) = g(t, t'), \quad t, t' \in \mathbb{Z}. \quad (2.3.2)$$

Il est facile de vérifier par induction que les propriétés (2.3.1) et (2.3.2) sont équivalentes à

$$\begin{aligned} \mu(t + ks) &= \mu(t), \\ g(t + ks, t' + ks) &= g(t, t'), \quad k \in \mathbb{Z}. \end{aligned}$$

Dans la définition (2.3.1), le sens "périodique" est le même que dans Shao et Lund (2004). Posons $t = rs + m$, $r \in \mathbb{Z}$ et $m \in \{1, \dots, s\}$. Avec des données mensuelles, $s = 12$, r correspond à l'année de la date t (par rapport à la date de la première observation) et m représente le mois. Par exemple, avec des données commençant en janvier, $t = 43$ s'écrit $t = 3s + 7$ et correspond au mois de juillet de la 4^e année. Afin d'alléger la notation, posons

$$E[X_{rs+m}] = E[X_m] = \mu(m) = \mu_m. \quad (2.3.3)$$

et

$$Cov(X_t, X_{t-h}) = Cov(X_{rs+m}, X_{rs+m-h}) = g(m, m-h) = \gamma_m(h) \quad (2.3.4)$$

où $m \in \{1, \dots, s\}$ et $h \in \mathbb{Z}$. Les notations $g(m, m-h)$ et $\gamma_m(h)$ sont équivalentes car la valeur de $g(m, m-h)$ est complètement déterminée par les valeurs de m et de h . Ainsi, la fonction de moyenne d'un processus périodique est complètement

déterminée par les s moyennes saisonnières μ_1, \dots, μ_s . La fonction de covariance est spécifiée par les s fonctions de covariances saisonnières $\{\gamma_m(h) : h \in \mathbb{Z}\}$, $m = 1, \dots, s$. Bien sûr, un processus périodique $\{X_t\}$ n'est pas stationnaire ni en moyenne ni en covariance. Notons cependant que d'après (2.3.4), $Cov(X_t, X_{t-h})$ dépend du mois de la date t mais ne dépend pas de l'année. Si $s = 1$, (2.3.3) et (2.3.4) correspondent à

- $E[X_t] = \mu_1$,
- $Cov(X_t, X_{t-h}) = \gamma_1(h)$, $t, h \in \mathbb{Z}$,

et nous retrouvons les conditions de stationnarité au sens large.

Un processus périodique tel que défini ici est aussi appelé "périodiquement corrélé" (Gladyshev, 1961; Hurd, 1989), "périodiquement stationnaire" (Monin, 1963) et "cyclostationnaire" (Gardner et Franks, 1975).

Dans la suite, nous noterons

$$\begin{aligned} \rho_m(h) &= Corr(X_{rs+m}, X_{rs+m-h}) \\ &= \frac{\gamma_m(h)}{\sqrt{\gamma_m(0)\gamma_{m-h}(0)}} \end{aligned} \quad (2.3.5)$$

l'autocorrélation de délai h à la saison m .

Considérons le processus multivarié dans lequel les composantes représentent les 12 observations mensuelles d'une année :

$$\mathbf{X}_r = (X_{rs+1}, \dots, X_{rs+s})', \quad r \in \mathbb{Z}.$$

Nous avons

$$E[\mathbf{X}_r] = (\mu_1, \dots, \mu_s)' = \boldsymbol{\mu}$$

et

$$Cov(\mathbf{X}_r, \mathbf{X}_{r-h}) = [\gamma_i((i-j) + hs)]_{i,j=1,\dots,s}$$

puisque

$$\begin{aligned}
 \text{Cov}(X_{rs+i}, X_{(rs-h)s+j}) &= \text{Cov}(X_{rs+i}, X_{rs+j-hs}) \\
 &= g(i, j - hs) \\
 &= g(i, i - [(i - j) + hs]) \\
 &= \gamma_i((i - j) + hs).
 \end{aligned}$$

Nous constatons que $\text{Cov}(\mathbf{X}_r, \mathbf{X}_{r-h})$ dépend de h mais ne dépend pas de r et donc le processus multivarié $\{\mathbf{X}_r\}$ est stationnaire au sens large (au sens usuel du terme) et cette dernière propriété sera fort utile dans l'étude des processus périodiques.

La notion de bruit blanc périodique va intervenir dans la suite.

Définition 2.3.2. *Le processus $\{\epsilon_t : t \in \mathbb{Z}\}$ est un bruit blanc périodique de période s si les variables aléatoires ϵ_t sont non corrélées en t avec $E[\epsilon_t] = 0$ et $\text{Var}(\epsilon_{rs+m}) = \sigma_{\epsilon, m}^2 > 0$ pour toutes les saisons $m \in \{1, \dots, s\}$.*

Un bruit blanc périodique est un cas particulier de processus périodique tel que décrit dans la définition (2.3.1). Aussi, avec $s = 1$, nous retrouvons la notion usuelle de bruit blanc faible.

Un processus périodique purement non déterministe admet une représentation $MA(\infty)$ (représentation de Wold) de la forme suivante :

$$X_{rs+m} = \sum_{k=0}^{\infty} \psi_{k,m} \epsilon_{rs+m-k} \quad (2.3.6)$$

où $\sum_{k=0}^{\infty} \psi_{k,m}^2 < \infty$, $m = 1, \dots, s$ et $\{\epsilon_t\}$ est un bruit blanc périodique $(\sigma_{\epsilon,1}^2, \dots, \sigma_{\epsilon,s}^2)$.

De (2.3.6), il découle que

$$\begin{aligned}
 \gamma_m(h) &= \text{Cov}(X_{rs+m}, X_{rs+m-h}) \\
 &= \sum_{k=0}^{\infty} \psi_{k+h,m} \psi_{k,m-h} \sigma_{\epsilon, m-k-h}^2
 \end{aligned} \quad (2.3.7)$$

Avec la notation utilisée en (2.3.6) et (2.3.7), il est implicite que les paramètres doivent être interprétés de manière périodique, c'est-à-dire que $\psi_{k,m+s} = \psi_{k,m}$ et $\sigma_{\epsilon,m+s}^2 = \sigma_{\epsilon,m}^2$, $m \in \{1, \dots, s\}$. Aussi, nous posons $\psi_{k,m} = 0$ si $k < 0$. Une classe de processus périodiques qui ne fait intervenir qu'un nombre fini de paramètres est la classe des processus *ARMA* périodiques.

2.4. PROCESSUS *ARMA* PÉRIODIQUES

D'un point de vu intuitif, un processus autorégressif-moyenne mobile périodique (*PARMA*) est un processus périodique dans lequel chaque saison est représentée par un processus *ARMA* stationnaire. Comme la fonction moyenne est périodique, nous pouvons écrire $\mu_{m+ks} = \mu_m$, $m \in \{1, \dots, s\}$ et $k \in \mathbb{Z}$.

Définition 2.4.1. *Le processus périodique $\{X_t, t \in \mathbb{Z}\}$ de période s et de vecteur moyenne $\mu = (\mu_1, \dots, \mu_s)'$ est un processus *ARMA* périodique d'ordre $(p_1, q_1; \dots; p_s, q_s)$, noté *PARMA* $(p_1, q_1; \dots; p_s, q_s)$, s'il est solution de l'équation aux différences*

$$(X_{rs+m} - \mu_m) - \phi_{1,m}(X_{rs+m-1} - \mu_{m-1}) - \dots - \phi_{p_m,m}(X_{rs+m-p_m} - \mu_{m-p_m}) = \epsilon_{rs+m} - \theta_{1,m}\epsilon_{rs+m-1} - \dots - \theta_{q_m,m}\epsilon_{rs+m-q_m}, \quad (2.4.1)$$

pour $t = rs + m \in \mathbb{Z}$, $m = 1, \dots, s$, et $\{\epsilon_t\}$ est un bruit blanc périodique $(\sigma_{\epsilon,1}^2, \dots, \sigma_{\epsilon,s}^2)$.

Les coefficients $\phi_{i,m}$, $i = 1, \dots, p_m$ et $\theta_{i,m}$, $i = 1, \dots, q_m$ sont les paramètres autorégressifs et moyennes mobiles respectivement de la m -ième saison, $m = 1, \dots, s$. En utilisant l'opérateur retard B qui agit comme à l'habitude sur $t = rs + m$ et donc sur m , l'équation (2.4.1) s'écrit de façon abrégée

$$\phi_m(B)(X_{rs+m} - \mu_m) = \theta_m(B)\epsilon_{rs+m} \quad (2.4.2)$$

où

$$\begin{aligned} \phi_m(B) &= 1 - \phi_{1,m}B - \dots - \phi_{p_m,m}B^{p_m}, \\ \theta_m(B) &= 1 - \theta_{1,m}B - \dots - \theta_{q_m,m}B^{q_m}, \end{aligned}$$

sont les opérateurs autorégressifs et moyennes mobiles respectivement correspondant à la m -ième saison, $m = 1, \dots, s$.

Posons

$$p^* = \max_{1 \leq j \leq s} \{[(p_j - j)/s] + 1\} \quad (2.4.3)$$

et

$$q^* = \max_{1 \leq j \leq s} \{[(q_j - j)/s] + 1\} \quad (2.4.4)$$

où $[x]$ désigne le plus grand entier inférieur ou égal à x . Si $p_1 = \dots = p_s = p$ et si $p \leq s$, alors $p^* = 1$; si $s < p \leq 2s$, $p^* = 2$ et de façon générale, si $(k-1)s < p \leq ks$ où $k \geq 1$ est un entier, alors $p^* = k$. Le comportement de q^* est similaire si $q_1 = \dots = q_s = q$. Dans la suite, $(A)_{i,j}$ désignera l'élément (i, j) de la matrice A . Le résultat suivant est présenté dans Vecchia (1985); voir aussi Gladyshev (1961).

Théorème 2.4.1. *Les équations (2.4.2) pour $m = 1, \dots, s$, sont équivalentes à la représentation ARMA multivariée (VARMA) suivante :*

$$\Phi_0(\mathbf{X}_r - \boldsymbol{\mu}) - \sum_{k=1}^{p^*} \Phi_k(\mathbf{X}_{r-k} - \boldsymbol{\mu}) = \sum_{k=0}^{q^*} \Theta_k \epsilon_{r-k} \quad (2.4.5)$$

où $\mathbf{X}_r = (X_{rs+1}, \dots, X_{rs+s})'$, $\epsilon_r = (\epsilon_{rs+1}, \dots, \epsilon_{rs+s})'$, $\boldsymbol{\mu} = (\mu_1, \dots, \mu_s)'$, p^* et q^* sont définis par (2.4.3) et (2.4.4). Les coefficients matriciels $\phi_i, i = 0, \dots, p^*$, et $\theta_j, j = 0, \dots, q^*$, sont donnés par

$$(\Phi_0)_{i,j} = \begin{cases} 1 & \text{si } i = j, \\ 0 & \text{si } i < j, \\ -\phi_{i-j,i} & \text{sinon,} \end{cases} \quad (\Theta_0)_{i,j} = \begin{cases} 1 & \text{si } i = j, \\ 0 & \text{si } i < j, \\ -\theta_{i-j,i} & \text{sinon,} \end{cases} \quad (2.4.6)$$

et

$$(\Phi_k)_{i,j} = \phi_{ks+i-j,i}, \quad k = 1, \dots, p^*, \quad (2.4.7)$$

$$(\Theta_l)_{i,j} = \theta_{ls+i-j,i}, \quad l = 1, \dots, q^*. \quad (2.4.8)$$

Démonstration

Pour simplifier l'écriture, supposons à nouveau que $p_1 = \dots = p_s = p$, $q_1 = \dots = q_s = q$ et que $p = p^*s$, $q = q^*s$. Les s équations (2.4.1) correspondant à $m = 1, \dots, s$ s'écrivent

$$\begin{aligned}
 \begin{pmatrix} X_{rs+1} - \mu_1 \\ X_{rs+2} - \mu_2 \\ \vdots \\ \vdots \\ X_{rs+m} - \mu_m \\ \vdots \\ X_{rs+s} - \mu_s \end{pmatrix} &= \begin{pmatrix} \phi_{1,1}(X_{rs+1-1} - \mu_s) + \dots + \phi_{p,1}(X_{rs+1-p} - \mu_{1-p}) \\ \phi_{1,2}(X_{rs+2-1} - \mu_1) + \dots + \phi_{p,2}(X_{rs+2-p} - \mu_{2-p}) \\ \vdots \\ \vdots \\ \phi_{1,m}(X_{rs+m-1} - \mu_{m-1}) + \dots + \phi_{p,m}(X_{rs+m-p} - \mu_{m-p}) \\ \vdots \\ \phi_{1,s}(X_{rs+s-1} - \mu_{s-1}) + \dots + \phi_{p,s}(X_{rs+s-p} - \mu_{s-p}) \end{pmatrix} \\
 &= \begin{pmatrix} \epsilon_{rs+1} \\ \epsilon_{rs+2} \\ \vdots \\ \vdots \\ \epsilon_{rs+m} \\ \vdots \\ \epsilon_{rs+s} \end{pmatrix} - \begin{pmatrix} \theta_{1,1}\epsilon_{rs+1-1} + \dots + \theta_{q,1}\epsilon_{rs+1-q} \\ \theta_{1,2}\epsilon_{rs+2-1} + \dots + \theta_{q,2}\epsilon_{rs+2-q} \\ \vdots \\ \vdots \\ \theta_{1,m}\epsilon_{rs+m-1} + \dots + \theta_{q,m}\epsilon_{rs+m-q} \\ \vdots \\ \theta_{1,s}\epsilon_{rs+s-1} + \dots + \theta_{q,s}\epsilon_{rs+s-q} \end{pmatrix}
 \end{aligned}$$

La preuve du résultat (2.4.5) repose sur l'écriture de chaque ligne de ce système comme une somme de $(p^* + 1)$ blocs de s -éléments (partie AR) et $(q^* + 1)$ blocs de s -éléments (partie MA). Ces deux décompositions étant similaires, nous présentons en détail uniquement la décomposition de la partie AR .

La première ligne de ce système est donnée par

$$\begin{aligned}
 (X_{rs+1} - \mu_1) - \phi_{p,1}(X_{rs+1-p} - \mu_{1-p}) - \dots - \phi_{1,1}(X_{rs+1-1} - \mu_s) \\
 = \epsilon_{rs+1} - \theta_{p,1}\epsilon_{rs+1-p} - \dots - \theta_{1,1}\epsilon_{rs+1-1}
 \end{aligned} \tag{2.4.9}$$

De cette ligne, la décomposition de la partie AR en $(p^* + 1)$ blocs de s -éléments nous donne

$$\begin{aligned}
& (X_{rs+1} - \mu_1) - \phi_{sp^*,1}(X_{rs+1-p^*s} - \mu_{1-p^*s}) - \dots - \phi_{1,1}(X_{rs+1-1} - \mu_s) \\
= & (X_{rs+1} - \mu_1) - \phi_{sp^*,1}(X_{rs+1-p^*s} - \mu_{1-p^*s}) - \dots - \phi_{(s-1)p^*+1,1}(X_{rs-(p^*-1)s} - \mu_s) - \\
& \dots - \phi_{s,1}(X_{(r-1)s+s} - \mu_{1-s}) \dots - \phi_{1,1}(X_{rs+1-1} - \mu_s) \quad (2.4.10)
\end{aligned}$$

On peut écrire l'équation (2.4.10) sous la forme

$$\begin{aligned}
\begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix} & \begin{pmatrix} X_{rs+1} - \mu_1 \\ X_{rs+2} - \mu_2 \\ \vdots \\ \vdots \\ X_{rs+s} - \mu_s \end{pmatrix} - \begin{pmatrix} \phi_{s,1}, & \dots, & \phi_{1,1} \end{pmatrix} \begin{pmatrix} X_{(r-1)s+1} - \mu_1 \\ X_{(r-1)s+2} - \mu_2 \\ \vdots \\ \vdots \\ X_{(r-1)s+s-1} - \mu_{s-1} \\ X_{(r-1)s+s} - \mu_s \end{pmatrix} \\
& - \begin{pmatrix} \phi_{2s,1}, & \dots, & \phi_{s+1,1} \end{pmatrix} \begin{pmatrix} X_{(r-2)s+1} - \mu_1 \\ X_{(r-2)s+2} - \mu_2 \\ \vdots \\ \vdots \\ X_{(r-2)s+s} - \mu_s \end{pmatrix} \\
& \vdots \\
& - \begin{pmatrix} \phi_{sp^*,1}, & \dots, & \phi_{(s-1)p^*+1,1} \end{pmatrix} \begin{pmatrix} X_{(r-p^*)s+1} - \mu_1 \\ \vdots \\ \vdots \\ \vdots \\ X_{(r-p^*)s+s} - \mu_s \end{pmatrix}.
\end{aligned}$$

De la même façon, la partie AR pour la m -ième ligne s'écrit

$$\begin{aligned}
& (X_{rs+m} - \mu_m) - \phi_{p^*s,m}(X_{rs+m-p^*s} - \mu_{m-p^*s}) - \dots - \phi_{1,m}(X_{rs+m-1} - \mu_{m-1}) \\
= & (X_{rs+m} - \mu_m) - \phi_{p^*s,m}(X_{rs+m-p^*s} - \mu_{m-p^*s}) - \dots - \phi_{m,m}(X_{rs+m-m} - \mu_{m-m})
\end{aligned}$$

$$-\phi_{m-1,m}(X_{rs+m-(m-1)} - \mu_{m-(m-1)}) - \dots - \phi_{2,m}(X_{rs+m-2} - \mu_{m-2}) - \phi_{1,m}(X_{rs+m-1} - \mu_{m-1})$$

$$= \begin{pmatrix} -\phi_{m-1,m}, & \dots, & -\phi_{1,m}, & 1, 0, \dots, 0 \end{pmatrix} \begin{pmatrix} X_{rs+1} - \mu_1 \\ X_{rs+2} - \mu_2 \\ \vdots \\ X_{rs+m-1} - \mu_{m-1} \\ X_{rs+m} - \mu_m \\ X_{rs+m+1} - \mu_{m+1} \\ \vdots \\ X_{rs+s} - \mu_s \end{pmatrix}$$

$$- \phi_{p^*s,m}(X_{(r-p^*)s+m} - \mu_m) - \dots - \phi_{m,m}(X_{rs} - \mu_s)$$

$$= \begin{pmatrix} -\phi_{m-1,m}, & \dots, & -\phi_{1,m}, & 1, 0, \dots, 0 \end{pmatrix} \begin{pmatrix} X_r - \mu \\ X_{r-1} - \mu \\ X_{r-2} - \mu \\ \vdots \\ X_{(r-p^*)s+1} - \mu_1 \\ X_{(r-p^*)s+2} - \mu_2 \\ \vdots \\ X_{(r-p^*)s+m} - \mu_m \\ \vdots \\ X_{(r-p^*)s+s} - \mu_s \end{pmatrix} \\ - \begin{pmatrix} \phi_{s+m-1,m}, & \dots, & \phi_{m,m} \end{pmatrix} \begin{pmatrix} X_{r-1} - \mu \\ X_{r-2} - \mu \\ \vdots \\ X_{(r-p^*)s+1} - \mu_1 \\ X_{(r-p^*)s+2} - \mu_2 \\ \vdots \\ X_{(r-p^*)s+m} - \mu_m \\ \vdots \\ X_{(r-p^*)s+s} - \mu_s \end{pmatrix} \\ - \begin{pmatrix} \phi_{2s+m-1,m}, & \dots, & \phi_{s+m+1,m} \end{pmatrix} \begin{pmatrix} X_{r-2} - \mu \\ \vdots \\ X_{(r-p^*)s+1} - \mu_1 \\ X_{(r-p^*)s+2} - \mu_2 \\ \vdots \\ X_{(r-p^*)s+m} - \mu_m \\ \vdots \\ X_{(r-p^*)s+s} - \mu_s \end{pmatrix} \\ + \dots$$

$$- \begin{pmatrix} 0, 0, \dots, 0, \phi_{p^*s+m-m,m}, \dots, \phi_{p^*s+m-s,m} \end{pmatrix} \begin{pmatrix} X_{(r-p^*)s+1} - \mu_1 \\ X_{(r-p^*)s+2} - \mu_2 \\ \vdots \\ X_{(r-p^*)s+m} - \mu_m \\ \vdots \\ X_{(r-p^*)s+s} - \mu_s \end{pmatrix}$$

Par hypothèse, $\phi_{j,m} = 0$ si $j > p^*s$. Ainsi au niveau de la m -ième ligne, $m = 1, \dots, s$, la partie autorégressive peut se décomposer comme suit :

$$= \begin{pmatrix} -\phi_{m-1,m}, & \dots, & -\phi_{1,m}, & 1, 0, \dots, 0 \end{pmatrix} \begin{pmatrix} X_r - \mu \\ X_{r-1} - \mu \\ X_{r-2} - \mu \\ \vdots \\ X_{(r-p^*)s+1} - \mu_1 \\ X_{(r-p^*)s+2} - \mu_2 \\ \vdots \\ X_{(r-p^*)s+m} - \mu_m \\ \vdots \\ X_{(r-p^*)s+s} - \mu_s \end{pmatrix}$$

$$\begin{aligned}
& - \left(\phi_{s+m-1,m}, \dots, \phi_{m,m} \right) \left(\mathbf{X}_{r-1} - \boldsymbol{\mu} \right) \\
& - \left(\phi_{ks+m-1,m}, \dots, \phi_{ks+m-s,m} \right) \left(\mathbf{X}_{r-k} - \boldsymbol{\mu} \right) \\
& + \dots \\
& - \left(\phi_{p^*s+m-1,m}, \dots, \phi_{p^*s+m-m,m}, \dots, \phi_{p^*s+m-s,m} \right) \left(\mathbf{X}_{r-p^*} - \boldsymbol{\mu} \right).
\end{aligned}$$

Par analogie, la partie moyenne mobile de la m -ième ligne admet la décomposition suivante :

$$\begin{aligned}
& \left(-\theta_{m-1,m}, \dots, -\theta_{1,m}, 1, 0, \dots, 0 \right) \boldsymbol{\epsilon}_r \\
& - \left(\theta_{s+m-1,m}, \dots, \theta_{m,m} \right) \boldsymbol{\epsilon}_{r-1} \\
& + \dots \\
& - \left(\theta_{q^*s+m-1,m}, \dots, \theta_{q^*s+m-m,m}, \dots, \theta_{q^*s+m-s,m} \right) \boldsymbol{\epsilon}_{r-q^*}.
\end{aligned}$$

En rassemblant les deux représentations AR et MA , les équations (2.4.2) s'écrivent

$$\Phi_0(\mathbf{X}_r - \boldsymbol{\mu}) + \sum_{k=1}^{p^*} \Phi_k(\mathbf{X}_{r-k} - \boldsymbol{\mu}) = \Theta_0 \boldsymbol{\epsilon}_r - \sum_{l=1}^{q^*} \Theta_l \boldsymbol{\epsilon}_{r-l}.$$

Les coefficients matriciels $\phi_0, \theta_0, \phi_1, \dots, \phi_{p^*}, \theta_1, \dots, \theta_{q^*}$ sont donnés par (2.4.6), (2.4.7) et (2.4.8) respectivement. Ce qui complète la démonstration.

De la présentation vectorielle (2.4.5), nous pouvons décrire les conditions sous lesquelles un processus *PARMA* admet une représentation causale et inversible.

En multipliant (2.4.5) à gauche par Φ_0^{-1} , on a

$$\mathbf{X}_r - \sum_{k=1}^{p^*} \Phi_0^{-1} \Phi_k \mathbf{X}_{r-k} = \Phi_0^{-1} \Theta_0 \boldsymbol{\epsilon}_r - \sum_{k=0}^{q^*} \Phi_0^{-1} \Theta_k \boldsymbol{\epsilon}_{r-k}$$

et en remplaçant $\Phi_0^{-1}\Phi_k$ par $\tilde{\Phi}_k$ et $\Phi_0^{-1}\Theta_0\epsilon_r$ par η_r , nous obtenons

$$\mathbf{X}_r - \sum_{k=1}^{p^*} \tilde{\Phi}_k \mathbf{X}_{r-k} = \eta_r - \sum_{k=0}^{q^*} \Phi_0^{-1} \Theta_k \Theta_0^{-1} \Phi_0 \eta_{r-k}$$

où $\{\eta_t\}$ est un bruit blanc multivarié. Posons $\tilde{\Theta}_k = \Phi_0^{-1} \Theta_k \Theta_0^{-1} \Phi_0$, alors l'équation (2.4.5) est équivalente à

$$\mathbf{X}_r - \sum_{k=1}^{p^*} \tilde{\Phi}_k \mathbf{X}_{r-k} = \eta_r - \sum_{k=0}^{q^*} \tilde{\Theta}_k \eta_{r-k} \quad (2.4.11)$$

ou encore

$$\tilde{\Phi}(B)\mathbf{X}_t = \tilde{\Theta}(B)\eta_t. \quad (2.4.12)$$

Théorème 2.4.2. *La représentation (2.4.5) est causale si*

$$\det \Phi(z) \neq 0 \quad \forall z \in \mathbb{C} \text{ tel que } |z| \leq 1$$

et est inversible si

$$\det \Theta(z) \neq 0 \quad \forall z \in \mathbb{C} \text{ tel que } |z| \leq 1$$

Démonstration

Sans perte de généralité, nous pouvons poser $\boldsymbol{\mu} = \mathbf{0}$. Selon le théorème 11.3.1 de Brockwell et Davis (1991), le processus $\{\mathbf{X}_t\}$ est stationnaire causal si

$\det \tilde{\phi}(z) \neq 0, \forall z \in \mathbb{C} \text{ tel que } |z| \leq 1$. Mais

$$\tilde{\phi}(z) = \phi_0^{-1} \phi(z) \text{ et } \det \phi_0^{-1} = (\det \phi_0)^{-1} = 1$$

par (2.4.6). Donc, nous avons

$$\det \tilde{\phi}(z) \neq 0, \iff \det \phi(z) \neq 0$$

et la condition pour la causalité est établie.

De façon similaire, par le théorème 11.3.2 de Brockwell et Davis (1991),

le processus $\{X_t\}$ est inversible si $\det \tilde{\theta}(z) \neq 0, \forall z \in \mathbb{C} \text{ tel que } |z| \leq 1$. Mais

$$\tilde{\theta}(z) = \phi_0^{-1} \theta(z) \theta_0^{-1} \phi_0 \text{ et } \det \tilde{\theta}(z) = \det \theta(z).$$

La condition pour l'inversibilité s'ensuit et la démonstration est complétée.

Ainsi nous pouvons déduire de (2.4.5), les conditions devant s'appliquer aux coefficients du *PARMA* défini par (2.4.1) afin qu'il admette une représentation causale de la forme (2.3.6) et qu'il soit inversible.

Shao et Lund (2004) donne la relation entre les coefficients $\psi_{k,m}$ de (2.3.6) et les paramètres *AR* et *MA* de (2.4.1). Dans la suite, nous supposons que $p_m = p, q_m = q, m = 1, \dots, s$. Shao et Lund (2004) montrent aussi que les autocovariances $\gamma_m(h) = Cov(X_{rs+m}, X_{rs+m-h})$ d'un *PARMA* défini par (2.4.1) satisfont une relation de récurrence similaire à celle dans le cas *ARMA* non périodique, à savoir

$$\gamma_m(h) = \sum_{k=1}^p \phi_{k,m} \gamma_{m-k}(k-h), \quad (2.4.13)$$

pour $h > \max\{p, q\}$.

2.4.1. Condition pour la causalité d'un *PAR(1)*

Nous nous intéressons maintenant aux conditions pour la causalité d'un modèle *PAR(1)* pour s quelconque :

$$X_{rs+m} = \phi_{1,m} X_{rs+m-1} + \epsilon_{rs+m}, \quad m = 1, 2, \dots, s. \quad (2.4.14)$$

Les équations (2.4.14) peuvent s'écrire sous la forme matricielle suivante

$$\begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ -\phi_{1,2} & 1 & 0 & \dots & 0 & 0 \\ 0 & -\phi_{1,3} & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -\phi_{1,s} & 1 \end{bmatrix} \begin{bmatrix} X_{rs+1} \\ X_{rs+2} \\ X_{rs+3} \\ \vdots \\ \vdots \\ X_{rs+s} \end{bmatrix} =$$

$$\begin{bmatrix} 0 & 0 & 0 & \dots & 0 & \phi_{1,1} \\ 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 0 \end{bmatrix} \begin{bmatrix} X_{(rs-1)s+1} \\ X_{(rs-1)s+2} \\ X_{(rs-1)s+3} \\ \vdots \\ \vdots \\ X_{(rs-1)s+s} \end{bmatrix} + \begin{bmatrix} \epsilon_{rs+1} \\ \epsilon_{rs+2} \\ \epsilon_{rs+3} \\ \vdots \\ \vdots \\ \epsilon_{rs+s} \end{bmatrix}, \quad (2.4.15)$$

$$\phi_0 \mathbf{X}_r = \phi_1 \mathbf{X}_{r-1} + \epsilon_r.$$

Il s'en suit que

$$\phi(z) = \Phi_0 - \Phi_1 z = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & -\phi_{1,1}z \\ -\phi_{1,2} & 1 & 0 & \dots & 0 & 0 \\ 0 & -\phi_{1,3} & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -\phi_{1,s} & 1 \end{bmatrix}$$

et en développant le déterminant selon la première ligne, nous obtenons

$$\det \phi(z) = 1 + (-1)^{s-1} (-\phi_{1,1}z) (-\phi_{1,2}) (-\phi_{1,3}) \dots (-\phi_{1,s})$$

$$= 1 - \phi_{1,1} \phi_{1,2} \phi_{1,3} \dots \phi_{1,s} z = 1 - az$$

où $a = \prod_{m=1}^s \phi_{1,m}$. Les valeurs des coefficients $\phi_{1,m}$ pour lesquelles il existe une solution causale sont

$$|a| < 1 \Leftrightarrow |\phi_{1,1} \phi_{1,2} \phi_{1,3} \dots \phi_{1,s}| < 1. \quad (2.4.16)$$

Nous déduisons de (2.4.16) qu'une représentation causale existe même si certains coefficients $\phi_{1,m}$ sont supérieurs à 1 en valeur absolue.

2.5. PROCESSUS AUTOREGRESSIFS PÉRIODIQUES

2.5.1. Motivation

Le modèle autorégressif $AR(1)$ n'est pas adéquat pour décrire des séries d'apports mensuels ou hebdomadaires car il ne tient pas compte de la variation saisonnière de la moyenne, de l'écart type, et de la fonction d'autocorrélation. C'est pour cela que le modèle autorégressif périodique $PAR(p)$ (par rapport aux autres modèles définis dans la section 2.2) est mieux adapté pour représenter ces séries d'apports. En vertu de sa paramétrisation, il permet de représenter la structure d'autocorrélation qui varie selon les saisons (Rasmussen et al., 1998). Ce type de modèle est très utilisé en hydrologie. Il est devenu populaire à la suite des travaux de Thomas et Fiering (1962) sur les modèles saisonniers et de Matalas (1967) qui a étendu le modèle de Thomas et Fiering au cas multivarié. Les principales caractéristiques des modèles PAR sont présentées plus en détails dans cette section.

2.5.2. Équations de Yule-Walker pour un $PAR(p)$

En suivant l'approche utilisée pour un modèle AR non saisonnier, on peut retrouver les équations théoriques de Yule-Walker pour un modèle PAR . Ainsi, pour $k = 1, \dots, p$ de l'équation (2.4.13), on obtient les équations périodiques de Yule-Walker pour la saison m :

$$\begin{aligned} \gamma_m(1) &= \phi_{1,m}\gamma_{m-1}(0) + \phi_{2,m}\gamma_{m-2}(1) + \dots + \phi_{p,m}\gamma_{m-p}(p-1) \\ \gamma_m(2) &= \phi_{1,m}\gamma_{m-1}(1) + \phi_{2,m}\gamma_{m-2}(0) + \dots + \phi_{p,m}\gamma_{m-p}(p-2) \\ &\vdots \\ \gamma_m(p) &= \phi_{1,m}\gamma_{m-1}(p-1) + \phi_{2,m}\gamma_{m-2}(p-2) + \dots + \phi_{p,m}\gamma_{m-p}(0) . \end{aligned}$$

Ces équations périodiques peuvent s'écrire sous la forme

$$\gamma_m(p) = \Phi_m \Gamma_{p,m} \quad (2.5.1)$$

où

$$\Phi_m = \left(\phi_{1,m} \phi_{2,m} \dots \phi_{p,m} \right)', \quad \gamma_m(p) = \left(\gamma_m(1) \gamma_m(2) \dots \gamma_m(p) \right)',$$

et

$$\Gamma_{p,m} = \begin{bmatrix} \gamma_{m-1}(0) & \gamma_{m-2}(1) & \dots & \gamma_{m-p}(p-1) \\ \gamma_{m-1}(1) & \gamma_{m-2}(0) & \dots & \gamma_{m-p}(p-2) \\ \vdots & \vdots & \vdots & \vdots \\ \gamma_{m-1}(p-1) & \gamma_{m-2}(p-2) & \dots & \gamma_{m-p}(0) \end{bmatrix}.$$

Pour $k = 0$, la variance $\gamma_m(0)$, $m = 1, \dots, s$, satisfait la relation

$$\gamma_m(0) = \phi_{1,m} \gamma_{m-1}(1) + \phi_{2,m} \gamma_{m-2}(2) + \dots + \phi_{p,m} \gamma_{m-p}(p) + \sigma_{\epsilon,m}^2$$

où $\sigma_{\epsilon,m}^2 = \text{var}(\epsilon_{rs+m})$.

2.5.3. Estimation par la méthode des moments

Nous présentons dans ce qui suit la formule générale, permettant d'estimer les paramètres d'un modèle *PAR* quelconque.

Cas général PAR(p)

De l'équation (2.5.1) nous avons

$$\Phi_m = \Gamma_{p,m}^{-1} \gamma_m(p). \quad (2.5.2)$$

Des conditions pour l'inversibilité de $\Gamma_{p,m}$ sont fournies par la proposition 5.1.1 de Brockwell et Davis (1991). L'estimateur par la méthode des moments $\hat{\phi}_m$ est obtenu en remplaçant dans (2.5.2) $\gamma_m(h)$ par $\hat{\gamma}_m(h)$, $h = 0, 1, \dots, p-1$. La variance du bruit $\sigma_{\epsilon,m}^2$ est obtenue de (2.4.13) en posant $k = 0$:

$$\sigma_{\epsilon,m}^2 = \gamma_m(0) - \sum_{k=1}^p \phi_{k,m} \gamma_m(k).$$

Modèle PAR(1)

Le modèle *PAR(1)* centré, aussi appelé le modèle périodique de Thomas et

Fiering (1962), est décrit par l'équation suivante

$$X_{rs+m} = \phi_{1,m} X_{rs+m-1} + \epsilon_{rs+m}.$$

Puisque ce modèle ne contient que 2 paramètres où m est fixée, nous considérons l'équation (2.4.13) pour $k = 0, 1$:

$$\gamma_m(0) = \phi_{1,m} \gamma_m(1) + \sigma_m^2,$$

et

$$\gamma_m(1) = \phi_{1,m} \gamma_{m-1}(0).$$

Nous trouvons

$$\phi_{1,m} = \frac{\gamma_m(1)}{\gamma_{m-1}(0)},$$

et

$$\begin{aligned} \sigma_m^2 &= \gamma_m(0) - \phi_{1,m} \gamma_m(1) \\ &= \gamma_m(0) - \frac{\gamma_m(1)}{\gamma_{m-1}(0)} \gamma_m(1) \\ &= \gamma_m(0) [1 - \rho_m^2(1)]. \end{aligned}$$

Modèle PAR(2)

Le modèle PAR(2) centré est décrit par l'équation suivante :

$$X_{rs+m} = \phi_{1,m} X_{rs+m-1} + \phi_{2,m} X_{rs+m-2} + \epsilon_{rs+m}.$$

En résolvant ces deux équations

$$\begin{cases} \phi_{1,m} \gamma_{m-1}(0) + \phi_{2,m} \gamma_{m-2}(1) = \gamma_m(1), \\ \phi_{1,m} \gamma_{m-1}(1) + \phi_{2,m} \gamma_{m-2}(0) = \gamma_m(2), \end{cases} \quad (2.5.3)$$

nous trouvons

$$\phi_{1,m} = \frac{\gamma_m(1) \gamma_{m-2}(0) - \gamma_m(2) \gamma_{m-2}(1)}{\gamma_{m-1}(0) \gamma_{m-2}(0) - \gamma_{m-1}(1) \gamma_{m-2}(1)},$$

et

$$\phi_{2,m} = \frac{\gamma_m(2) \gamma_{m-1}(0) - \gamma_m(1) \gamma_{m-1}(1)}{\gamma_{m-1}(0) \gamma_{m-2}(0) - \gamma_{m-1}(1) \gamma_{m-2}(1)}.$$

Après avoir estimé les paramètres $\phi_{1,m}$ et $\phi_{2,m}$, la variance du bruit est obtenue à partir de l'équation suivante :

$$\gamma_m(0) = \phi_{1,m}\gamma_m(1) + \phi_{2,m}\gamma_m(2) + \sigma_{\epsilon,m}^2.$$

2.5.4. Validation d'un modèle *PAR*

Pagano (1978) a démontré que les estimateurs des paramètres du modèle *PAR* pour les différentes saisons sont asymptotiquement indépendants dans le cas gaussien. La validation d'un modèle périodique peut donc être établie en examinant les propriétés des résidus pour chaque saison. Ils devraient être approximativement indépendants, normalement distribués et homoscédastiques.

2.6. MODÈLES *PAR* AVEC UNE VARIABLE EXPLICATIVE

Les résultats obtenus avec un bon modèle devraient normalement répondre au principal objectif qu'on s'est fixé au début de cette étude : construire des modèles capables de reproduire les principales caractéristiques statistiques des séries historiques. Or, il s'avère que les modèles *PAR* ne reproduisent pas de façon satisfaisante l'autocorrélation interannuelle. Pour cette raison, nous avons décidé d'inclure dans le modèle *PAR* une variable explicative qui est l'apport annuel de l'année précédente. Il s'agit alors d'un modèle de fonction de transfert dans le contexte des modèles périodiques, que nous noterons *PARX*. Il prend la forme suivante

$$X_{rs+m} = \sum_{k=1}^p \phi_{k,m} X_{rs+m-k} + B_m V_r + \epsilon_{rs+m} \quad (2.6.1)$$

V_r est l'apport total de l'année précédente, $r = 0, \dots, n-1$.

L'équation (2.6.1) peut être interprétée comme une équation de régression linéaire multiple dans laquelle les variables explicatives sont les apports des p mois précédents $X_{rs+m-1}, \dots, X_{rs+m-p}$ et V_r l'apport total de l'année $r-1$. Les

estimations des moindres carrés de ϕ_m et B_m peuvent être obtenus à l'aide d'un logiciel de régression linéaire multiple en minimisant

$$g(\phi, B_m) = \sum_{r=0}^{n-1} \sum_{m=1}^s (\epsilon_{rs+m})^2 = \sum_{r=1}^n \sum_{m=1}^s \left[X_{rs+m} - \sum_{k=1}^p \phi_{k,m} X_{rs+m-k} - B_m V_r \right]^2. \quad (2.6.2)$$

Nous ne pouvons cependant calculer les p premières erreurs $\epsilon_1, \dots, \epsilon_p$. Nous avons choisi de les poser égales à leurs moyennes, c'est-à-dire 0. Il s'agit alors de moindres carrés conditionnels. Cette approche est équivalente à utiliser les p premiers apports X_1, \dots, X_p comme valeurs initiales afin de calculer les erreurs $\epsilon_{p+1}, \dots, \epsilon_{ns}$.

De plus, afin d'obtenir des erreurs homoscédastiques, l'estimation se fera plutôt à partir des données standardisées

$$Z_{rs+m} = \frac{X_{rs+m} - \mu_{x,m}}{\sigma_{x,m}}$$

où $\mu_{x,m} = E[X_{rs+m}]$ et $\sigma_{x,m}^2 = \text{var}(X_{rs+m})$. Dans la pratique, $\mu_{x,m}$ et $\sigma_{x,m}^2$ sont estimées par $\hat{\mu}_{x,m}$ et $\hat{\sigma}_{x,m}^2$ définies respectivement par (1.2.1) et (1.2.2).

2.7. MODÈLE *PAR* MULTISITE ET CORRÉLATION SPATIALE

Dans la section 2.3, nous avons présenté des modèles *PAR* permettant de représenter une seule séquence d'apports naturels, mais l'objectif de cette étude est de modéliser les séries d'apports de plusieurs sites différents. Nous voulons avoir un modèle *PAR* multisite qui traduit aussi la cohérence spatiale des apports entre les différents sites.

2.7.1. Modèle *PAR*(p) multisite

- *Modèle général*

Nous pouvons généraliser le modèle unidimensionnel au cas multidimensionnel. Pour cela, on considère le vecteur standardisé des apports aux I sites $\mathbf{Z}_{rs+m} =$

$(Z_{rs+m}^{(1)}, \dots, Z_{rs+m}^{(I)})'$ de dimension $I \times 1$. Ainsi, le modèle $PAR(p)$ multidimensionnel s'exprime de la façon suivante :

$$Z_{rs+m} = \sum_{k=1}^p \Phi_{k,m} Z_{rs+m-k} + \epsilon_{rs+m} \quad (2.7.1)$$

où les $\Phi_{k,m}$ sont les coefficients autorégressifs matriciels. La matrice $\Phi_{k,m}$ est de dimension $I \times I$:

$$\Phi_{k,m} = \begin{pmatrix} \Phi_{k,m}^{(1)} & \Phi_{k,m}^{(1,2)} & \dots & \Phi_{k,m}^{(1,I)} \\ \Phi_{k,m}^{(2,1)} & \Phi_{k,m}^{(2)} & \dots & \Phi_{k,m}^{(2,I)} \\ \vdots & \vdots & \vdots & \vdots \\ \Phi_{k,m}^{(m,1)} & \Phi_{k,m}^{(m,2)} & \dots & \Phi_{k,m}^{(I)} \end{pmatrix} \quad (2.7.2)$$

et $\epsilon_{rs+m} = (\epsilon_{rs+m}^{(1)}, \dots, \epsilon_{rs+m}^{(I)})'$ est le vecteur des erreurs de dimension $I \times 1$.

Dans un modèle PAR multidimensionnel, la corrélation entre deux sites provient d'une part des éléments non diagonaux différents de zéro des matrices $\Phi_{k,m}$ et d'autre part des éléments non diagonaux non nuls de la matrice de variance-covariance $\sum_m = \text{var}(\epsilon_{rs+m})$. Dans un modèle $PAR(p)$, il y a potentiellement $p \times I^2$ paramètres autorégressifs différents de zéro. Une façon d'obtenir un modèle plus parcimonieux est de supposer que les coefficients $\Phi_{k,m}$ sont des matrices diagonales. Dans ce cas, la corrélation entre les sites est complètement déterminée par la matrice \sum_m .

- *Modèle $PAR(1)$ contemporain*

Nous supposons maintenant que le coefficient matriciel $\Phi_{k,1}$ de (2.7.1) est diagonal :

$$\Phi_{1,m} = \begin{pmatrix} \Phi_{1,m}^{(1)} & 0 & \dots & 0 \\ 0 & \Phi_{1,m}^{(2)} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & \dots & \Phi_{1,m}^{(I)} \end{pmatrix}. \quad (2.7.3)$$

Avec ce modèle, non seulement le nombre de paramètres est réduit de beaucoup mais le problème d'estimation du modèle multidimensionnel par les moindres carrés ordinaires revient à effectuer I estimations unidimensionnelles. Ce modèle réduit, étudié en particulier par Salas et al. (1980), est appelé modèle *PAR* contemporain ou *CPAR*. Dans un contexte comme celui-ci, la méthode des moindres carrés généralisés fournit en général des estimateurs plus efficaces mais ils n'ont pas été considérés dans ce travail. Les moindres carrés généralisés permettraient possiblement de mieux reproduire la corrélation spatiale.

- Estimation des paramètres du modèle *CPAR*

De manière analogue au cas unidimensionnel, les moyennes et les écarts types pour chaque saison m et chaque site i sont estimés par (1.2.1) et (1.2.2), respectivement.

Pour le site i , la moyenne est

$$\hat{\mu}_{x,m}^{(i)} = \frac{1}{n} \sum_{r=0}^{n-1} X_{rs+m}^{(i)} \quad (2.7.4)$$

et l'écart type est

$$\hat{\sigma}_{x,m}^{(i)} = \left[\frac{1}{n} \sum_{r=0}^{n-1} (X_{rs+m}^{(i)} - \hat{\mu}_m^{(i)})^2 \right]^{\frac{1}{2}}. \quad (2.7.5)$$

Pour les paramètres autorégressifs de la matrice $\Phi_{k,m}$, ils sont obtenus en appliquant indépendamment la méthode des moments site par site. A l'issue de cet exercice, nous disposons donc pour chaque saison m de l'estimation

$$\hat{\Phi}_{1,m} = \begin{pmatrix} \hat{\Phi}_{1,m}^{(1)} & 0 & \dots & 0 \\ 0 & \hat{\Phi}_{1,m}^{(2)} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & \dots & \hat{\Phi}_{1,m}^{(I)} \end{pmatrix},$$

$m = 1, \dots, s.$

2.7.2. Estimation de la matrice de covariance spatiale

Le vecteur des erreurs pour l'ensemble des sites $\epsilon_{rs+m} = (\epsilon_{rs+m}^{(1)}, \dots, \epsilon_{rs+m}^{(I)})'$ peut être estimé par le vecteur des résidus du modèle

$$\begin{aligned}\hat{\epsilon}_{rs+m} &= (\hat{\epsilon}_{rs+m}^{(1)}, \dots, \hat{\epsilon}_{rs+m}^{(I)})' \\ &= \mathbf{Z}_{rs+m} - \sum_{k=1}^p \hat{\Phi}_{k,m} \mathbf{Z}_{rs+m-k}\end{aligned}\quad (2.7.6)$$

où $\hat{\epsilon}_{rs+m}^{(i)}$, $r = 0, \dots, n-1$, est le résidu du site i pour la saison m de l'année $r+1$.

Pour l'estimation de \sum_m , nous avons retenu l'approche utilisée par Perreault et Latraverse (2001) qui consiste, pour chaque saison m , à appliquer la transformation logarithmique à chaque composante du vecteur. Le vecteur des résidus est noté $\mathbf{R}_{rs+m} = (\log \hat{\epsilon}_{rs+m}^{(1)}, \dots, \log \hat{\epsilon}_{rs+m}^{(I)})'$. Cette transformation s'appuie sur le fait que les données d'apport ne suivent pas en général une loi normale comme on l'a vu au chapitre 1 et plusieurs auteurs ont préféré la loi lognormale à la loi normale pour la distribution des erreurs. La matrice \sum_m est alors estimée par

$$\hat{\sum}_m = \frac{1}{n} \sum_{r=0}^{n-1} (\mathbf{R}_{rs+m} - \bar{\mathbf{R}}_m)(\mathbf{R}_{rs+m} - \bar{\mathbf{R}}_m)' \quad (2.7.7)$$

$$\text{où } \bar{\mathbf{R}}_m = \frac{1}{n} \sum_{r=0}^{n-1} \mathbf{R}_{rs+m}.$$

Sous des conditions générales, nous savons que $\hat{\sum}_m$ est un estimateur convergent de \sum_m , même si les observations ne suivent pas des lois normales ou lognormales. Dans le modèle *CPAR*, la corrélation spatiale (entre les sites) est entièrement décrite par les corrélations croisées dans \sum_m .

2.8. GÉNÉRATION DE RÉALISATIONS DU MODÈLE *CPARX*

Aucune transformation n'est appliquée aux données historiques. Ceci est dû, comme nous l'avons déjà mentionné, à la contrainte de linéarité imposée par SDDP lors de l'optimisation. Nous pensons qu'il est judicieux d'effectuer des transformations sur les résidus au lieu des apports. Cette modification de l'approche traditionnelle (transformer les observations) permet d'éviter l'emploi d'une

relation non linéaire tout en préservant approximativement la distribution marginale des apports (Perreault et Latraverse, 2001).

Dans cette étude, nous allons utiliser deux approches pour générer des vecteurs d'erreurs. L'approche paramétrique, basée sur la loi lognormale à trois paramètres et une approche non paramétrique en utilisant une technique de bootstrap. L'approche paramétrique utilisée dans Perreault et Latraverse (2001) repose sur l'hypothèse que le vecteur \mathbf{R}_{r_s+m} suit une loi multinormale $N_I(\mathbf{0}, \Sigma_m)$. Des vecteurs d'erreurs indépendants et identiquement distribués sont alors générés de la loi multinormale $N_I(\mathbf{0}, \Sigma_m)$.

La deuxième approche consiste à rééchantillonner de façon aléatoire et avec remise les vecteurs d'erreurs parmi les n vecteurs de résidus ϵ_{r_s+m} , $r = 0, 1, \dots, n-1$. Cette approche est tout à fait non paramétrique dans le sens qu'elle ne repose sur aucune hypothèse distributionnelle sur les erreurs. Elle est conceptuellement plus simple et aussi plus facile à implanter.

2.8.1. Transformation des erreurs

On emploie souvent la transformation logarithmique parce qu'elle prend en compte la non-linéarité de la relation entre les apports mais aussi préserve lors de la simulation la distribution de probabilité qui peut être non normale. Dans la présente étude, nous allons considérer la loi lognormale à 3 paramètres pour ϵ_{r_s+m} dont la fonction de densité de probabilité est donnée par l'équation suivante :

$$f(x : m_m, \alpha_m, \beta_m) = \frac{1}{(x - m_m)\beta_m\sqrt{2\pi}} \exp \left[-\frac{1}{2} \left[\frac{\log(x - m_m) - \alpha_m}{\beta_m} \right]^2 \right].$$

On dispose donc de n réalisations à partir desquelles il est possible d'estimer les paramètres α_m , β_m et m_m de la loi lognormale pour la saison m . Les méthodes des moments et du maximum de vraisemblance sont les plus employées pour estimer ces paramètres. Dans le cas de la méthode des moments utilisée dans notre étude, on peut montrer (Stedinger, 1980) que ces estimateurs sont donnés par

$$\hat{m}_m = \bar{r}_m - \frac{s_m}{\sqrt{\hat{\lambda} - 1}}, \quad \hat{\alpha}_m = \log \left(\frac{s_m}{\sqrt{\hat{\lambda}(\hat{\lambda} - 1)}} \right), \quad \hat{\beta}_m = \sqrt{\log(\hat{\lambda})}$$

où

$$\hat{\lambda} = \left[1 + \frac{C_s}{2} \left(C_s + \sqrt{4 + C_s^2} \right) \right]^{\frac{1}{3}} + \left[1 + \frac{C_s}{2} \left(C_s - \sqrt{4 + C_s^2} \right) \right]^{\frac{1}{3}} - 1,$$

\bar{r}_m est la moyenne arithmétique des résidus $\hat{\epsilon}_{rs+m}$, $r = 0, \dots, n-1$, pour la saison m , s_m l'écart-type et C_s désigne le coefficient d'asymétrie.

Aussi, on peut estimer l'espérance \mathbb{E}_{LN} et la variance \mathbb{V}_{LN} de cette distribution lognormale à partir des expressions suivantes :

$$\hat{\mathbb{E}}_{LN} = \frac{1}{n} \sum_{r=1}^n \log(\hat{\epsilon}_{rs+m} - \hat{n}_m), \quad \hat{\mathbb{V}}_{LN} = \frac{1}{n} \sum_{r=1}^n \left[\log(\hat{\epsilon}_{rs+m} - \hat{n}_m) - \hat{\mathbb{E}}_{LN} \right]^2. \quad (2.8.1)$$

Ce sont ces deux quantités que nous allons utiliser lors de la génération des apports. L'algorithme utilisé est décrit à la figure 3.2 dans Perreault et Latraverse (2001).

2.8.2. Rééchantillonnage des vecteurs de résidus

L'article d'Efron (1979) a été un évènement majeur en statistique. C'est ce dernier qui a introduit les méthodes de rééchantillonnage appelées aussi bootstrap. La méthode de rééchantillonnage utilisée dans cette étude est simple. Elle suit une approche totalement différente de celle présentée auparavant. C'est une méthode d'estimation non paramétrique basée sur des techniques de rééchantillonnage des observations, c'est-à-dire qu'on effectue des tirages aléatoires avec remise parmi les vecteurs de résidus afin de constituer un nouvel échantillon de vecteurs de résidus pour les I sites. En rééchantillonnant des vecteurs de résidus, nous espérons préserver la dépendance spatiale entre les apports.

Cette méthode est mise en oeuvre afin de vérifier si elle permet de mieux reproduire la corrélation spatiale et la corrélation interannuelle des apports simulés. Efron et Tibshirani (1993) présentent une description générale des méthodes bootstrap.

Afin d'illustrer la mise en oeuvre de la méthode, considérons un modèle $PAR(1)$ pour chacun des I sites en utilisant les apports standardisés. Pour le

site i à la saison m , les résidus sont donnés par

$$\hat{\epsilon}_{rs+m}^{(i)} = Z_{rs+m}^{(i)} - \hat{\phi}_{1,m} Z_{rs+m-1}^{(i)},$$

$r = 0, 1, \dots, n-1$. L'estimation du modèle $PAR(1)$ à chacun des I sites permet de former les vecteurs de résidus $\hat{\epsilon}_{rs+m} = (\hat{\epsilon}_{rs+m}^{(1)}, \dots, \hat{\epsilon}_{rs+m}^{(I)})'$, $r = 0, 1, \dots, n-1$.

S'il y a de la corrélation spatiale entre les apports aux différents sites, on peut espérer que ces corrélations ressortent à travers les composantes des vecteurs de résidus $\hat{\epsilon}_{rs+m}$. Au lieu de générer des vecteurs d'erreurs de la loi $N_I(\mathbf{0}, \hat{\Sigma}_m)$, nous choisisons au hasard et avec remise n vecteurs parmi $\hat{\epsilon}_{rs+m}$, $r = 0, 1, \dots, n-1$, que nous notons

$$\epsilon_m^*, \epsilon_{s+m}^*, \dots, \epsilon_{(n-1)s+m}^*.$$

Ainsi, pour le site i , en utilisant les erreurs $\epsilon_{rs+m}^{(i)*}$, $r = 0, 1, \dots, n-1$, nous obtenons une réalisation bootstrap à l'aide de l'équation

$$Z_{rs+m}^{(i)*} = \hat{\phi}_m Z_{rs+m-1}^{(i)*} + \epsilon_{rs+m}^{(i)*}, \quad (2.8.2)$$

$r = 0, 1, \dots, n-1$. Comme valeurs initiales, nous avons utilisés les apports standardisés de l'année 1960. Comme dans la méthode basée sur la loi lognormale, l'approche bootstrap permet de générer autant de réalisations qu'on le souhaite. Il reste toutefois à établir théoriquement la validité d'une telle approche. Ceci déborde du cadre de ce mémoire.

Chapitre 3

APPLICATION DES MODÈLES *PAR* ET *PARX* AUX APPORTS

Dans ce chapitre, nous allons mettre en oeuvre les modèles *PAR* et *PARX* présentés dans le chapitre précédent pour reproduire les caractéristiques statistiques des apports observés. Nous présentons seulement les résultats issus d'un *PAR*(1) et *PARX*(1) puisque ce sont les modèles qui ont donné les meilleurs résultats lors des simulations, par rapport aux modèles *PAR*(2) et *PAR*(3). Notons aussi que notre programme de simulation fonctionne quel que soit l'ordre p .

En séries chronologiques, la validation peut se faire par l'analyse des résidus, en utilisant dans ce cas là une multitude de tests ainsi que des procédures graphiques. En hydrologie, l'approche est différente. Elle consiste à simuler un certain nombre de scénarios et à vérifier si ceux-ci reproduisent en moyenne le comportement de l'échantillon observé. Le modèle retenu sera celui qui reproduit le mieux les caractéristiques statistiques, la dépendance interannuelle et la corrélation spatiale des apports observés. Ce sont ces deux approches que nous allons utiliser dans cette étude.

Dans la section 2.8, nous avons présenté deux approches pour le traitement des résidus, une approche paramétrique basée sur la loi lognormale des résidus et une autre approche purement non paramétrique qui consiste à rééchantillonner

les résidus.

Dans la suite de notre étude, nous utiliserons l'expression "méthode transformée", pour désigner la série d'apports simulés, obtenue après la transformation basée sur la loi lognormale des résidus et l'expression "méthode de rééchantillonnage" pour désigner la série d'apports simulés, obtenue après rééchantillonnage des résidus. Nous allons comparer ces deux méthodes, et voir celle qui reproduit le mieux les corrélations spatiales. Notons que le but de notre étude n'est pas de faire des études de prévision comme c'est généralement le cas en séries chronologiques, mais l'objectif consiste ici à choisir des modèles qui permettent de reproduire au mieux les caractéristiques statistiques des observations historiques.

Lors de la mise en oeuvre des modèles périodiques, nous allons utiliser des données mensuelles. La seule raison est que pour les données hebdomadaires, nous avons des résultats sur 52 semaines donc en considérant les 17 réservoirs, le modèle périodique contient $17 \times 52 \times p$ paramètres. Ainsi, pour ne pas alourdir le texte nous avons choisi les séries mensuelles. Notons cependant que le programme en Matlab est standard, il fonctionne aussi bien avec des données mensuelles que des données hebdomadaires.

L'estimation des paramètres et la génération des apports ont été réalisées à l'aide du progiciel Matlab version 7. Pour les représentations graphiques, nous avons utilisé le visualisateur d'Hydro-Québec, ce logiciel est très convivial et permet de visualiser les résultats des simulations sous forme de tableaux et de représentations graphiques.

Toutes les routines Matlab utilisées dans notre mémoire sont fournies en annexe ou sur une disquette.

3.1. PLAN DES SIMULATIONS

Dans cette section, nous présentons les algorithmes qui ont été mis en oeuvre pour estimer les paramètres ainsi que le plan de génération des apports.

3.1.1. Algorithme d'estimation pour le modèle *PAR*

- Pour un réservoir i avec $i = 1, \dots, 17$, le programme se compose des étapes suivantes :

- (1) Lire les paramètres nécessaires à la simulation : $n = 42$, le nombre d'années, $s = 12$, le nombre de saisons, $p = 1$, le nombre de termes autorégressifs et $I = 17$, le nombre de réservoirs.
- (2) Traiter la série d'apports mensuels observés en la mettant sous forme de vecteurs structurés.
- (3) Calculer pour $m = 1, \dots, s$, les moyennes $\hat{\mu}_{x,m}^{(i)}$ et les écarts types $\hat{\sigma}_{x,m}^{(i)}$ définis par (2.7.4) et (2.7.5).
- (4) Standardiser les données afin d'éliminer les variations saisonnières dans les distributions marginales. Ainsi pour le réservoir i , $i = 1, \dots, I$, $Z_{rs+m}^{(i)}$ représente l'apport standardisé de l'année r , $r = 0, 1, \dots, n - 1$ et de la saison m , $m = 1, \dots, s$:

$$Z_{rs+m}^{(i)} = \frac{X_{rs+m}^{(i)} - \hat{\mu}_{x,m}^{(i)}}{\hat{\sigma}_{x,m}^{(i)}}.$$

- (5) Estimer les ps paramètres autorégressifs par la méthode des moments définie par (2.5.2).
- (6) Calculer les ns résidus ϵ_{rs+m} , $r = 0, 1, \dots, n - 1$, $m = 1, \dots, s$, définis par (2.7.6)
- (7) Transformer les résidus :
 - a.1 - Transformation basée sur la loi lognormale.
 - a.2 - Calculer la matrice de covariance des résidus $\hat{\Sigma}_m$ définie par (2.7.7).
 - b - Construire des échantillons bootstrap en effectuant des tirages aléatoires avec remise sur les vecteurs des résidus définis par (2.8.2).

3.1.2. Algorithme d'estimation pour le modèle *PARX*

Le programme se compose des mêmes étapes que celui du modèle *PAR*. En plus de cela on doit :

- (1) Calculer l'apport total V_r de l'année précédente, défini par

$$V_r = \sum_{\tau=0}^{n-1} X_{(r-1)s+m}, \quad r = 0, \dots, n-1, \quad m = 1, \dots, s.$$

- (2) Estimer les ps paramètres autorégressifs par la méthode des moindres carrés et les s paramètres B_m définis par (2.6.2).

3.1.3. Algorithme de génération d'apports pour le modèle *PAR* et *PARX*

Dépendamment du modèle choisi, on effectue les étapes suivantes lors de la génération des apports :

- (1) Période de chauffe : 1 année
 - (a) Générer 1 scénario de résidus pour la saison courante m , pour les I réservoirs.
 - (b) Calculer les apports simulés correspondants.
- (2) Pour chaque scénario k , $k = 2, \dots, N$, et $N = 100$:
 - (a) Générer n vecteurs multinormaux indépendants de résidus.
 - (b) Utiliser l'une des deux méthodes de transformation des résidus pour la prise en compte de la dépendance spatiale.
 - (i) Dans le cas de la transformation lognormale
 - Introduire la corrélation spatiale $\hat{\Sigma}_m$ définie par (2.7.7) dans la fonction génératrice.
 - Transformation des résidus normaux en résidus lognormaux à 3 paramètres. Les termes d'erreurs $\epsilon_{rs+m}^{(i)}$ d'une saison m peuvent être approchés par les résidus du modèle $\hat{\epsilon}_{rs+m}^{(i)}$, c'est-à-dire

$$\hat{\epsilon}_{rs+m}^{(i)} = \exp(\hat{E}_{LN}^{(i)} + (\hat{V}_{LN}^{(i)})^{1/2} \mu_m^{(i)}) + \hat{m}_m^{(i)},$$
 où $\hat{E}_{LN}^{(i)}$ et $(\hat{V}_{LN}^{(i)})^{1/2}$ sont définies par (2.8.1).
 - (ii) Dans le cas du rééchantillonnage des résidus :
 - Introduire les vecteurs obtenus par rééchantillonnage définie par (2.8.2) dans la fonction génératrice.

- (3) Calculer l'apport standardisé en fonction du ou des apports standardisés aux périodes précédentes, c'est-à-dire

$$Z_{rs+m}^{(i)} = \sum_{k=1}^p \hat{\phi}_{k,m}^{(i)} Z_{rs+m-k}^{(i)} + \hat{\epsilon}_{rs+m}^{(i)}$$

- (4) Déstandardiser les apports générés, c'est-à-dire

$$X_{rs+m}^{(i)} = \hat{\mu}_{x,m}^{(i)} + \hat{\sigma}_{x,m}^{(i)} Z_{rs+m}^{(i)}$$

- (5) Nouveau vecteur d'apports du réservoir i pour la saison m .
 (6) Générer le scénario suivant, etc.

3.2. MISE EN OEUVRE DU MODÈLE $PAR(1)$

Pour alléger le texte et bien cibler à quel niveau se feront nos comparaisons, nous présentons seulement les résultats des simulations pour les réservoirs Caniapiscou, Churchill Falls et Gouin. Les données utilisées lors de la mise en oeuvre du modèle PAR sont celles fournies par l'unité PQDH (Prévisions et Qualité des Données Hydroélectriques). Les caractéristiques de ces données observées ont déjà été présentées dans la section 1.1 du chapitre 1.

3.2.1. Modèle estimé $PAR(1)$ et test d'adéquation de Ljung et Box

Les principaux résultats de l'application du test de Ljung et Box sur les résidus sont présentés dans les tableaux 3.1 et 3.2. On remarque que le modèle est adéquat pour les réservoirs Caniapiscou et Gouin, puisque le test d'adéquation de Ljung et Box pour un modèle $PAR(1)$ donne des valeurs- p toutes supérieures à 5% pour les délais $h = 1, 3, 5, 10$. Cependant, pour Churchill Falls, aux délais 5 et 10 pour les mois de février, mars et juillet, le test de Ljung et Box ($Q_{LB}(m)$) donne des niveaux de signification expérimentaux inférieurs à 5%; ce qui veut dire que le modèle choisi n'est pas adéquat. Par contre pour les délais 1 et 3 pour toutes les périodes, le test d'adéquation de Ljung et Box montre que le modèle $PAR(1)$ est adéquat pour ce type de séries.

TABLEAU 3.2. Estimation et validation du modèle $PAR(1)$ pour le réservoir Gouin

Gouin	Coefficients AR	0,52	0,66	0,45	0,31	-0,22	0,57	0,37	0,58	0,57	0,52	0,61	0,60	
	Variances résiduelles	0,000	0,044	0,026	0,090	0,007	0,049	0,122	0,059	0,006	0,002	0,087	0,032	
	Test Ljung-Box :													
	h=1	1,75	1,24	0,00	0,10	0,23	0,00	1,85	4,17	0,31	0,14	0,013	0,10	
		Q_{LB}												
		Valeur-p	0,19	0,27	0,95	0,75	0,63	0,99	0,17	0,04	0,58	0,71	0,91	0,75
	h=3	Q_{LB}	3,95	1,3	0,90	4,33	1,63	0,02	2,36	5,60	1,09	0,43	2,23	3,31
		Valeur-p	0,27	0,73	0,83	0,23	0,65	0,99	0,50	0,13	0,78	0,93	0,52	0,35
	h=5	Q_{LB}	4,44	1,60	2,07	4,52	2,47	0,10	2,84	9,61	4,33	4,32	7,09	5,44
		Valeur-p	0,49	0,90	0,84	0,48	0,78	0,99	0,72	0,09	0,50	0,50	0,21	0,36
	h=10	Q_{LB}	5,56	12,10	5,18	11,36	4,56	2,76	8,60	12,91	8,13	5,63	10,14	7,26
		Valeur-p	0,85	0,28	0,88	0,33	0,92	0,99	0,57	0,23	0,62	0,84	0,43	0,70

3.2.2. Analyse des hydrogrammes simulés

Les figures 3.1 à 3.3 donnent une représentation graphique des hydrogrammes mensuels historiques et simulés des 3 réservoirs. Ces hydrogrammes ont été construits à partir des apports mensuels (en $m^3/seconde$) des réservoirs au cours des 12 mois. Les données historiques ont été observées sur 42 années (1961-2002). Le premier mois de l'année est janvier et le dernier est décembre. On dispose donc de 42 hydrogrammes historiques. Par contre, les hydrogrammes simulés ont été obtenus à partir de 100 réalisations de 42 années. Ce qui donne 4200 hydrogrammes.

Deux types d'hydrogrammes simulés sont présentés dans cette étude, ceux calculés à partir des résidus transformés et ceux calculés à partir du rééchantillonnage des résidus. Notons que le visualisateur n'affiche pas les hydrogrammes au complet. Pour chaque période $m = 1, \dots, s$, il prend le minimum et le maximum de l'apport pour chaque période et il affiche l'hydrogramme au complet, donc pour chaque saison, nous avons 2 hydrogrammes. Ce qui donne 24 hydrogrammes pour les 12 saisons. Notons que les hydrogrammes extrêmes intéressent beaucoup les hydrologues puisque, par exemple, cela leur permet d'estimer de façon conservatrice les dimensions des évacuateurs de crue lors de la construction de nouveaux barrages hydro-électriques.

L'analyse des hydrogrammes comparatifs mensuels des figures 3.1 à 3.3 pour les réservoirs Caniapiscau, Churchill Falls et Gouin montre une certaine variation entre les hydrogrammes simulés et les hydrogrammes observés. Aussi, malgré que les pointes de crue diffèrent d'une année à une autre, elles sont très bien reproduites pour les 3 réservoirs où on voit bien que la crue printanière survient généralement entre le 5^e et 7^e mois comme c'est le cas pour les hydrogrammes historiques.

Si on compare les deux méthodes, on observe bien que les hydrogrammes obtenus à partir de la méthode de rééchantillonnage reproduisent assez bien les formes des hydrogrammes observés comparativement à la méthode transformée et ceci pour les 3 réservoirs. La période couvrant les 9^e, 10^e et 11^e mois qui fait référence

à l'automne, est caractérisée par une augmentation des débits due aux fortes précipitations. Cette période représente la crue automnale. Nous remarquons durant cette période une grande variabilité dans le cas de la méthode transformée alors que dans la méthode par rééchantillonnage, la variation est moins importante.

Ainsi, en se basant sur ces résultats, on peut dire que les hydrogrammes obtenus à partir de la méthode de rééchantillonnage sont plus près des hydrogrammes historiques. Il semble donc que l'approche par rééchantillonnage permet de reproduire des hydrogrammes plus réalistes pour décrire les apports en eau aux différents réservoirs.

Nous remarquons dans les figures 3.1, 3.2 et 3.3 que les deux méthodes de génération d'hydrogrammes conduisent à l'occasion à des valeurs négatives. Cette carence des hydrogrammes simulés est vraisemblablement due aux estimations négatives de certains coefficients autorégressifs. En effet, nous remarquons dans les tableaux 3.1 et 3.2 la présence de trois coefficients négatifs : celui de juin pour Caniapiscau, celui de juillet pour Churchill Falls et celui de mai pour le réservoir Gouin. Une façon d'éviter les coefficients négatifs est de faire de l'estimation sous contraintes, une approche fort complexe d'un point de vue numérique.

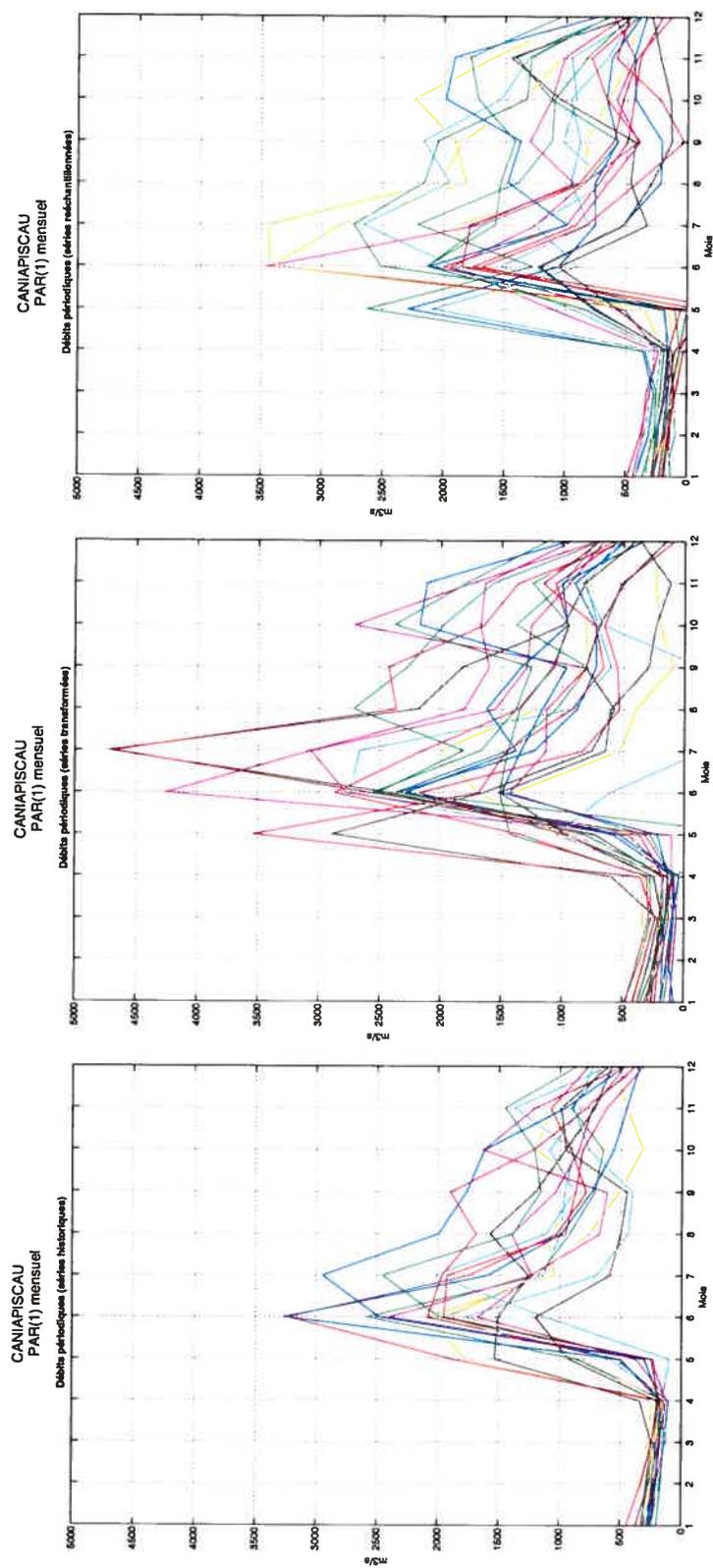


FIGURE 3.1. Hydrogrammes comparatifs pour les séries historiques, transformées et rééchantillonnées au réservoir Caniapiscau.

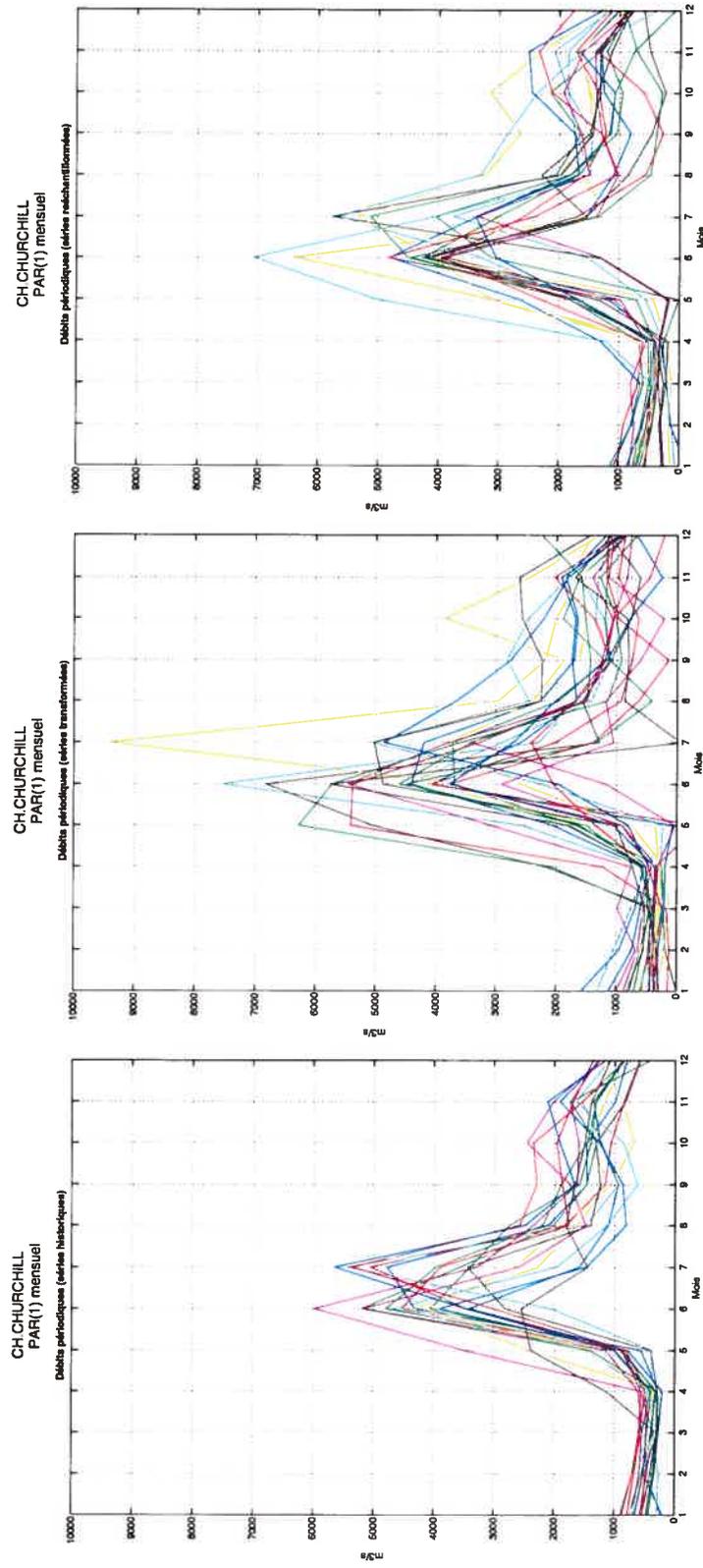


FIGURE 3.2. Hydrogrammes comparatifs pour les séries historiques, transformées et rééchantillonnées au réservoir Churchill Falls.

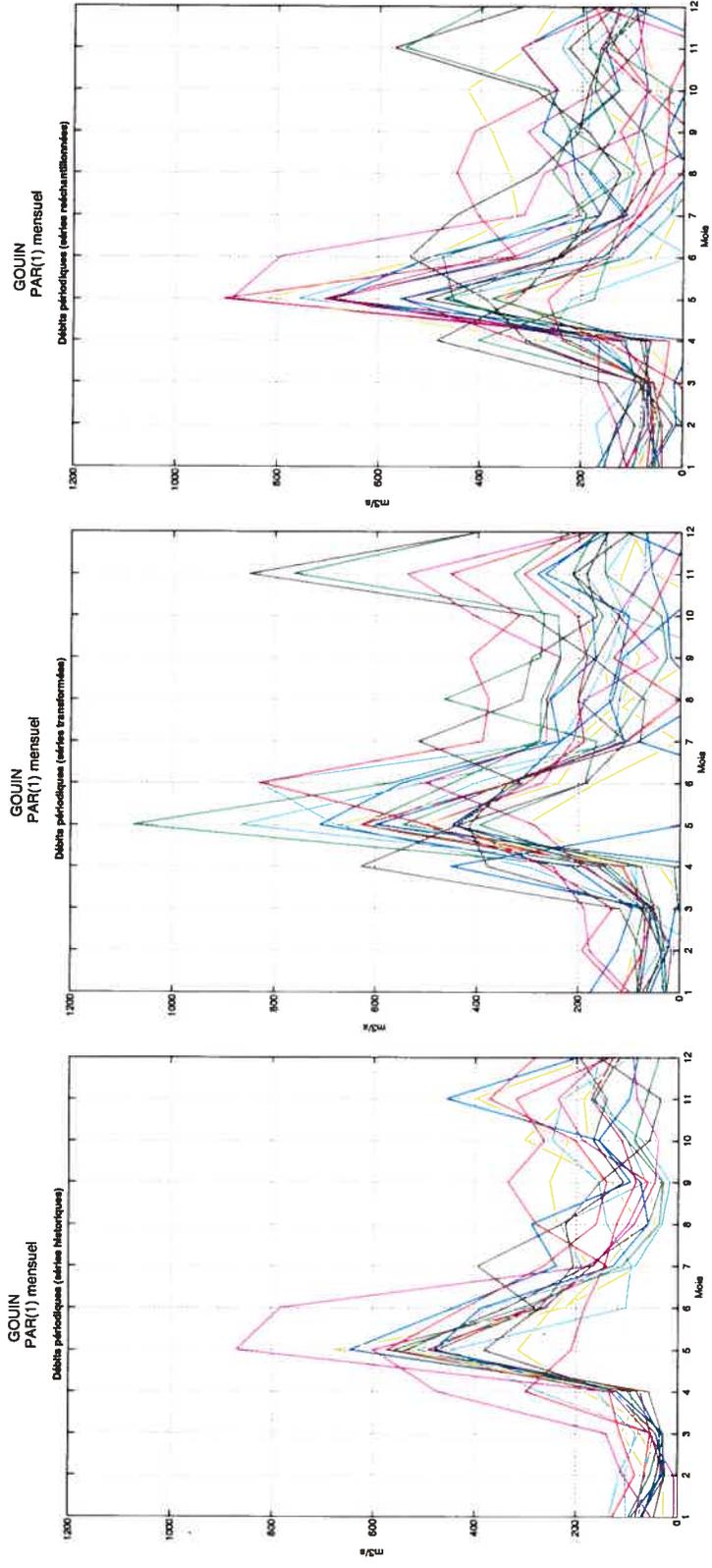


FIGURE 3.3. Hydrogrammes comparatifs pour les séries historiques, transformées et rééchantillonnées au réservoir Gouin.

3.2.3. Discussion des résultats de simulation dans le cas d'un modèle $PAR(1)$

- Moyennes

Le tableau 3.3 donne un sommaire de la comparaison des deux méthodes étudiées pour le modèle $PAR(1)$. Pour la méthode de rééchantillonnage, il ressort des résultats du tableau 3.3 que l'apport moyen annuel observé est bien reproduit. En effet, l'apport moyen obtenu par simulation pour le réservoir Caniapiscou (767, 52) est proche de l'apport moyen observé (768, 03), de même que pour Churchill Falls (1419, 26) et Gouin (175, 70) pour un apport moyen observé respectif de 1422, 4 et 175, 26.

Malgré une légère surestimation de la moyenne annuelle pour les réservoirs Caniapiscou (771, 57), Churchill Falls (1430, 02) et Gouin (174, 67), la méthode transformée performe aussi bien que la méthode par rééchantillonnage. Ainsi d'après le tableau 3.3, nous pouvons dire que les deux méthodes reproduisent assez bien la moyenne annuelle des apports.

Les diagrammes en boîte des figures 3.4 à 3.6, ont été construits à partir de 100 scénarios, où pour chaque scénario k , pour une saison m , l'apport moyen défini par l'équation (1.2.1) est calculé à partir des 42 années. Ainsi, pour chaque scénario k , nous calculons la moyenne sur les 42 années et nous obtenons \bar{x}_m^k , $k = 1, 2, \dots, 100$, et le diagramme en boîte est obtenu à partir de ces 100 moyennes.

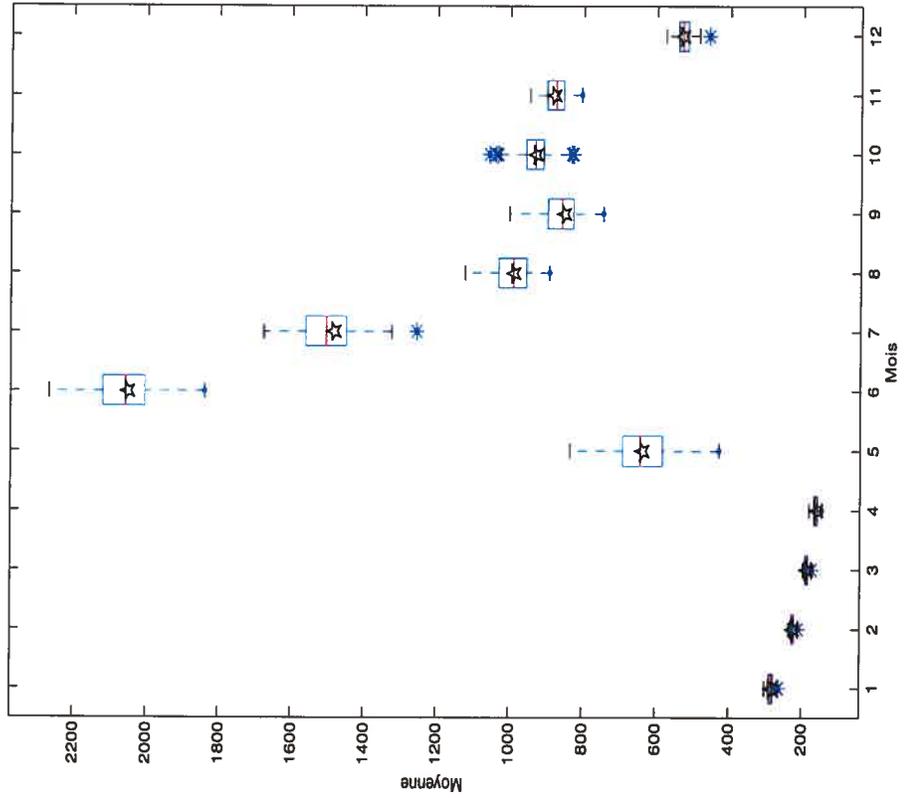
L'étoile (★) qui apparaît à chaque saison dans les figures 3.4 à 3.6 constitue l'apport moyen observé. D'après les figures 3.4 à 3.6, on remarque que la moyenne observée est proche de la médiane des 100 moyennes simulées pour chaque période, ce qui indique que l'on reproduit bien les moyennes à chaque saison puisque l'étoile qui constitue l'historique est à l'intérieur de l'intervalle interquartile et ceci pour les 3 réservoirs.

TABLEAU 3.3. Résultats comparatifs des statistiques annuelles observées et simulées d'un modèle PAR(1) pour les différents réservoirs d'Hydro-Québec.

Code	Sites	Historique		Transformée		Réchantionnée	
		moyenne	écart type	moyenne	écart type	moyenne	écart type
1	Caniapiscau	768,03	131,93	771,57	139,73	767,52	136,04
2	Laforge	163,73	27,11	164,18	28,61	164,13	27,68
3	LaGrande4	556,24	91,64	557,77	100,44	164,13	96,81
4	EOL Total	850,32	118,54	851,96	127,45	850,62	123,44
5	Bersimis	318,10	40,30	318,83	48,30	317,91	43,83
6	Churchill Falls	1422,44	223,41	1430,02	223,14	1419,26	214,96
7	Manic 5	646,89	96,40	646,30	94,62	646,25	94,33
8	Lac Ste-Anne	191,63	25,12	191,72	28,11	191,70	27,85
9	Outardes 4	364,46	46,62	363,94	51,62	365,10	50,90
10	Ste-Marguerite	108,93	19,34	108,96	16,89	109,52	16,66
11	Gouin	175,26	29,15	174,67	35,31	175,70	34,78
12	Manouane	45,67	9,23	46,07	10,72	45,94	9,93
13	Rapide Blanc	167,82	32,32	173,84	44,77	168,05	35,13
14	Mattawin	73,88	13,74	74,38	15,43	73,95	14,70
15	TR-BE-LT2	173,88	28,26	180,26	35,86	174,22	26,68
16	GM-SH-Lagab	106,92	29,92	107,87	31,97	107,20	31,07
17	LG2-SAK	977,58	172,58	981,01	186,19	979,09	176,23

CANIAPISCAU

Moyenne mensuelle (séries transformées)



Moyenne mensuelle (séries rééchantillonnées)

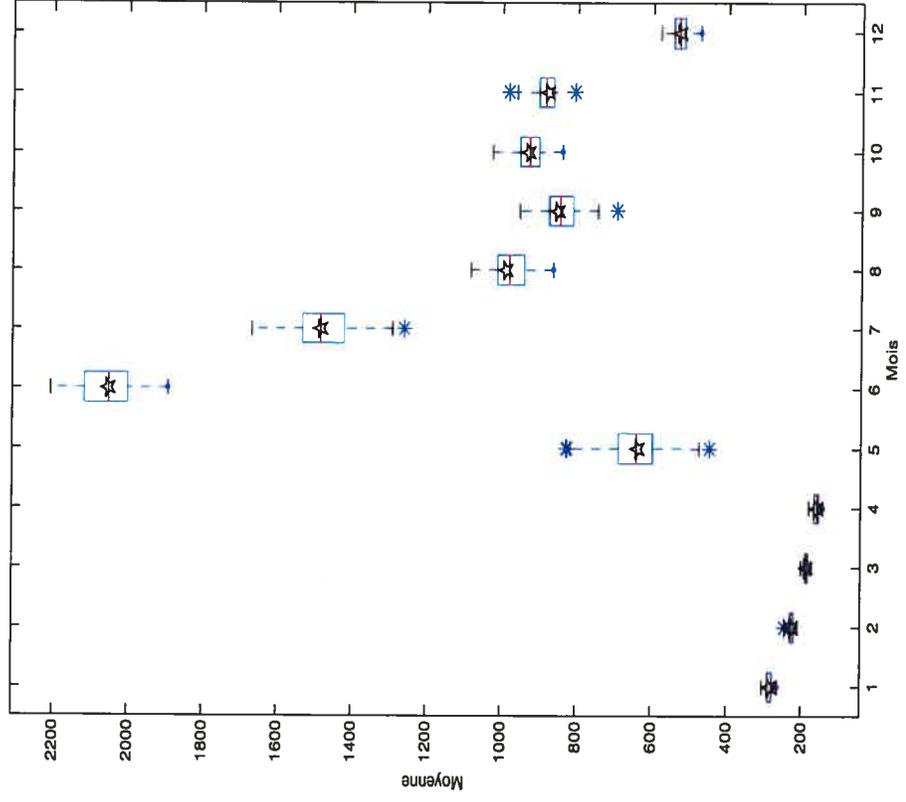
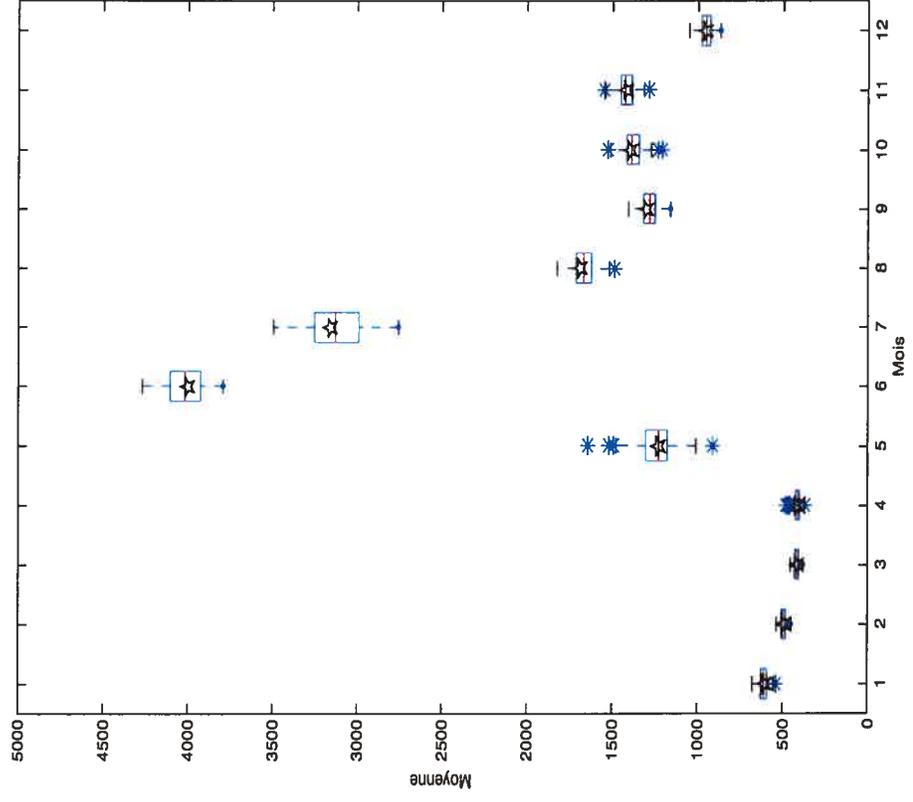


FIGURE 3.4. Moyenne mensuelle historique (★) au réservoir Caniapiscau et diagrammes en boîte des moyennes mensuelles de 100 séries générées avec un modèle $PAR(1)$ suite à la transformation logarithmique ou suite au rééchantillonnage des résidus.

Churchill Falls

Moyenne mensuelle (series reechantillonnees)



Churchill Falls

Moyenne mensuelle (series transformees)

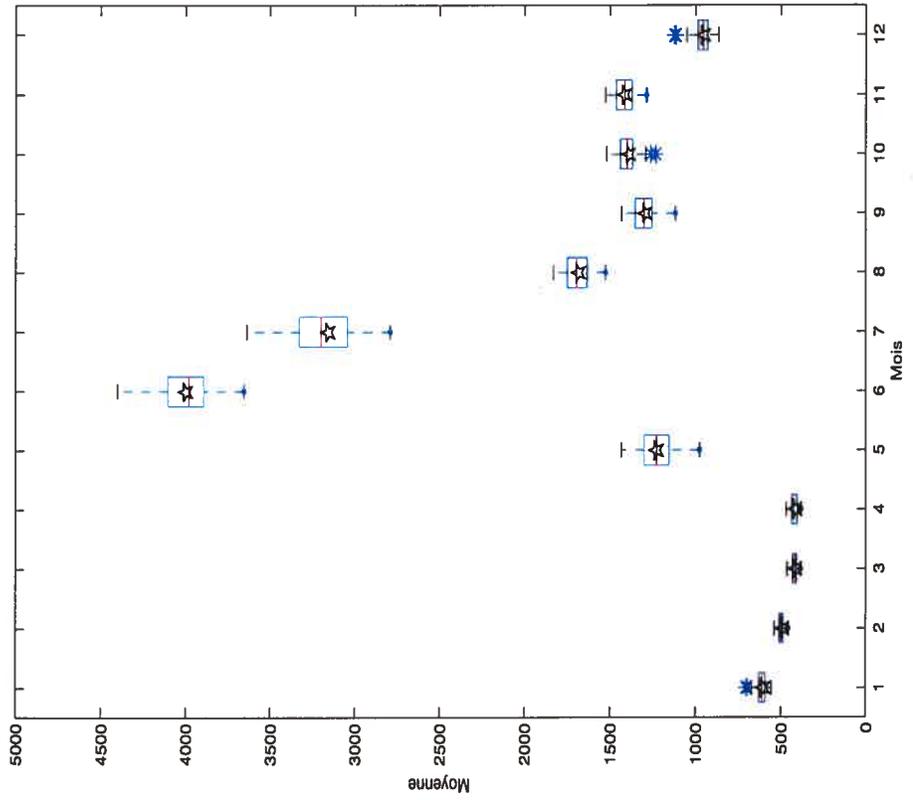
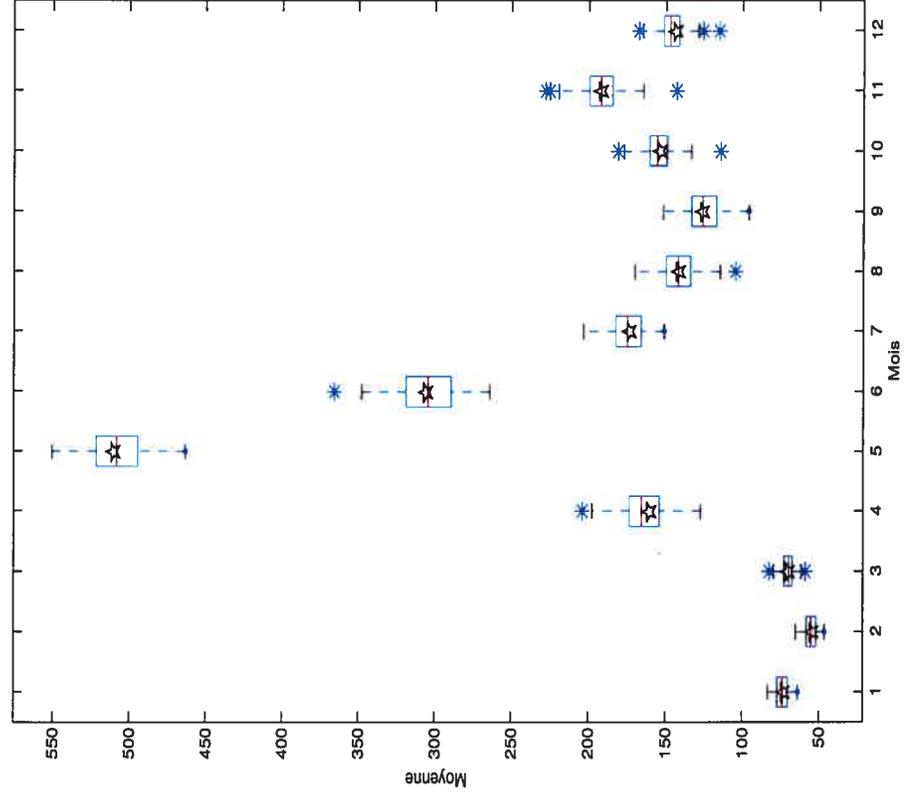


FIGURE 3.5. Moyenne mensuelle historique (★) au réservoir Churchill Falls et diagrammes en boîte des moyennes mensuelles de 100 séries générées avec un modèle *PAR(1)* suite à la transformation logarithmique ou suite au rééchantillonnage des résidus.

GOUJIN

Moyenne mensuelle (séries rééchantillonnées)



GOUJIN

Moyenne mensuelle (séries transformées)

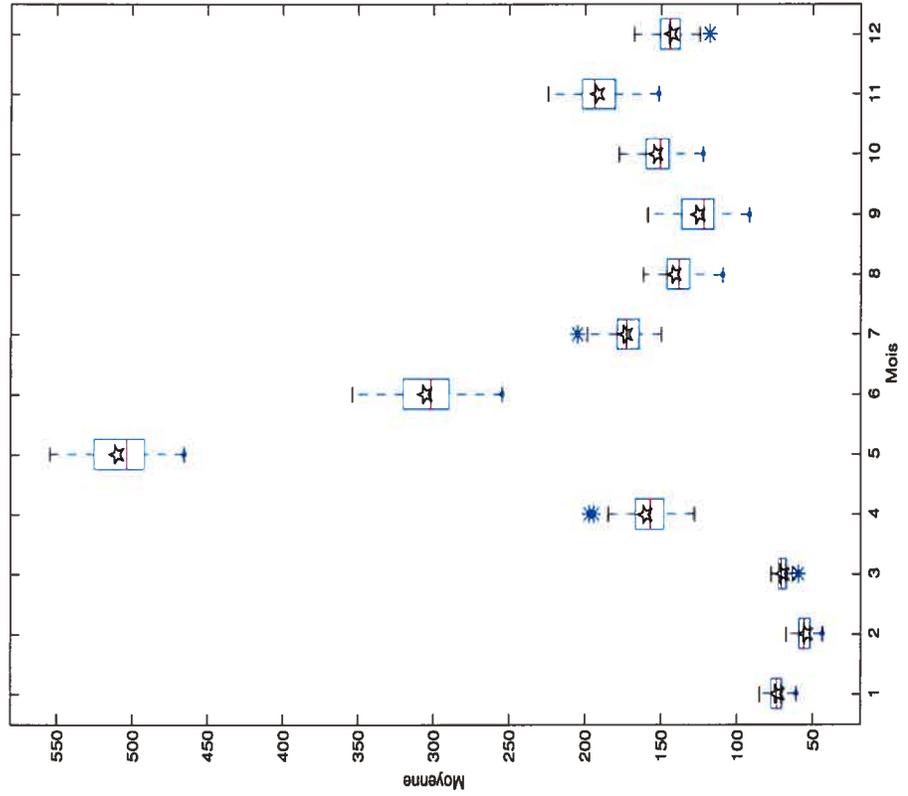


FIGURE 3.6. Moyenne mensuelle historique (★) au réservoir Gouin et diagrammes Gouin et diagrammes en boîte des moyennes mensuelles de 100 séries générées avec un modèle $PAR(1)$ suite à la transformation logarithmique ou suite au rééchantillonnage des résidus.

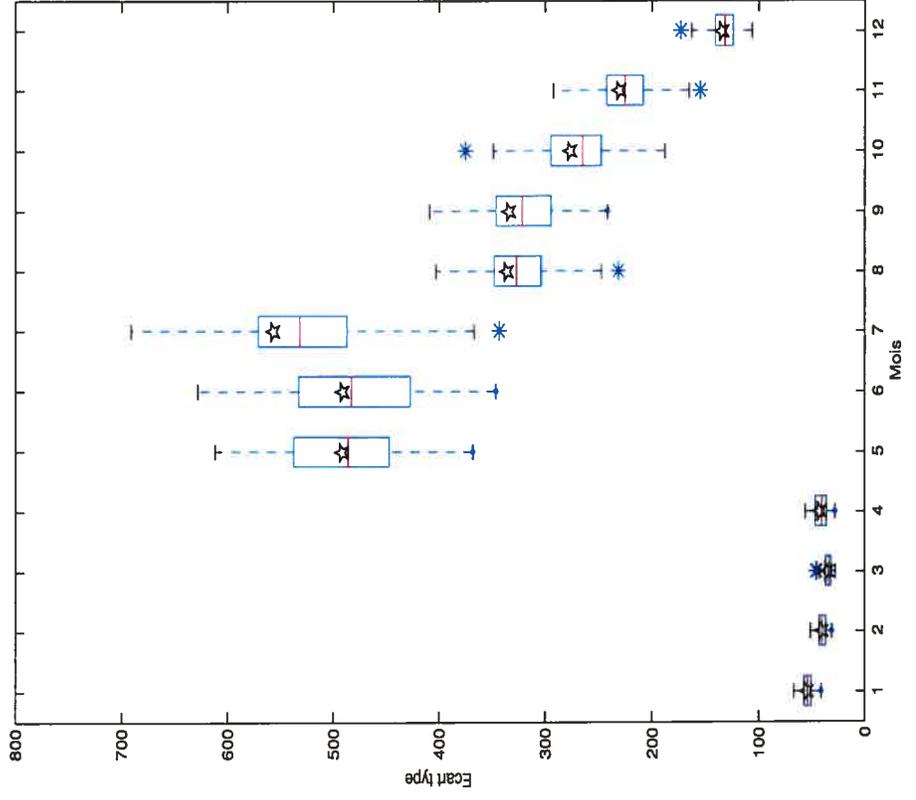
- Écart types

Les diagrammes en boîte des figures 3.7 à 3.9, ont aussi été construits à partir de 100 scénarios, où pour chaque scénario k , l'écart type $\sigma_{x,m}^{(k)}$ défini par l'équation (1.2.2) est calculé à partir des 42 années. Nous calculons l'écart type simulé pour chaque saison m , c'est-à-dire $\sigma_{x,m}^{(1)}, \sigma_{x,m}^{(2)}, \dots, \sigma_{x,m}^{(100)}$, et le diagramme en boîte est obtenu à partir de ces 100 écarts types, $m = 1, 2, \dots, 12$.

Les étoiles (★) dans les figures 3.7 à 3.9 représentent les écarts types observés. Pour chaque saison m , dans le cas de la méthode par rééchantillonnage, on remarque que cet écart type est proche de la médiane des 100 écarts types simulés, ce qui illustre qu'on reproduit assez bien les écarts types à chaque saison puisque l'étoile est à l'intérieur de l'intervalle interquartile pour les 3 réservoirs considérés. On observe aussi d'après les figures 3.7 à 3.9 que la méthode par rééchantillonnage reproduit mieux les écarts types à chaque saison. De plus, la méthode transformée conduit à davantage de valeurs à l'écart surtout durant les crues printanières.

CANIAPISCAU

Écart type mensuel (séries rééchantillonnées)



CANIAPISCAU

Écart type mensuel (séries transformées)

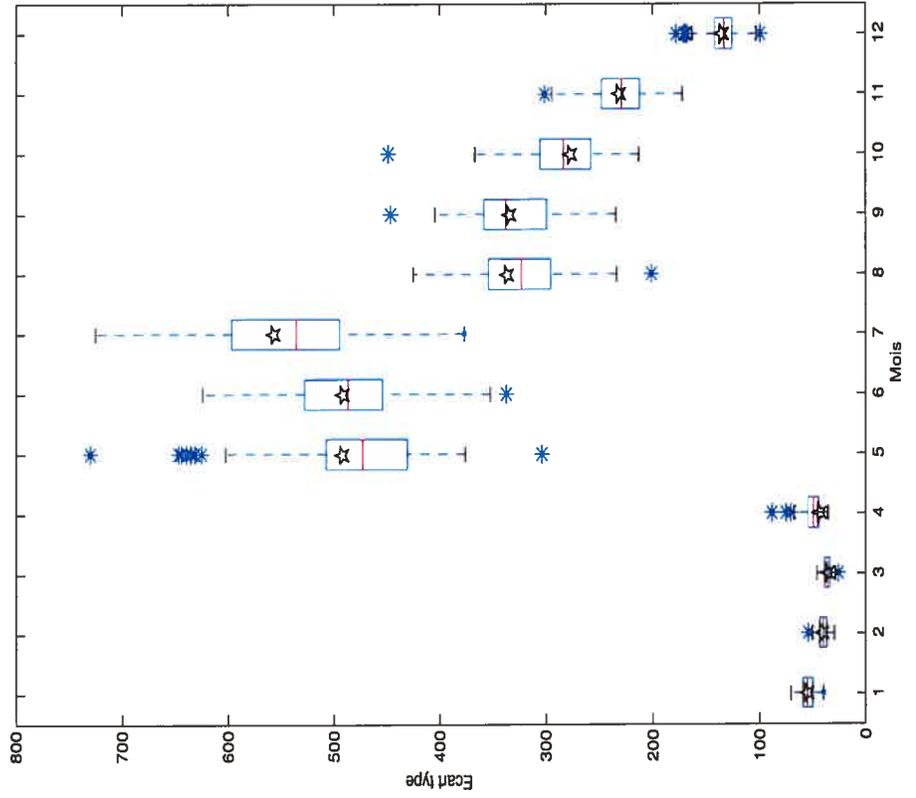
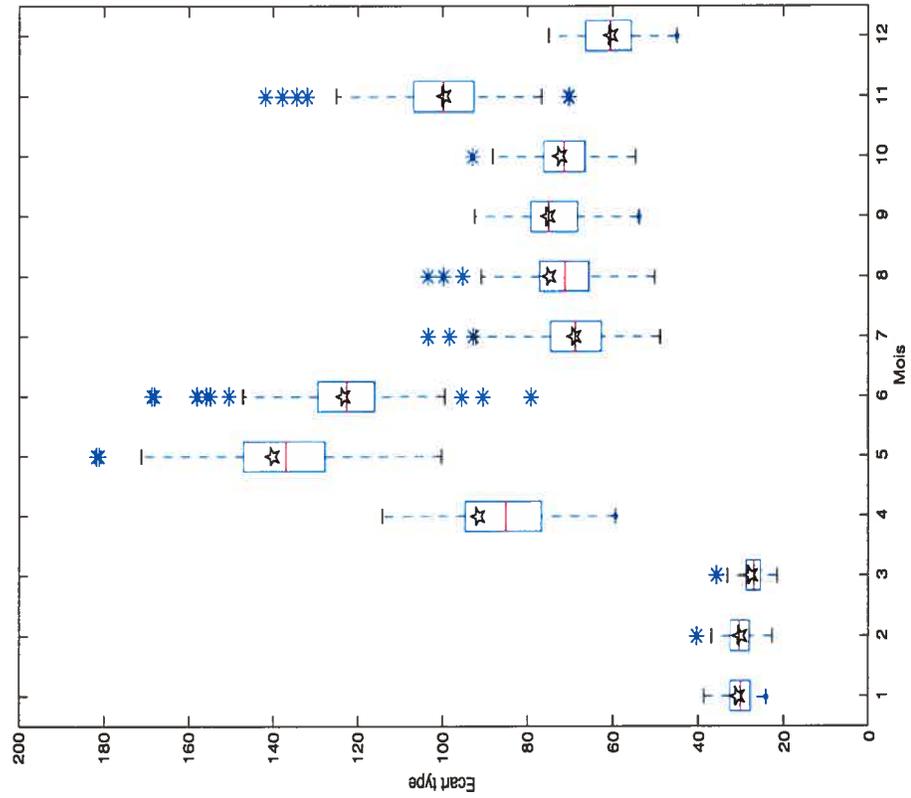


FIGURE 3.7. Écart type mensuel historique (★) au réservoir Caniapiscau et diagrammes en boîte des écarts types mensuels de 100 séries générées avec un modèle $PAR(1)$ suite à la transformation logarithmique ou suite au rééchantillonnage des résidus.

GOUJIN
Écart type mensuel (séries transformées)



GOUJIN
Écart type mensuel (séries rééchantillonnées)

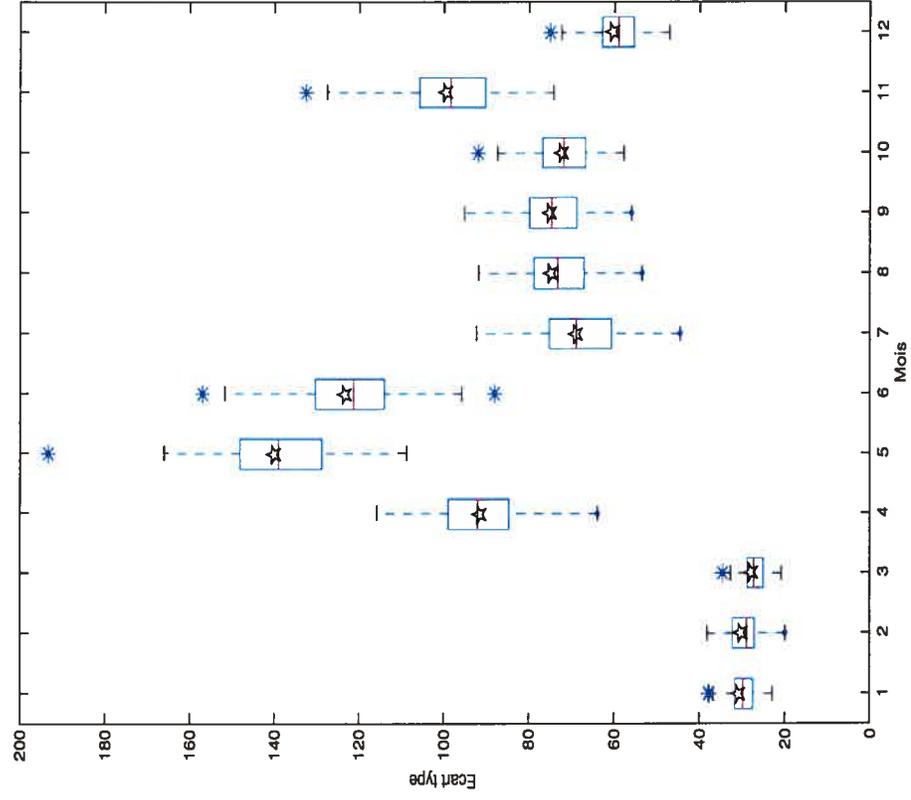


FIGURE 3.9. Écart type mensuel historique (★) au reservoir Gouin et diagrammes Gouin et diagrammes en boîte des écarts types mensuels de 100 séries générées avec un modèle $PAR(1)$ suite à la transformation logarithmique ou suite au rééchantillonnage des résidus.

- Autocorrélations mensuelles

Les résultats des autocorrélations périodiques mensuelles pour les 3 réservoirs sont présentés dans les figures 3.10 à 3.12. Les diagrammes en boîte des figures 3.10 à 3.12, ont été construits à partir de 100 scénarios, où pour chaque scénario k , l'autocorrélation de délai 1 définie par l'équation (1.3.1) est calculée à partir des 42 années. Nous calculons l'autocorrélation mensuelle simulée de délai 1 pour chaque saison m c'est-à-dire $\rho_m^{(1)}, \rho_m^{(2)}, \dots, \rho_m^{(100)}$ et le diagramme en boîte est obtenu à partir de ces 100 valeurs, $m = 1, \dots, 12$. Les étoiles (★) dans les figures 3.10 à 3.12 représentent les autocorrélations mensuelles observées. Des calculs similaires ont aussi été faits pour les délais 2 et 3.

L'analyse de ces figures montre des autocorrélations très élevées pour la majorité des périodes sauf pour certains mois où on remarque une diminution des autocorrélations. Cette diminution prononcée des autocorrélations correspond à la crue printanière.

Les graphiques 3.10 à 3.12 montrent également que la structure des corrélations est bien préservée puisque nous observons des autocorrélations élevées pour les 3 premiers mois. Après cette période, on observe une décroissance graduelle des autocorrélations, engendrée par une augmentation rapide du débit d'eau dans les réservoirs due surtout à la fonte des neiges. On observe aussi lors de la crue automnale le même phénomène à partir du 10^e mois, mais de façon moins prononcée.

Pour les diagrammes en boîte des figures 3.10 à 3.12, on remarque que l'autocorrélation observée est proche de la médiane des 100 autocorrélations simulées et ceci pour chaque saison m , ce qui indique qu'on reproduit bien l'autocorrélation pour chaque période m . En effet, on observe bien l'étoile à l'intérieur de l'intervalle interquartile et ceci pour les 3 réservoirs.

On reproduit aussi assez bien les autocorrélations de délai 2 pour les 3 réservoirs. Cependant ce n'est plus le cas au délai 3.

Finalement, on peut dire à partir des figures 3.10 à 3.12 que la structure des autocorrélations saisonnières pour le modèle $PAR(1)$ est bien préservée puisque les autocorrélations obtenues par simulation pour les deux méthodes sont similaires à celles que nous avons obtenues avec les données historiques.

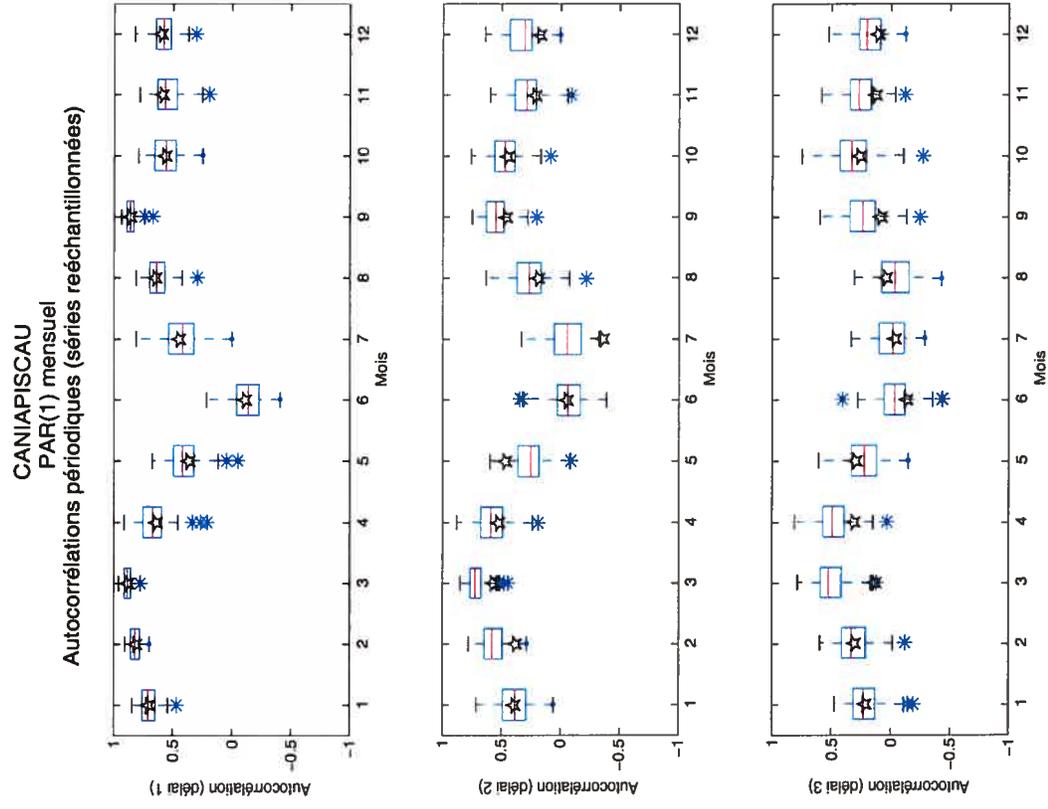
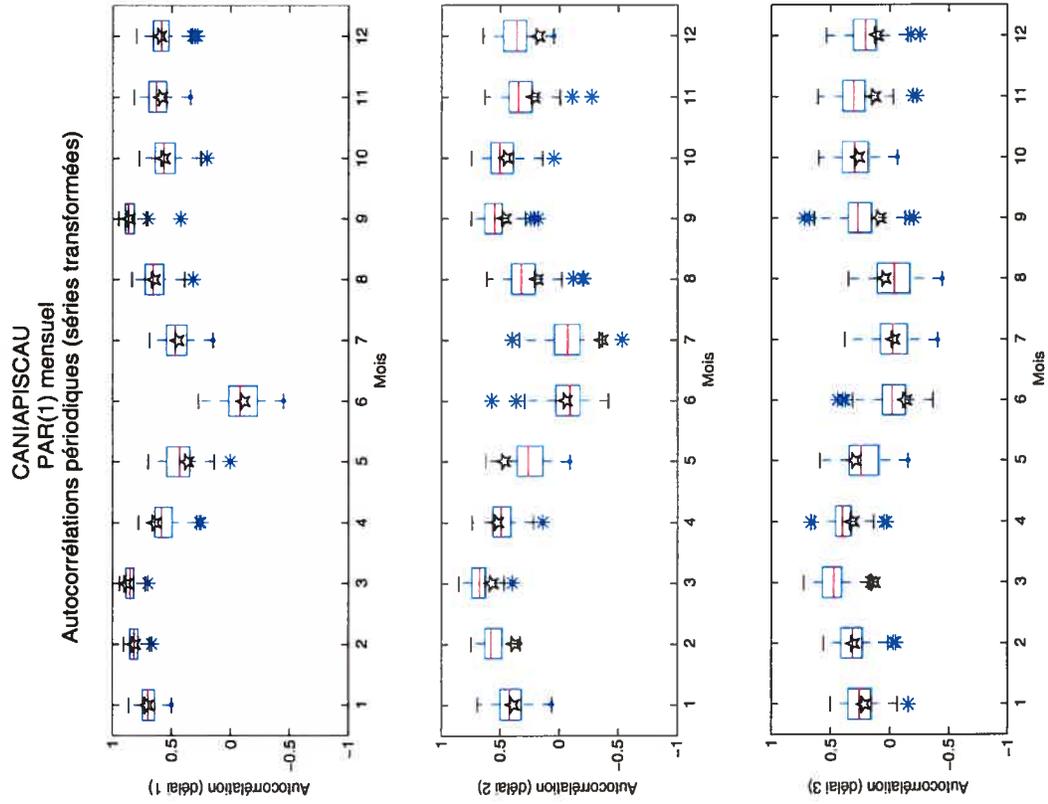


FIGURE 3.10. Autocorrélations périodiques de délai 1, 2 et 3 au réservoir Caniapiscau et diagrammes en boîte des autocorrélations mensuelles de 100 séries générées avec un modèle $PAR(1)$ suite à une transformation logarithmique ou suite au rééchantillonnage des résidus.

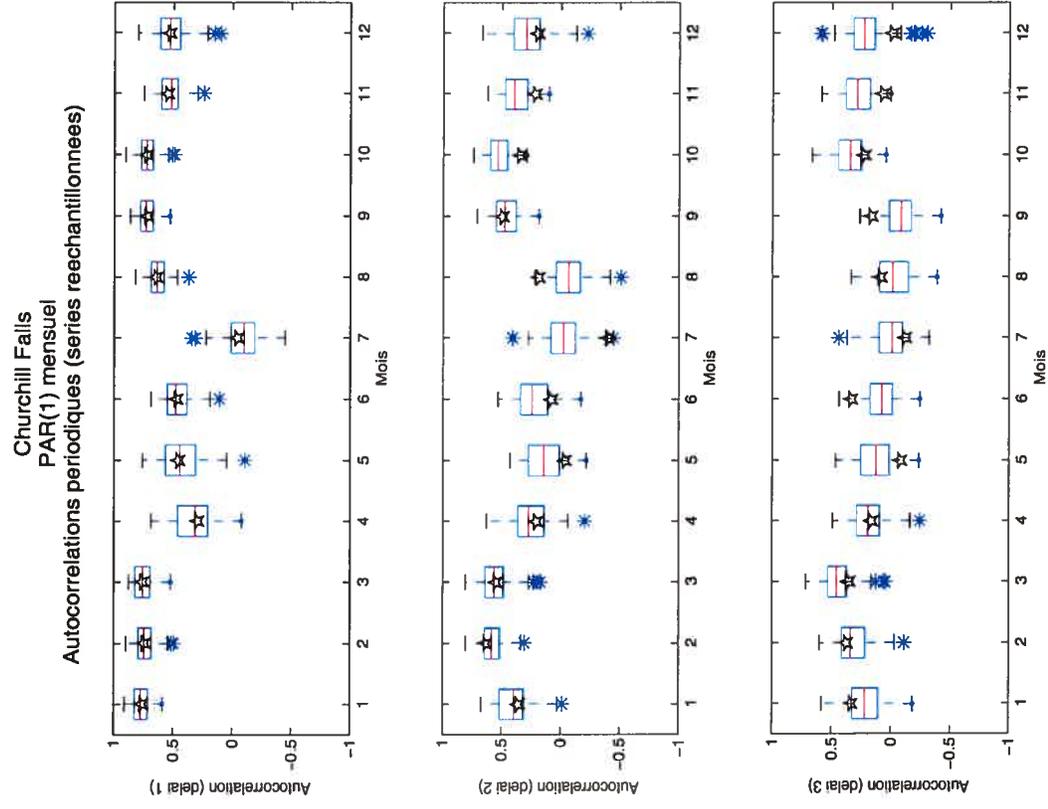
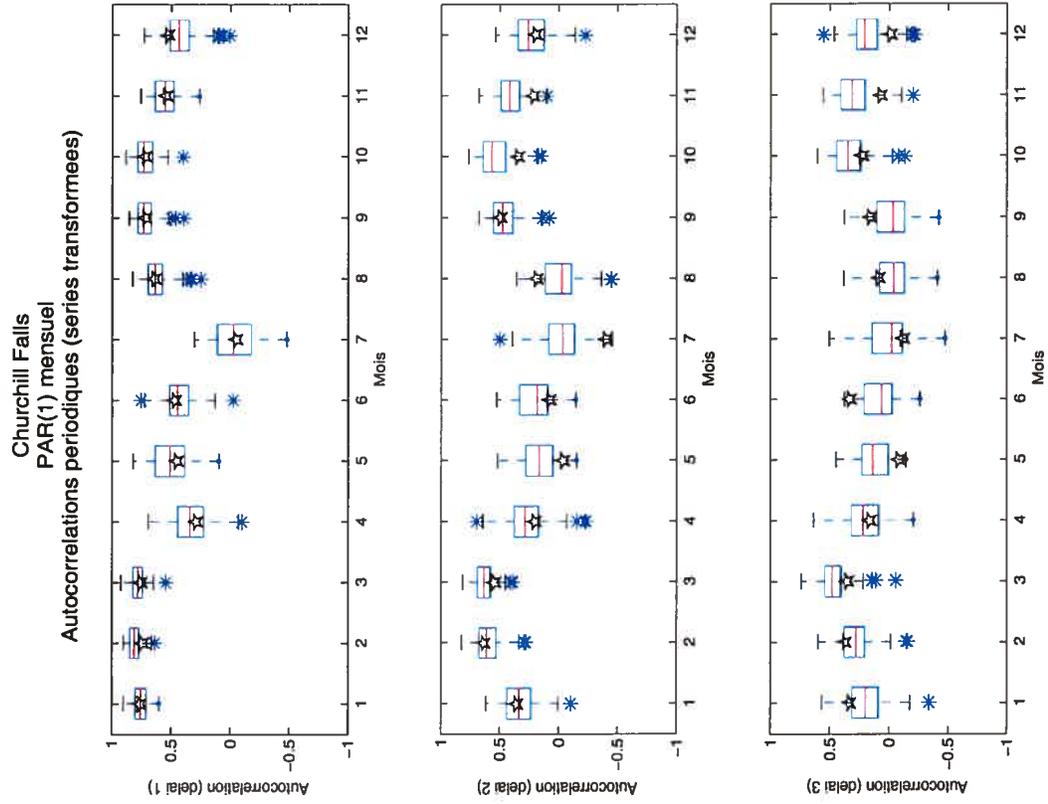


FIGURE 3.11. Autocorrélations périodiques de délai 1, 2 et 3 au réservoir Churchill Falls et diagrammes en boîte des autocorrélations mensuelles de 100 séries générées avec un modèle $PAR(1)$ suite à une transformation logarithmique ou suite au rééchantillonnage des résidus.

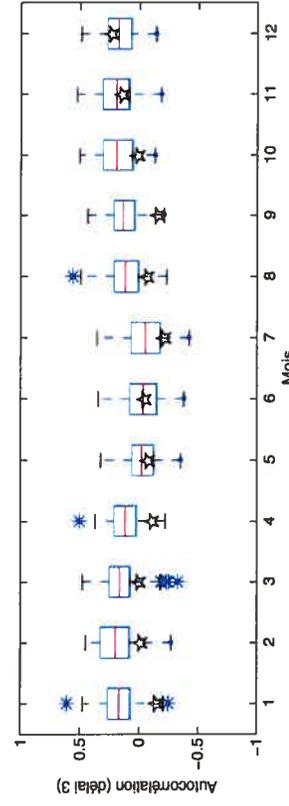
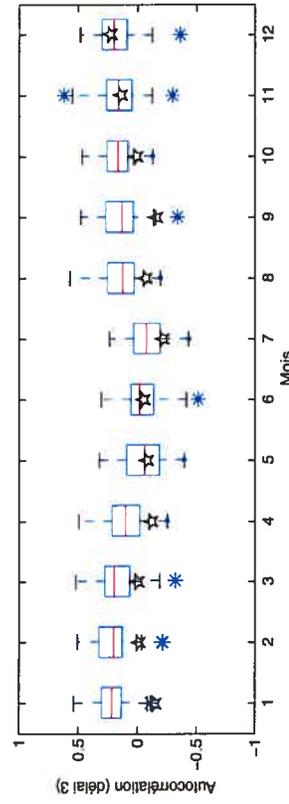
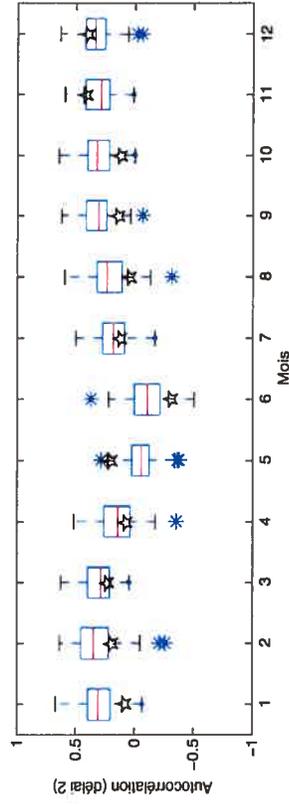
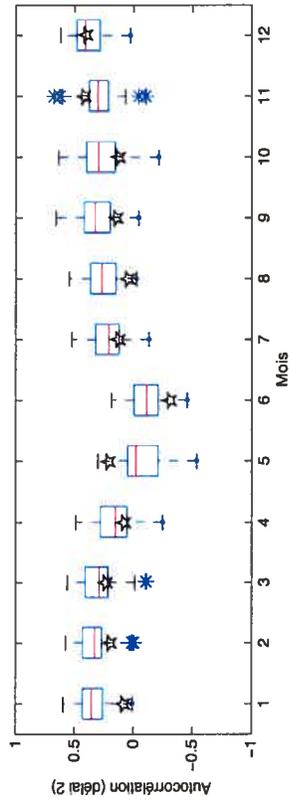
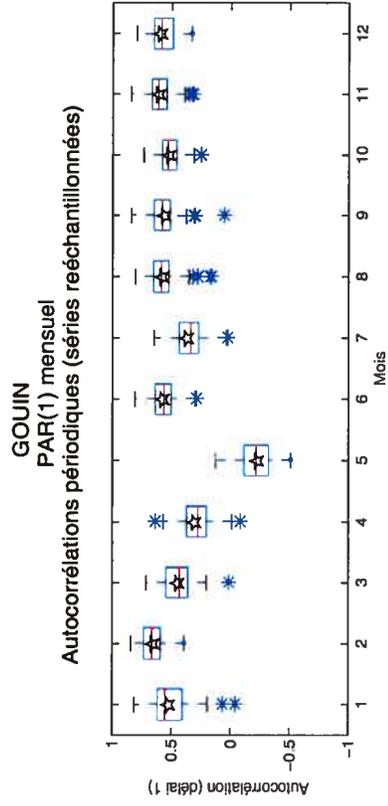
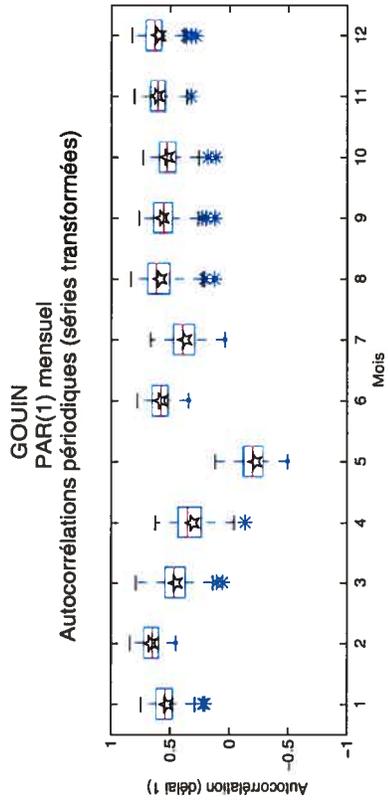


FIGURE 3.12. Autocorrélations périodiques de délai 1, 2 et 3 au réservoir Gouin et diagrammes en boîte des autocorrélations mensuelles de 100 séries générées avec un modèle $PAR(1)$ suite à une transformation logarithmique ou suite au rééchantillonnage des résidus.

3.2.4. Discussion des résultats des corrélations interannuelles simulées avec un $PAR(1)$

Dans le tableau 3.4, nous présentons les résultats comparatifs des corrélations interannuelles observées et simulées de délai 1. Il s'agit des autocorrélations $\rho^{(i)}(1) = corr(V_r^{(i)}, V_{r-1}^{(i)})$ où $V_r^{(i)}$ représente l'apport total d'eau durant l'année r au site i , $r = 0, \dots, n - 1$, $i = 1, \dots, I$. Sous l'hypothèse de bruit blanc fort, l'écart type est $\frac{1}{\sqrt{n}}$ avec $n = 42$ et les limites de signification au niveau 5% sont approximativement $\pm 0,30$. Les étoiles dans le tableau 3.4 identifient les valeurs significatives.

Nous remarquons que les corrélations interannuelles simulées au réservoir Caniapiscau ne sont pas élevées. En effet, les corrélations estimées à partir des méthodes transformées $(-0,01)$ et par rééchantillonnage $(0,02)$ sont très faibles par rapport à la corrélation observée $(0,26)$. Nous observons aussi pour le réservoir Churchill Falls une corrélation significative dans le cas de données historiques $(0,42)$ alors qu'elle est très faible pour la méthode transformée $(0,03)$ et la méthode de rééchantillonnage $(0,00)$ où on n'observe aucune corrélation significative. Le même constat est fait pour les autres réservoirs tels que Laforge, LaGrande4 ou Ste-Marguerite.

À partir des résultats du tableau 3.4, on peut dire que l'autocorrélation interannuelles de délai 1 observée dans l'historique n'est pas du tout reproduite par aucune des deux méthodes et pour l'ensemble des réservoirs. Ceci peut s'expliquer par le fait que dans le modèle $PAR(1)$, il n'existe aucun terme qui inclue la dépendance interannuelle.

TABLEAU 3.4. Résultats comparatifs des corrélations interannuelles observées et simulées d'un modèle PAR(1) pour les différents réservoirs d'Hydro-Québec.

Code	Sites	Historique	Transformée	Rééchantionnée
1	Caniapiscau	0,26	0,02	-0,01
2	Laforge	0,33 *	0,02	0,01
3	LaGrande4	0,37 *	0,05	0,05
4	EOL Total	0,25	0,00	0,01
5	Bersimis	0,24	-0,02	-0,02
6	Churchill Falls	0,42 *	0,03	0,00
7	Manic 5	0,28	-0,01	0,02
8	Lac Ste-Anne	0,10	-0,01	-0,03
9	Outardes 4	0,29	-0,02	-0,01
10	Ste-Marguerite	0,51 *	0,00	0,01
11	Gouin	-0,14	0,01	-0,01
12	Manouane	-0,05	0,03	0,00
13	Rapide Blanc	-0,07	-0,01	-0,00
14	Mattawin	0,10	0,01	0,03
15	TR-BE-LT2	0,19	0,00	0,00
16	GM-SH-Lagab	0,13	0,00	-0,02
17	LG2-SAK	0,31 *	0,01	-0,01

3.2.5. Discussion des corrélations spatiales simulées avec un $PAR(1)$

Les principaux résultats des autocorrélations spatiales sont présentés dans les tableaux 3.5 et 3.7. Les corrélations entre les réservoirs Caniapiscau (1), Churchill Falls (6) et Gouin (11) sont en gras. Notons que ces autocorrélations ont été calculées à partir de données standardisées. Elles ont été estimées à partir de la matrice de covariance usuelle \sum_m définie par l'équation (2.7.7).

Les résultats du tableau 3.5 montrent dans le cas de la méthode transformée que les corrélations spatiales estimées sont faibles par rapport aux corrélations obtenues à partir de données observées. En effet, la corrélation spatiale entre Caniapiscau (1) et Churchill Falls (6) est 0,44 alors que dans le tableau 1.3, la corrélation spatiale historique est 0,83. Le même constat a été fait pour Caniapiscau (1) et Gouin (11) avec une corrélation négative de $-0,02$ alors qu'en historique, la corrélation observée est 0,35. Pour la corrélation entre Churchill Falls (6) et Gouin (11), elle est de 0,01 alors que la corrélation historique est 0,32. À partir de ces résultats, on voit que le modèle $PAR(1)$ ne reproduit pas de façon satisfaisante la corrélation spatiale.

Dans le tableau 3.7 relatif aux corrélations spatiales obtenues à partir de la méthode de rééchantillonnage, nous observons une nette amélioration des corrélations estimées entre certains sites mais nous sommes encore loin des corrélations observées du tableau 1.3. Plus précisément, la corrélation entre Caniapiscau (1) et Churchill Falls (6) est 0,64 alors que dans le tableau 1.3, elle est 0,83. Pour Caniapiscau (1) et Gouin (11), elle est 0,01 alors qu'en historique elle est 0,35. De même pour Churchill Falls (6) et Gouin (11), elle est 0,05 alors que dans le tableau 1.3, on a une corrélation de 0,32.

La comparaison des tableaux 3.5 et 3.7 fait ressortir une amélioration intéressante des corrélations spatiales simulées avec la méthode par rééchantillonnage mais pour plusieurs sites, nous sommes encore loin des corrélations historiques. Dans les tableaux 3.6 et 3.8, nous présentons les écarts types des corrélations

spatiales pour les différents réservoirs obtenus par simulation. Pour chaque corrélation, l'écart type présenté correspond à l'écart type expérimental des N corrélations obtenus des N scénarios simulés.

TABLEAU 3.7. Corrélations spatiales entre les 17 réservoirs obtenus avec un modèle PAR(1) et en rééchantillonnant les résidus.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	1,00	0,67	0,72	0,55	0,26	0,64	0,56	0,48	0,43	0,41	0,01	0,00	0,03	-0,03	0,07	-0,01	0,61
2		1,00	0,74	0,63	0,25	0,41	0,41	0,34	0,36	0,27	0,05	0,02	0,06	-0,01	0,03	0,01	0,75
3			1,00	0,70	0,21	0,36	0,50	0,39	0,40	0,34	0,01	-0,03	0,01	-0,02	0,01	-0,04	0,75
4				1,00	0,39	0,33	0,50	0,39	0,48	0,33	0,13	0,01	0,15	0,05	0,02	0,05	0,77
5					1,00	0,31	0,60	0,62	0,78	0,53	0,43	0,30	0,51	0,28	0,33	0,27	0,27
6						1,00	0,50	0,46	0,34	0,32	0,05	0,05	0,07	-0,01	0,11	0,04	0,34
7							1,00	0,78	0,77	0,64	0,17	0,09	0,21	0,06	0,14	0,07	0,48
8								1,00	0,74	0,62	0,13	0,09	0,22	0,06	0,19	0,05	0,41
9									1,00	0,61	0,29	0,16	0,33	0,17	0,20	0,16	0,39
10										1,00	0,15	0,12	0,23	0,14	0,17	0,15	0,30
11											1,00	0,58	0,76	0,46	0,56	0,48	0,01
12												1,00	0,71	0,68	0,73	0,68	0,00
13													1,00	0,60	0,70	0,57	0,05
14														1,00	0,55	0,68	0,00
15															1,00	0,63	0,01
16																1,00	0,01
17																	1,00

3.2.6. Conclusion

Finalement, à partir de ces résultats, on conclut que si le modèle $PAR(1)$ reproduit assez bien certaines caractéristiques statistiques, il n'en demeure pas moins qu'il n'est pas capable de modéliser de façon satisfaisante ce type de séries d'apports. En effet, ce modèle ne reproduit pas la dépendance interannuelle entre les apports hydrologiques ni les corrélations spatiales entre les sites. C'est pourquoi l'ajout d'une nouvelle variable exogène au modèle PAR nous paraît indispensable.

3.3. MISE EN OEUVRE DU MODÈLE $PAR(1)$ AVEC VARIABLE EXPLICATIVE

Pour remédier aux insuffisances du modèle $PAR(1)$, on a introduit le modèle périodique $PARX(1)$ avec une nouvelle variable explicative, qui est l'apport total de l'année précédente. Nous espérons ainsi mieux reproduire les corrélations spatiales et les corrélations interannuelles.

3.3.1. Modèle estimé $PARX(1)$ et test d'adéquation de Ljung et Box

Les principaux résultats sur l'estimation et la validation du modèle $PARX(1)$ sont présentés dans les tableaux 3.9 et 3.10. Les valeurs- p du test de Ljung-Box nous permettent de constater qu'au niveau de signification 5% pour les délais 1, 3, 5 et 10 le modèle estimé est adéquat pour les réservoirs Caniapiscou et Gouin. Cependant, pour Churchill Falls, le modèle est rejeté avec $h = 1$ au mois de février et septembre, $h = 5$ et $h = 10$ au mois de mars, puisque les valeurs- p sont inférieures à 5%.

TABLEAU 3.9. Estimation et validation du modèle $PARX(1)$ pour les réservoirs Caniapiscou et Churchill Falls.

Sites	Mois	1	2	3	4	5	6	7	8	9	10	11	12
Caniapiscou	Coefficients AR	0,68	0,82	0,88	0,62	0,40	-0,11	0,44	0,62	0,85	0,58	0,59	0,57
	Coefficients B_m	0,08	-0,02	-0,04	0,16	-0,13	0,12	0,12	0,14	0,02	-0,13	0,11	0,18
	Variance résiduelle	0,000	0,002	0,104	0,206	0,101	0,018	0,090	0,057	0,090	0,095	0,030	0,003
	Test Ljung-Box :												
	h=1	0,08	2,98	0,58	0,91	0,30	0,27	0,59	0,11	0,99	0,19	2,03	12,73
	Valeur- p	0,77	0,08	0,45	0,34	0,58	0,60	0,44	0,74	0,31	0,67	0,16	0,24
	h=3	4,69	2,98	1,36	5,58	0,98	4,25	0,71	0,16	3,04	0,31	4,90	2,84
	Valeur- p	0,20	0,39	0,72	0,13	0,81	0,24	0,87	0,98	0,38	0,96	0,18	0,42
	h=5	9,35	4,71	3,21	6,81	1,29	11,02	2,29	3,11	3,64	0,49	5,33	4,28
	Valeur- p	0,096	0,45	0,67	0,23	0,93	0,04	0,81	0,68	0,60	0,99	0,38	0,51
h=10	18,87	10,22	9,93	10,39	3,07	18,77	5,12	8,69	7,28	5,41	12,73	9,87	
Valeur- p	0,02	0,42	0,44	0,41	0,98	0,03	0,88	0,56	0,70	0,86	0,24	0,45	
Churchill Falls	Coefficients AR	0,71	0,67	0,71	0,25	0,44	0,45	-0,16	0,60	0,76	0,73	0,57	0,53
	Coefficients B_m	0,20	0,19	0,10	0,10	0,06	0,17	0,37	0,14	-0,13	-0,12	0,09	0,03
	Variance résiduelle	0,063	0,030	0,145	0,114	0,136	0,032	0,001	0,040	0,021	0,039	0,001	0,062
	Test Ljung-Box :												
	h=1	0,57	5,05	1,00	0,49	0,04	0,76	0,83	1,06	4,02	0,05	0,13	3,41
	Valeur- p	0,45	0,02	0,32	0,48	0,85	0,38	0,36	0,30	0,04	0,82	0,72	0,06
	h=3	4,62	7,51	6,07	1,65	3,20	5,49	5,63	1,38	4,41	4,17	0,53	3,98
	Valeur- p	0,20	0,05	0,11	0,65	0,36	0,14	0,13	0,71	0,22	0,24	0,91	0,26
	h=5	4,89	8,31	11,58	5,84	4,61	9,85	8,97	1,57	6,36	4,98	0,69	4,53
	Valeur- p	0,43	0,14	0,04	0,32	0,46	0,07	0,11	0,90	0,27	0,42	0,98	0,48
h=10	10,01	12,10	21,95	8,02	5,83	15,47	17,69	7,52	12,18	7,51	5,50	6,80	
Valeur- p	0,44	0,28	0,01	0,63	0,83	0,12	0,05	0,68	0,27	0,68	0,85	0,74	

TABLEAU 3.10. Estimation et validation du modèle $PARX(1)$ pour le réservoir Gouin.

Sites	Mois	1	2	3	4	5	6	7	8	9	10	11	12
	Coefficients AR	0,77	0,61	0,40	0,39	-0,20	0,58	0,35	0,59	0,57	0,52	0,65	0,61
	Coefficients B_m	-0,55	-0,20	-0,10	0,42	-0,14	0,06	-0,17	0,00	-0,01	0,10	-0,01	-0,01
	Variance résiduelle	0,001	0,044	0,017	0,064	0,005	0,039	0,153	0,062	0,005	0,011	0,100	0,037
	Test Ljung-Box :												
	h=1	2,69	0,29	0,00	0,06	0,01	0,01	1,47	3,86	0,65	0,80	0,24	0,03
	Valeur-p	0,10	0,59	0,96	0,81	0,91	0,93	0,23	0,04	0,42	0,37	0,62	0,85
Gouin	h=3	8,58	0,58	0,53	3,56	0,61	0,05	3,03	5,56	1,82	1,87	3,46	3,23
	Valeur-p	0,03	0,90	0,91	0,31	0,89	0,99	0,39	0,13	0,61	0,60	0,32	0,36
	h=5	9,40	0,89	2,03	4,27	1,23	0,22	3,61	9,72	5,22	6,46	7,84	4,60
	Valeur-p	0,09	0,97	0,85	0,51	0,94	0,99	0,61	0,08	0,39	0,26	0,17	0,47
	h=10	10,91	8,44	5,17	12,36	2,31	3,71	9,60	13,35	8,14	8,45	16,47	6,61
	Valeur-p	0,36	0,59	0,88	0,26	0,99	0,96	0,48	0,20	0,61	0,59	0,08	0,76

3.3.2. Analyse des hydrogrammes simulés

Une analyse visuelle des figures 3.13 à 3.15 ne permet pas de déceler des changements significatifs par rapport aux hydrogrammes générés avec le modèle $PAR(1)$. À nouveau, le rééchantillonnage donne des résultats plus près des hydrogrammes historiques que la transformation logarithmique.

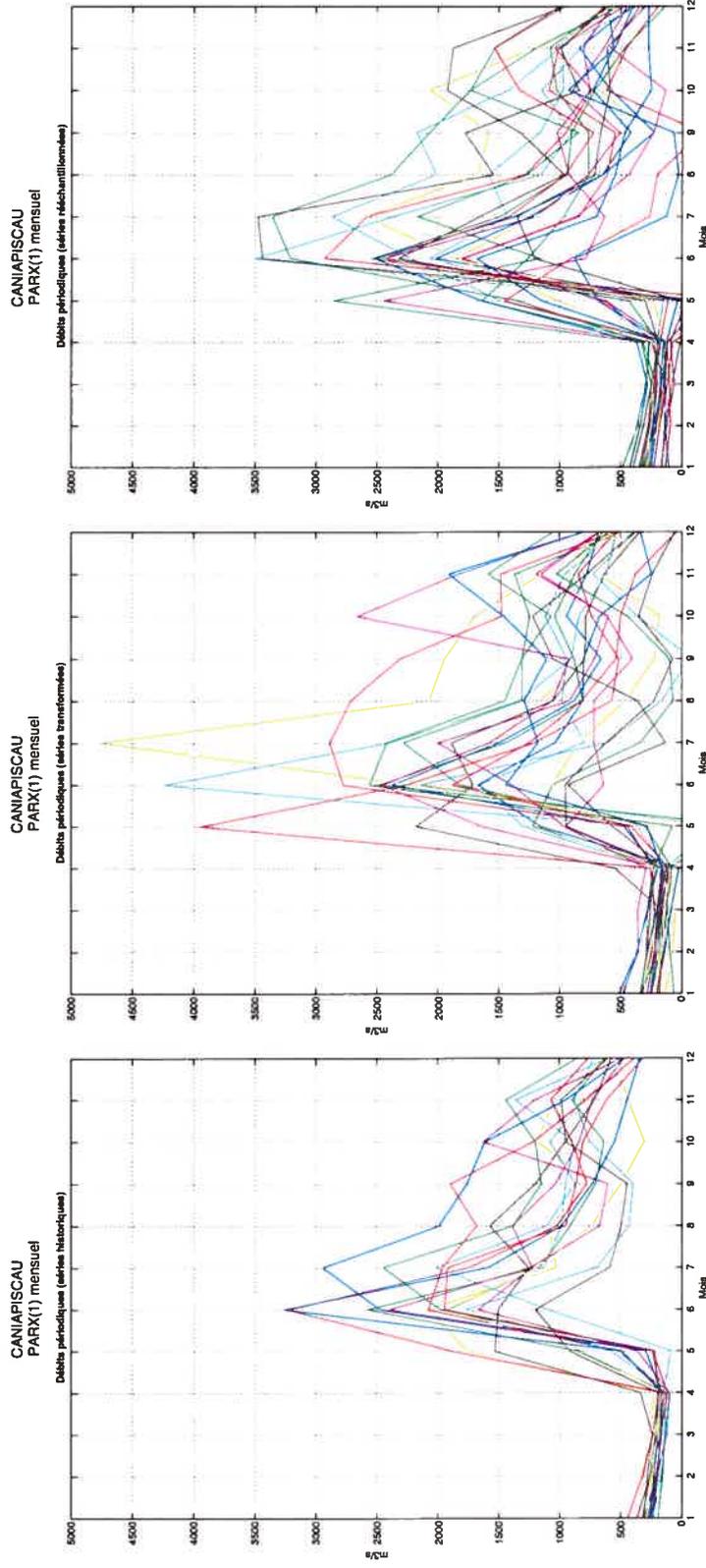


FIGURE 3.13. Hydrogrammes comparatifs pour les séries historiques, transformées et rééchantillonnées au réservoir Caniapiscau.

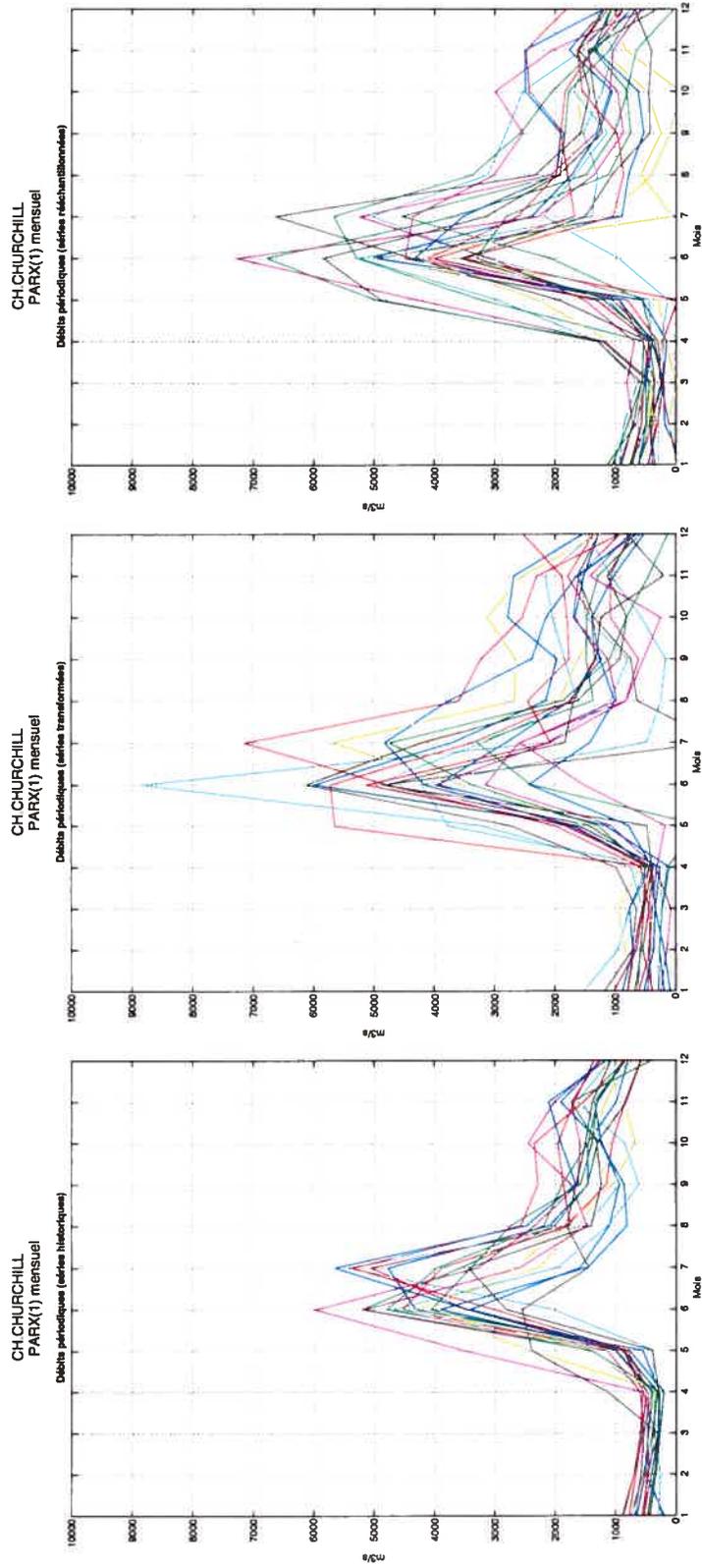


FIGURE 3.14. Hydrogrammes comparatifs pour les séries historiques, transformées et rééchantillonnées au réservoir Churchill Falls.

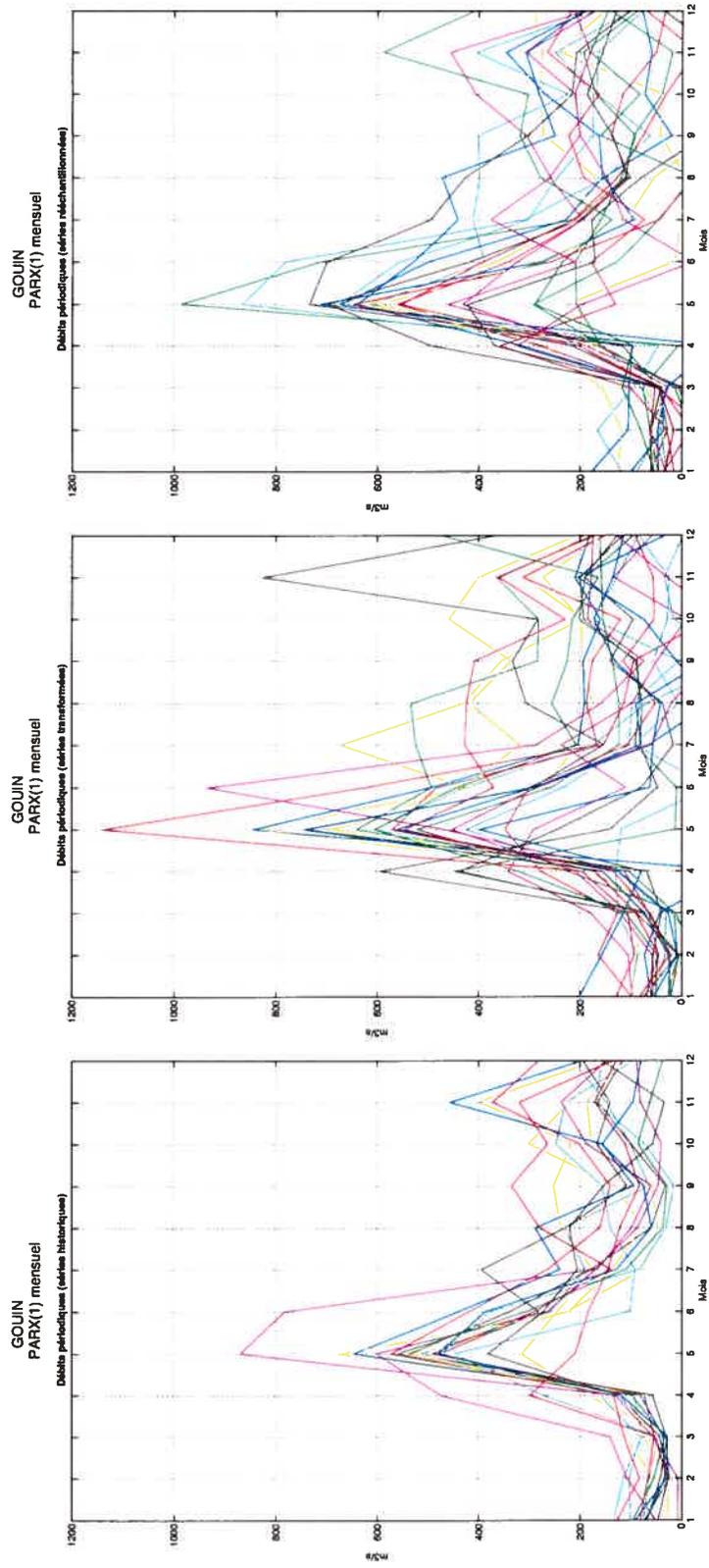


FIGURE 3.15. Hydrogrammes comparatifs pour les séries historiques, transformées et rééchantillonnées au réservoir Gouin.

3.3.3. Discussion des résultats de simulation dans le cas d'un modèle $PARX(1)$

-Moyennes et écarts types

Dans le tableau 3.7, nous retrouvons les résultats comparatifs des statistiques annuelles observées et simulées d'un modèle $PARX(1)$ pour les 17 réservoirs. Au niveau de l'estimation de la moyenne, on voit bien que la méthode par rééchantillonnage donne de bons résultats puisque la moyenne annuelle observée pour Caniapiscau est de 768,03 alors que celle de la moyenne annuelle estimée est de 770,44. De même que pour l'écart type annuel estimé (138,78), qui est proche du vrai écart type (131,93). Le même constat est fait pour Churchill Falls avec une moyenne annuelle estimée de 1425,18 qui se rapproche de la moyenne annuelle historique qui est de 1422,44 et avec un écart type estimé de 219,02 alors que le vrai écart type est de 223,41. Pour Gouin la moyenne estimée (176,18) est presque similaire à la moyenne observée (175,26) alors que l'écart type estimé (34,61) est plus grand que le vrai écart type (29,15). Pour la méthode transformée, on observe une surestimation de la moyenne notamment pour le réservoir Churchill Falls (1436,42) alors que pour le réservoir Gouin (176,31), l'autocorrélation est similaire à la valeur obtenue pour la méthode par rééchantillonnage.

Les diagrammes en boîte des figures 3.16 à 3.21, ont été obtenus de la même façon que pour le modèle $PAR(1)$. En analysant les figures 3.16 à 3.18, nous remarquons que l'étoile qui apparaît à chaque période m , $m = 1, \dots, 12$, qui constitue l'apport moyen observé est proche de la médiane des 100 moyennes simulées pour chaque période m et ceci pour les deux méthodes. En effet, l'étoile est à l'intérieur de l'intervalle interquartile, ce qui prouve qu'on reproduit assez bien les moyennes observées pour les réservoirs Caniapiscau, Churchill Falls et Gouin.

Pour les figures 3.19 à 3.21, nous remarquons que l'écart type observé est bien reproduit aussi bien pour la méthode de rééchantillonnage que la méthode transformée. Cependant pour cette dernière, on observe beaucoup de valeurs à

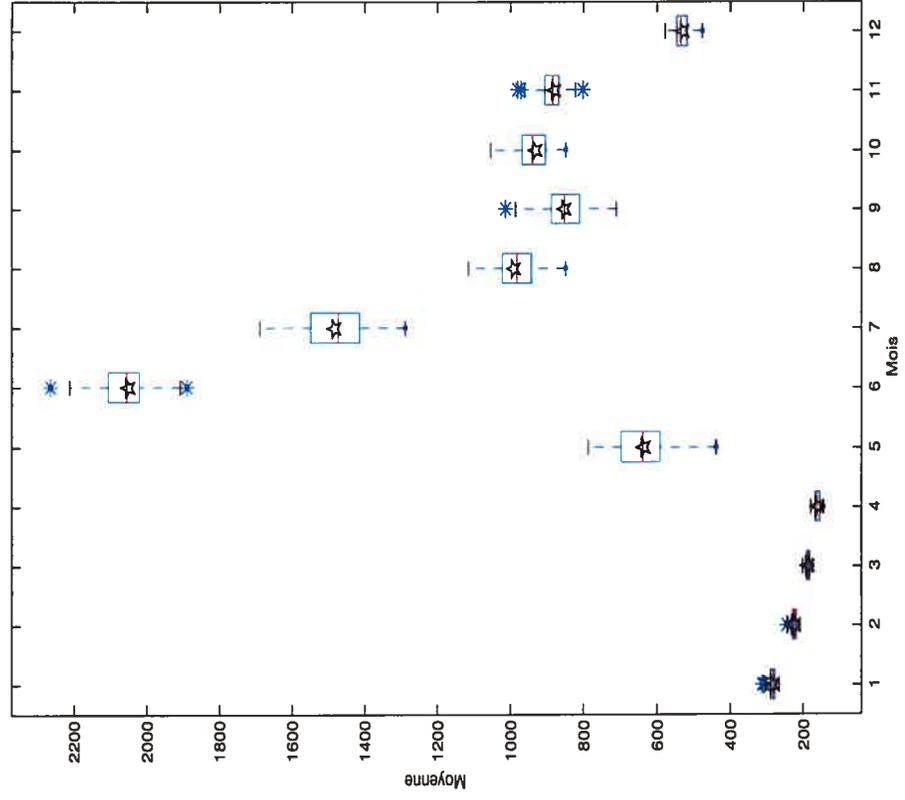
l'écart surtout pour les réservoirs Caniapiscau et Churchill Falls. On remarque aussi dans la figure 3.21 pour le réservoir Gouin, qu'il y a moins de valeurs à l'écart par rapport à ce qui a été observé dans la figure 3.9 du modèle $PAR(1)$ surtout dans le cas de la méthode transformée. On se basant sur ces résultats, on peut dire que le modèle $PARX$ reproduit aussi bien la moyenne et l'écart type que le modèle PAR et ceci pour les deux méthodes.

TABLEAU 3.11. Résultats comparatifs des statistiques annuelles observées et simulées d'un modèle *PARX*(1) pour les différents réservoirs d'Hydro-Québec.

Code	Sites	Historique		Transformée		Rééchantionnée	
		moyenne	écart type	moyenne	écart type	moyenne	écart type
1	Caniapiscau	768,03	131,93	771,05	140,40	770,44	138,78
2	Laforge	163,73	27,11	165,60	28,67	165,99	27,71
3	LaGrande4	556,24	91,64	559,91	99,20	560,30	90,21
4	EOL Total	850,32	118,54	855,63	123,94	859,63	125,48
5	Bersimis	318,10	40,30	322,41	47,84	319,04	43,73
6	Churchill Falls	1422,44	223,41	1436,42	214,05	1425,18	219,02
7	Manic 5	646,89	96,40	650,53	95,46	649,75	94,73
8	Lac Ste-Anne	191,63	25,12	193,11	28,06	192,62	27,79
9	Outardes 4	364,46	46,62	368,22	52,99	367,12	51,44
10	Ste-Marguerite	108,93	19,34	110,64	16,89	111,00	16,66
11	Gouin	175,26	29,15	176,31	35,12	176,18	34,61
12	Manouane	45,67	9,23	46,25	10,21	45,85	9,76
13	Rapide Blanc	167,82	32,32	169,34	35,78	167,98	34,79
14	Mattawin	73,88	13,74	74,46	15,21	73,85	14,90
15	TR-BE-LT2	173,88	28,26	177,63	28,00	174,19	26,04
16	GM-SH-Lagab	106,92	29,92	108,64	31,22	107,01	30,92
17	LG2-SAK	977,58	172,58	986,59	177,08	982,21	177,21

CANIAPISCAU

Moyenne mensuelle (séries rééchantillonnées)



CANIAPISCAU

Moyenne mensuelle (séries transformées)

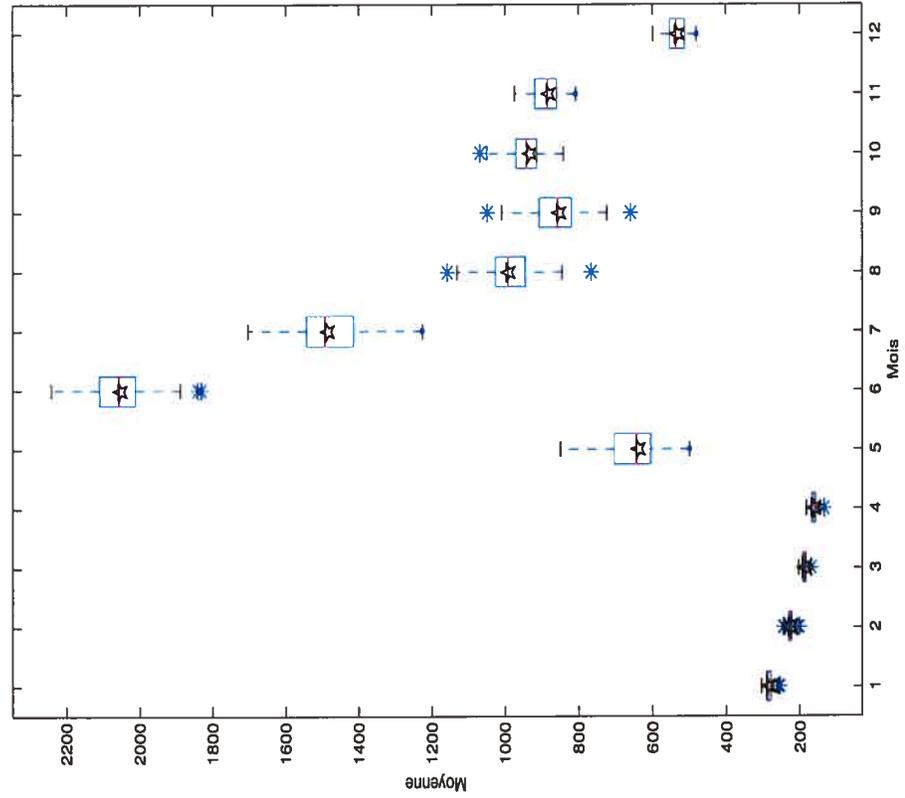
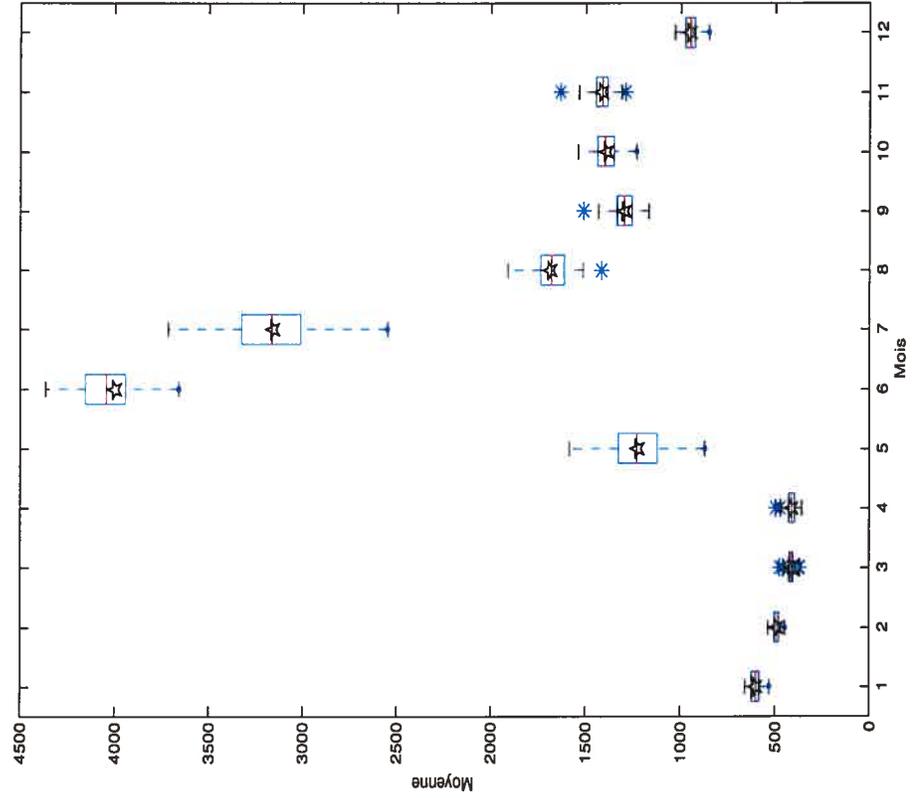


FIGURE 3.16. Moyenne mensuelle historique (★) au réservoir Caniapiscau et diagrammes en boîte des moyennes mensuelles de 100 séries générées avec un modèle *PARX*(1) suite à la transformation logarithmique ou suite au rééchantillonnage des résidus.

Churchill Falls
Moyenne mensuelle (series reechantillonnees)



Churchill Falls
Moyenne mensuelle (series transformees)

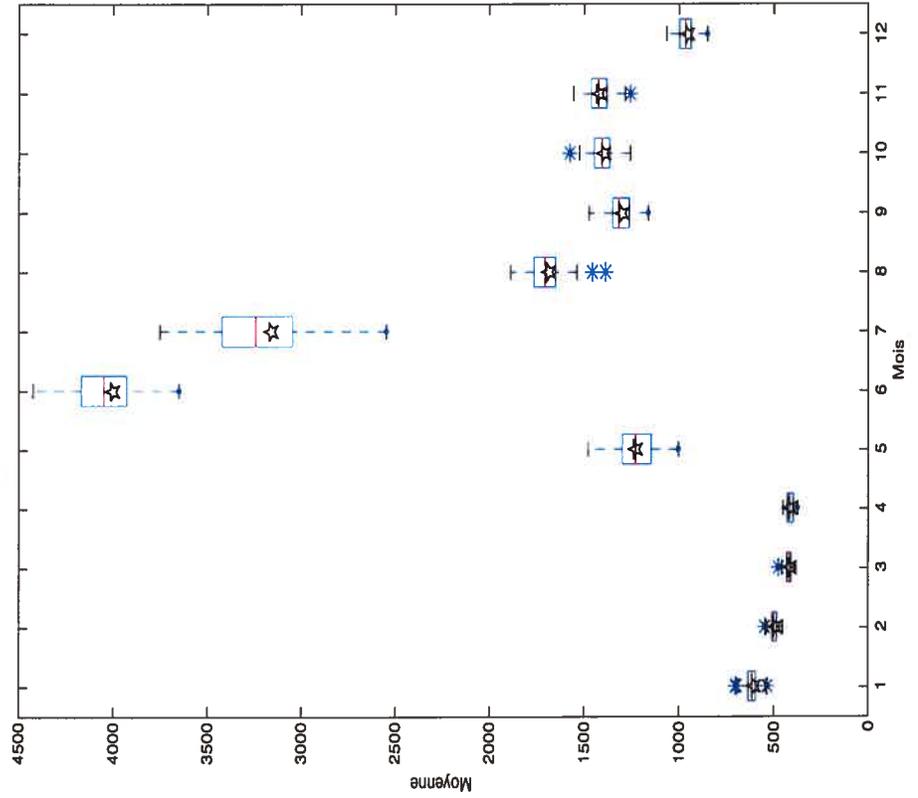


FIGURE 3.17. Moyenne mensuelle historique (★) au réservoir Churchill Falls et diagrammes en boîte des moyennes mensuelles de 100 séries générées avec un modèle *PARX*(1) suite à la transformation logarithmique ou suite au rééchantillonnage des résidus.

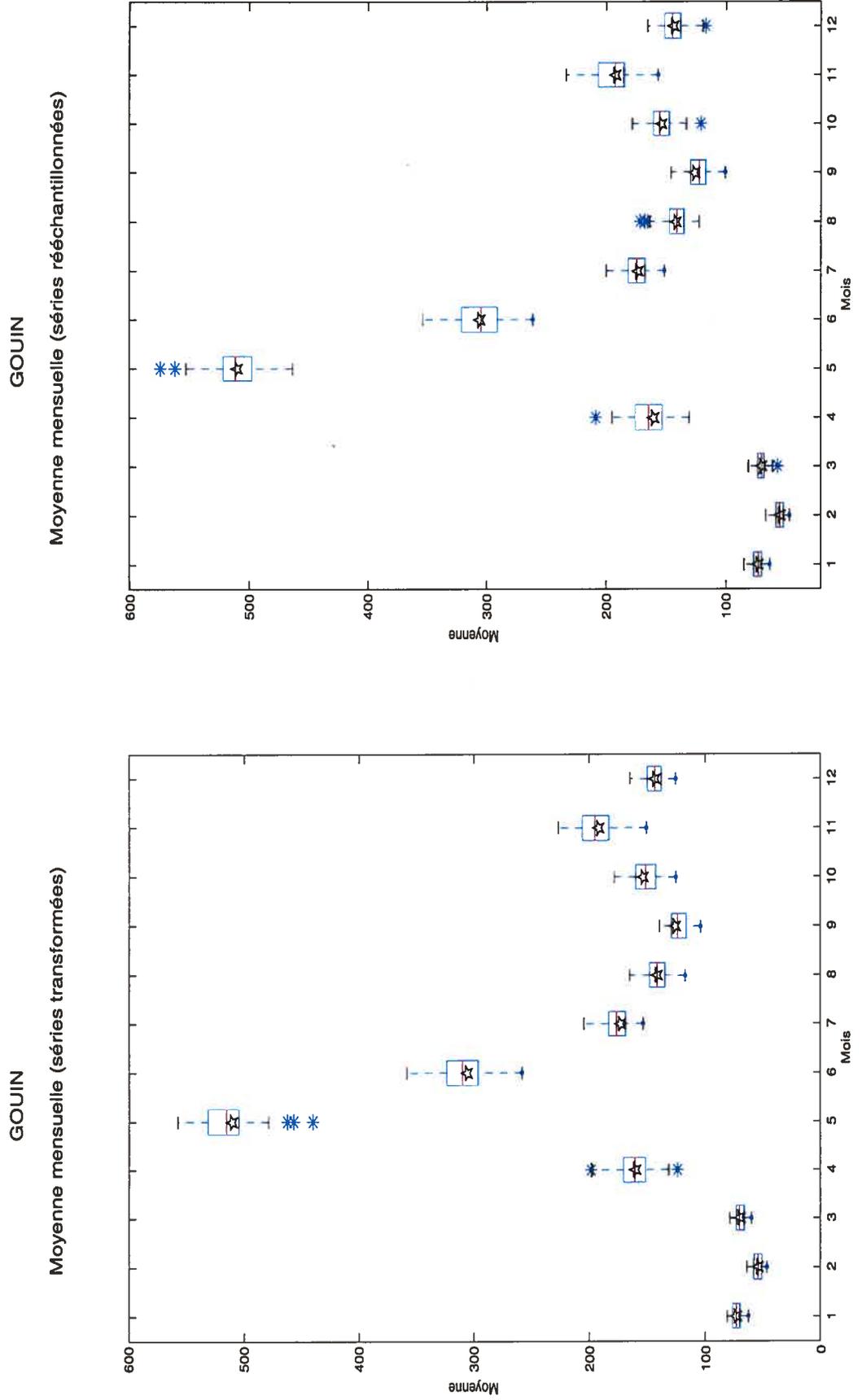
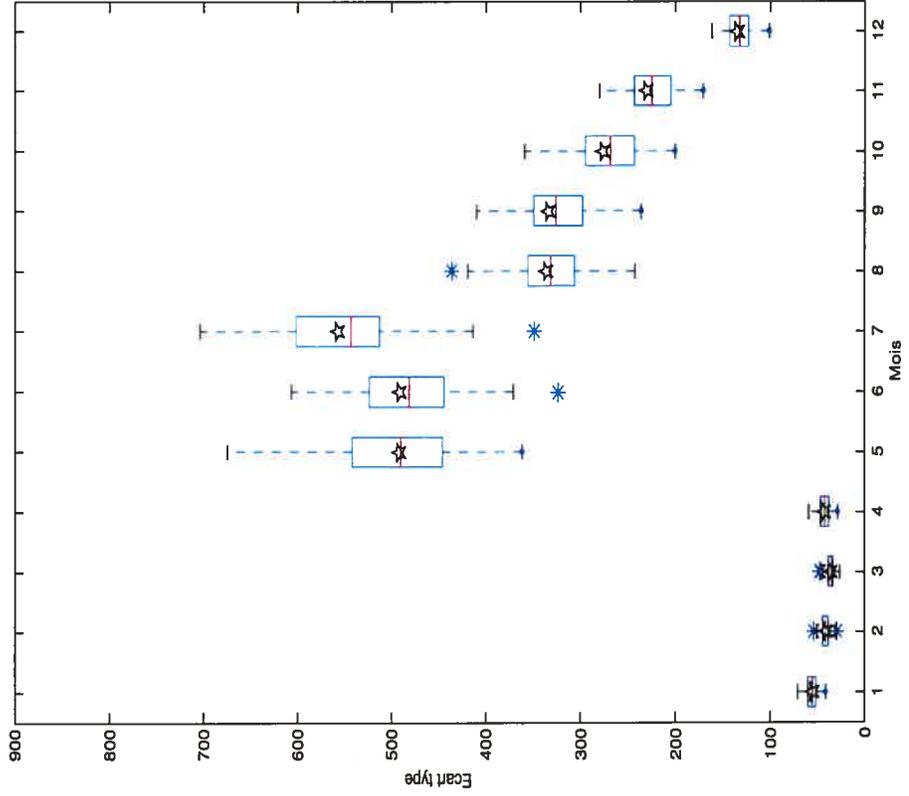


FIGURE 3.18. Moyenne mensuelle historique (★) au réservoir Goujin et diagrammes en boîte des moyennes mensuelles de 100 séries générées avec un modèle *PARX(1)* suite à la transformation logarithmique ou suite au rééchantillonnage des résidus.

CANIAPISCAU

Écart type mensuel (séries rééchantillonnées)



CANIAPISCAU

Écart type mensuel (séries transformées)

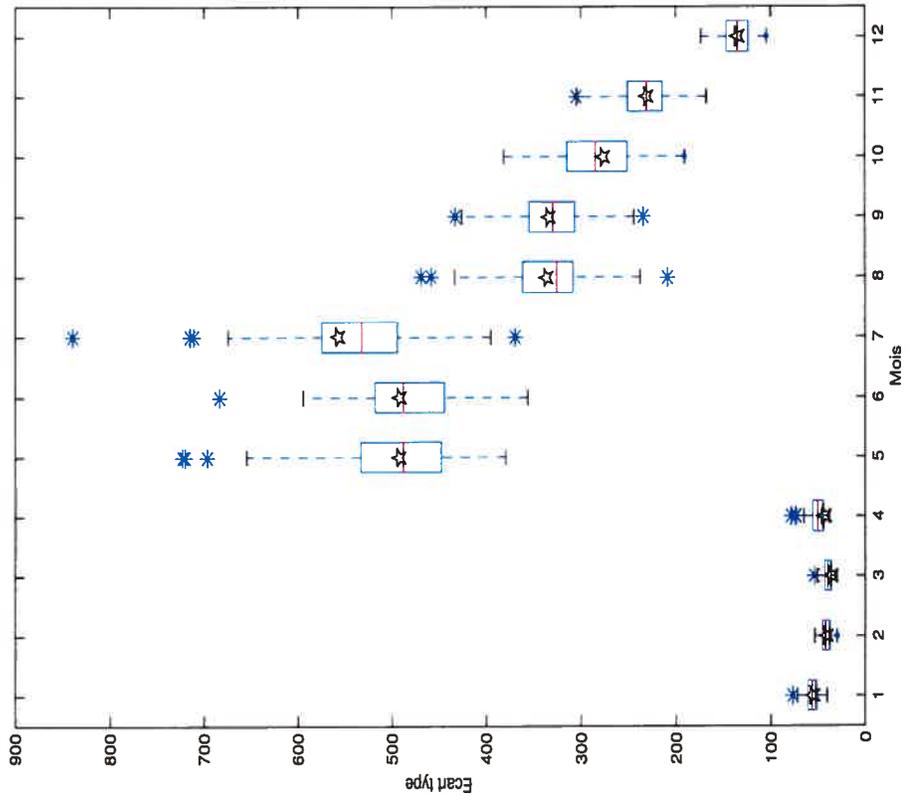


FIGURE 3.19. Écart type mensuel historique (★) au réservoir Caniapiscau et diagrammes en boîte des écarts types mensuels de 100 séries générées avec un modèle $PARX(1)$ suite à la transformation logarithmique ou suite au rééchantillonnage des résidus.

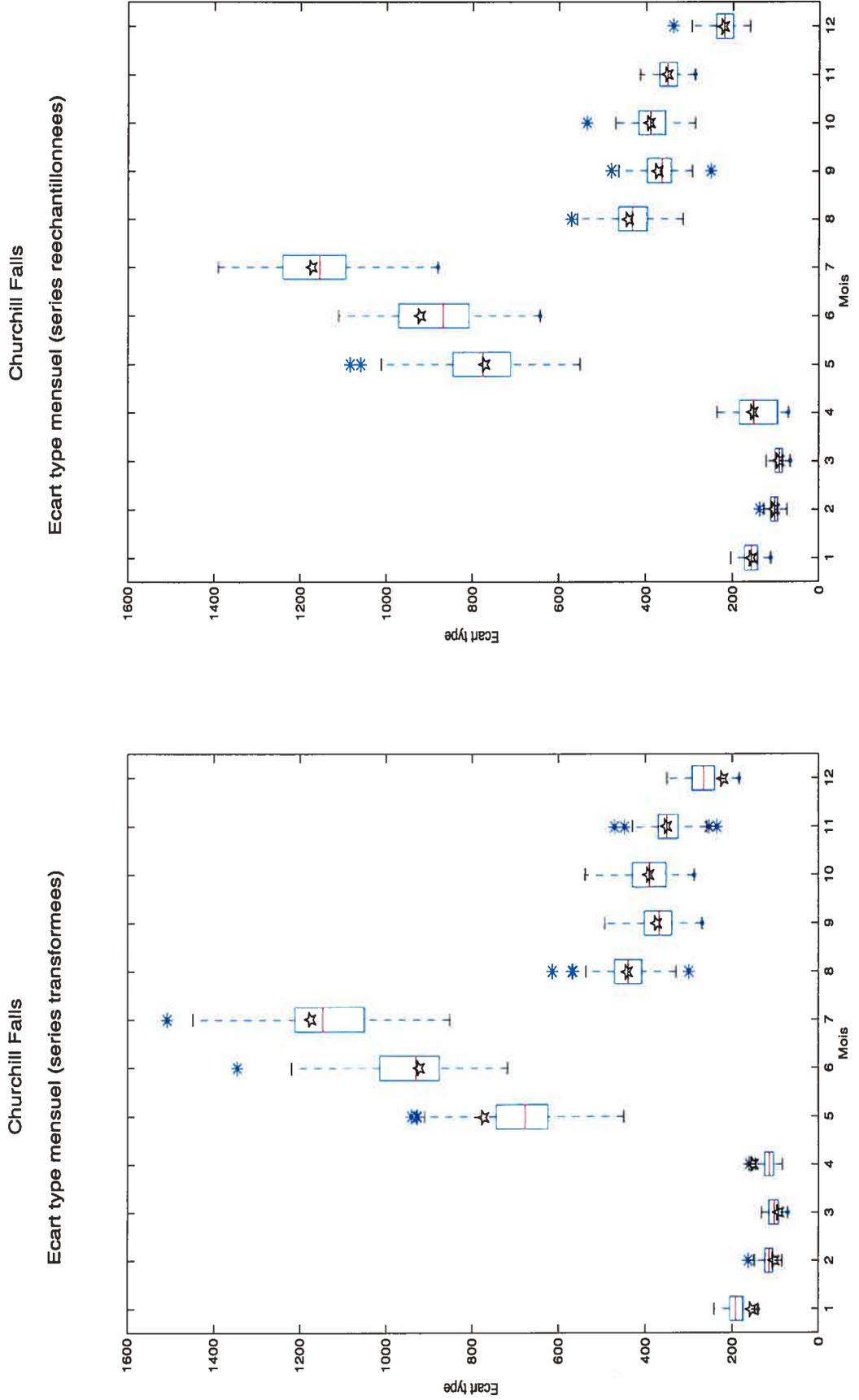


FIGURE 3.20. Écart type mensuel historique (★) au réservoir Churchill Falls et diagrammes en boîte des écarts types mensuels de 100 séries générées avec un modèle *PARX*(1) suite à la transformation logarithmique ou suite au rééchantillonnage des résidus.

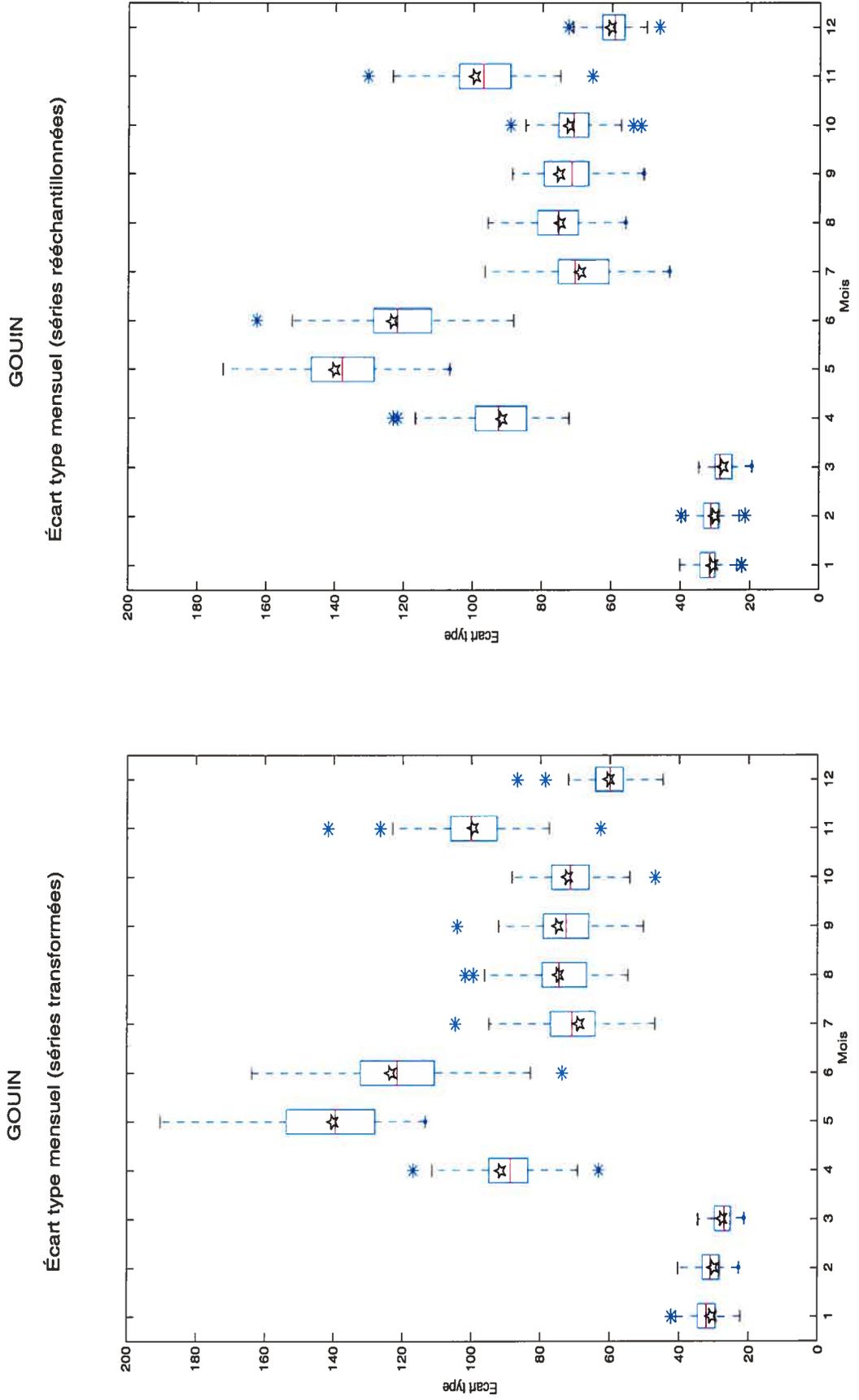


FIGURE 3.21. Écart type mensuel historique (★) au réservoir Gouin et diagrammes Gouin et diagrammes en boîte des écarts types mensuels de 100 séries générées avec un modèle $PARX(1)$ suite à la transformation logarithmique ou suite au rééchantillonnage des résidus.

-Autocorrélations mensuelles

Les résultats des autocorrélations périodiques mensuelles pour les 3 réservoirs dans le cas d'un modèle $PARX(1)$ sont présentés dans les figures 3.22 à 3.24. Les diagrammes en boîte de ces figures, ont été construits à partir de 100 scénarios, où pour chaque scénario k , l'autocorrélation de délai 1 définie par l'équation 1.3.1 du chapitre 1 est calculée avec les 42 années. Nous calculons l'autocorrélation mensuelle simulée de délai 1 pour chaque saison $m = 1, \dots, s$, c'est-à-dire $\rho_m^1, \rho_m^2, \dots, \rho_m^{100}$ et c'est à partir des 100 autocorrélations que nous construisons les diagrammes en boîte pour chaque période m . L'étoile (★) qui apparaît à chaque saison dans les figures 3.22 à 3.24 représente l'autocorrélation mensuelle observée.

Les autocorrélations mensuelles estimées ont été obtenues cette fois-ci par la méthode des moindres carrés. En examinant les figures 3.22 à 3.24, qui contiennent les diagrammes en boîte des paramètres estimés pour chaque période, nous remarquons dans le cas de la méthode par rééchantillonnage, que les autocorrélations mensuelles de délai 1 pour le modèle $PARX(1)$ sont très semblables à celles du modèle $PAR(1)$. En effet, on observe bien pour les 3 réservoirs que l'étoile qui constitue la vraie valeur est à l'intérieur de l'intervalle interquartile pour les différentes périodes m . Ce qui signifie qu'on reproduit assez bien l'autocorrélation mensuelle observée. Même la tendance qui veut que les autocorrélations décroissent lors de la crue printanière, surtout à partir du 5^e ou 6^e mois, est respectée. En analysant les figures 3.22 à 3.24, nous remarquons aussi pour la méthode transformée que l'étoile est à l'extérieur de l'intervalle interquartile surtout pour la période 3 et 5.

Concernant les autocorrélations de délais 2 et 3, il est difficile de dire si une méthode fait mieux que l'autre.

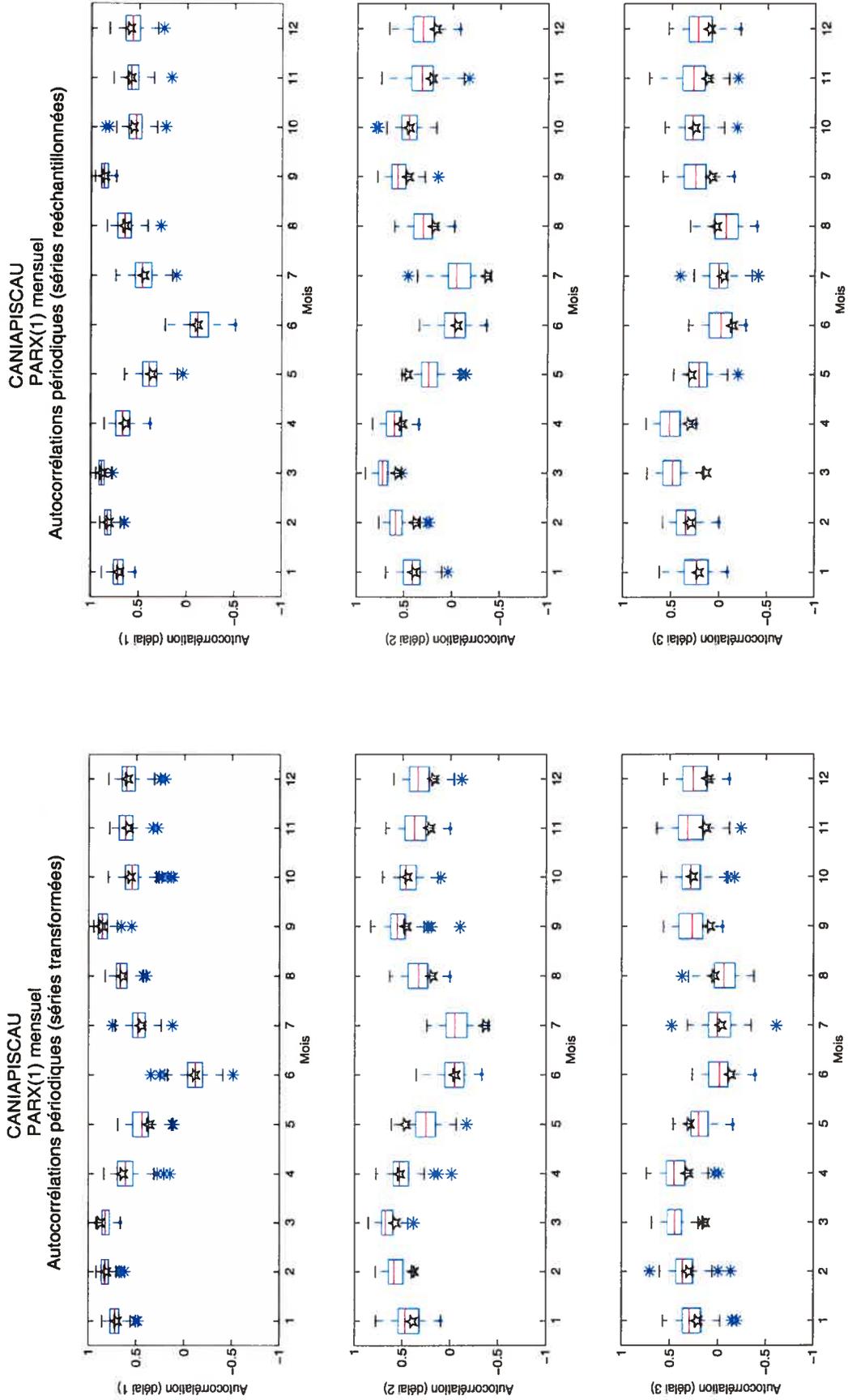


FIGURE 3.22. Autocorrélations périodiques au réservoir Caniapiscau et diagrammes en boîte des autocorrélations mensuelles de 100 séries générées avec un modèle *PARX(1)* suite à la une transformation logarithmique ou suite au rééchantillonnage des résidus.

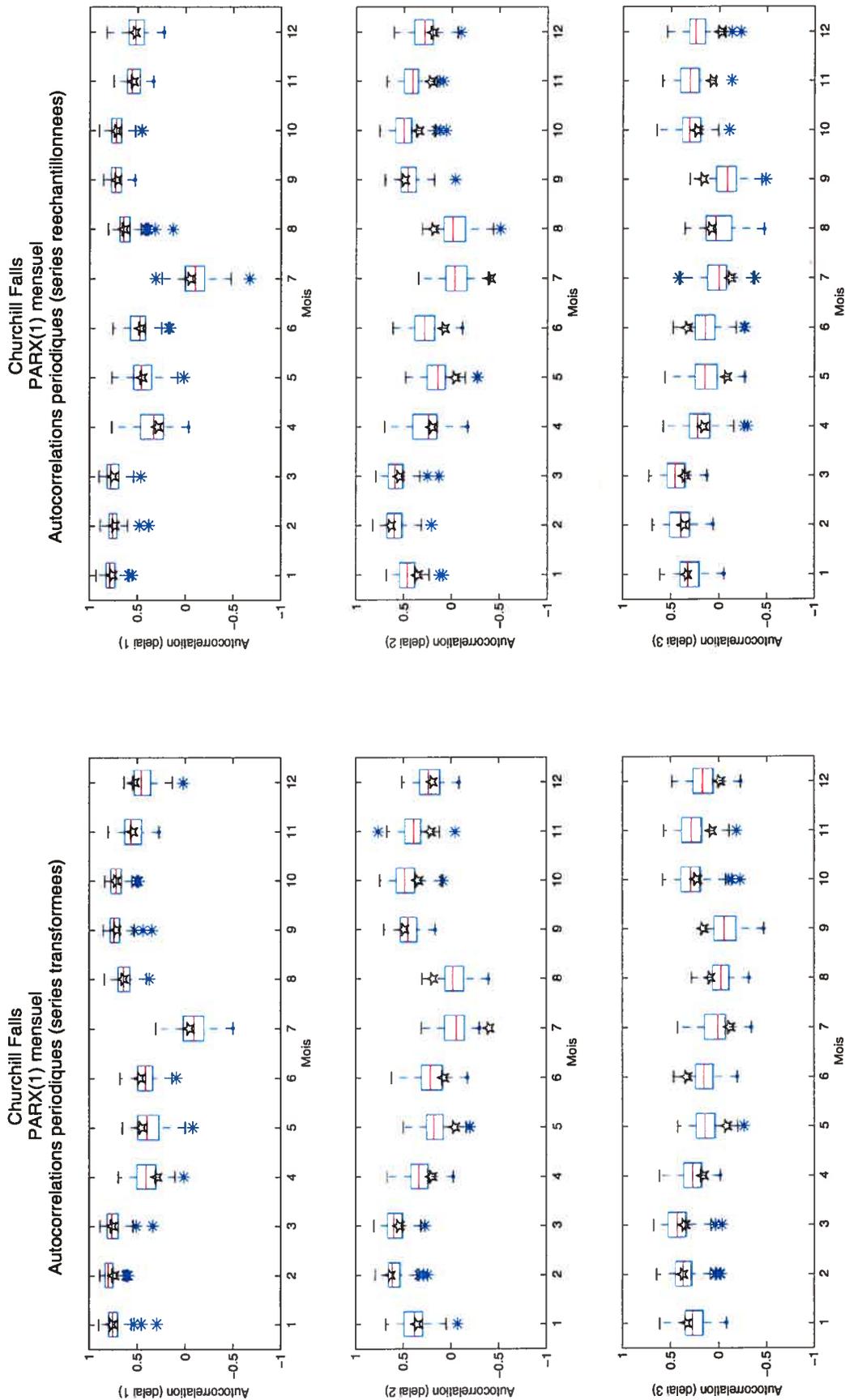


FIGURE 3.23. Autocorrelations périodiques au réservoir Churchill Falls et diagrammes en boîte des autocorrelations mensuelles de 100 séries générées avec un modèle $PARX(1)$ suite à la transformation logarithmique ou suite au rééchantillonnage des résidus.

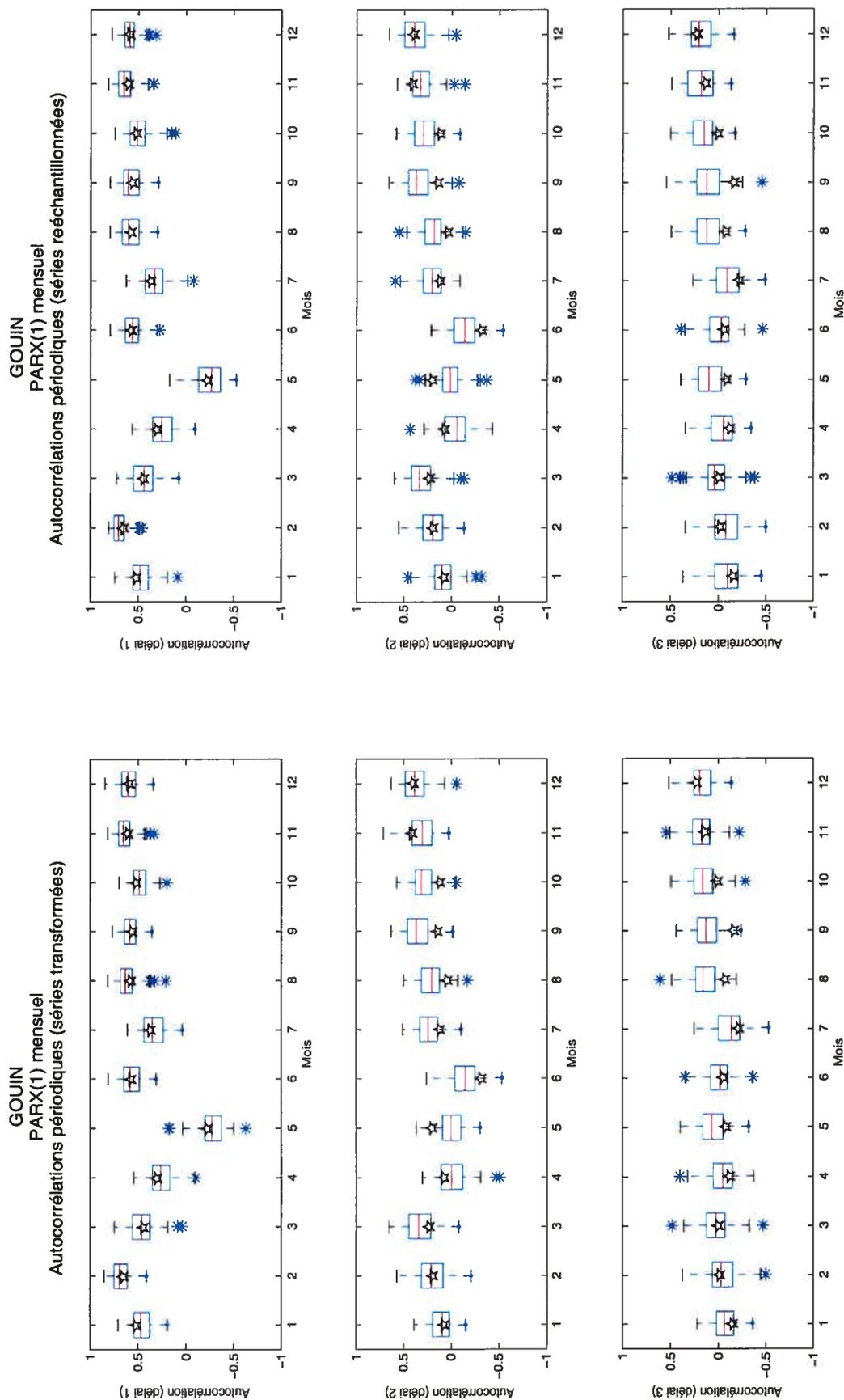


FIGURE 3.24. Autocorrélations périodiques au réservoir Gouin et diagrammes en boîte des autocorrélations mensuelles de 100 séries générées avec un modèle $PARX(1)$ suite à la transformation logarithmique ou suite au rééchantillonnage des résidus.

3.3.4. Discussion des résultats des corrélations interannuelles simulées avec un $PARX(1)$

L'un des objectifs de l'introduction d'une variable explicative dans le modèle $PARX(1)$ est de remédier aux insuffisances du modèle $PAR(1)$. En effet, avec ce type de modèle nous étions incapable de reproduire la corrélation interannuelle, parce qu'il n'y avait aucun terme qui mettait en évidence la dépendance interannuelle existante entre les apports. L'ajout d'une nouvelle variable exogène dans le modèle permet de prendre en compte la persistance interannuelle des apports de certains réservoirs. Le tableau 3.12 donne les résultats comparatifs des corrélations interannuelles pour les différents réservoirs obtenus par simulation. Sous l'hypothèse de bruit blanc fort, l'écart type est $\frac{1}{\sqrt{n}}$ avec $n = 42$ et les limites de signification au niveau 5% sont approximativement $\pm 0,30$. Les étoiles dans le tableau 3.12 identifient les valeurs significatives.

À partir de ce tableau, nous observons généralement que les corrélations interannuelles sont assez bien reproduites pour les deux méthodes, surtout pour le réservoir Churchill Falls. Si on compare avec les vraies valeurs, on remarque que la corrélation estimée pour le réservoir Caniapiscau est de 0,21 dans le cas de la méthode de rééchantillonnage et de 0,23 dans le cas de la méthode transformée alors que la vraie valeur est de 0,26. Même si cette valeur n'est pas significative, nous remarquons que les corrélations obtenues pour les deux méthodes sont proches de la corrélation interannuelle observée. Pour Churchill Falls avec une corrélation de 0,37 dans le cas de la méthode de rééchantillonnage et de 0,39 dans le cas de la méthode transformée, nous sommes assez près de la corrélation interannuelle observée (0,42) qui est significative. Concernant le réservoir Gouin, on remarque que les résultats sont proches de la valeur observée qui est non significative. En effet, nous observons une corrélation interannuelle négative estimée à $-0,15$ dans le cas de la méthode transformée et $-0,17$ dans le cas de la méthode de rééchantillonnage alors que la vraie valeur est $-0,14$.

TABLEAU 3.12. Résultats comparatifs des corrélations interannuelles observées et simulées d'un modèle *PARX*(1) pour les différents réservoirs d'Hydro-Québec.

Code	Sites	Historique	Transformée	Rééchantionnée
1	Caniapiscau	0,26	0,22	0,21
2	Laforge	0,33 *	0,28	0,26
3	LaGrande4	0,37 *	0,31 *	0,31 *
4	EOL Total	0,25	0,18	0,21
5	Bersimis	0,24	0,19	0,19
6	Churchill Falls	0,42 *	0,39 *	0,37 *
7	Manic 5	0,28	0,24	0,26
8	Lac Ste-Anne	0,10	0,08	0,07
9	Outardes 4	0,29	0,25	0,24
10	Ste-Marguerite	0,51 *	0,41 *	0,44 *
11	Gouin	-0,14	-0,15	-0,17
12	Manouane	-0,05	-0,04	-0,03
13	Rapide Blanc	-0,07	-0,07	-0,10
14	Mattawin	0,10	0,09	0,12
15	TR-BE-LT2	0,19	0,17	0,16
16	GM-SH-Lagab	0,13	0,12	0,10
17	LG2-SAK	0,31 *	0,27	0,26

3.3.5. Discussion des corrélations spatiales simulées avec un *PARX(1)*

Les principaux résultats des autocorrélations spatiales sont présentés dans les tableaux 3.13 et 3.15. Nous remarquons d'après ces tableaux, que la méthode de rééchantillonnage préserve mieux les corrélations spatiales existantes entre les différents réservoirs. En effet, si nous comparons les résultats des deux tableaux à ceux du tableau 1.3, nous remarquons que la corrélation entre Caniapiscau (1) et Churchill Falls (6) est de 0,45 dans le cas de la méthode transformée et de 0,64 dans le cas de la méthode de rééchantillonnage alors que la corrélation spatiale observée est de 0,83. Pour Caniapiscau (1) et Gouin (11), la corrélation est nulle avec la transformation et $-0,04$ avec le rééchantillonnage alors qu'elle est de 0,35 pour la corrélation observée. Concernant Churchill Falls (6) et Gouin (11), on remarque aussi que la corrélation est 0,02 pour la transformation et 0,04 pour le rééchantillonnage alors que la corrélation observée est 0,32.

Dans les tableaux 3.14 et 3.16, nous présentons les écarts types des corrélations spatiales pour les différents réservoirs obtenus par simulation.

Ainsi, à partir de ces résultats, nous pouvons dire que le modèle *PARX* ne fait pas mieux que le modèle *PAR* mais par contre la méthode de rééchantillonnage donne de meilleurs résultats que la transformation basée sur la loi lognormale.

TABLEAU 3.16. Écart types des corrélations spatiales pour les 17 sites obtenus avec un modèle *PARX*(1) et en rééchantillonnant les résidus.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	0,00	0,10	0,08	0,11	0,15	0,12	0,12	0,13	0,12	0,13	0,14	0,15	0,14	0,15	0,15	0,13	0,11
2		0,00	0,07	0,12	0,15	0,16	0,15	0,15	0,14	0,18	0,15	0,16	0,15	0,15	0,15	0,15	0,08
3			0,00	0,09	0,15	0,16	0,13	0,13	0,13	0,17	0,14	0,15	0,13	0,15	0,14	0,14	0,09
4				0,00	0,14	0,17	0,12	0,13	0,11	0,17	0,16	0,17	0,17	0,17	0,15	0,15	0,07
5					0,00	0,14	0,11	0,10	0,08	0,14	0,13	0,15	0,12	0,14	0,15	0,14	0,14
6						0,00	0,16	0,16	0,17	0,18	0,14	0,15	0,15	0,15	0,17	0,15	0,19
7							0,00	0,08	0,07	0,12	0,16	0,17	0,16	0,16	0,17	0,15	0,14
8								0,00	0,07	0,12	0,15	0,16	0,15	0,14	0,16	0,15	0,13
9									0,00	0,13	0,15	0,16	0,16	0,15	0,16	0,15	0,12
10										0,00	0,15	0,15	0,14	0,15	0,15	0,16	0,18
11											0,00	0,09	0,06	0,13	0,10	0,12	0,14
12												0,00	0,08	0,08	0,08	0,09	0,17
13													0,00	0,10	0,08	0,10	0,14
14														0,00	0,12	0,07	0,16
15															0,00	0,10	0,16
16																0,00	0,15
17																	0,00

CONCLUSION

Dans ce mémoire, nous nous sommes intéressés à l'analyse statistique et à la modélisation stochastique des séries chronologiques. Le but de notre étude était de construire des modèles stochastiques performants capables de modéliser les apports en eau à pas de temps inférieur à l'année pour les principaux réservoirs d'Hydro-Québec.

Dans un premier temps, nous avons exposé les caractéristiques statistiques des apports naturels. Nous avons observé à partir des hydrogrammes et des différentes figures que ces apports naturels présentaient une périodicité, des corrélations saisonnières importantes et une corrélation interannuelle et spatiale. Nous avons remarqué une augmentation importante du niveau des apports d'eau dans les différents réservoirs due surtout à la crue printanière. Cette modification des apports se traduit généralement par une diminution des autocorrélations périodiques. De plus, à partir des hydrogrammes observés, nous avons constaté que les pointes de crue sont variables d'une année à une autre. Cette caractéristique fait ressortir toute la difficulté entourant la modélisation de ce type de données.

Nous avons ensuite concentré nos efforts sur les modèles de séries chronologiques susceptibles de modéliser ces données. Nous avons conclu que ce type de données ne pouvaient être modélisées par des modèles *ARMA* classiques, du fait que ces derniers ne tiennent pas compte de la variation saisonnière présente dans les séries d'apports hebdomadaires et mensuelles. C'est ce qui nous a amené à utiliser les modèles *PARMA*. Ce type de modèles a été conçu spécifiquement pour prendre en compte la variation saisonnière de la fonction d'autocorrélation.

Par la suite, nous avons présenté les aspects théoriques des modèles *PAR* univariés et nous avons fait le lien avec les modèles *ARMA* multivariés. Nous avons discuté de l'estimation et de la validation de tels modèles. Pour tenir compte du fait que les apports en eau sont en général plus près d'une distribution lognormale que d'une distribution normale, il est d'usage en hydrologie de générer des bruits blancs gaussiens et d'appliquer la transformation exponentielle afin d'obtenir des bruits lognormaux. Une approche qui semble nouvelle en hydrologie consiste à rééchantillonner les résidus des modèles *PAR* estimés. Il s'agit d'une approche non paramétrique et en plus, elle est très simple à implémenter.

À l'aide de simulations, nous avons étudié de près le modèle *PAR*(1). Il se dégage que la méthode de rééchantillonnage donne de meilleurs résultats que la transformation reliée à la distribution lognormale car elle permet de mieux reproduire les hydrogrammes, les corrélations saisonnières et les corrélations spatiales. Cependant avec les modèles *PAR*(1), nous avons obtenu des résultats décevants en ce qui a trait à l'interdépendance annuelle, surtout que cette dépendance est bien présente dans plusieurs séries, notamment pour les réservoirs de la Côte-Nord (Churchill Falls (6), LaGrande 4 (3), ou Ste-Marguerite (10)). Nous avons estimé que l'une des causes vient du fait que dans les modèles *PAR*, il n'existe aucun terme pour mettre en évidence ce type de corrélation. C'est pourquoi, nous avons proposé un modèle plus complet en incorporant dans le modèle *PAR* une nouvelle variable explicative, qui n'est rien d'autre que l'apport de l'année précédente. Les résultats obtenus à partir des simulations avec les modèles *PARX* surpassent ceux obtenus avec le modèle *PAR*(1). En effet, en plus de reproduire les caractéristiques statistiques de base, la corrélation interannuelle est maintenant reproduite de façon satisfaisante.

Le champ d'application des modèles périodiques peut être élargi aux données journalières. En effet, même si les modèles pluie - débit les plus couramment utilisés ont été développés avec des données hebdomadaire ou mensuelles, il y a

actuellement un besoin de poursuivre des recherches afin d'obtenir des modèles parcimonieux, c'est-à-dire qui comportent peu de paramètres. Aussi, de nouvelles perspectives sont ouvertes pour la modélisation des apports surtout en utilisant l'estimation bayésienne. Cette partie fait l'objet de travaux en cours par Angers, Merleau et Perreault (2004).

ANNEXE : CODE MATLAB

Vu la taille des fichiers, nous présentons seulement les programmes d'estimation et de simulation d'un modèle $PAR(1)$ lorsque les résidus sont transformés. Les autres programmes sont disponible sur la disquette ci-jointe.

3.4. PROGRAMMES DES SIMULATIONS AVEC UN MODÈLE $PAR(1)$ POUR LA MÉTHODE TRANSFORMÉE

3.4.1. Programme principal de modélisation des apports

```
*****
*Définition des 17 sites selon leurs codes
*****
NomSITES = char('0B00031','0B00032', '0B00033','0B00068',
'0B00457','0B00030', '0B00024','0B00026','0B00021', ...
'0B00335', '0B00013','0B00014','0B00015','0B00017','0B00454',
'0B00455','0B00456');
*****
*Lecture et traitement de la série
*****
FichierESPHebdo = 'fichier de données';
annee = 42;
NbPeriode = 52;
[NbreStation  a]= size(NomSITES);
nbresemaines = 52;
nbremois = 12;
```

```

NomSite = NomSITES;
*****
*Appel de la fonction 'Gestionhebdo' pour la mise en forme      *
*des séries mensuelles                                          *
*****
EspHebdo = GestionHebdo(FichierESPHebdo,NomSite,nbresemaines,annee);
nscenarios = 100;
nyears = 42;
p = 1;
for k = 1 :NbreStation
    x = EspHebdo(k).histose;
*****
*Appel de la fonction Agregat                                  *
*****
    g=agregat(x);
    Visualisateur(k).Q = g;
    Visualisateur(k).an-min =min(EspHebdo(k).annee);
    Visualisateur(k).an-max = max(EspHebdo(k).annee);
    Visualisateur(k).id = NomSite(k, :);
    Visualisateur(k).nom = EspHebdo(k).nom;
    Visualisateur(k).aire = NaN;
    Visualisateur(k).titre='PAR(1)mensuel';
*****
*Appel de la fonction Standard                                *
*****
    [z,Moy(k, :),Std(k, :),Stat(k, :, :)] = Standard(g);
*****
*Appel de la fonction moment                                  *
*****
    PHI(k, :, :) = Moment(z,p);
*****

```

```

* Appel de la fonction résidus *
*****
R = Residual(z,(PHI(k, :, :)),[]);
*****
* Appel de la fonction transforme *
*****
for i = 1 :nbremois
[y( :,i), yStand( :,i), alpha, beta, m(k,i), my(k,i), sy(k,i)] = Transforme(R( :,i));
end
ResidTransStand(k, :, :) = yStand;
end
*****
* Appel de la fonction Covar *
*****
for i = 1 :nbremois
[At( :, i, :),Covt( :, i, :)] = Covar(ResidTransStand( :, :,i)');
end
*****
* Appel de la fonction Generatrice *
*****
[Q]=Generatrice(At,sy,my,m,PHI,Moy,Std,nyears,nscenarios,NbreStation,nbremois,p);

*****
*Visualisateur des données générées *
*****
for i=1 :NbreStation
for j=1 : nscenarios
a = squeeze(Q(j, i, :));
QQ( :,j, :) = reshape(a,nbremois,nyears)';
end
Visualisateur(i).MQ = QQ;

```

```

    for ii=1 :nscenarios
        for jj=1 :nyears
            H(jj, i, ii)=sum(QQ(jj, ii, :));
        end
    end
end
close(h)

```

3.4.2. Les fonctions

```

*****
*Fonction 'Gestionhebdo' permettant
*la mise en forme des séries hebdomadaires
*****
function [SerieHebdo] = GestionHebdo(data, NomSite, nsemaines, annee);
[nsite a] = size(NomSite);
y = textread(data,'s','delimiter',' ');
l = length(y);
x = reshape(y, nsemaines+3, l/(nsemaines+3))';
Bassin = x( :, 1);
Cle = x( :, 2);
Annee = str2double(x( :, 3));
ESP-hebdo = str2double(x( :, 4 :end));
h = waitbar(0,'Mise en forme des séries ESP hebdo de PDA')
for i=1 :nsite
    waitbar(i/nsite,h);
    Ind = strmatch(NomSite(i, :),Cle);
    Datahebdo = ESP-hebdo(Ind, :);
    Heannee = Année(Ind);
    SerieHebdo(i).id = NomSite(i, :);
    SerieHebdo(i).nom = Bassin(annee*(i-1)+1);
    SerieHebdo(i).annee = Heannee;

```

```

SerieHebdo(i).histose = Datahebdo;
SerieHebdo(i).aire = [];
end
close(h)
*****
*Fonction 'Agregat' permettant d'agrégér les données
* hebdomadaires en données mensuelles*
*****
function [xag] = Agregat(x)
for i=1 :42
for j=1 :11
xag(i,j)=sum(x(i,1+4*(j-1) :4*j))/4;
end
xag(i,12)=sum(x(i,45 :52));
end
*****
*Fonction 'Standard' permettant de standardiser les données
*et de calculer les moyennes et les écarts types périodiques.
*****
function [xStand,mx,sx,Stat] =Standard(x)
[n p] = size(x);
mx = mean(x);
sx = std(x);
xStand = (x - ones(n,1)*mx)./(ones(n,1)*sx);
Stat = [mx;sx];
*****
*Fonction 'Moment' permettant d'estimer les paramètres autorégressifs
*****
function [phi] = Moment(x,p)
[ny np]=size(x);
if p>0

```

```

m = pacfat(x,p);
phi = [ ];
for t=1 :np
clear a;
for k=1 : p
for i=1 : p
ix = abs(i-k);
jx = max(t-i, t-k);
if jx<=0
jx = jx + np * (1+fix(jx/np));
end
if ix>0
a(k, i) = m(ix, jx);
else
a(k, i) = 1;
end
end
y(k, 1) = m(k, t);
end
phi = [phi inv(a)*y];
end end;

*****
*Fonction 'Residual' permettant de calculer les résidus
*****

function e = Residual(x, phit, tht)
[ny np] = size(x);
p = size(phit, 2);
q = size(tht, 1);
if p==0, phit=zeros(1, np);
end
if q==0, tht=zeros(1, np);

```

```

end
if p==1
phi=squeeze(phit)';
else
phi=squeeze(phit);
end
x = reshape(x', 1, ny*np);
ndata = length(x);
e=zeros(1, ndata);
for i=p+1 :q
t=rem(i, np); if t==0,t=np;
end
e(i) = x(i) - sum( phi(1 :p, t)'.*x(i-1 :-1 :i-p) );
end
for i=max(p, q)+1 :ndata
t=rem(i, np); if t==0,t=np;
end
e(i) = x(i) - sum( phi(1 :p, t)'.*x(i-1 :-1 :i-p) ) ...
+ sum( tht(1 :q, t)'.*e(i-1 :-1 :i-q) );
end;
e = reshape(e,np,ny)';
*****
*Fonction 'Transforme' permettant de transformer les
*résidus (transformation lognormale des résidus)
*****
function [y, yStand, alpha, beta, m, my, sy] = Transforme(x)
n = length(x);
mx = mean(x);
sx = std(x);
Cs = (n/((n - 1) * (n - 2))) * sum((x - mx)^3)/(sx^3);
omega1 = 1 + .5. * Cs^2;

```

```

omega2 = sqrt(Cs^2 + .25.*Cs^4);
theta = (omega1+omega2)^(1/3)+(omega1-omega2)^(1/3)-1;
alpha = log(sx/(theta*(theta-1)));
beta = sqrt(log(theta));
m = mx - sx./sqrt(theta - 1);
if min(x)<m
m=min(x)-1;
end;
y = log(x-m);
my = mean(y);
sy = std(y);
yStand = (y-my)/sy;
*****
*Fonction 'Covar' permettant de calculer la matrice de covariance
*****
function [B, Sigmat]=Covar(x)
R=x;
Sigmat=cov(R);
[eval eval]=eig(Sigmat);
B=eval*sqrt(eval);
*****
*Fonction 'Generatrice'. Programme permettant la génération des apports
*****
function [Q]=Generatrice(At,se,me,m,Phi,MuPeriode,SigPeriode,...
nyears,nscenarios,NbreStation,NbPeriode,p)
for i=1 :NbPeriode
vt = squeeze(At( :, i, :))*randn(NbreStation,1);
Et( :,i) = exp( vt.*se( :, i) + me( :, i) ) + m( :, i);
end
x=zeros(NbreStation, 2*NbPeriode);
for j=1 : NbreStation

```

```

for i=(p+1) :NbPeriode
x(j,i) = squeeze(Phi(j, :, i))*x(j, i-1 :-1 :i-p)' + Et(j, i);
end;
end
for ks=1 : nscenarios
ks
for ky=1 : nyears
for i=1 : NbPeriode
vt = squeeze(At( :, i, :))*randn(NbreStation,1);
Et( :, NbPeriode+i) = exp( vt.*se( :, i) + me( :, i) ) + m( :,i);
end
for j=1 : NbreStation
for i=NbPeriode+1 :2*NbPeriode
x(j,i) = squeeze(Phi(j, :,i-NbPeriode))*x(j,i-1 :-1 :i-p)' + Et(j,i);
end;
Q(ks,j,NbPeriode*(ky-1)+1 :ky*NbPeriode) = x(j,NbPeriode+1 :2*NbPeriode)...
.*SigPeriode(j, :) + MuPeriode(j, :);
end
x(1 :NbreStation, 1 :NbPeriode)=x(1 :NbreStation, NbPeriode+1 :2*NbPeriode);
end
end
Qsimul(j).sim=Q( :, :, :);

```

BIBLIOGRAPHIE

ABRAHAM, B. et LEDOLTER, J. (1983). *Statistical Methods for Forecasting*. John Wiley and Sons, Inc., New York.

ANGERS, J.F, MERLEAU, J. et PERREAULT, L. (2004). Landmark registration of hydrographs and Bayesian estimation of a mean hydrograph, *Proceedings of the International Sri Lankan Statistical Conference*. eds : deSilva, BM. et Mukhopadhyay, N.

BASAWA, I.V, LUND, R. et QIN SHAO. (2004). First-order seasonal autoregressive processes with periodically varying parameters. *Statistics and Probability Letters*, **67**, 299-306.

BASAWA, I.V. et LUND, R. (2001). Large sample properties of parameter estimates for periodic ARMA models. *Journal of Time Series Analysis*, **22**, 651-663.

BROCKWELL, P.J. et DAVIS, R.A. (1991). *Time Series : Theory and Methods*. 2^e éd. Springer-Verlag, New York.

BROCKWELL, P.J. et DAVIS, R.A. (2002). *Introduction to Time Series and Forecasting*. 2^e éd. Springer-Verlag, New York.

BOX, G.E.P. et JENKINS, G.M. (1970). *Time Series Analysis : Forecasting and Control*. San Francisco : Holden-Day.

BOX, G.E.P, JENKINS, G.M. et REINSEL, G.C. (1994). *Time Series Analysis Forecasting and Control*. 3^e éd. Englewood Cliffs, New Jersey : Prentice Hall.

DAVIS, N., TRIGGS, C.M. et NEWDOLD, P. (1977). Significance levels of the Box-Pierce portmanteau statistic in finite samples. *Biometrika*, **64**, 517-22.

EFRON, B. (1979). Bootstrap methods : another look at the jackknife. *Annals of statistics*, **7**, 1-26.

- EFRON, B. et TIBSHIRANI, R.J. (1993). *An Introduction to the Bootstrap*. Chapman and Hall.
- FRANSES, P.H et PAAP, R. (2005). *Periodic Time Series Models*. Oxford University Press.
- HIPEL, K.W. et MCLEOD.A.I.(1994). *Time Series Modelling of Water Resources and Environmental Systems*. Elsevier Science B.V, Amsterdam.
- HURD, H.L. (1989). Representation of strongly harmonizable periodically correlated processes and their covariances. *Journal of Multivariate Analysis*, **29**, 53-67.
- GARDNER, W. et FRANKS, L.E.(1975). Characterization of cyclostationary random signal processes. *IEEE Transactions on Information Theory*, **21**, 4-14.
- GLADYSHEV, E.G. (1961). Periodically Correlated random Sequences. *Soviet Mathematics*, **2**, 385-388.
- KOTTEGODA, N.T. (1980). Stochastic Water Resources Technology. Department of Civil Engineerig, University of Birmingham.
- LJUNG, G.M. et BOX, G.E.P. (1978). On a mesure of lack of fit in Time Series Models. *Biometrika*, **65**, 297-303.
- LOUCKS, D.P, STEDINGER, J.R. et HAITH, D.A. (1981). Water Resource Systems Planning and Analysis. Cornell University, Prentice-Hall Inc, New Jersey.
- LUND, R. et BASAWA, I.V. (2000). Recursive prediction and likelihood evaluation for periodic ARMA models. *Journal of Time Series Analysis*, **20**, 75-93.
- MATALAS, N.C. (1967). Mathematical assessment of synthetic hydrology. *Water Resources Research*, **4**, 937-945.
- MONIN, A.S. (1963). Stationnary and periodic time series in the general circulation of the atmosphere. In : Proc. Symp. on Time Series Analysis, Rosenblatt, M. Ed, Wiley, New York, pp.144-151.
- PAGANO, M. (1978). On periodic and multiple autoregressions. *Annals of Statistics*, **6**, 1310-1317.
- PERREAULT, L. et LATRAVERSE, M. (2001). Modélisation des apports naturels pour la prise en compte de leur aléa dans la méthode SDDP de planification de la production. Rapport technique, Hydro-Québec.

PROTHERO, D.L. et WALLIS, K.F. (1976). Modelling macroeconomic time series (with discussion). *Journal of the Royal Statistical Society, A*, 139, 468-500.

ROY, R. (1980). Analyse et prévision de séries chronologiques. *Département d'informatique et de recherche opérationnelle*, Université de Montréal.

SALAS, J.D, TABIOS, G.Q. et BARTOLINI, P. (1980). Approaches to multivariate modelling of water resources time series. *Water Resources Bulletin*, 21, 683-708.

SHAO, Q. et LUND, R. (2004). Computation and characterization of autocorrelations and partial autocorrelations in periodic ARMA models. *Journal of Time Series Analysis*, 25, 1-14.

STEDINGER, J.R. (1980). Fitting log normal distributions to hydrologic data. *Water Resources Research*, 16, 481-490.

THOMAS, H.A, et FIERING, M.B. (1962). Mathematical synthesis of streamflow sequences for the analysis of river basins by simulation. In : *Design of Water Resource Systems*. Mass, A., Hufschmidt, M.M., Dorfman, R., Thomas, H., Marglin, S.A, and Fair, G.M. (Editors). Harvard University, Cambridge, MA, 459-493.

VECCHIA, A.V. (1985). Periodic autoregressive-moving average (PARMA) modelling with applications to water resources. *Water Resources Research*, 21, 721-730.