POLYPUBLIE

Génération d'arbres de scénarios d'apports d'eau

Khadija Dhouib

Document en libre accès dans PolyPublie
Open Access document in PolyPublie

Date	2008
Type	Mémoire ou thèse / Dissertation or Thesis Mémoire de maîtrise/Master's thesis
URL de PolyPublie PolyPublie URL	https://publications.polymtl.ca/8210/
Conditions d'utilisation Terms of use	© Tous droits réservés / All rights reserved
Référence Citation	Dhouib, K. (2008). Génération d'arbres de scénarios d'apports d'eau [Mémoire de maîtrise, École Polytechnique de Montréal]. PolyPublie. https://publications.polymtl.ca/8210/

Détails de la thèse
Thesis details

Directeurs de recherche Advisors	André Turgeon
Programme Program	Non spécifié

Ce fichier a été téléchargé à partir de PolyPublie, le dépôt institutionnel de Polytechnique Montréal
This file has been downloaded from PolyPublie, the institutional repository of Polytechnique Montréal
https://publications.polymtl.ca

UNIVERSITÉ DE MONTRÉAL

GÉNÉRATION D'ARBRES DE SCÉNARIOS D'APPORTS D'EAU

KHADIJA DHOUIB
DÉPARTEMENT DE MATHÉMATIQUES ET DE GÉNIE INDUSTRIEL
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION
DU DIPLÔME DE MAÎTRISE EN MATHÉMATIQUES APPLIQUÉES
(MATHÉMATIQUES APPLIQUÉES)
DÉCEMBRE 2008

© Khadija Dhouib, 2008.

Library and Archives Canada

Published Heritage Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et Archives Canada

Direction du Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-48912-3
Our file Notre référence
ISBN: 978-0-494-48912-3

NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Ce mémoire intitulé :

GÉNÉRATION D'ARBRES DE SCÉNARIOS D'APPORTS D'EAU

présenté par : DHOUIB Khadija

en vue de l'obtention du diplôme de : Maîtrise ès sciences appliquées

a été dûment accepté par le jury d'examen constitué de :

M. ADJENGUE Luc-Désiré, Ph.D., président

M. TURGEON André, Ph.D., membre et directeur de recherche

M. LABIB Richard, ing., Ph.D., membre

À ma fille Asmaa

REMERCIEMENTS

Tout d'abord, je voudrais remercier mon directeur de recherche, le professeur André Turgeon, pour m'avoir admise comme étudiante à la maîtrise et m'avoir fait confiance et soutenue durant mes années d'études. Mes remerciements vont aussi à Stéphane Krau qui m'a guidée tout au long de mes travaux et qui a toujours été compréhensif et ouvert. Je veux aussi remercier les membres du jury, messieurs Adjengue et Labib, d'avoir accepté d'évaluer ce travail.

Je ne voudrais pas oublier les membres de ma famille : toute ma gratitude va vers mes parents et mes soeurs qui m'ont soutenue durant toutes mes études et mon mari, Abderrezak, qui a été d'une aide inestimable durant la rédaction de ce mémoire. Finalement, je n'oublierai pas de remercier mes amis Imed Laouini, Alain Régnier, Mounira Groiez, Amira Dems et Shadi SharifAzadeh avec qui j'ai partagé les plus beaux moments de ma maîtrise.

RÉSUMÉ

Ce mémoire porte sur la modélisation des apports d'eau à des réservoirs alimentant des centrales hydroélectriques. Cette modélisation n'est pas évidente en raison du fait que les apports d'eau sont aléatoires et corrélés dans le temps et dans l'espace. Une telle modélisation est nécessaire pour bien représenter le comportement aléatoire des apports d'eau dans les modèles d'optimisation qui déterminent les règles de gestion des installations hydroélectriques.

Le problème d'optimiser la gestion des installations hydroélectriques est non convexe et stochastique. Ce problème est difficile à solutionner lorsqu'il y a plusieurs réservoirs dans le système. En fait, il ne peut pas être résolu dans un temps raisonnable pour un système comprenant plus de trois réservoirs. Des problèmes de beaucoup plus grande taille peuvent cependant être résolus lorsque le nombre de pas de temps pour lequel la règle de gestion doit être déterminée d'avance est très petit (3 ou 4 périodes). Dans ce cas, il suffit de représenter les aléas des apports par des arbres de scénarios puis de faire appel à des logiciels de programmation linéaire ou non linéaire pour résoudre le problème d'optimisation. Ce mémoire porte exclusivement sur la construction d'arbres de scénarios d'apports.

Nous disposons de séries historiques d'apports provenant de différents sites sur la rivière la Gatineau. Le modèle Periodic-Auto-Regressive d'ordre 2 est choisi pour modéliser la corrélation temporelle et l'analyse en composantes principales est retenue pour transformer les variables corrélées en variables indépendantes.

La méthode des moments et les méthodes de simulations (parallèle et séquentielle), qui se basent sur la transformation cubique, ont été utilisées pour générer des arbres ayant une seule variable. Deux approches sont proposées pour générer des arbres ayant plusieurs variables. La première utilise les résultats de l'analyse en composantes principales. Elle

génère d'abord des arbres indépendants de composantes principales et les recombine pour obtenir un arbre ayant plusieurs variables mutuellement corrélés. La deuxième approche recourt à l'algorithme proposé par Høyland, Kaut et Wallace [10], qui est basé sur la décomposition de Cholesky. L'évaluation de ces différentes méthodes se fait en comparant les propriétés statistiques (les quatre premiers moments et la matrice de corrélation) de l'historique avec celles des variables générés.

Les résultats révèlent que la méthode des moments est meilleure que la méthode de simulation, surtout que cette dernière nécessite une classification qui détériore les résultats. La méthode de Høyland, Kaut et Wallace [10] est nettement supérieure à celle qui utilise les composantes principales et reproduit mieux les moments et la matrice de corrélation. Mais ses résultats se dégradent quand on utilise la classification.

ABSTRACT

The aim of this thesis is to present a procedure to generate inflow scenario trees. The inflows to the reservoirs in a river are spatially and temporally correlated. These properties are used to construct scenario trees. By using Periodic-Auto-Regressive model, the temporal correlation is explained and Principal Component Analysis is used to model spatial correlation and reduce the dimension of variables.

Moment matching method is used to generate a limited number of discrete outcomes that satisfy specified statistical properties (four first moments). The basic idea is to minimize the distance between the statistical properties of the generated outcomes and the specified properties. The second method is based on simulations generated sequentially or in parallel. Clustering is applied to reduce the number of outcomes. The third method presented is based on Høyland, Kaut et Wallace [10] algorithm which is based on Cholesky decomposition.

TABLE DES MATIÈRES

DÉDICACE iv

REMERCIEMENTS v

RÉSUMÉ vi

ABSTRACT viii

TABLE DES MATIÈRES viii

LISTE DES TABLEAUX xiii

LISTE DES FIGURES xv

LISTE DES ABBRÉVIATIONS xvii

LISTE DES ANNEXES xviii

CHAPITRE 1 : INTRODUCTION 1

1.1 : Mise en contexte 1

1.2 : Gestion des réservoirs 2

1.3 : Objectif de recherche 3

CHAPITRE 2 : MODÉLISATION 4

2.1 : Présentation des données 4

2.1.1 : Traitement préliminaire des données 5

2.2 : Synchronisation des séries 5

2.2.1 : Première méthode 7

2.2.2 : Deuxième méthode 9

2.2.3 : Comparaison des deux méthodes 11

2.3 : Modélisation temporelle 15

2.3.1 : Modèles PAR(p) 15

2.4 : Modélisation spatiale 17

2.4.1 : Analyse en composantes principales 17

2.5 : Résumé 19

CHAPITRE 3 : GÉNÉRATION D'ARBRES DE SCÉNARIOS 21

3.1 : Définition d'un arbre de scénarios 22

3.2 : Construction d'arbres unidimensionnels 23

3.2.1 : Méthode des moments 23

3.2.1.1 : Approche séquentielle 24

3.2.1.2 : Approche globale 25

3.2.2 : Méthode basée sur les simulations 28

3.2.2.1 : Simulations en parallèle 28

3.2.2.2 : Simulations séquentielles 29

3.2.2.3 : Transformation cubique 30

3.2.2.4 : Classification non-hiérarchiques des scénarios 32

3.2.2.5 : Le logiciel SCENRED 2.0 36

3.2.3 : Méthodes hybrides 37

3.3 : Construction d'arbres multidimensionnels 37

3.3.1 : Recombinaison des composantes principales 38

3.3.1.1 : Méthode des moments 39

3.3.1.2 : Méthode de simulation parallèle 40

3.3.1.3 : Méthode de simulation séquentielle 41

3.3.2 : L'heuristique de Høyland et al. 42

3.3.2.1 : Transformation matricielle 42

3.3.2.2 : L'algorithme noyau 44

3.3.2.3 : L'heuristique 45

3.4 : Discussion 46

CHAPITRE 4 : CONCLUSION 49

BIBLIOGRAPHIE 51

ANNEXES 55

LISTE DES TABLEAUX

Tableau 3.1 : Erreurs sur les moments de ξ_t^1 (méthode des moments) 26

Tableau 3.2 : Erreurs sur les moments de ξ_t^2 (méthode des moments) 27

Tableau 3.3 : Erreurs sur les moments de ξ_t^1 (transformation cubique) 33

Tableau 3.4 : Erreurs sur les moments de ξ_t^2 (transformation cubique) 34

Tableau A.1 : Résultats de la modélisation PAR 55

Tableau A.1 : Résultats de la modélisation PAR (suite et fin) 56

Tableau A.2 : Pourcentage de variance expliquée par ξ_t^i 57

Tableau A.3 : Erreurs sur les moments des résidus générés de Baskatong (Méthode des moments) 58

Tableau A.4 : Erreurs sur les moments des résidus générés de Cabonga (Méthode des moments) 59

Tableau A.5 : Erreurs sur les moments des résidus générés de Paugan (Méthode des moments) 60

Tableau A.6 : Comparaison des matrices de corrélation $CORR_t$ et $CORR_t^*$ (Méthode des moments) 61

Tableau A.7 : Erreurs sur les moments des résidus générés de Baskatong (Méthode séquentielle) 62

Tableau A.8 : Erreurs sur les moments des résidus générés de Cabonga (Méthode séquentielle) 63

Tableau A.9 : Erreurs sur les moments des résidus générés de Paugan (Méthode séquentielle) 64

Tableau A.10 : Comparaison des matrices de corrélation $CORR_t$ et $CORR_t^*$ (Méthode séquentielle) 65

Tableau A.11 : Erreurs sur les moments des résidus générés de Baskatong (Høyland) 66

Tableau A.12 : Erreurs sur les moments des résidus générés de Cabonga (Høyland) 67

Tableau A.13 : Erreurs sur les moments des résidus générés de Paugan (Høyland) 68

Tableau A.14 : Comparaison des matrices de corrélation $CORR_t$ et $CORR_t^*$ (Høyland) 69

LISTE DES FIGURES

Figure 2.1 : Les apports d'eau aux trois sites de 1954 à 1956 5

Figure 2.2 : Les apports hebdomadaires superposés à Baskatong de 1950 à 2006 et la moyenne hebdomadaire 6

Figure 2.3 : 1^{er}(a), 3^{e}(b), 5^{e} et dernier (c) pas de la première méthode de synchronisation pour les trois sites 10

Figure 2.4 : Fonction de synchronisation pour l'année 1950 pour le site de Baskatong et effet de la synchronisation sur la fonction des apports 12

Figure 2.5 : Fonction de synchronisation pour l'année 2005 pour le site de Baskatong et effet de la synchronisation sur la fonction des apports 13

Figure 2.6 : Les hydrogrammes de Paugan avant et après la synchronisation (2^{e} méthode) 14

Figure 2.7 : Erreur relative causée par les deux méthodes de synchronisation. . 16

Figure 3.1 : Simulation parallèle 29

Figure 3.2 : Simulation séquentielle 30

Figure 3.3 : Construction de l'arbre d'apports à partir des résidus 38

Figure 3.4 : Un exemple du comportement de l'erreur en utilisant l'heuristique 46

Figure C.1 : Les prévisions d'apports pour les semaines 6, 7 et 8 (Méthode des moments) 72

Figure C.2 : Les prévisions d'apports pour les semaines 9, 10 et 11 (Méthode des moments) 73

Figure C.3 : Les prévisions d'apports pour les semaines 6, 7 et 8 (Simulation séquentielle) 74

Figure C.4 : Les prévisions d'apports pour les semaines 9, 10 et 11 (Simulation séquentielle) 75

Figure C.5 : Les prévisions d'apports pour les semaines 6, 7 et 8 (Simulation parallèle) 76

Figure C.6 : Les prévisions d'apports pour les semaines 9, 10 et 11 (Simulation parallèle) 77

Figure C.7 : Les prévisions d'apports pour les semaines 6, 7 et 8 (Méthode de Høyland) 78

Figure C.8 : Les prévisions d'apports pour les semaines 9, 10 et 11 (Méthode de Høyland) 79

LISTE DES ABBRÉVIATIONS

N	Nombre d'années de l'historique.
S	Nombre de sites.
$hyd_n^s(t)$	Hydrogramme de l'année n au site s.
$\widetilde{hyd_n^s}(t)$	Hydrogramme issu de $hyd_n^s(t)$ ayant subi une transformation dans le temps.
$hyd_{ref}^s(t)$	Hydrogramme de référence au site s.
$h_n^s()$	Fonction de transformation du temps pour l'année n au site s.
δ_n^s	Paramètre de décalage pour l'hydrogramme de l'année n au site s.
$hmoy^s(t)$	Hydrogramme moyen au site s.
$hmoy(t)$	Hydrogramme moyen des trois sites.
$t_{max}^s(n)$	Le temps de l'apport maximal de l'année n au site s.
τ_{max}^s	La médiane des $t_{max}^s(n)$ au site s, où $n = 1, \ldots, N$.
APP_t^s	Vecteur d'apports pour la semaine t au site s.
R_t^s	Vecteur de résidus de la semaine t au site s.
R_t^{s*}	Vecteur de résidus générés pour la semaine t au site s.
$\phi_{j,t}^s$	Paramètres de régression du modèle PAR(p) dans la semaine t au site s, où $j = 0, \ldots, p$.
ξ_t^i	i^{ieme} composante principale de la semaine t.
ξ_t^{s*}	Composante principale générée pour la semaine t.
X_t^s	Vecteur de nombres aléatoires.
Y_t^s	Vecteur de nombres aléatoires issu de la transformation de X_t^s.
$\mathbf{X}_t$	Matrice définie par $\left(X_t^1 \cdots X_t^S\right)$.
$\mathbf{Y}_t$	Matrice définie par $\left(Y_t^1 \cdots Y_t^S\right)$.
$\mathbf{R}_t$	Matrice définie par $\left(R_t^1 \cdots R_t^S\right)$.
$\mathbf{CORR}_t$	Matrice de corrélation des résidus $\left[R_t^1 \cdots R_t^S\right]$.
$\mathbf{L}$	Matrice triangulaire issue de $\mathbf{CORR}_t = \mathbf{LL}'$.
$M_{k,t}^s$	Moments des résidus de la semaine t au site s où $k = 1, 2, 3, 4$.
M_k	Moments cibles.
m_k	Moments d'une variable générée.

LISTE DES ANNEXES

ANNEXE A : TABLEAUX **55**

A.1 : Modélisation PAR(2) 55

A.2 : Analyse en composantes principales 57

A.3 : Méthode des moments 58

A.4 : Méthode séquentielle 62

A.5 : Høyland (algorithme noyau) 66

ANNEXE B : THÉORIE **70**

B.1 : Heuristique de Høyland et al. 70

ANNEXE C : LES ARBRES **72**

C.1 : Méthode des moments 72

C.2 : Simulation séquentielle 74

C.3 : Simulation parallèle 76

C.4 : Méthode de Høyland 78

CHAPITRE 1 : INTRODUCTION

1.1 Mise en contexte

Le Québec est une région riche en eau douce avec plus de 130 000 cours d'eau et un million de lacs. Près de 96% de l'électricité vendue au Québec provient de l'hydroélectricité. En fait, Hydro-Québec est le plus grand producteur d'hydroélectricité en Amérique du Nord. Son parc hydroélectrique compte présentement 58 centrales et 24 grands réservoirs. Sa production annuelle se situe autour de 165 TWh.

Les installations hydroélectriques d'Hydro-Québec sont gérées de façon à maximiser les revenus tout en s'assurant que la demande soit satisfaite. Lorsque la production hydroélectrique n'est pas suffisante, des centrales thermiques sont mises en opération et/ou de l'énergie électrique est achetée des producteurs voisins. Ces alternatives sont plus coûteuses que de produire de l'hydroéléctricité. D'autre part, en périodes hors-pointe, comme durant l'été, Hydro-Québec exporte ses excédents d'électricité aux États-Unis où la demande est souvent très élevée à cause de la climatisation. Durant l'hiver (période de pointe), Hydro-Québec soutire beaucoup d'eau de ses réservoirs parce que la demande est très forte et les apports en eau très faibles puisque les précipitations sont en neige. Hydro-Québec doit cependant maintenir ses réservoirs à des niveaux sécuritaires pour prévenir les inondations et à des niveaux suffisamment élevés pour assurer la navigation et garantir l'approvisionnement en eau potable.

En résumé, la meilleure gestion est un compromis entre le besoin de générer des bénéfices et le contrôle des risques d'inondations et de pénurie d'eau sachant que la demande en électricité et les apports futurs sont incertains. En 2003, sous la direction de M. André Turgeon, la chaire de recherche industrielle Hydro-Québec/CRSNG sur la gestion des

systèmes hydriques a été créee pour promouvoir le développement de nouvelles méthodes d'optimisation stochastique pour la gestion des réservoirs. Plusieurs projets sont menés par le groupe de travail dont la modélisation des apports d'eau naturels aux réservoirs. Cette maîtrise s'inscrit dans ce projet.

1.2 Gestion des réservoirs

Le problème de gestion des réservoirs est un problème dynamique, stochastique, non linéaire et de grande taille. Le véritable défi réside dans la taille et dans l'aspect stochastique du problème.

Labadie [12] et Yeh [22] présentent une classification détaillée des modèles d'optimisation et de simulation des systèmes de gestion des réservoirs. Labadie [12] sépare les méthodes d'optimisation en deux groupes : implicites et explicites. L'optimisation stochastique implicite (OSI) consiste à trouver une solution au problème de gestion à partir d'un grand nombre de scénarios d'apports provenant de l'historique ou générés synthétiquement. L'optimisation stochastique explicite (OSE) requiert en revanche une formulation probabiliste du problème. Les apports sont représentés par des distributions de probabilités plutôt que par les scénarios eux-mêmes. Les techniques d'optimisation explicites sont nombreuses : Stochastic Dynamic Programming (SDP), Chance-constrained Programming, Stochastic Linear Programming (SLP), Stochastic Optimal Control (OCP). Les méthodes explicites demandent généralement plus de mémoire et de temps de calcul que les méthodes implicites.

Malgré le grand nombre de méthodes d'optimisation, leur application dans le domaine de la gestion des réservoirs reste limitée (Yeh [22]). Les modèles d'optimisation sont souvent incapables de bien prendre en compte tous les éléments d'un problème. De plus, la stochasticité des apports est souvent ignorée. La programmation dynamique traite le côté aléatoire, mais seulement pour des problèmes de petite taille. Pour les problèmes de grande

taille, la programmation dynamique est difficilement applicable en raison du fait que le temps de calcul et l'espace mémoire requis augmentent exponentiellement avec le nombre de variables d'état. Pour arriver à résoudre les problèmes, il faut alors avoir recours à des techniques d'approximation, de décomposition ou d'agrégation de façon à réduire la taille des problèmes. À cause des limites des techniques d'optimisation, les gestionnaires des réservoirs préfèrent souvent se fier aux modèles de simulation qui collent mieux à la réalité.

En résumé, une bonne gestion des réservoirs nécessite une bonne représentation du phénomène aléatoire des apports d'eau. Ceci a souvent pour effet d'augmenter considérablement la taille du problème, ce qui rend les techniques d'optimisation difficilement applicables. Une façon de contrecarrer le problème lorsque le nombre de pas de temps est petit est de générer des arbres de scénarios d'apports. Cette approche permet une représentation du comportement probable des apports d'eau sans nécessiter beaucoup d'espace mémoire.

1.3 Objectif de recherche

L'objectif de ce mémoire est de construire des arbres de scénarios d'apports à partir des historiques des apports aux différents sites du bassin versant de la Gatineau. Pour construire ces arbres, il faut tout d'abord choisir un modèle statistique pour caractériser les séries hydrologiques puis faire appel aux méthodes de génération d'arbres de scénarios. Par après, il faut construire un outil informatique qui générera les arbres de scénarios à partir des modèles retenus.

CHAPITRE 2 : MODÉLISATION

Nous considérons les données que nous traitons, soient les apports hebdomadaires, comme la somme d'une partie déterministe et d'une partie aléatoire dont la distribution est à déterminer. Les méthodes de modélisation qui permettent de prédire la partie déterministe seront abordées dans ce chapitre. Celui-ci est divisé en deux parties. La première concerne la transformation des données où deux méthodes de synchronisation sont exposées. La deuxième partie présente la procédure choisie pour modéliser la corrélation spatiale et temporelle des données.

2.1 Présentation des données

Nous avons travaillé avec l'historique des apports hebdomadaires moyens du 2 janvier 1950 au 17 juillet 2006 à trois sites différents et plus précisément aux réservoirs Cabonga et Baskatong et à la centrale Paugan. Le réservoir Cabonga a une superficie de 434 km^2 et Baskatong une superficie de 413 km^2. Le réservoir Cabonga se déverse dans le réservoir Baskatong et ce dernier alimente la centrale Paugan située en aval sur la rivière Gatineau. Cette centrale au fil de l'eau a une puissance installée de 202 MW.

Les apports diffèrent beaucoup d'un site à l'autre (figure (2.1)) et sont nettement plus grands à Baskatong qu'à Cabonga. Ils sont naturellement élevés au printemps lors de la crue. Celle-ci peut survenir entre les mois de février et avril. Historiquement la rivière Gatineau est reconnue comme étant une rivière qui cause des inondations. La prévision des apports au printemps est critique pour les gestionnaires du bassin. Les données des années 2005 et 2006 ont été utilisées pour tester les modèles de prévision.

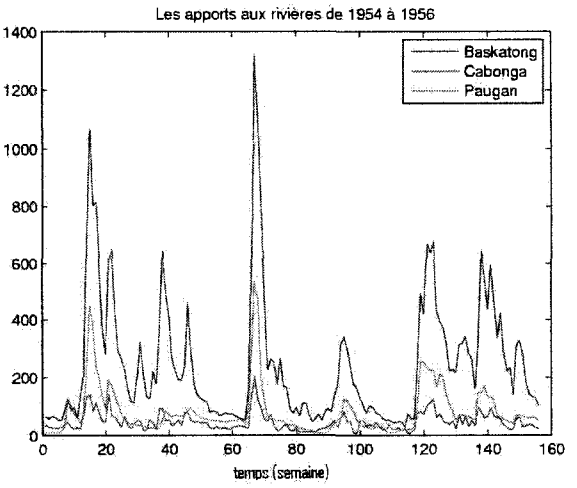


Figure 2.1 Les apports d'eau aux trois sites de 1954 à 1956

2.1.1 Traitement préliminaire des données

Les données disponibles sont les apports hebdomadaires moyens. Or, dans une année il y a 52 semaines plus 1 ou 2 jours, selon que l'année est bissextile ou non. Pour pouvoir comparer les apports d'une semaine d'une année donnée à ceux de la même semaine d'une autre année, il faut enlever le 29 février des années bissextiles et le dernier jour de l'année de façon à avoir des années de 364 jours seulement, et donc de 52 semaines. De cette manière, il est possible de bien déterminer les caractéristiques des apports dans une semaine donnée à partir des données de l'historique.

2.2 Synchronisation des séries

L'analyse de la crue printanière est une étape cruciale dans ce travail. Comme il est illustré dans la figure (2.2), où sont superposés les apports hebdomadaires des 55 années de l'historique, il est difficile de discerner le début et la fin de la crue. En fait, la crue

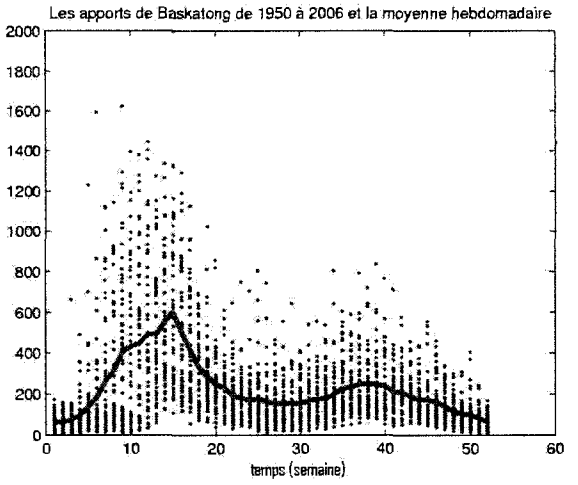


Figure 2.2 Les apports hebdomadaires superposés à Baskatong de 1950 à 2006 et la moyenne hebdomadaire

printanière ne débute pas au même moment d'une année à l'autre. On constate aussi que les maxima des 55 hydrogrammes se trouvent entre les semaines 6 et 18. Comme Halldin [6] et Lindell [13], nous émettons l'hypothèse que le début des crues est indépendant du volume des crues. De cette manière, l'analyse des séries hydrologiques se fait en deux étapes. On prévoit d'abord le moment du début de la crue puis l'amplitude de celle-ci.

Pour prévoir le début de la crue, Lindell [13] suggère d'enregistrer le début des crues de toutes les années de l'historique. Il définit le début d'une crue par la valeur de l'apport qui dépasse un multiple de la moyenne annuelle. Il note que ces valeurs sont normalement distribuées. Quant à Halldin [13], il a développé une méthode qui utilise la température pour prévoir le début de la crue. Dans ce mémoire, nous supposons que le début de la crue est connu.

Une fois le début de la crue fixé, le problème est de prévoir le volume d'eau de cette crue. Pour l'obtenir, il faut transformer les séries de manière à obtenir un hydrogramme de base. On définit un hydrogramme de base comme un hydrogramme référence des apports

naturels moyens à un certain site de prévision (Merleau et al. [16]). Référence signifie que l'hydrogramme de base reproduit les caractéristiques moyennes des hydrogrammes historiques. Parmi ces caractéristiques, Merleau et al. [16] mentionnent la forme de la crue, sa pointe (son maximum) ainsi que les volumes de crues. D'un point de vue pratique, on définit l'hydrogramme de base comme la moyenne des hydrogrammes historiques synchronisés.

Cette section décrit deux méthodes de synchronisation des séries. La première, développée par Halldin [6], déplace les séries de manière à ce que la forme en cloche de la crue arrive approximativement au même moment pour toutes les séries historiques. La seconde, développée par Merleau et al. [16], synchronise mais selon le maximum de la crue plutôt que selon sa forme.

2.2.1 Première méthode

Dans cette première méthode, on cherche à synchroniser les crues en déplaçant les courbes d'apports de chacune des années de manière à faire superposer les crues. Il s'agit d'effectuer une transformation du temps t pour que la crue de chacune des années arrive plus ou moins au même moment. Halldin [6] a utilisé la méthode proposée par Ramsay et Silverman [17].

Soit $hyd_n^s(t)$ l'hydrogramme de l'année n où $n = 1, \ldots, N$, et $\widetilde{hyd_n^s}(t)$ la courbe issue de $hyd_n^s(t)$ suivant une transformation du temps :

$$\widetilde{hyd_n^s}(t) = hyd_n^s\big(h_n^s(t)\big) \tag{2.1}$$

où $h_n^s(t)$ le temps transformé (appelé aussi temps synchrone). Un cas particulier est de choisir une transformation linéaire du temps $h_n^s(t) = t + \delta_n^s$ où δ_n^s est le paramètre de décalage de l'année n et du site s. Le problème consiste à trouver le paramètre de décalage δ_n^s qui permet de minimiser la différence entre un hydrogramme de référence $hyd_{ref}^s(t)$,

dont la forme et l'amplitude sont connues, et l'hydrogrammes $hyd_n^s(t)$.

$$\min_{\delta_n^s} \sum_{t=1}^{52} \left(hyd_n^s(t+\delta_n^s) - hyd_{ref}^s(t) \right)^2 \quad \forall n = 1, \ldots, N \tag{2.2}$$

Comme la forme de la crue n'est pas connue, on doit procéder itérativement. L'hydrogramme de référence est choisi tout d'abord comme étant la moyenne de tous les hydrogrammes.

$$hyd_{ref}^s(t) = hmoy^s(t) = \frac{1}{N} \sum_{n=1}^{N} hyd_n^s(t) \qquad \forall t \tag{2.3}$$

Ainsi à la première itération, les paramètres de décalage δ_n^s pour les N années sont calculés avec l'équation (2.2) et les n hydrogrammes $hyd_n^s(t+\delta_n^s)$ sont décalés ou synchronisés pour devenir $\widetilde{hyd_n^s}(t)$. À l'itération suivante, l'hydrogramme de référence devient la moyenne des hydrogrammes synchronisés $\widetilde{hyd_n^s}(t)$, ce qui permet de calculer les nouveaux δ_n^s. On procède ainsi jusqu'à ce que les δ_n^s soient tous nuls. L'algorithme peut être résumé comme suit :

1: Calculer la courbe moyenne $hmoy^s(t) = \frac{1}{N} \sum_{n=1}^{N} hyd_n^s(t)$ pour tout t.
2: **si** $\delta_n^s \neq 0$ pour au moins un n où $n = 1, \ldots, N$ **alors**
3: **pour** $n = 1, \ldots, N$ années dans l'historique ;
4: $\delta_n^s = argmin\{\sum_{t=1}^{52} (hyd_n^s(t+\delta_n^s)^s - hmoy^s(t))^2\}$ et $\delta_n^s \in Z$
5: $\widetilde{hyd_n^s}(t) = hyd_n^s(t+\delta_n^s)$
6: Recalculer la moyenne $hmoy^s(t) = \frac{1}{N} \sum_{n=1}^{N} \widetilde{hyd_n^s}(t)$ pour tout t

Généralement, quelques itérations suffisent pour que cet algorithme converge. Cette procédure est souvent appelée analyse procustéenne (Ramsay et Silverman [17]).

Halldin [6] propose de synchroniser les apports de tous les sites en même temps. La courbe de référence devient alors la courbe moyenne des S sites.

$$hmoy(t) = \frac{1}{N \cdot S} \sum_{s=1}^{S} \sum_{n=1}^{N} hyd_n^s(t) \quad \forall t \tag{2.4}$$

Puisque les apports aux trois sites sont très différents (figure 2.1), l'échelle des apports a été modifiée pour que ceux-ci varient entre 0 et 100.

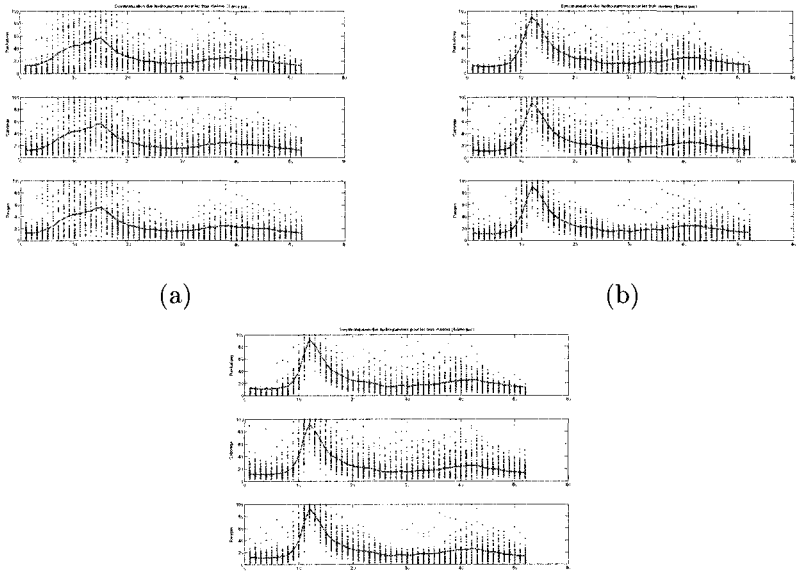
On peut montrer que l'estimation de δ_n^s, pour n donné, dans l'équation (2.2) est la valeur qui maximise la corrélation entre $hmoy^s(t)$ et $hyd_n^s(t)$. En d'autres termes, cela consiste à calculer la fonction de corrélation entre $hmoy^s(t)$ et $hyd_n^s(t)$ et de prendre la valeur qui maximise cette fonction comme δ_n^s. Pour résoudre ce problème, la fonction de corrélation croisée *xcorr* de Matlab a été utilisée. Cette fonction a comme paramètres les vecteurs $hmoy^s(t)$ et $hyd_n^s(t)$ et détermine c, le vecteur des corrélations croisées et *lag* le vecteur des décalages associés à c. On retient pour δ_n^s la valeur du *lag* pour laquelle la corrélation croisée c est maximale.

La figure (2.3) montre les résultats de cette méthode. Les apports aux trois sites ont été mis à l'échelle. Le maximum est établi à 100. À mesure que les itérations augmentent, on constate que la moyenne (en trait plein) représente mieux les caractéristiques moyennes des volumes d'eau dans une crue.

2.2.2 Deuxième méthode

Au lieu de considérer la forme de la courbe d'apports, Merleau et al.[16] proposent de considérer l'instant où l'apport maximal se produit. Soit $t_{max}^s(n)$ cet instant pour une année n et un site s et soit τ_{max}^s le temps de référence pris comme médiane des $t_{max}^s(n)$ pour toutes les années n, $\tau_{max}^s \in Z$. La fonction de synchronisation de cette méthode étire le temps au début et le compresse à la fin si le maximum d'apport survient avant τ_{max}^s (figure 2.5). Si le maximum survient après, c'est l'inverse (figure 2.4).

Comme pour la méthode précédente, on cherche à faire une transformation du temps t. On cherche la fonction de transformation du temps $h_n^s(t)$ qui fera en sorte que la courbe d'apport transformée $\widetilde{hyd_n^s}(t) = hyd_n^s(h_n^s(t))$ ait son maximum au temps τ_{max}^s.



(a) (b)

(c)

Figure 2.3 1^{er}(a), 3^{e}(b), 5^{e} et dernier (c) pas de la première méthode de synchronisation pour les trois sites

Merleau et al. [16] définissent la fonction de synchronisation $h_n^s(t)$ comme une fonction linéaire par partie :

$$h_n^s(t) = \sum_{j=1}^{P} L_j(t) I_{D_j}(t) = \sum_{j=1}^{P} (b_j + m_j t) I_{D_j}(t) \tag{2.5}$$

où $I_{D_j}(t)$ est une fonction qui indique quand la fonction linéaire $L_j(t)$ est non-nulle. D_j est le domaine où $L_j(t)$ est non-nulle. Les contraintes suivantes doivent être respectées :

$$\begin{aligned} & h_n^s(1) = 1 \\ & h_n^s\left(t_{max}^s(n)\right) = \tau_{max}^s \\ & h_n^s(52) = 52 \end{aligned} \tag{2.6}$$

Dans ces conditions, P doit être égal à 2 et les ordonnées à l'origine b_j et les pentes m_j se calculent comme suit :

$$m_1 = \frac{\tau_{max}^s - 1}{t_{max}^s(n) - 1} \tag{2.7}$$

$$b_1 = 1 - m_1 \tag{2.8}$$

$$m_2 = \frac{52 - \tau_{max}^s}{52 - t_{max}^s(n)} \tag{2.9}$$

$$b_2 = 52 - 52 m_2 \tag{2.10}$$

En modifiant le temps, la valeur de l'apport associé à la semaine 10 peut se retrouver à la semaine 10.6. Il faut alors faire une interpolation linéaire pour avoir des valeurs d'apports qui se situent sur l'échelle du temps $t = 1, 2, \ldots, 52$. Sur les figures (2.4) et (2.5) le résultat de l'interpolation est représenté par une courbe en trait pointillé.

2.2.3 Comparaison des deux méthodes

Avec la première méthode, les courbes d'apports subissent un déplacement vers la droite ou vers la gauche. La forme de la crue reste alors inchangée. Les caractéristiques des données réelles sont préservées. Ce ne sont que les extrémités des courbes qui sont changées.

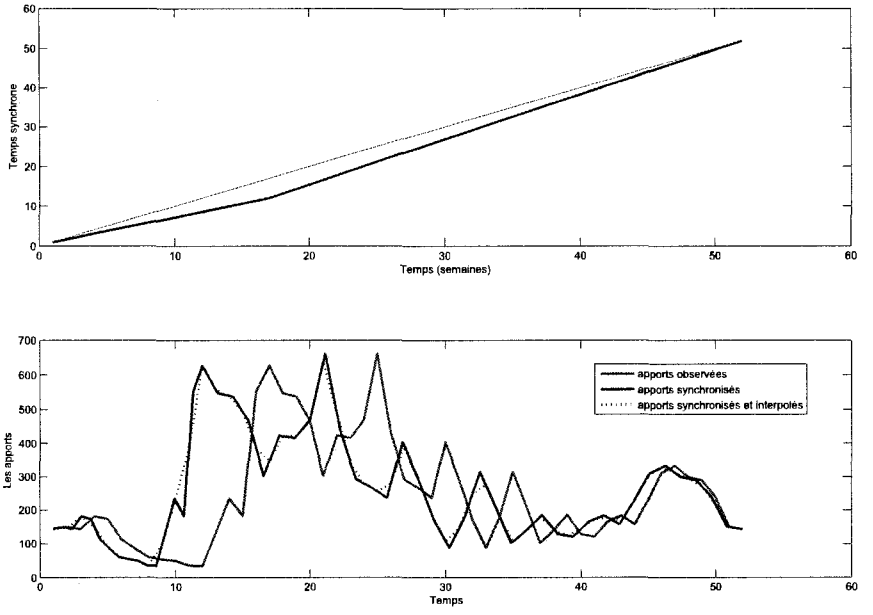


Figure 2.4 Fonction de synchronisation pour l'année 1950 pour le site de Baskatong et effet de la synchronisation sur la fonction des apports

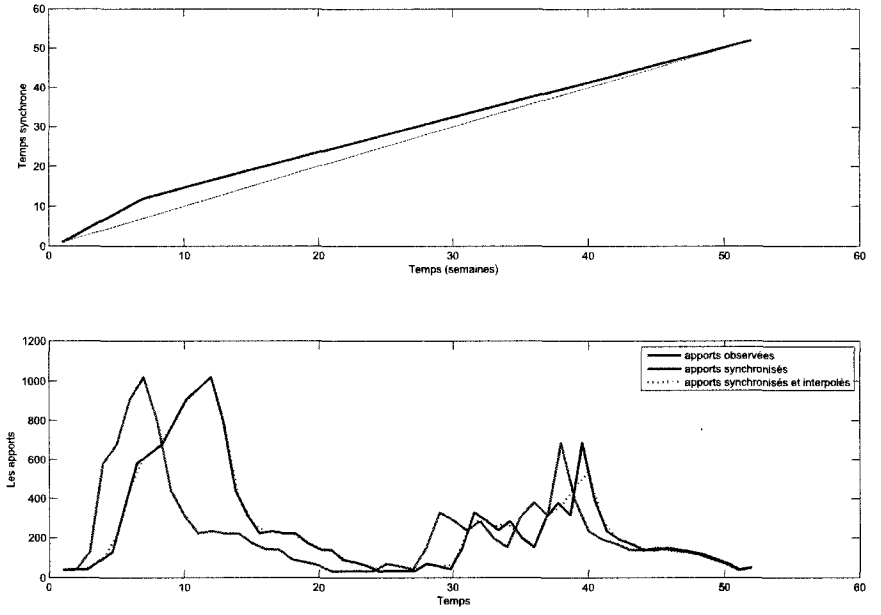


Figure 2.5 Fonction de synchronisation pour l'année 2005 pour le site de Baskatong et effet de la synchronisation sur la fonction des apports

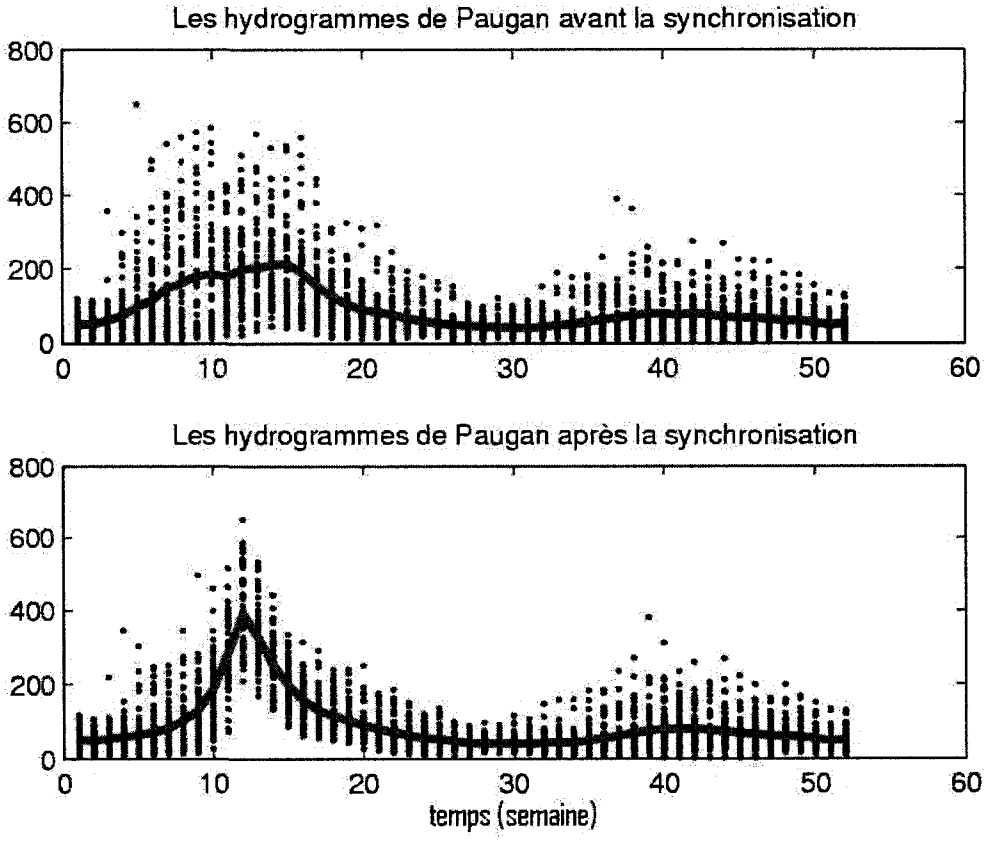


Figure 2.6 Les hydrogrammes de Paugan avant et après la synchronisation (2^e méthode)

Comme les apports sont à leur plus bas niveau aux extrémités, cette transformation ne génère pas beaucoup d'erreurs. Avec la deuxième méthode les courbes sont contractées puis décontractées. La forme de la crue se trouve alors modifiée. La première méthode est donc plus appropriée puisqu'elle préserve la forme de la crue.

Lorsque l'on compare les effets de la synchronisation sur la somme des apports annuels on trouve que la première méthode préserve mieux cette somme. En effet, la figure (2.7) montre le pourcentage d'erreur causé par les deux méthodes de synchronisation pour les 55 hydrogrammes des trois sites. On constate que la première méthode est toujours meilleure, surtout pour les hydrogrammes des premières et dernières années de l'historique des sites Baskatong et Paugan. Dans ces cas-là, la pointe de la crue arrive soit très tôt dans l'année (semaine 6) soit très tard (semaine 18). Avec la méthode de Merleau et al. [16], la forme de la crue est alors déformée de manière importante. Il s'en suit que l'erreur relative est souvent importante et peut même atteindre 15%.

2.3 Modélisation temporelle

2.3.1 Modèles PAR(p)

La modélisation PAR (*Periodic-Auto-Regressive*) est la méthode privilégiée par les ingénieurs d'Hydro-Québec pour modéliser la corrélation temporelle. D'ailleurs, Halldin [6] a opté pour cette méthode après en avoir essayé plusieurs autres.

Soit APP_t^s un vecteur de taille N tel que chaque élément $app_{t,n}^s$ représente l'apport dans la semainte t de l'année n au site s.

$$APP_t^s = \left\{app_{t,1}^s, \ldots, app_{t,N}^s\right\} \tag{2.11}$$

On cherche à déterminer la relation linéaire entre les apports de la semaine t et les apports

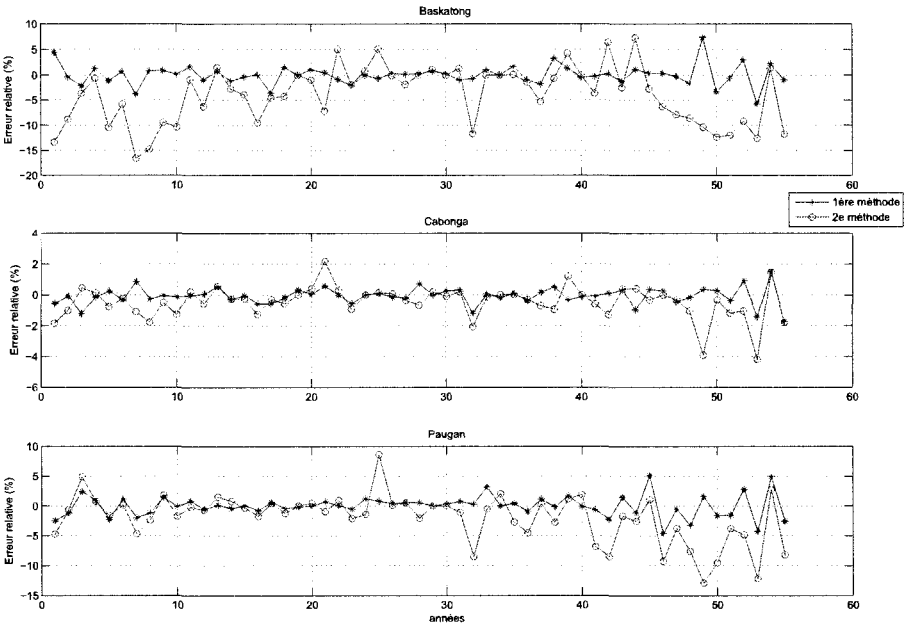


Figure 2.7 Erreur relative causée par les deux méthodes de synchronisation.

des semaines $t-j$ où $j=1,\ldots,p$. Nous avons :

$$APP_t^s = \phi_{0,t}^s + \sum_{j=1}^{p} \phi_{j,t}^s APP_{t-j}^s + R_t^s \tag{2.12}$$

où les paramètres $\phi_{j,t}^s$ sont des constantes à déterminer et R_t^s le vecteur des résidus de la semaine t au site s. Ainsi, nous n'avons pas un seul modèle pour toute l'année mais un modèle pour chaque semaine.

En plus de vérifier la validité du modèle de régression linéaire par le test de Fisher, le test de Student est utilisé pour valider chacun des paramètres $\phi_{j,t}^s$. Il faut tester l'hypothèse $H_0 : \phi_{j,t}^s = 0$ contre $H_1 : \phi_{j,t}^s \neq 0$. Si on ne rejette pas H_0 on supprime la variable APP_{t-j}^s du modèle. L'apport de la semaine $t-j$ n'influence pas de façon significative l'apport de la

semaine t. Le test de Student est un test marginal. Plus précisément, il teste la contribution de la j^{ieme} variable étant donné que toutes les autres variables $i \neq j$ sont dans le modèle.

Pour décider du nombre p de semaines à utiliser dans la régression, l'erreur quadratique moyenne (EQM) a été observée. Tant qu'il y a une diminution significative de cette erreur le nombre de variables est augmenté. À aucun moment, l'erreur n'a cessé de diminuer. Toutefois, lorsque p est très grand la valeur de plusieurs paramètres $\phi^s_{j,t}$ sont insignifiants. Après plusieurs essais et par souci de parcimonie nous avons finalement opté pour un modèle à deux variables. Dans la majorité des cas, les variables ont un effet significatif. Les résultats sont présentés à l'annexe (A.1). Les paramètres $\phi^s_{j,t}$ considérés statistiquement nuls sont en gras. Les apports APP^s_{t-j} associés à ces paramètres sont donc exclus du modèle.

2.4 Modélisation spatiale

En plus de modéliser la dépendance temporelle des séries, la dépendance entre les séries aux sites avoisinants doit être prise en compte (Salas, Delleur, Yevjevich et Lane [18]).

Vitoriano, Cerisola et Ramos [21] utilisent la régression linéaire pour modéliser la dépendance entre les apports aux différents sites. Halldin [6], tout comme Dupačovà, Consigli et Wallace [1], proposent d'utiliser l'analyse en composantes principales pour expliquer la dépendance spatiale.

L'analyse en composantes principales, qui est appliquée aux résidus R^s_t des S sites, est présentée dans la section qui suit.

2.4.1 Analyse en composantes principales

L'analyse en composantes principales est utilisée pour réduire le nombre de variables aléatoires lorsque celles-ci sont corrélées. L'analyse est effectuée sur les résidus R^s_t de la

semaine t aux S sites fortement corrélés entre eux.

L'analyse en composantes principales (ACP) est une méthode mathématique qui transforme un nombre de variables aléatoires mutuellement corrélées en un nombre de variables aléatoires indépendantes appellées composantes principales. L'idée est de trouver des combinaisons linéaires des variables aléatoires $(R_t^1, \ldots, R_t^S)$. Ces combinaisons linéaires doivent toutes être indépendantes les unes des autres et exprimer par ordre croissant la variance contenue. Ce sont ces combinaisons linéaires, qu'on désigne par $\xi_t^1, \ldots, \xi_t^Q$ $(Q \leq S)$, qui sont appelées composantes principales.

$$\begin{aligned} \xi_t^1 &= b_{11}R_t^1 + b_{21}R_t^2 + \ldots + b_{S1}R_t^S \\ &\vdots \\ \xi_t^Q &= b_{1Q}R_t^1 + b_{2Q}R_t^2 + \ldots + b_{SQ}R_t^S \end{aligned} \tag{2.13}$$

La première composante principale ξ_t^1 doit être choisie de manière à avoir la plus grande variance, la deuxième composante doit avoir la deuxième plus grande variance tout en étant orthogonale à la première, et ainsi de suite ...

Dans le cadre de ce travail, la fonction *princomp* de Matlab a été utilisée :

$$[COEFS, SCORES, VAR] = princomp(R_t^1, R_t^2, \ldots, R_t^S) \tag{2.14}$$

Cette fonction génère COEFS, une matrice $(S \times S)$:

$$COEFS = \begin{bmatrix} b_{11} & b_{12} & \ldots & b_{1S} \\ b_{21} & b_{22} & \ldots & b_{2S} \\ \vdots & \vdots & \vdots & \vdots \\ b_{S1} & b_{S2} & \ldots & b_{SS} \end{bmatrix} \tag{2.15}$$

et SCORES, la nouvelle représentation des variables dans le nouvel espace généré par les composantes principales :

$$\begin{aligned} SCORES &= \begin{bmatrix} \xi_t^1 & \xi_t^2 & \ldots & \xi_t^S \end{bmatrix} \\ &= \begin{bmatrix} R_t^1 & R_t^2 & \ldots & R_t^S \end{bmatrix} COEFS \end{aligned} \tag{2.16}$$

puis VAR, un vecteur qui contient les valeurs propres :

$$VAR = \begin{bmatrix} \lambda_t^1 \\ \lambda_t^1 \\ \vdots \\ \lambda_t^S \end{bmatrix} \tag{2.17}$$

où $\lambda_t^1 \geq \cdots \geq \lambda_t^S \geq 0$. Chaque valeur propre donne la part de la variance expliquée par la composante principale correspondante. Le pourcentage expliqué par ξ_t^j est égal à $\frac{\lambda_t^j}{\sum_{i=1}^{S} \lambda_t^i}$. Les dernières composantes principales qui ne contribuent pas significativement à l'explication de la variance sont ignorées. De cette manière, il est possible de réduire le nombre de variables explicatives.

Nous avons appliqué l'analyse en composantes principales aux résidus des trois séries de Baskatong, Cabonga et Paugan dans chaque semaine. Le tableau à l'annexe (A.2) présente les pourcentages de variance des composantes principales dans chaque semaine. Pour la plus part des semaines une composante suffit pour représenter toute la variance contenue dans les variables. Dans les autres semaines deux composantes sont nécessaires. Le pourcentage de variance expliquée par deux composantes principales varie entre 86% et 100%[1].

Ceci nous permet de réduire le nombre de variables : de trois variables corrélées à deux variables indépendantes. Même si la réduction des variables ne semble pas énorme, elle sera utile lors de la génération d'arbres de résidus. Au lieu de générer trois arbres pour les trois sites, il suffira de générer indépendamment deux arbres de composantes principales qu'il faudra recombiner pour obtenir trois arbres de résidus corrélés.

2.5 Résumé

Ce chapitre a été consacré à la transformation et la modélisation des données. Deux méthodes de synchronisation ont été testées. La première, proposée par Ramsay et Silverman [17], a été retenue parce qu'elle préserve mieux les séries initiales. La modélisation

[1] 100% est un arrondi à deux décimales près.

Periodic-Auto-Regressive a été utilisée pour modéliser la dépendance temporelle. Un modèle PAR(2) a été retenu pour chaque semaine de l'année et pour chaque site. L'analyse en composantes principales a été choisie pour modéliser la dépendance spatiale des résidus des trois sites pour chaque semaine. L'analyse en composantes principales permet de transformer les résidus mutuellement corrélés en des variables indépendantes. Ces variables indépendantes représentent la partie aléatoire des apports qui doivent être générés. Pour ce faire, nous verrons dans le prochain chapitre des méthodes de génération d'arbres de scénarios.

CHAPITRE 3 : GÉNÉRATION D'ARBRES DE SCÉNARIOS

Un arbre de scénarios est une manière de représenter un processus stochastique discret, et plus précisément les valeurs que peuvent prendre une variable aléatoire discrète, dans le temps. Dans ce mémoire, les valeurs possibles des apports aux sites Baskatong, Cabonga et Paugan pour un horizon de trois ou quatre périodes, sont représentées par un arbre de scénarios.

Il existe plusieurs approches pour générer des arbres de scénarios. Celles-ci sont décrites dans l'article de Gülpinar, Rustem et Settergren [5]. La classification de ces auteurs est utilisée dans ce chapitre. Une distinction est faite entre les méthodes d'optimisation, telle que la méthode des moments, et les méthodes de simulations. Ces dernières sont décrites par Dupačovà, Consigli et Wallace [1].

La méthode des moments consiste à générer un arbre d'apports qui minimise les différences entre les moments de l'arbre et ceux de l'historique. Elle peut être utilisée de façon globale, qui détermine les apports pour toutes les périodes de l'arbre en même temps, ou de manière séquentielle qui détermine les apports d'une période à la fois. On trouve dans Vitoriano, Cerisola et Ramos [21] les détails de l'approche séquentielle.

Par ailleurs, la méthode des simulations consiste à générer tout d'abord un très grand nombre de branches. Ensuite, à l'aide des techniques de réduction, un arbre équivalent (avec moins de branches) est construit. Pour générer un grand nombre de branches, Høyland, Kaut et Wallace [10] recommandent la transformation cubique. Celle-ci permet de générer un grand nombre de valeurs qui reproduisent les quatre premiers moments d'une variable aléatoire. Heitsch et Römisch [7] et Dupačovà, Gröwe-Kuska et Römisch [2] ont développé des outils mathématiques et des algorithmes pour réduire le nombre de branches

dans les arbres de scénarios. SCENRED 2.0 est d'ailleurs un outil développé par Heitsch[1]. Notons que Gülpinar, Rustem et Settergren [5] proposent des méthodes hybrides qui utilisent l'optimisation et la simulation pour générer un arbre.

Outre ces approches sophistiquées, il existe des approches plus faciles et plus directes pour générer des arbres de scénarios. Dupačovà, Consigli et Wallace [1] décrivent une méthode qui construit les arbres à partir des séries historiques auxquelles sont assignées des probabilités. Lindell [13] présente dans son mémoire les arbres de Römisch qui sont construits directement à partir de la moyenne et de la variance des données historiques.

Les méthodes citées ci-dessus sont utilisées pour générer des arbres unidimensionnels, c'est-à-dire ayant une seule variable aléatoire. Pour la construction d'arbres multidimensionnels, Halldin [6] propose de générer des arbres unidimensionnels de composantes principales et de les recombiner pour obtenir des arbres multidimensionnels de résidus. Høyland, Kaut et Wallace [10] proposent une heuristique basée sur la décomposition de Choleksy et la transformation cubique pour générer des nombres aléatoires pour des variables mutuellement corrélées.

La première section de ce chapitre est consacrée à la définition d'un arbre de scénarios. Les méthodes de construction d'arbres unidimensionnels et plus spécifiquement la méthode des moments et les méthodes de simulations sont décrites dans la deuxième section. La procédure de construction d'arbres multidimensionnels est présentée à la troisième section. La méthode de Høyland, Kaut et Wallace [10] y est entre autres présentée. La dernière section est dédiée à la description et á la discussion des résultats.

3.1 Définition d'un arbre de scénarios

Un scénario désigne ici une séquence d'apports hebdomadaires pour un horizon donné. Lorsqu'il n'y a qu'un scénario, le problème de gestion des réservoirs est déterministe. Lors-

[1] www.math.hu-berlin.de/ heitsch/

qu'il y a un grand nombre de scénarios, le problème est stochastique et un arbre de scénarios est utilisé pour représenter la stochasticité des apports.

Chaque période de l'arbre représente un pas de temps de l'horizon. Chaque nœud (sauf le premier nœud de l'arbre) a un unique ancêtre qui appartient à la période précédente. Chaque nœud a un ensemble de successeurs. Il y a généralement une structure de dépendance dans le temps : ce qui arrive dans une période est fonction de ce qui est arrivé dans la période précédente, quoiqu'il est possible d'avoir un arbre sans structure de dépendance (Dupačovà, Consigli et Wallace [1]).

Les arbres peuvent, ou pas, avoir le même nombre de successeurs à chaque nœud. Le pas de temps associé à chaque période peut être très court au début (jour, semaine) et plus long à la fin (mois, trimestre). Le nombre de branches peut aussi varier selon la période considérée : beaucoup de branches aux premières périodes et moins à la fin de l'horizon. Le nombre de branches peut aussi varier dans la même période selon les scénarios.

3.2 Construction d'arbres unidimensionnels

Dans cette section, les méthodes de construction d'arbres unidimensionnels, c'est-à-dire ayant une seule variable, sont décrites. Ces méthodes sont utilisées pour générer des arbres de composantes principales ξ_t^s.

3.2.1 Méthode des moments

La méthode des moments permet de générer un petit nombre de réalisations (les branches de l'arbre) d'une variable aléatoire dont on connaît certaines propriétés statistiques. On cherche à minimiser les différences entre les propriétés des réalisations aléatoires et les propriétés de la variable ξ_t^s. On choisit comme propriétés les quatre premiers moments, soient

l'espérance mathématique, la variance, l'asymétrie et le kurtosis (mesure de l'applatissement ou des pics de la distribution par rapport à une distribution normale).

3.2.1.1 Approche séquentielle

L'approche séquentielle de la méthode des moments est décrite dans l'article de Vitoriano, Cerisola et Ramos [21]. Dans ce travail, cette approche a été utilisée pour générer des valeurs des composantes principales ξ_t^s. Le problème d'optimisation est résolu à chaque période de l'arbre et consiste à :

$$\begin{aligned} \min \quad & \sum_{k=1}^{4} w_k (m_k - M_k)^2 \\ \text{sujet à} \quad & \sum_{j=1}^{n} p_j = 1 \\ & p_j \geq 0 \qquad\qquad j = 1, \ldots, n \\ \text{où} \quad & m_1 = \sum_{j=1}^{n} x_j p_j \\ & m_k = \sum_{j=1}^{n} (x_j - m_1)^k p_j \quad k = 2, 3, 4 \end{aligned} \tag{3.1}$$

M_k et m_k où $k = 1, 2, 3, 4$, sont les quatre premiers moments de ξ_t^s et des valeurs générées $X = \{x_1, \ldots, x_n\}$ et w_k est le poids (importance relative) qu'on désire donner à la k^e propriété. Pour enlever l'effet de l'ordre de grandeur des quatre moments, on définit $w_k = \frac{w'_k}{M_k^2}$ où $\sum_k w'_k = 1$.

Le problème est non-linéaire et généralement non-convexe. Pour un même problème, plusieurs solutions existent. Dans notre cas, un minimum local est suffisant. Lorsque l'on obtient une valeur de l'objectif proche ou égale à zéro on peut déduire que l'objectif est atteint puisque les valeurs qui ont été générées ont les mêmes propriétés statistiques que celles de la variable originale. Si on ne réussit pas à obtenir une correspondance parfaite, cela peut vouloir dire que le problème est non réalisable (les propriétés statistiques ne sont pas cohérentes). Dans ce cas, on peut faire varier les poids w'_k et rechercher une solution approximative (Dupačovà, Consigli et Wallace [1]).

Le problème (3.1) a été solutionné avec la fonction *fmincon* de Matlab qui trouve un minimum local à un problème avec contraintes. Le problème consiste à trouver les valeurs de x_j et p_j, pour $j = 1, \ldots, n$, qui solutionne le problème (3.1). La solution initiale $\{x_{1,0}, \ldots, x_{n,0}\}$ fournie à l'algorithme sont des nombres pris au hasard compris entre les valeurs minimale et maximale de ξ_t^s. Les valeurs $p_{j,0}$, pour $j = 1, \ldots, n$, sont initialisées comme étant équiprobables et fixées égale à $\frac{1}{n}$.

L'algorithme *fmincon* s'arrête lorsqu'un minimum local est trouvé ou lorsque le nombre d'itérations maximales est atteint. Lorsque le nombre maximum d'itérations est atteint l'algorithme est réinitialisé avec de nouvelles valeurs $x_{j,0}$. La procédure est répétée jusqu'à ce que la valeur de la fonction objectif soit inférieure à un seuil fixé *tol* ou que le nombre de réinitialisations soit égal à dix. En pratique, l'algorithme converge généralement après 3 à 5 initialisations. La valeur de *tol* a été fixée à 0,01.

En choisissant de donner le même poids w'_k à tous les moments et de générer cinq réalisations ($n = 5$) pour chacune des composantes principales ξ_1^t et ξ_2^t, pour $t = 10, \ldots, 17$, qui représente la période de crue, la méthode donne de bons résultats. Si l'on compare les moments cibles M_k et les moments des valeurs générés m_k on constate que l'erreur relative $\left|\frac{M_k - m_k}{M_k}\right|$ est très faible et qu'elle est en moyenne inférieure à 1% (voir tableaux 3.1 et 3.2).

3.2.1.2 Approche globale

Contrairement à l'approche séquentielle, l'approche globale considère tous les nœuds de l'arbre de scénarios en même temps. Selon Gülpinar, Rustem et Settergren [5], un arbre de scénarios généré avec l'approche globale est plus réaliste qu'un arbre généré avec l'approche séquentielle. De plus, selon eux l'approche séquentielle échoue souvent à construire un arbre qui soit globalement optimal.

Le problème de l'approche globale réside dans l'implémentation. En plus du nombre de variables et du nombre de propriétés statistiques qui sont multipliées par le nombre

Tableau 3.1 Erreurs sur les moments de ξ_t^1 (méthode des moments)

Semaine t	k	M_k	m_k	$\lvert M_k - m_k \rvert$	$\left\lvert \frac{M_k - m_k}{M_k} \right\rvert$
10	1	0	0.0001	0.0001	—
	2	19378	19456	78	0.0040
	3	-2915071	-2926511	11440	0.0039
	4	1608635830	1594969353	13666478	0.0085
11	1	0	-0.0001	0.0001	—
	2	60271	60255	16	0.0003
	3	15292399	15297736	5337	0.0003
	4	14885686279	14876503594	9182685	0.0006
12	1	0	-0.0005	0.0005	—
	2	54096	55032	936	0.0173
	3	5287663	5295461	7798	0.0015
	4	7545278662	7444375814	100902848	0.0134
13	1	0	0.0001	0.0001	—
	2	55842	56472	630	0.0113
	3	19891698	20059407	167709	0.0084
	4	20923205017	20560103945	363101072	0.0174
14	1	0	0.0001	0.0001	—
	2	22466	22926	460	0.0205
	3	2386214	2421112	34898	0.0146
	4	1898412051	1827144964	71267086	0.0375
15	1	0	-0.0002	0.0002	—
	2	10195	10194	1	0.0001
	3	867135	867102	33	0.0000
	4	318875920	318870136	5784	0.0000
16	1	0	0	0	—
	2	9986	9985	1	0.0001
	3	-2100862	-2100985	123	0.0001
	4	983999589	983801198	198391	0.0002
17	1	0	0.0005	0.0005	—
	2	11493	11791	298	0.0259
	3	-2262493	-2338995	76502	0.0338
	4	1275371423	1194606579	80764844	0.0633

Tableau 3.2 Erreurs sur les moments de ξ_t^2 (méthode des moments)

Semaine t	k	M_k	m_k	$\lvert M_k - m_k \rvert$	$\left\lvert \frac{M_k - m_k}{M_k} \right\rvert$
10	1	0	-0.0001	0.0001	—
	2	2271	2271	1	0.0003
	3	-76674	-76677	4	0.0000
	4	15632859	15638046	5187	0.0003
11	1	0	0.0004	0.0004	—
	2	6113	6113	0	0.0000
	3	-306029	-306091	62	0.0002
	4	108944640	108952882	8242	0.0001
12	1	0	0.001	0.001	—
	2	10075	10102	26	0.0026
	3	1242348	1245987	3639	0.0029
	4	523970071	521178047	2792024	0.0053
13	1	0	-0.0005	0.0005	—
	2	3471	3471	0	0.0001
	3	228613	228750	137	0.0006
	4	58982887	58954444	28444	0.0005
14	1	0	0.0003	0.0003	—
	2	2211	2260	49	0.0222
	3	-17310	-17324	13	0.0008
	4	21864866	21407511	457355	0.0209
15	1	0	-0.0003	0.0003	—
	2	2152	2139	13	0.0061
	3	244301	256268	11967	0.0490
	4	58220914	55914615	2306299	0.0396
16	1	0	0.0003	0.0003	—
	2	777	781	4	0.0049
	3	-21304	-21454	150	0.0070
	4	2459152	2428393	30759	0.0125
17	1	0	-0.0003	0.0003	—
	2	1806	1807	0	0.0001
	3	141214	141205	9	0.0001
	4	32794446	32801557	7110	0.0002

de périodes considérées, il faut ajouter au problème (3.1) les contraintes de corrélation entre les périodes. De plus, le degré de non-convexité augmente avec le nombre de variables et le nombre de contraintes. Par conséquent, même si cette approche semble plus solide théoriquement, elle requiert un très grand effort dans la programmation et est de loin la méthode la plus lente. Selon Gülpinar, Rustem et Settergren [5], les résultats obtenus par l'approche globale ne semblent pas justifier tout l'effort et tout le temps fourni.

3.2.2 Méthode basée sur les simulations

Comme dans la méthode des moments, on cherche à générer des réalisations aléatoires qui reproduisent les caractéristiques statistiques de la variable ξ_t^s. En revanche, on simule dans ce cas un grand nombre de réalisations aléatoires qui sont ensuite classifiées pour en réduire le nombre.

Il existe deux approches de simulation : l'approche parallèle et l'approche séquentielle (Gülpinar, Rustem et Settergren [5]). Celles-ci sont décrites ci-dessous de même que la transformation cubique qui permet de simuler un grand nombre de réalisations aléatoires d'une variable dont les quatre premiers moments sont connus. Finalement, les méthodes qui permettent de réduire le nombre de branches de l'arbre, et plus précisément la méthode SCENRED 2.0 de Römisch et la classification non-hiérarchique, sont présentées.

3.2.2.1 Simulations en parallèle

La méthode parallèle commence par simuler un éventail de scénarios, ou d'un ensemble de chemins (Kaut et Wallace [11]), pour l'horizon étudié. Par la suite, une classification est appliquée pour construire un arbre de scénarios.

Comme montré dans la figure (3.1), la classification des scénarios est d'abord appliquée à la première période. Les apports de la première période sont groupés dans différentes

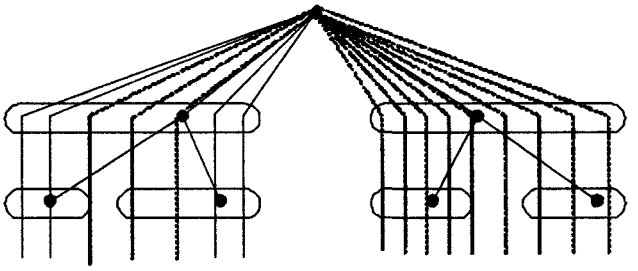


Figure 3.1 Simulation parallèle

classes selon leurs valeurs. Chaque classe de scénarios est remplacée par la valeur de son centroïde qui peut être la moyenne, la médiane ou le centre de gravité des données. Ces classes de scénarios créées à la première période sont à leur tour divisées en sous-classes selon les valeurs des apports à la période suivante et ainsi de suite... Il est important que la classe créée contienne suffisamment de scénarios pour qu'elle puisse toujours être divisée en sous-classes dans les périodes suivantes. On appelle cette procédure *forward method.* Quand la réduction de l'arbre commence à la fin de l'horizon et procède à rebours pour se terminer au nœud initial, on l'appelle *backward method* (Heitsch et Römisch [8]).

3.2.2.2 Simulations séquentielles

Contrairement à la méthode précèdente, où toutes les périodes étaient simulées en même temps, la méthode séquentielle[2](voir figure (3.2)) simule d'abord pour une période, puis ces simulations sont organisées en classes qui sont remplacées par leur centroïde. À partir de chaque centroïde on effectue d'autres simulations pour la période suivante et on classe de nouveau ces simulations, ainsi de suite jusqu'à la fin de l'horizon.

Comme seul le centroïde est gardé en mémoire et que les autres scénarios (les plus extrêmes) sont supprimés ceci donne des arbres plus homogènes (Gülpinar, Rustem et

[2]Cette méthode est appelée *sequential sampling algorithm* par Dupačovà, Gröwe-Kuska et Römisch [2]

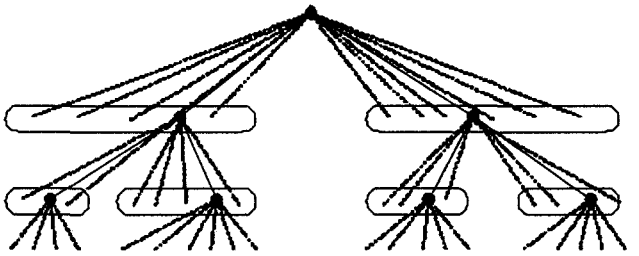


Figure 3.2 Simulation séquentielle

Settergren [5]). Puisqu'il n'est pas nécessaire de garder en mémoire tous les scénarios, cette méthode requiert généralement moins d'espace mémoire que la méthode des simulations en parallèle.

Dans un même arbre, il est possible d'utiliser les simulations parallèles et séquentielles à la fois. Par exemple, on peut d'abord utiliser la méthode parallèle et faire le classement. Puis quand les classes deviennent trop petites relancer les simulations.

3.2.2.3 Transformation cubique

Høyland, Kaut et Wallace [10] ont utilisé la transformation cubique pour générer des réalisations aléatoires à partir d'une distribution définie par ses quatre premiers moments. Soit X, un vecteur de nombres générés aléatoirement (au moyen d'un générateur de nombre aléatoires) dont la distribution est quelconque[3] et Y, un vecteur de nombres aléatoires dont la distribution est définie par ses quatre premiers moments $E(Y^k) = M_k$ pour $k = 1, \ldots, 4$. Au moyen d'une transformation cubique, il est possible d'obtenir Y en fonction de X :

$$Y = a + bX + cX^2 + dX^3 \tag{3.2}$$

où a, b, c, d sont les paramètres à déterminer.

[3]Høyland, Kaut et Wallace [10] ont généralisé les travaux de Fleishman [4] qui présente cette méthode mais avec X ayant une distribution normale

On peut développer l'équation (3.2) et écrire $E(Y^k)$ en fonction de $E(X^k)$, $k = 1, \ldots, 12$ et de a, b, c, d :

$$
\begin{aligned}
E[Y] &= a + bE[X] + cE[X^2] + dE[X^3] \\
E[Y^2] &= a^2 + 2abE[X] + (2ac + b^2)E[X^2] + (2ad + 2bc)E[X^3] \\
&\quad + (2bd + c^2)E[X^4] + 2cdE[X^5] + d^2E[X^6] \\
E[Y^3] &= a^3 + 3a^2bE[X] + (3a^2c + 3ab^2)E[X^2] + (3a^2d + 6abc + b^3)E[X^3] \\
&\quad + (a(6bd + 3c^2) + 3b^2c)E[X^4] + (6acd + 3b^2d + 3bc^2)E[X^5] \\
&\quad + (3ad^2 + 6bcd + c^3)E[X^6] + (3bd^2 + 3c^2d)E[X^7] \\
&\quad + 3cd^2E[X^8] + d^3E[X^9] \qquad (3.3) \\
E[Y^4] &= a^4 + 4a^3bE[X] + (4a^3c + 6a^2b^2)E[X^2] \\
&\quad + (4a^3d + 12a^2bc + 4ab^3)E[X^3] \\
&\quad + (a^2(12bd + 6c^2) + 12ab^2c + b^4)E[X^4] \\
&\quad + (12a^2cd + a(12b^2d + 12bc^2) + 4b^3c)E[X^5] \\
&\quad + (3a^2d^2 + a(24bcd + 4c^3) + 4b^3d + 6b^2c^2)E[X^6] \\
&\quad + (a(12bd^2 + 12c^2d) + 12b^2cd + 4bc^3)E[X^7] \\
&\quad + (12acd^2 + 6b^2d^2 + 12bc^2d + c^4)E[X^8] \\
&\quad + (4ad^3 + 12bcd^2 + 4c^3d)E[X^9] \\
&\quad + (4bd^3 + 6c^2d^2)E[X^{10}] + 4cd^3E[X^{11}] + d^4E[X^{12}]
\end{aligned}
$$

Il suffit de résoudre ce système d'équations non-linéaires pour obtenir les valeurs des paramètres a, b, c, d.

Autrement dit, pour générer n réalisations de la variable aléatoire Y, il faut :

- Générer n réalisations de la variable X. Nous avons utilisé la fonction *randn* de Matlab.

- Calculer $E[X^k], k = 1, \ldots, 12$.
- Résoudre le système d'équations (3.3) avec $E[Y^k] = M_k$ puis déterminer les valeurs de a, b, c, d.
- Calculer les valeurs obtenues pour Y avec (3.2).

La fonction *fsolve* de Matlab a été utilisée pour trouver les racines d'un système d'équations non-linéaires. Les valeurs initiales a, b, c, d fournies à *fsolve* sont désignées par a_0, b_0, c_0 et d_0 et sont des nombres générés aléatoirement par une loi uniforme(0,1).

L'algorithme *fsolve* s'arrête lorsqu'un minimum local est trouvé ou lorsque le nombre d'itérations maximales est atteint (fixé à 10 fois la valeur par défaut de Matlab). Si l'algorithme converge vers un minimum local qui n'est pas une racine, on recommence avec de nouvelles valeurs de $[a_0, b_0, c_0, d_0]$.

En pratique, l'algorithme échoue souvent à trouver les racines du système d'équations non-linéaires (équation (3.3)) mais une solution approximative où on approche les valeurs cibles $M_k, k = 1, \ldots, 4$ suffit. On prend alors la solution la plus proche de la racine. De plus, pour enlever l'effet de l'ordre de grandeur des quatre moments, $M_k - E((a + bX + cX^2 + dX^3)^k)$ est divisée par M_k^2.

La transformation cubique a été utilisée pour simuler des composantes principales (voir tableaux 3.3 et 3.4). La méthode a été évaluée en comparant les moments cibles M_k et les moments des nombres générés $m_k = E(Y^k)$. Cette méthode donne de bons résultats mais elle est moins précise que la méthode des moments. L'erreur relative est en moyenne de 9% et l'erreur maximale de 28%.

3.2.2.4 Classification non-hiérarchiques des scénarios

Puisqu'il y a un grand nombre de données à classer, les méthodes de classification non-hiérarchiques ont été choisies. Ces méthodes supposent que le nombre de classes k est connu

Tableau 3.3 Erreurs sur les moments de ξ_t^1 (transformation cubique)

Semaine t	k	M_k	m_k	$\lvert M_k - m_k \rvert$	$\left\lvert \frac{M_k - m_k}{M_k} \right\rvert$
10	1	0	0.2267	0.2267	—
	2	19378	18841	537	0.0277
	3	-2915071	-2652807	262264	0.0900
	4	1608635830	1511511613	97124218	0.0604
11	1	0	-1.8573	1.8573	—
	2	60271	65460	5189	0.0861
	3	15292399	18840529	3548130	0.2320
	4	14885686279	17062637911	2176951632	0.1462
12	1	0	-1.4567	1.4567	—
	2	54096	55641	1545	0.0286
	3	5287663	6584468	1296805	0.2453
	4	7545278662	7920911206	375632544	0.0498
13	1	0	-0.8531	0.8531	—
	2	55842	54110	1731	0.0310
	3	19891698	18219027	1672671	0.0841
	4	20923205017	19478554562	1444650455	0.0690
14	1	0	-0.1162	0.1162	—
	2	22466	22142	323	0.0144
	3	2386214	2120430	265784	0.1114
	4	1898412051	1828407155	70004896	0.0369
15	1	0	-1.2028	1.2028	—
	2	10195	10800	605	0.0593
	3	867135	985839	118704	0.1369
	4	318875920	345903529	27027609	0.0848
16	1	0	-0.7839	0.7839	—
	2	9986	11441	1455	0.1457
	3	-2100862	-2540123	439262	0.2091
	4	983999589	1205116598	221117009	0.2247
17	1	0	-0.9063	0.9063	—
	2	11493	12032	539	0.0469
	3	-2262493	-2489955	227462	0.1005
	4	1275371423	1380264569	104893145	0.0822

Tableau 3.4 Erreurs sur les moments de ξ_t^2 (transformation cubique)

Semaine t	k	M_k	m_k	$\lvert M_k - m_k \rvert$	$\left\lvert \frac{M_k - m_k}{M_k} \right\rvert$
10	1	0	-0.3113	0.3113	—
	2	2271	2211	59	0.0261
	3	-76674	-56697	19976	0.2605
	4	15632859	14466321	1166538	0.0746
11	1	0	0.447	0.447	—
	2	6113	6303	189	0.0310
	3	-306029	-357826	51797	0.1693
	4	108944640	115048735	6104095	0.0560
12	1	0	-0.0576	0.0576	—
	2	10075	10284	209	0.0207
	3	1242348	1305246	62898	0.0506
	4	523970071	543102245	19132174	0.0365
13	1	0	-0.0077	0.0077	—
	2	3471	3579	108	0.0311
	3	228613	245986	17373	0.0760
	4	58982887	62103327	3120440	0.0529
14	1	0	-0.0189	0.0189	—
	2	2211	2216	5	0.0022
	3	-17310	-19866	2556	0.1477
	4	21864866	21916458	51592	0.0024
15	1	0	-0.3659	0.3659	—
	2	2152	1975	177	0.0823
	3	244301	215897	28404	0.1163
	4	58220914	50795142	7425771	0.1275
16	1	0	-0.0246	0.0246	—
	2	777	742	34	0.0440
	3	-21304	-15414	5890	0.2765
	4	2459152	2147040	312111	0.1269
17	1	0	-0.0837	0.0837	—
	2	1806	1789	17	0.0096
	3	141214	138373	2841	0.0201
	4	32794446	32143423	651023	0.0199

à priori. La plupart des algorithmes de classification non-hiérarchiques se distinguent par la manière d'identifier les centroïdes des classes et par la manière de réassigner les observations.

Voici quelques méthodes employées pour obtenir les centroïdes initiaux de k classes (tirées de Sharma [19]) :

1. Sélectionner les k premières observations.
2. Sélectionner de façon aléatoire k observations.
3. Sélectionner la première observation comme le centroïde de la première classe. Sélectionner un deuxième centroïde de manière à ce que la distance qui le sépare du premier ne dépasse pas une certaine valeur (établie par l'utilisateur), et ainsi de suite ...
4. Utiliser une heuristique qui permet d'identifier k centroïdes qui soient le plus éloignés possible.
5. Utiliser des centroïdes fournis par l'utilisateur

Les méthodes pour réassigner les observations dans les classes se distinguent par leur manière de mettre à jour les centroïdes. Deux techniques sont décrites dans la littérature :

Mise à jour par lot : Appelée en anglais *batch update* elle consiste à réassigner chaque observation au centroïde le plus proche. Une fois que toutes les observations ont été assignées, on recalcule les nouveaux centroïdes puis on réassigne toutes les observations aux nouveaux centroïdes, et ainsi de suite...

Mise à jour au-fur-et-à-mesure : Appelée en anglais *online update* elle consiste à actualiser les centroïdes au fur et à mesure que les observations sont réassignées.

La fonction *kmeans* de Matlab a été utilisée pour effectuer le classement. Cette fonction permet de minimiser la distance totale de chaque observation au centroïde de la classe auquelle il appartient. La fonction *kmeans* utilise la mise à jour par lot et la mise à jour au-fur-et-à-mesure l'une à la suite de l'autre. La première étape permet d'identifier rapidement

mais approximativement les centroïdes. La seconde étape permet de raffiner la solution. Matlab garantit que *kmeans* converge au moins vers un minimum local. La distance utilisée est la distance euclidienne et les centroïdes initiaux sont choisis au hasard (options par défaut de la fonction *kmeans*).

Des contraintes supplémentaires peuvent être ajoutées pour raffiner la création des classes. On peut exiger, par exemple, un nombre d'observations maximal et minimal par classe ou fixer le rapport entre la plus grande et la plus petite classe. Plus les contraintes sont sévères, plus le temps de résolution est grand. Il s'agit de trouver un équilibre entre la rapidité d'exécution du programme et la qualité de la solution. Dans ce travail, aucune contrainte supplémentaire n'a été ajoutée. La fonction *kmeans* de Matlab a été utilisée et le nombre de classes a été fixé égal à 5.

3.2.2.5 Le logiciel SCENRED 2.0

L'autre méthode de réduction d'arbres qui a été utilisée dans le cadre de ce travail a été développée par Heitsch[4]. Ses recherches ont conduit à l'élaboration d'un logiciel nommé SCENRED. Celui-ci traite de deux types d'arbres TREE et FAN. Pour le type TREE, il faut définir dans le fichier d'entrée le nombre de nœuds de l'arbre et le nombre de variables à chaque nœud. Chaque nœud est décrit par son prédécesseur, sa probabilité et les valeurs des variables aléatoires. Pour le type FAN (éventail de scénarios), il faut définir le nombre de périodes de l'arbre, le nombre de scénarios (nombre de branches) et le nombre de variables à chaque noeud (les apports aux trois sites). Dans le fichier d'entrée, il faut donner la probabilité associée à chaque scénario ainsi que les valeurs des variables aléatoires à chaque période. Le programme SCENRED 2.0 offre deux possibilités : soit de construire ou de réduire un arbre. La construction peut s'appliquer uniquement aux arbres de type FAN alors que la réduction s'applique aux deux types d'arbres. Le programme offre le choix entre la méthode *forward* et *backward*. De plus, il faut déterminer le pourcentage de réduction désiré ($0 < p < 1$) et la norme de réduction (Maximum, L1, L2).

[4]www.math.hu-berlin.de/ heitsch/

3.2.3 Méthodes hybrides

D'autres manières de construire des arbres existent. Les méthodes hybrides combinent les principales idées de l'optimisation et celles de la simulation. Gülpinar, Rustem et Settergren [5] proposent de générer les apports d'eau par des simulations puis d'utiliser l'optimisation pour calculer les probabilités associées. Lindell [13] suggère aussi une méthode hybride où à chaque nœud des simulations sont lancées pour la période suivante. La moyenne et l'écart-type de ces simulations sont calculées et un (sous-)arbre, composé de moins de branches mais qui a les mêmes caractéristiques statistiques des simulations, est trouvé. Aucune méthode hybride n'a été programmée dans le cadre de ce travail.

3.3 Construction d'arbres multidimensionnels

Dans cette section, la construction d'arbres multidimensionnels, c'est-à-dire ayant plusieurs variables mutuellement corrélées, est présentée. Plusieurs approches existent. Vitoriano, Cerisola et Ramos [21] suggèrent de construire un premier arbre puis d'utiliser celui-ci pour construire les autres arbres par régression linéaire. Une forte corrélation doit exister entre les différents arbres. Gülpinar, Rustem et Settergren [5] suggèrent d'ajouter au problème d'optimisation (3.1) des contraintes pour modéliser la corrélation. Halldin [6] propose de construire des arbres unidimensionnels de composantes principales puis de les recombiner pour obtenir des arbres multidimensionnels. Finalement, Høyland, Kaut et Wallace [10] proposent une heuristique, basée sur la décomposition de Cholesky, qui permet de construire un arbre multidimensionnel. Dans ce mémoire, seulement ces deux dernières méthodes sont considérées. Tout d'abord, la procédure de recombinaison des arbres de composantes principales est décrite. Ensuite, l'approche de Høyland, Kaut et Wallace [10] est présentée.

Les résidus R_t^s, pour $t = 1, 2, 3, \ldots, 52$, issus de la modélisation PAR(2) sont indépendants des résidus des semaines passées ou futures. Ceci permet de générer d'abord un arbre

de résidus avec un horizon de plusieurs semaines sans structure de dépendance, puis avec le modèle PAR(2), défini au chapitre 2, de calculer les valeurs des apports à la période t :

$$APP_t^{s*} = \phi_{0,t}^s + \phi_{1,t}^s app_{t-1}^s + \phi_{2,t}^s app_{t-2}^s + R_t^{s*} \tag{3.4}$$

où app_{t-1}^s et app_{t-2}^s sont les valeurs générées ou réelles des apports dans les deux périodes précédentes, R_t^{s*} le vecteur des résidus générés et APP_t^{s*} le vecteur des apports générés.

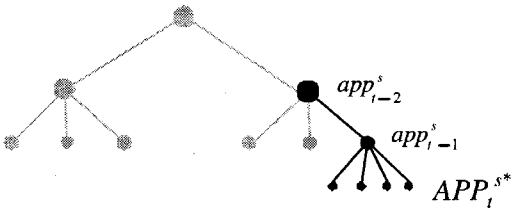


Figure 3.3 Construction de l'arbre d'apports à partir des résidus

3.3.1 Recombinaison des composantes principales

Dans cette section, la procédure de construction d'arbres de résidus mutuellement corrélés à partir des arbres indépendants de composantes principales (générés par la méthode des moments, la méthode de simulation parallèle ou la méthode de simulation séquentielle) sera présentée. Mais avant de donner les algorithmes qui ont été implantés, voici comment déduire les résidus à partir des composantes principales. Nous savons du chapitre 2 que :

$$\begin{bmatrix} \xi_t^1 & \xi_t^2 & \dots & \xi_t^S \end{bmatrix} = \begin{bmatrix} R_t^1 & R_t^2 & \dots & R_t^S \end{bmatrix} COEFS$$

Ce qui peut aussi s'écrire :

$$\begin{bmatrix} \xi_t^1 & \xi_t^2 & \dots & \xi_t^S \end{bmatrix} \cdot (COEFS)^T = \begin{bmatrix} R_t^1 & R_t^2 & \dots & R_t^S \end{bmatrix} COEFS \cdot (COEFS)^T$$

puisque $COEFS \cdot (COEFS)^T = I$ (contrainte d'orthogonalité), il est possible de reconstruire les résidus $R_t^1, R_t^2, \dots, R_t^S$ à partir des composantes principales $\xi_t^1, \xi_t^2, \dots, \xi_t^S$.

$$\begin{bmatrix} R_t^1 & R_t^2 & \dots & R_t^S \end{bmatrix} = \begin{bmatrix} \xi_t^1 & \xi_t^2 & \dots & \xi_t^S \end{bmatrix} \cdot (COEFS)^T \tag{3.5}$$

On peut aussi les approximer en utilisant seulement les Q premières composantes :

$$\begin{bmatrix} R_t^1 & R_t^2 & \dots & R_t^S \end{bmatrix} \cong \begin{bmatrix} \xi_t^1 & \xi_t^2 & \dots & \xi_t^Q \end{bmatrix} \cdot (COEFS(1:Q))^T \tag{3.6}$$

3.3.1.1 Méthode des moments

L'algorithme qui permet de construire l'arbre d'apports en utilisant la méthode des moments procède comme suit :

1: **pour** $t = t_{debut}$ **à** t_{fin} ;
2: Faire l'analyse en composantes principales (ACP) des résidus $R_t^1, R_t^2, \dots, R_t^S$ (voir section 2.4.1) pour obtenir ξ_t^s où $s = 1, \dots, Q$ et $Q \leq S$.
3: **pour** $s = 1, \dots, Q$;
4: Calculer les moments M_k de ξ_t^s, où $k = 1, 2, 3, 4$.
5: Générer par la méthode des moments ξ_t^{s*} des réalisations aléatoires qui ont les mêmes moments M_k.
6: Calculer R_t^{s*} à partir de ξ_t^{s*}(équation 3.6).
7: Construire l'arbre des résidus.
8: Construire l'arbre d'apports (équation (3.4))

Les résultats obtenus par cette procédure sont présentés à l'annexe (A.3). La validation des résultats s'appuie sur la comparaison des moments M_k des résidus R_t^s et des moments m_k des résidus générés R_t^{s*}. L'erreur relative $\left|\frac{M_k - m_k}{M_k}\right|$ pour le site Baskatong est très faible et plus précisèment inférieure à 6% pour les quatre moments. Pour Cabonga, l'erreur relative est généralement supérieure à 50% et atteint parfois les 100%. Ce résultat ne surprend pas puisque les apports à Cabonga sont très faibles par rapport aux apports des autres sites et peuvent être considérés comme du bruit. Pour le site Paugan, la moyenne et la variance sont bien reproduites mais le troisième et quatrième moment ont une erreur relative moyenne de 18% et 7%.

Quant à la corrélation qui existe entre les résidus R_t^s, nous l'avons évaluée en calculant la statistique RMSE (*Root Mean Squared Error*) qui est la racine carrée de la moyenne des carrés de la différence entre chacun des éléments de la matrice $CORR_t$ et $CORR_t^*$. Selon cette statistique, l'erreur est en moyenne de 23%. Le tableau des résultats à l'annexe (A.3) qui présente les matrices de corrélations $CORR_t$ et $CORR_t^*$ révèle que c'est principalement la corrélation entre les résidus de Baskatong et Paugan qui est reproduite adéquatement.

3.3.1.2 Méthode de simulation parallèle

L'algorithme qui permet de construire l'arbre d'apports en utilisant la méthode de simulation parallèle procède comme suit :

1: **pour** $t = t_{debut}$ **à** t_{fin} ;
2: Faire l'ACP des résidus $R_t^1, R_t^2, \ldots, R_t^S$ (voir section 2.4.1) pour obtenir ξ_t^s où $s = 1, \ldots, Q$ et $Q \leq S$.
3: **pour** $s = 1, \ldots, Q$;
4: Calculer les moments M_k de ξ_t^s où $k = 1, 2, 3, 4$.
5: Utiliser la transformation cubique pour simuler un grand nombre de réalisations aléatoires ξ_t^{s*} qui ont les mêmes moments M_k.
6: Piger au hasard n réalisations parmi chacune des ξ_t^{s*}.
7: Calculer R_t^{s*} à partir de ξ_t^{s*}(équation 3.6).
8: Construire l'éventail de scénarios de résidus.
9: Construire l'éventail de scénarios d'apports (équation (3.4)).
10: Utiliser SCENRED 2.0 pour réduire l'éventail de scénarios en arbre de scénarios.

On ne peut pas évaluer la méthode simulation parallèle comme les méthodes précèdentes puisque son but n'est pas de reproduire les moments cibles. Plus loin dans la discussion une évaluation qualitative de l'arbre d'apports sera effectuée.

3.3.1.3 Méthode de simulation séquentielle

L'algorithme qui permet de construire l'arbre d'apports en utilisant la méthode de simulation séquentielle procède comme suit :

1: **pour** $t = t_{debut}$ **à** t_{fin} ;
2: Faire l'ACP des résidus $R_t^1, R_t^2, \ldots, R_t^S$ (voir section 2.4.1) pour obtenir ξ_t^s, $s = 1, \ldots, Q$ et $Q \leq S$.
3: **pour** $s = 1, \ldots Q$;
4: Calculer les moments M_k de ξ_t^s où $k = 1, 2, 3, 4$.
5: Utiliser la transformation cubique pour simuler un grand nombre de réalisations aléatoires ξ_t^{s*} qui ont les mêmes moments M_k que ξ_t^s.
6: Calculer R_t^{s*} à partir de ξ_t^{s*}(équation3.6).
7: Faire k classes avec les réalisations des résidus R_t^{s*}. Identifier chaque centroïde et attribuer une probabilité selon le nombre d'éléments dans chaque classe.
8: Construire l'arbre de résidus
9: Construire l'arbre d'apports (équation (3.4)).

À l'annexe (A.4) les moments obtenus des résidus R_t^{s*}, avant d'effectuer le classement (étape 6), sont présentés. Pour le site Baskatong, l'erreur relative est inférieure à 11% et est en moyenne de 4%. Pour le site Paugan, la moyenne des erreurs relatives de la moyenne et de la variance est faible mais est légèrement élevée pour l'asymétrie et le kurtosis (22% et 11%). De plus, comme pour la méthode des moments les résultats obtenus pour le site Cabonga sont insatisfaisants. Il en est de même pour la corrélation. Elle a été reproduite avec une erreur moyenne de 23%. D'ailleurs, comme la méthode des moments, seulement la corrélation entre Paugan et Baskatong tend vers les valeurs cibles.

Finalement, la classification effectuée à l'étape 7 modifie les résultats. La moyenne est conservée pour les trois sites, mais la variance, l'asymétrie et le kurtosis ne sont pas préservés pour les sites Paugan et Cabonga. Pour Baskatong, l'erreur relative moyenne est près de 10%, 21% et 40% pour la variance, l'asymétrie et le kurtosis.

3.3.2 L'heuristique de Høyland et al.

Høyland, Kaut et Wallace [10] présentent une heuristique qui génère des variables mutuellement corrélées dont ont connaît la matrice de corrélation et les quatre premiers moments. Dans le cadre de ce travail il s'agit de $CORR_t$, la matrice de corrélation des résidus des trois sites, et des quatre premiers moments $M^s_{k,t}$, $k = 1, 2, 3, 4, s = 1, \ldots, S$ de ces résidus pour chacun des sites. Leur algorithme se résume comme suit :

1. Générer S variables aléatoires indépendantes dont les moments correspondent à $M^s_{k,t}$.
2. Transformer les S variables aléatoires pour que la matrice de corrélation corresponde à $CORR_t$. Ceci a pour effet d'altérer la valeur des troisièmes et quatrièmes moments des S variables.
3. Suite aux altérations, transformer de nouveau les S variables pour obtenir les bons moments cibles.

Si les variables à l'étape 1 sont indépendantes, cette procédure, qu'on nommera algorithme noyau, est suffisante pour produire des variables avec les moments et la matrice de corrélation spécifiée. L'indépendance est possible seulement si les réalisations des variables aléatoires sont infinies et équiprobables. Puisque les réalisations sont finies, il faut procéder itérativement et essayer d'approcher les spécifications des moments et de la matrice de corrélation, d'où l'heuristique (Høyland, Kaut et Wallace [10]).

Cet algorithme utilise la transformation cubique (voir section 3.2.2.3) aux étapes 1 et 3. À l'étape 2, il utilise une transformation matricielle, basée sur la décomposition de Cholesky, présentée dans la section suivante.

3.3.2.1 Transformation matricielle

Le but de cette transformation est de convertir un ensemble de variables aléatoires $\mathbf{X}_t$ mutuellement indépendantes en un ensemble de variables $\mathbf{Y}_t$ mutuellement dépendantes,

dont la matrice de corrélation est $CORR_t$.

$$\mathbf{Y}_t = L\mathbf{X}_t \tag{3.7}$$

où L est la matrice triangulaire qui provient de la décomposition de Cholesky

$$CORR_t = LL^T \tag{3.8}$$

Si les S variables aléatoires X_t proviennent de population normales $N(0,1)$ et sont mutuellement indépendantes, alors les S variables aléatoires $\mathbf{Y}_t = L\mathbf{X}_t$ proviennent aussi d'une population normale ayant la matrice de corrélation $CORR_t$.

On suppose que X_t^s possède ces propriétés :

i. $E[(X_t^s)^k]$ existe pour $k = 1, \ldots, 4$

ii. $E[X_t^s] = 0$ et $E[(X_t^s)^2] = 1$

iii. les variables aléatoires sont indépendantes : $R = I$

Si $\mathbf{Y}_t = L\mathbf{X}_t$ et L provient de la décomposition de Cholesky de $CORR_t$, alors Y_t^s possède les propriétés suivantes :

i. $E[(Y_t^s)^k]$ existe pour $k = 1, \ldots, 4$

ii. $E[Y_t^s] = 0$ et $E[(Y_t^s)^2] = 1$

iii. Y_t^s a une matrice de corrélation $CORR_t = LL^T$

iv. $E[(Y_t^s)^3] = \sum_{i=1}^{s} L_{si}^3 E[(X_t^i)^3]$

v. $E[(Y_t^s)^4] = \sum_{i=1}^{s} \left(L_{si}^4 (E[(X_t^i)^4] - 3) \right) + 3$

Les propriétés iv et v ont comme réciproque :

$$E[(X_t^s)^3] = \frac{1}{L_{ss}^3} \left(E[(Y_t^s)^3] - \sum_{i=1}^{s-1} L_{si}^3 E[(X_t^i)^3] \right) \tag{3.9}$$

$$E[(X_t^s)^4] = \frac{1}{L_{ss}^4} \left[E[(Y_t^s)^4] - 3 - \sum_{i=1}^{s-1} L_{si}^4 (E[(X_t^i)^3] - 3) \right] + 3 \tag{3.10}$$

Puisque la transformation matricielle s'effectue sur des variables de moyenne nulle et de variance égale à 1, il faut centrer et réduire les résidus R_t^s :

$$Y_t^s = \frac{R_t^s - M_{1,t}^s}{\sqrt{M_{2,t}^s}} \tag{3.11}$$

ainsi les moments cibles de Y_t^s sont :

$$E[Y_t^s] = 0 \qquad E[(Y_t^s)^2] = 1 \tag{3.12}$$

$$E[(Y_t^s)^3] = \frac{M_{3,t}^s}{(M_{2,t}^s)^{\frac{3}{2}}} \qquad E[(Y_t^s)^4] = \frac{M_{4,t}^s}{(M_{2,t}^s)^2} \tag{3.13}$$

3.3.2.2 L'algorithme noyau

L'algorithme noyau est l'algorithme initial, sans heuristique, que proposent Høyland, Kaut et Wallace [10]. Il est composé d'une première phase de traitement préliminaire des données (les trois premières étapes) et d'une seconde de génération des variables (les trois dernières étapes).

Algorithme 3.1 Algorithme noyau

1: **pour** $t = t_{debut}$ **à** t_{fin} ;
2: Spécifier les moments cibles $M_{k,t}^s$ et la matrice de corrélation cible $CORR_t$ des résidus $\mathbf{R}_t$.
3: Normaliser $\mathbf{R}_t$ (3.11) et trouver les moments cibles $E[(Y_t^s)^k]$ de $\mathbf{Y}_t$ (3.12 & 3.13).
4: Calculer L (3.8) et déduire les moments cibles $E[(X_t^s)^k]$ de $\mathbf{X}_t$ (3.9 & 3.10).
5: Générer X_t^{s*}, avec les moments cibles $E[(X_t^s)^k]$ pour $i = 1, \ldots, S$.
6: Déduire Y_t^{s*} : $\mathbf{Y}_t = L\mathbf{X}_t$.
7: Calculer les résidus $R_t^{s*} = \sqrt{M_{2,t}^s} Y_t^{s*} + M_{1,t}^s$.

À l'étape 1, on transforme les moments cibles pour qu'il soit possible d'utiliser la transformation matricielle. On obtient $E[(Y_t^s)^k]$ qui sont les moments cibles des résidus mutuellement corrélés centrés et réduits. Ensuite on calcule L, la matrice triangulaire qui provient de la décomposition de Cholesky et les moments cibles $E[(X_t^s)^k]$ des résidus indépendants.

Une fois que ces cibles ont été identifiées, on procède à la génération des variables. D'abord on génère de manière indépendante les S variables ayant les moments spécifiés $E[(X_t^s)^k]$. Puis à l'aide de l'équation (3.7), on transforme les variables de sorte que la matrice de corrélation soit égale à $CORR_t$. Finalement la dernière étape permet de passer des variables centrées et réduites à des variables dont les spécifications correspondent à celles de $\mathbf{R}_t$.

3.3.2.3 L'heuristique

L'algorithme noyau est suffisant lorsque les variables générées X_t^{s*}, $s = 1, \ldots, S$, sont mutuellement indépendantes. Or selon Høyland, Kaut et Wallace [10], l'indépendance n'est pas possible puisqu'un nombre fini de réalisations aléatoires sont générés. Pour atteindre itérativement cette indépendance, ils proposent une heuristique. Les modifications sont apportées aux étapes 4 et 5 de l'algorithme noyau. À l'étape 4 on essaie d'atteindre itérativement l'indépendance des variables à ϵ_x près, et à l'étape 5 on essaie d'atteindre la corrélation $CORR_t$ à ϵ_y près. Le détail de ces modifications se trouve à l'annexe (B.1). Les auteurs n'ont pu prouver la convergence de l'heuristique. Si l'heuristique échoue à trouver une solution appropriée, ils préconisent de répéter l'exercice ou d'augmenter le nombre de scénarios.

Nous avons programmé l'heuristique mais la convergence à une solution a été difficile, surtout à l'étape 5. La figure (3.4) illustre un exemple où la statistique $RMSE$ ne converge pas d'une itération à l'autre vers le seuil d'arrêt ϵ_y fixé à 0.01 (en ligne pointillée).

Par conséquent, nous avons décidé de ne pas recourir à l'heuristique puisque la solution initiale est satisfaisante. Les résultats, obtenus uniquement grâce à l'algorithme noyau (3.1), sont présentés à l'annexe (A.5). L'erreur relative $\left|\frac{M_k - m_k}{M_k}\right|$ obtenue par l'algorithme noyau, est très faible pour les trois sites. Elle se situe en moyenne à 4%, 5% et 7% pour Baskatong, Paugan et Cabonga. La corrélation est aussi très bien reproduite : la statistique RMSE donne une erreur moyenne de 3%.

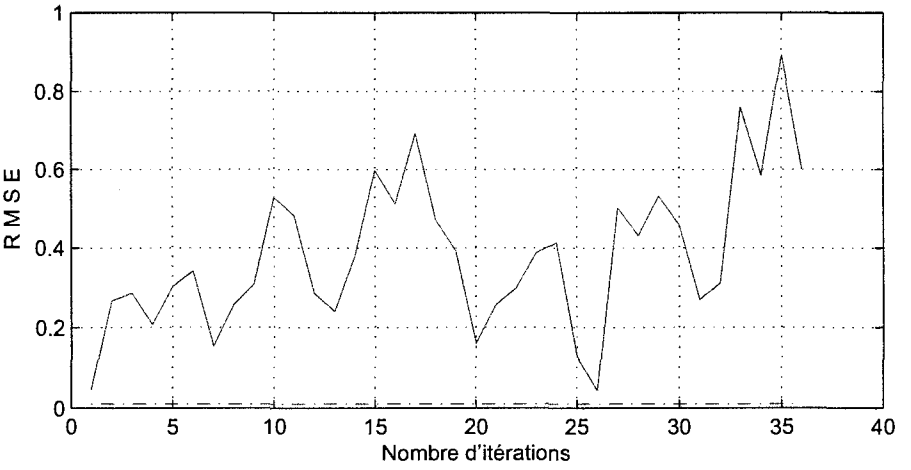


Figure 3.4 Un exemple du comportement de l'erreur en utilisant l'heuristique

Après avoir procédé à la classification, on constate une détérioration des résultats qui est similaire à la simulation séquentielle. En effet, la moyenne est conservée pour les trois sites mais la variance, l'asymétrie et le kurtosis sont altérés pour les sites de Paugan et Cabonga. La classification ne semble pas avoir le même effet sur le site Baskatong : l'erreur relative se situe en moyenne autour de 12%, 22% et 43% pour la variance, l'asymétrie et le kurtosis.

3.4 Discussion

Ce chapitre a tout d'abord porté sur les méthodes de génération d'arbres unidimensionnels, c'est-à-dire ayant une seule variable. La méthode des moments et la méthode des simulations, qui utilise la transformation cubique, ont été utilisées pour générer un arbre dont la variable est une composante principale. Ces deux méthodes permettent de générer des nombres aléatoires qui ont les mêmes caractéristiques statistiques (les quatre premiers moments) que chacune des composantes principales. La méthode génère un nombre res-

treint de valeurs (5 dans ce travail), tandis que la méthode des simulations génère un nombre plus important de valeurs (environ 100). La méthode des moments s'est avérée être plus performante car elle reproduit mieux les caractéristiques statistiques. De plus, elle ne nécessite pas une classification des nombres aléatoires qui détériore les résultats.

La deuxième section du chapitre a été consacrée aux méthodes de génération d'arbres multidimensionnels, c'est-à-dire ayant plusieurs variables. La première approche présentée a été la construction d'arbres de résidus à partir de la recombinaison des arbres de composantes principales générées par la méthode des moments ou la méthode séquentielle. Les résultats obtenus avec les deux méthodes sont semblables : en général les résidus générés au site Baskatong sont les plus fidèles à l'historique suivis de près par les résidus du site Paugan. Les résidus générés au site Cabonga, obtenus par les deux méthodes, ne reflètent pas l'historique. Ceci n'est pas sans lien avec le fait que les apports de Cabonga sont presque considérés comme du bruit en comparaison avec les deux autres sites. D'autre part, la corrélation obtenue par les deux méthodes ne traduit pas bien les données de l'historique. L'analyse en composantes principales a donc permis de générer des résidus proches de l'historique pour les sites Baskatong et Paugan mais a échoué à reproduire correctement les résidus du site Cabonga et la corrélation entre les résidus.

La méthode de Høyland, Kaut et Wallace [10] a été capable de résoudre ce problème : l'erreur relative moyenne pour le site de Cabonga, tout comme pour les deux autres sites, est très faible. De plus, elle a aussi été plus performante pour la génération de résidus corrélés. La statistique RMSE est très faible, autour de 3% en comparaison à 23% pour la méthode de simulation séquentielle.

La classification des valeurs générées par la méthode de simulation séquentielle et par la méthode de Høyland, Kaut et Wallace [10] avaient le même effet. Indépendamment de la méthode, le classement dégradait les résultats. Seule la moyenne était conservée. Ce sont surtout les moments des résidus des sites de Paugan et de Cabonga qui ont été fortement modifiés.

Une évaluation qualitative des arbres de scénarios permet de compléter la discussion sur les résultats. Dans l'annexe (C), les arbres générés par les quatre méthodes sont présentés. Pour chacune des méthodes, les apports réels ont été superposés à l'arbre de prédictions. Cela a été fait pour les trois sites et pour les semaines $6, 7, 8$ d'abord et $9, 10, 11$ ensuite. On constate que pour toutes les méthodes, sauf pour celle de la simulation parallèle, les arbres des prédictions offrent tous des performances équivalentes : ils sous-estiment les apports au début de la crue (semaines $6, 7, 8$) et surestiment les apports pour les semaines $9, 10$ et 11. Toutefois, les arbres indiquent bien que les volumes d'eau augmentent entre les semaines 6 à 8 et commencent à décroître entre les semaines 9 à 10. La méthode de simulation parallèle semble celle qui performe le mieux.

CHAPITRE 4 : CONCLUSION

Il est important de tenir compte du comportement aléatoire des apports d'eau aux installations hydroélectriques lorsque l'on cherche la règle optimale de gestion de ces installations. Le problème de trouver cette règle de gestion est cependant difficile à résoudre lorsqu'il y a plusieurs réservoirs à gérer. Ce problème d'optimisation ne peut en fait être résolu dans un temps raisonnable que pour des problèmes de petite taille (moins de 4 variables d'état).

Des problèmes de plus grande taille peuvent cependant être résolus lorsque le nombre de pas de temps pour lequel la règle de gestion doit être déterminée d'avance est très petit (3 ou 4 périodes). Dans ce cas, il suffit de représenter les aléas des apports par des arbres de scénarios puis de faire appel à des logiciels de programmation linéaire ou non linéaire pour résoudre le problème d'optimisation. Ce mémoire décrit différentes méthodes pour construire des arbres de scénarios.

Les données dont nous disposons sont des historiques d'apports prélevés aux trois sites du bassin versant de la rivière Gatineau. Ces historiques montrent que les apports sont fortement corrélés dans le temps et dans l'espace.

Après avoir effectué quelques transformations aux séries historiques dans le but de mieux représenter les apports probables lors d'une crue, nous avons construit un modèle de prédiction Periodic-Auto-Regressive d'ordre 2 (PAR(2)) par site et par semaine. Les résidus générés par ces modèles représentent en fait la partie aléatoire des apports qui seront générés. Toutes les techniques de génération d'arbres que nous avons présentées sont appliquées aux résidus. Puis les équations de régression des modèles PAR(2) sont utilisées pour retrouver l'arbre d'apports.

Nous avons présenté deux différentes approches pour générer des arbres de scénarios ayant plusieurs variables. La première consiste à utiliser l'analyse en composantes prin-

cipales (ACP) pour transformer les résidus corrélés en variables indépendantes (appelées composantes principales). Par la suite, il s'agit de générer des arbres indépendants entre eux et ayant une variable qui représente une composante principale. La dernière étape consiste à les recombiner pour obtenir des arbres ayant plusieurs variables corrélées. La deuxième approche est celle proposée par Høyland, Kaut et Wallace [10] et qui permet de générer directement des arbres de variables aléatoires corrélées.

Les différentes méthodes, citées ci-dessus pour générer des arbres à une variable, sont la méthode des moments, la méthode de simulation séquentielle et la méthode de simulation parallèle. Il ressort de notre travail que la méthode des moments est la meilleure. Elle reproduit mieux les quatre premiers moments (la moyenne, la variance, l'asymétrie et le kurtosis) et ne requiert pas d'étape de classification comme les méthodes de simulation. La classification a comme principal effet d'éloigner encore d'avantage les variables générés des données réelles qu'elles sont censées représenter.

Pour générer des arbres avec plusieurs variables, les résultats obtenus avec l'algorithme de Høyland, Kaut et Wallace [10] sont meilleurs que ceux obtenus avec l'approche basée sur les composantes principales. L'algorithme de Høyland, Kaut et Wallace [10] reproduit nettement mieux les quatre premiers moments et la matrice de corrélation. Mais étant donné qu'il requiert une étape de classification des nombres générés, les valeurs de l'asymétrie et du kurtosis sont moins précises.

Selon nous, il y aurait intérêt à comparer les méthodes de génération d'arbres dites directes (citées dans le mémoire) que nous n'avons pas testé à celles que nous avons implantées. Ces méthodes directes sont moins compliquées à implémenter et de plus donnent des arbres ayant les moyenne et variance souhaitées.

Ce travail a permis de faire un survol des méthodes de générations d'arbres de scénarios d'apports, d'en programmer un certain nombre et de discuter de leurs avantages

et limites. Cependant, il serait nécessaire de les implanter dans un modèle d'optimisation pour connaître l'impact sur la politique de gestion des réservoirs. Comme mentionné par Kaut et Wallace [11], il est possible d'avoir une méthode de génération de scénarios qui ne converge pas vers les distributions originales mais qui répond néanmoins très bien aux exigences de cas pratiques. L'objectif de générer un arbre de scénarios ne se réduit pas à une approximation de la distribution de probabilités du processus stochastique. Un bon arbre de scénarios est celui qui contribue à une bonne prise de décision.

BIBLIOGRAPHIE

[1] DUPAČOVÀ, J., CONSIGLI, G. et WALLACE, S.W. (2000), *Scenarios for multistage stochastic programs*, Annals of Operations Research, 100, pp. 25-53.

[2] DUPAČOVÀ, J. and GRÖWE-KUSKA, N. et RÖMISCH, W. (2003), *Scenario reduction in stochastic programming : an approach using probability metrics*, Mathematical Programming, A95, pp.493-511.

[3] FILLIBEN, J.J. (1975), *The probability plot correlation coefficient test for normality*, Technometrics, 1(17), pp.111-117.

[4] FLEISHMAN, A.I. (1978), *A method for simulating non-normal distributions*, Psychometrika, 43(1), pp.521-532.

[5] GÜLPINAR, F., RUSTEM, B. et SETTERGREN, R. (2004), *Simulation and optimization approaches to scenario tree generation*, Journal of Economic Dynamics et Control, 28, pp. 1291-1315.

[6] HALLDIN, R. (2005), *Stochastic modeling and optimization under uncertainty of a hydro power system*, Centre for Mathematical Sciences, Lund Institute of Technology.

[7] HEITSCH, H., RÖMISCH, W. (2003), *Scenario reduction algorithms in stochastic programming*, Computational Optimization and Applications, 24, pp. 187-206.

[8] HEITSCH, H., RÖMISCH, W. (Pre-print), *Scenario tree modelling for multistage stochastic programs.*

[9] HIGLE, J.L. (2005), *Stochastic programming : Optimization when uncertainty matters*, Tutorials in operations research, INFORMS.

[10] HØYLAND, K., KAUT, M. et WALLACE, S.W.(2003), *A heuristic for moment-matching scenario generation*, Computational Optimization and Applications, 24, pp.169-185.

[11] KAUT, M. et WALLACE, S.W. (2003), *Evaluation of Scenario-Generation Methods for Stochastic Programming*, citeseer.ist.psu.edu/kaut03evaluation.html.

[12] LABADIE, J.W. (2004), *Optimal operation of multireservoir systems : state-of-the-art review*, Journal of water resources planning and management, 130, pp. 93-111.

[13] LINDELL, F. (2000), *On generation of scenario trees for stochastic optimization*, Centre for Mathematical Sciences, Lund Institute of Technology.

[14] LOUCKS, D.P., STEDINGER, J.R. et HAITH, D.A. (1981), *Water resource systems planning and analysis*, Prentice-Hall.

[15] MAIDMENT, D.R. (1993), *Handbook of Hydrology*, McGraw-Hill.

[16] MERLEAU, J., EVIN, G., PERREAULT, L., FAVRE, A.C, TREMBLAY, D. et ANGERS, J.F. (2005), *Analyse descriptive des prévisions hydrologiques d'ensemble et*

modélisation par un mélange de deux lois gamma. Création d'hydrogrammes de base d'hydrogrammes prévisionnels à pas de temps journaliers, INRS-Eau, terre et environnement.

[17] RAMSAY, J. et SILVERMAN, B. (1997), *Functional data analysis*, Springer.

[18] SALAS, J.D., DELLEUR, J.W., YEVJEVICH, V. et LANE, W.L.(1980), *Applied modeling of hydrologic time series*, Water Ressources publications.

[19] SHARMA, S. (1996), *Applied multivariate techniques*, John Wiley et Sons

[20] TURGEON, A. (1981), *A decomposition method for the long term scheduling of reservoirs in series*, Water resources research, 17, pp. 1565-1570.

[21] VITORIANO, B., CERISOLA, S. et RAMOS, A. (2000), *Generating scenarios trees for hydro inflows*, 6th International Conference on probabilistic methods applied to power systems.

[22] YEH, W.W.G. (1985), *Reservoir management and operations models : a state-of-the-art review*, Water resources research, 21, pp. 1797-1818.

ANNEXE A : TABLEAUX

A.1 Modélisation PAR(2)

Tableau A.1: Résultats de la modélisation PAR

Semaine	Baskatong			Cabonga			Paugan		
	ϕ_0	ϕ_1	ϕ_2	ϕ_0	ϕ_1	ϕ_2	ϕ_0	ϕ_1	ϕ_2
1	21.42	0.42	0.17	12.74	0.44	0.07	17.94	0.88	-0.26
2	22.80	-0.15	0.65	9.56	0.16	0.45	18.95	**0.01**	0.57
3	1.90	0.59	0.24	8.03	0.36	0.20	**0.42**	1.15	-0.14
4	15.34	1.01	-0.29	9.81	0.58	**0.01**	-2.20	0.80	0.29
5	17.42	0.92	-0.15	7.97	0.44	0.18	23.19	1.01	-0.44
6	**-2.39**	2.78	-1.62	9.48	0.70	**0.02**	**-0.56**	1.33	-0.27
7	55.29	0.59	-0.12	12.17	0.34	0.19	13.50	0.76	**0.02**
8	30.63	1.29	-0.41	13.90	0.66	-0.05	22.94	0.77	**-0.01**
9	102.76	1.40	-0.63	23.27	0.79	-0.32	24.51	1.71	-0.71
10	226.50	1.09	-0.75	39.27	0.40	0.14	77.37	0.79	**0.01**
11	633.07	0.68	-0.69	90.49	0.30	-0.18	225.30	0.84	-0.79
12	793.98	0.32	-0.13	147.70	0.22	-0.48	242.18	0.54	-0.12
13	299.49	0.58	0.05	54.37	0.37	0.09	61.81	0.77	-0.03
14	287.86	0.50	-0.03	67.60	0.30	**-0.02**	46.75	0.85	-0.18
15	53.72	0.61	0.03	22.42	0.47	0.08	43.62	1.02	-0.30
16	164.38	0.72	-0.20	23.56	0.26	0.21	16.25	1.01	-0.23
17	116.36	0.64	-0.06	31.89	0.65	-0.21	58.01	0.84	-0.26
18	83.13	0.79	-0.11	32.19	0.23	0.08	38.38	0.74	-0.11
19	95.76	0.81	-0.26	17.70	0.44	0.04	14.72	0.83	-0.08
20	97.25	0.42	0.03	16.58	0.43	0.14	7.33	1.05	-0.21
21	82.84	0.67	-0.08	24.21	0.35	**0.01**	11.92	1.01	-0.21
22	12.90	1.04	-0.08	15.08	0.66	**-0.03**	10.81	1.03	-0.23
23	36.93	0.87	-0.07	22.73	0.28	0.08	17.99	0.79	-0.12
24	61.52	0.40	0.16	17.66	0.39	0.03	19.89	1.00	-0.29
25	42.21	0.82	-0.13	20.21	0.26	**-0.00**	6.92	0.69	0.04
26	54.01	1.16	-0.48	12.43	0.48	-0.04	13.35	0.70	-0.05
27	41.60	0.77	-0.15	15.73	0.88	-0.40	12.34	1.51	-0.72
28	50.40	0.61	0.04	9.98	0.55	0.18	15.18	0.45	0.11
29	21.02	1.50	-0.52	18.17	0.72	-0.33	7.46	0.93	-0.04

Suite à la page suivante

Tableau A.1: Résultats de la modélisation PAR (suite et fin)

Semaine	Baskatong			Cabonga			Paugan		
	ϕ_0	ϕ_1	ϕ_2	ϕ_0	ϕ_1	ϕ_2	ϕ_0	ϕ_1	ϕ_2
30	32.33	0.89	-0.21	12.94	0.39	**-0.03**	6.81	1.08	-0.22
31	42.11	0.53	0.16	18.96	0.31	**0.03**	4.80	1.11	-0.24
32	22.76	1.03	-0.09	15.91	0.62	0.10	8.81	0.78	-0.05
33	39.25	0.93	-0.06	23.68	0.53	-0.27	5.86	0.51	0.35
34	67.89	0.56	0.05	16.14	0.53	**-0.02**	6.00	0.87	**0.00**
35	19.83	1.26	-0.27	14.33	0.41	0.16	10.45	0.79	-0.05
36	31.50	0.87	**0.04**	17.69	0.47	**0.03**	-2.05	0.96	0.16
37	48.48	0.88	**0.04**	14.66	0.59	0.16	5.69	0.87	0.14
38	66.40	0.66	0.09	16.91	0.63	-0.17	6.17	1.03	-0.07
39	80.84	0.55	0.19	14.07	0.43	0.29	3.96	1.20	-0.10
40	75.30	0.74	-0.06	15.85	0.44	0.17	13.03	1.06	-0.17
41	78.73	0.75	-0.08	19.63	0.50	**-0.03**	9.39	0.83	0.19
42	23.80	0.74	0.10	19.35	0.63	**-0.02**	5.57	1.06	-0.11
43	107.59	0.96	-0.41	18.38	0.76	-0.14	8.79	1.03	-0.18
44	53.83	0.59	0.05	19.27	0.36	0.11	22.99	0.55	0.10
45	29.60	0.72	-0.04	14.46	0.58	-0.05	7.72	1.30	-0.37
46	37.24	0.70	-0.04	12.93	0.69	**-0.03**	-1.89	0.91	0.10
47	11.57	0.95	-0.08	23.09	0.37	**-0.00**	19.76	0.57	0.08
48	49.74	0.51	-0.06	18.87	0.28	0.07	5.20	0.95	-0.07
49	27.93	0.69	-0.07	8.69	0.69	**0.00**	7.17	0.61	0.19
50	41.63	0.78	-0.22	15.55	0.77	-0.31	5.03	1.15	-0.24
51	18.43	0.90	-0.19	17.37	0.26	0.11	7.37	0.77	0.02
52	10.70	0.69	0.05	17.28	0.40	-0.04	5.90	0.69	0.11

A.2 Analyse en composantes principales

Tableau A.2 Pourcentage de variance expliquée par ξ_t^i

Semaine t	% de variance expliquée par		Semaine	% de variance expliquée par	
	ξ_t^1	ξ_t^1 et ξ_t^2		ξ_t^1	ξ_t^1 et ξ_t^2
1	70	86	27	95	99
2	76	92	28	95	98
3	88	97	29	94	98
4	93	97	30	90	96
5	98	99	31	92	97
6	97	100	32	91	98
7	96	100	33	90	97
8	96	99	34	95	98
9	94	99	35	94	97
10	94	99	36	91	97
11	96	99	37	95	98
12	92	99	38	93	98
13	90	98	39	93	98
14	92	99	40	86	97
15	96	99	41	91	96
16	91	99	42	89	95
17	90	98	43	88	97
18	91	96	44	83	95
19	95	98	45	87	97
20	94	98	46	88	96
21	92	97	47	84	97
22	81	95	48	68	85
23	94	98	49	80	92
24	90	97	50	88	95
25	96	98	51	74	89
26	91	97	52	53	90

A.3 Méthode des moments

Tableau A.3 Erreurs sur les moments des résidus générés de Baskatong

Semaine	k	M_k	m_k	$\lvert M_k - m_k \rvert$	$\frac{M_k - m_k}{M_k}$
10	1	0	-0.001	0.001	—
	2	19183	19178	5	0.0003
	3	2862939	2866298	3358	0.0012
	4	1559747599	1574446580	14698981	0.0094
11	1	0	0	0	—
	2	59900	59903	4	0.0001
	3	15555398	15136258	419140	0.0269
	4	15050264644	14701639499	348625145	0.0232
12	1	0	0.0001	0.0001	—
	2	53320	56073	2753	0.0516
	3	5168591	5184746	16155	0.0031
	4	7277489480	7002037767	275451713	0.0378
13	1	0	-0.0003	0.0003	—
	2	55632	57173	1541	0.0277
	3	19692170	19723577	31407	0.0016
	4	20712814543	20150368913	562445630	0.0272
14	1	0	-0.0001	0.0001	—
	2	22127	22131	4	0.0002
	3	2357485	2330873	26612	0.0113
	4	1838101301	1840816815	2715514	0.0015
15	1	0	0.0009	0.0009	—
	2	10123	10126	4	0.0004
	3	860856	856868	3987	0.0046
	4	318308522	313774104	4534418	0.0142
16	1	0	-0.0001	0.0001	—
	2	9877	9880	3	0.0003
	3	2106651	2067001	39650	0.0188
	4	989409019	961571701	27837318	0.0281
17	1	0	0.0001	0.0001	—
	2	11207	11203	4	0.0004
	3	2084175	2171785	87609	0.042
	4	1168700609	1208968186	40267577	0.0345

Tableau A.4 Erreurs sur les moments des résidus générés de Cabonga

Semaine	k	M_k	m_k	$\lvert M_k - m_k \rvert$	$\left\lvert \frac{M_k - m_k}{M_k} \right\rvert$
10	1	0	0.0005	0.0005	—
	2	472	126	346	0.7333
	3	3074	931	2143	0.6971
	4	467904	54050	413854	0.8845
11	1	0	-0.0001	0.0001	—
	2	1223	401	822	0.6723
	3	26256	4377	21879	0.8333
	4	5836964	484990	5351974	0.9169
12	1	0	0.0001	0.0001	—
	2	1764	946	818	0.4637
	3	52306	14459	37847	0.7236
	4	10007212	2758440	7248772	0.7244
13	1	0	-0.0001	0.0001	—
	2	1080	355	725	0.671
	3	38599	6183	32416	0.8398
	4	5831539	536565	5294975	0.908
14	1	0	0	0	—
	2	719	360	359	0.4988
	3	9072	4206	4865	0.5363
	4	1929937	473561	1456376	0.7546
15	1	0	0.0001	0.0001	—
	2	628	62	566	0.9013
	3	28192	401	27791	0.9858
	4	3534051	11751	3522300	0.9967
16	1	0	0	0	—
	2	444	88	356	0.8015
	3	9908	1682	8226	0.8303
	4	771992	73934	698058	0.9042
17	1	0	0.0001	0.0001	—
	2	394	169	225	0.5701
	3	9064	3564	5500	0.6068
	4	810921	252633	558288	0.6885

Tableau A.5 Erreurs sur les moments des résidus générés de Paugan

Semaine	k	M_k	m_k	$\lvert M_k - m_k \rvert$	$\frac{M_k - m_k}{M_k}$
10	1	0	0.0031	0.0031	—
	2	2350	2352	2	0.0007
	3	121395	74250	47145	0.3884
	4	20639228	16583909	4055320	0.1965
11	1	0	-0.0004	0.0004	—
	2	6128	6084	45	0.0073
	3	366216	286405	79811	0.2179
	4	112755060	108329223	4425837	0.0393
12	1	0	0.0003	0.0003	—
	2	9976	9950	26	0.0027
	3	1057912	1157952	100040	0.0946
	4	416687707	496607479	79919771	0.1918
13	1	0	-0.0002	0.0002	—
	2	3371	3379	8	0.0022
	3	202363	218194	15831	0.0782
	4	50238608	51672479	1433871	0.0285
14	1	0	-0.0001	0.0001	—
	2	2200	2195	5	0.0025
	3	-24305	-16838	7467	0.3072
	4	21544143	21421564	122579	0.0057
15	1	0	-0.0012	0.0012	—
	2	2166	2167	1	0.0002
	3	257103	243438	13665	0.0531
	4	61929405	58221031	3708374	0.0599
16	1	0	0	0	—
	2	802	800	1	0.0014
	3	24522	21432	3090	0.126
	4	3039368	2568964	470403	0.1548
17	1	0	-0.0005	0.0005	—
	2	1928	1924	4	0.002
	3	163474	141130	22344	0.1367
	4	33715741	33439138	276603	0.0082

Tableau A.6 Comparaison des matrices de corrélation $CORR_t$ et $CORR_t^*$

Semaine	$CORR_t$	$CORR_t^*$	RMSE
10	$\begin{pmatrix} 1 & 0.3708 & 0.2175 \\ 0.3708 & 1 & 0.3721 \\ 0.2175 & 0.3721 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.7357 & 0.2197 \\ 0.7357 & 1 & 0.8283 \\ 0.2197 & 0.8283 & 1 \end{pmatrix}$	0.2754
11	$\begin{pmatrix} 1 & 0.2685 & 0.2013 \\ 0.2685 & 1 & 0.4774 \\ 0.2013 & 0.4774 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.4712 & 0.2011 \\ 0.4712 & 1 & 0.9595 \\ 0.2011 & 0.9593 & 1 \end{pmatrix}$	0.2464
12	$\begin{pmatrix} 1 & 0.5231 & 0.1444 \\ 0.5231 & 1 & 0.5175 \\ 0.1444 & 0.5175 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.7109 & 0.1411 \\ 0.7109 & 1 & 0.7671 \\ 0.1411 & 0.7671 & 1 \end{pmatrix}$	0.1473
13	$\begin{pmatrix} 1 & 0.4069 & 0.0861 \\ 0.4069 & 1 & 0.3366 \\ 0.0861 & 0.3366 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.7235 & 0.0855 \\ 0.7235 & 1 & 0.7417 \\ 0.0855 & 0.7417 & 1 \end{pmatrix}$	0.2424
14	$\begin{pmatrix} 1 & 0.6631 & 0.08 \\ 0.6631 & 1 & 0.2338 \\ 0.08 & 0.2338 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.9560 & 0.0795 \\ 0.9560 & 1 & 0.3810 \\ 0.0795 & 0.3810 & 1 \end{pmatrix}$	0.1545
15	$\begin{pmatrix} 1 & 0.2951 & 0.0732 \\ 0.2951 & 1 & 0.0011 \\ 0.0732 & 0.0011 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.9939 & 0.0741 \\ 0.9939 & 1 & -0.0206 \\ 0.0741 & -0.0206 & 1 \end{pmatrix}$	0.3296
16	$\begin{pmatrix} 1 & 0.4236 & 0.1728 \\ 0.4236 & 1 & 0.1157 \\ 0.1728 & 0.1157 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.9865 & 0.1728 \\ 0.9865 & 1 & 0.3364 \\ 0.1728 & 0.3364 & 1 \end{pmatrix}$	0.2850
17	$\begin{pmatrix} 1 & 0.6258 & 0.2399 \\ 0.6258 & 1 & 0.0302 \\ 0.2399 & 0.0302 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.9331 & 0.2345 \\ 0.9331 & 1 & 0.0184 \\ 0.2345 & 0.0184 & 1 \end{pmatrix}$	0.1450

A.4 Méthode séquentielle

Tableau A.7 Erreurs sur les moments des résidus générés de Baskatong

Semaine	k	M_k	m_k	$\lvert M_k - m_k \rvert$	$\frac{M_k - m_k}{M_k}$
10	1	0	0.3484	0.3484	—
	2	19183	19727	544	0.0284
	3	2862939	3045622	182683	0.0638
	4	1559747599	1652890564	93142965	0.0597
11	1	0	-0.2279	0.2279	—
	2	59900	58754	1146	0.0191
	3	15555398	14082092	1473306	0.0947
	4	15050264644	14088027049	962237595	0.0639
12	1	0	0.1529	0.1529	—
	2	53320	52932	387	0.0073
	3	5168591	5075377	93214	0.0180
	4	7277489480	7181637489	95851991	0.0132
13	1	0	-0.0035	0.0035	—
	2	55632	56174	542	0.0097
	3	19692170	20167181	475011	0.0241
	4	20712814543	21024358589	311544046	0.0150
14	1	0	-0.047	0.047	—
	2	22127	22019	107	0.0049
	3	2357485	2285086	72399	0.0307
	4	1838101301	1821335731	16765569	0.0091
15	1	0	-1.1717	1.1717	—
	2	10123	10603	480	0.0475
	3	860856	958070	97214	0.1129
	4	318308522	333845091	15536569	0.0488
16	1	0	0.4418	0.4418	—
	2	9877	10058	181	0.0183
	3	2106651	2125482	18831	0.0089
	4	989409019	980388518	9020500	0.0091
17	1	0	-1.0615	1.0615	—
	2	11207	10653	553	0.0494
	3	2084175	1888468	195707	0.0939
	4	1168700609	1074481269	94219340	0.0806

Tableau A.8 Erreurs sur les moments des résidus générés de Cabonga

Semaine	k	M_k	m_k	$\lvert M_k - m_k \rvert$	$\left\lvert \frac{M_k - m_k}{M_k} \right\rvert$
10	1	0	0.0291	0.0291	—
	2	472	115	358	0.7574
	3	3074	920	2154	0.7007
	4	467904	50862	417042	0.8913
11	1	0	0.0895	0.0895	—
	2	1223	369	855	0.6986
	3	26256	2754	23502	0.8951
	4	5836964	336507	5500457	0.9423
12	1	0	-0.0739	0.0739	—
	2	1764	945	819	0.4642
	3	52306	9710	42596	0.8144
	4	10007212	2513138	7494074	0.7489
13	1	0	-0.0065	0.0065	—
	2	1080	359	721	0.6676
	3	38599	5775	32824	0.8504
	4	5831539	516321	5315218	0.9115
14	1	0	-0.0046	0.0046	—
	2	719	356	363	0.5047
	3	9072	3582	5490	0.6052
	4	1929937	447039	1482898	0.7684
15	1	0	-0.0762	0.0762	—
	2	628	65	562	0.8962
	3	28192	452	27740	0.9840
	4	3534051	12631	3521420	0.9964
16	1	0	0.0415	0.0415	—
	2	444	92	352	0.7934
	3	9908	1864	8044	0.8119
	4	771992	85302	686690	0.8895
17	1	0	-0.1058	0.1058	—
	2	394	159	235	0.5969
	3	9064	3286	5778	0.6374
	4	810921	229877	581043	0.7165

Tableau A.9 Erreurs sur les moments des résidus générés de Paugan

Semaine	k	M_k	m_k	$\lvert M_k - m_k \rvert$	$\lvert\frac{M_k - m_k}{M_k}\rvert$
10	1	0	0.0791	0.0791	—
	2	2350	2215.8092	134	0.0572
	3	121395	71334.8945	50060	0.4124
	4	20639228	16005480.61	4633748	0.2245
11	1	0	0.4112	0.4112	—
	2	6128	5745.3315	383	0.0625
	3	366216	193267.8543	172948	0.4723
	4	112755060	88207000.75	24548059	0.2177
12	1	0	-0.4159	0.4159	—
	2	9976	9811.312	165	0.0165
	3	1057912	1047632.834	10279	0.0097
	4	416687707	453038517.5	36350810	0.0872
13	1	0	-0.0284	0.0284	—
	2	3371	3350.0686	21	0.0062
	3	202363	209743.2706	7380	0.0365
	4	50238608	53567548.41	3328940	0.0663
14	1	0	0.0079	0.0079	—
	2	2200	2183.5768	17	0.0076
	3	-24305	-17119.3701	7185	0.2956
	4	21544143	21583942.17	39799	0.0018
15	1	0	-0.9836	0.9836	—
	2	2166	1950.1671	216	0.0998
	3	257103	199299.5431	57803	0.2248
	4	61929405	46219870.27	15709535	0.2537
16	1	0	0.0286	0.0286	—
	2	802	795.7585	6	0.0072
	3	24522	22208.2026	2314	0.0943
	4	3039368	3114789.856	75422	0.0248
17	1	0	-0.4239	0.4239	—
	2	1928	1849.6789	78	0.0407
	3	163474	134286.5422	29188	0.1785
	4	33715741	32427886.26	1287855	0.0382

Tableau A.10 Comparaison des matrices de corrélation $CORR_t$ et $CORR_t^*$

Semaine	$CORR_t$	$CORR_t^*$	RMSE
10	$\begin{pmatrix} 1 & 0.3708 & 0.2175 \\ 0.3708 & 1 & 0.3721 \\ 0.2175 & 0.3721 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.7059 & 0.13070 \\ 0.7059 & 1 & 0.79450 \\ 0.1307 & 0.7945 & 1.00000 \end{pmatrix}$	0.2574
11	$\begin{pmatrix} 1 & 0.2685 & 0.2013 \\ 0.2685 & 1 & 0.4774 \\ 0.2013 & 0.4774 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.4312 & 0.1428 \\ 0.4312 & 1 & 0.9546 \\ 0.1428 & 0.9546 & 1 \end{pmatrix}$	0.2393
12	$\begin{pmatrix} 1 & 0.5231 & 0.1444 \\ 0.5231 & 1 & 0.5175 \\ 0.1444 & 0.5175 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.7523 & 0.1896 \\ 0.7523 & 1 & 0.7895 \\ 0.1896 & 0.7895 & 1 \end{pmatrix}$	0.1690
13	$\begin{pmatrix} 1 & 0.4069 & 0.0861 \\ 0.4069 & 1 & 0.3366 \\ 0.0861 & 0.3366 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.7357 & 0.1123 \\ 0.7357 & 1 & 0.7556 \\ 0.1123 & 0.7556 & 1 \end{pmatrix}$	0.2514
14	$\begin{pmatrix} 1 & 0.6631 & 0.08 \\ 0.6631 & 1 & 0.2338 \\ 0.08 & 0.2338 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.95150 & 0.0672 \\ 0.9515 & 1.00000 & 0.3709 \\ 0.0672 & 0.37090 & 1 \end{pmatrix}$	0.1507
15	$\begin{pmatrix} 1 & 0.2951 & 0.0732 \\ 0.2951 & 1 & 0.0011 \\ 0.0732 & 0.0011 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.9962 & 0.0472 \\ 0.9962 & 1 & -0.0403 \\ 0.0472 & -0.0403 & 1 \end{pmatrix}$	0.3313
16	$\begin{pmatrix} 1 & 0.4236 & 0.1728 \\ 0.4236 & 1 & 0.1157 \\ 0.1728 & 0.1157 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.9868 & 0.2467 \\ 0.9868 & 1 & 0.4006 \\ 0.2467 & 0.4006 & 1 \end{pmatrix}$	0.2996
17	$\begin{pmatrix} 1 & 0.6258 & 0.2399 \\ 0.6258 & 1 & 0.0302 \\ 0.2399 & 0.0302 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.975 & 0.2702 \\ 0.975 & 1 & 0.0493 \\ 0.2702 & 0.0493 & 1 \end{pmatrix}$	0.1654

A.5 Høyland (algorithme noyau)

Tableau A.11 Erreurs sur les moments des résidus générés de Baskatong

Semaine	k	M_k	m_k	$\lvert M_k - m_k \rvert$	$\lvert \frac{M_k - m_k}{M_k} \rvert$
10	1	0	0.1483	0.1483	—
	2	19183	19462	280	0.0146
	3	2862939	2986430	123491	0.0431
	4	1559747599	1606914038	47166439	0.0302
11	1	0	0.0358	0.0358	—
	2	59900	59214	686	0.0114
	3	15555398	15010471	544927	0.0350
	4	15050264644	14711199957	339064687	0.0225
12	1	0	-0.1421	0.1421	—
	2	53320	53496	176	0.0033
	3	5168591	5368575	199984	0.0387
	4	7277489480	7333101880	55612400	0.0076
13	1	0	0.7059	0.7059	—
	2	55632	55218	414	0.0074
	3	19692170	19688026	4144	0.0002
	4	20712814543	20912433737	199619194	0.0096
14	1	0	-0.0397	0.0397	—
	2	22127	22823	696	0.0315
	3	2357485	2763018	405533	0.1720
	4	1838101301	1958943217	120841916	0.0657
15	1	0	-0.776	0.776	—
	2	10123	10216	94	0.0093
	3	860856	857768	3088	0.0036
	4	318308522	310909926	7398597	0.0232
16	1	0	1.0611	1.0611	—
	2	9877	9463	415	0.0420
	3	2106651	2132380	25729	0.0122
	4	989409019	1041825429	52416411	0.0530
17	1	0	-2.0598	2.0598	—
	2	11207	10479	728	0.0649
	3	2084175	1704692	379483	0.1821
	4	1168700609	1016187525	152513084	0.1305

Tableau A.12 Erreurs sur les moments des résidus générés de Cabonga

Semaine	k	M_k	m_k	$\lvert M_k - m_k \rvert$	$\left\lvert \frac{M_k - m_k}{M_k} \right\rvert$
10	1	0	0.7036	0.7036	—
	2	472	459	13	0.0269
	3	3074	2934	139	0.0452
	4	467904	491150	23247	0.0497
11	1	0	-0.0407	0.0407	—
	2	1223	1205	18	0.0148
	3	26256	26268	12	0.0005
	4	5836964	6083319	246355	0.0422
12	1	0	-0.5732	0.5732	—
	2	1764	1788	25	0.0139
	3	52306	56996	4691	0.0897
	4	10007212	10695295	688084	0.0688
13	1	0	-0.3051	0.3051	—
	2	1080	1144	64	0.0592
	3	38599	39612	1014	0.0263
	4	5831539	5455310	376229	0.0645
14	1	0	0.5602	0.5602	—
	2	719	725	6	0.0087
	3	9072	8498	573	0.0631
	4	1929937	1990969	61033	0.0316
15	1	0	0.3631	0.3631	—
	2	628	673	46	0.0726
	3	28192	33993	5801	0.2058
	4	3534051	4342261	808211	0.2287
16	1	0	0.0976	0.0976	—
	2	444	454	11	0.0257
	3	9908	10564	657	0.0663
	4	771992	819704	47713	0.0618
17	1	0	-0.0659	0.0659	—
	2	394	392	1	0.0031
	3	9064	9090	26	0.0029
	4	810921	799864	11056	0.0136

Tableau A.13 Erreurs sur les moments des résidus générés de la centrale Paugan

Semaine	k	M_k	m_k	$\lvert M_k - m_k \rvert$	$\left\lvert \frac{M_k - m_k}{M_k} \right\rvert$
10	1	0	2.7736	2.7736	—
	2	2350	2397	47	0.0200
	3	121395	111288	10107	0.0833
	4	20639228	20680851	41623	0.0020
11	1	0	3.1761	3.1761	—
	2	6128	5881	248	0.0404
	3	366216	287734	78482	0.2143
	4	112755060	100651522	12103537	0.1073
12	1	0	0.5049	0.5049	—
	2	9976	9956	21	0.0021
	3	1057912	1156226	98315	0.0929
	4	416687707	473185501	56497794	0.1356
13	1	0	-0.2404	0.2404	—
	2	3371	3540	169	0.0502
	3	202363	224019	21656	0.1070
	4	50238608	55151068	4912460	0.0978
14	1	0	0.2944	0.2944	—
	2	2200	2260	60	0.0272
	3	-24305	-31373	7068	0.2908
	4	21544143	22354838	810695	0.0376
15	1	0	0.3814	0.3814	—
	2	2166	2187	21	0.0097
	3	257103	274429	17326	0.0674
	4	61929405	68727327	6797922	0.1098
16	1	0	0.0289	0.0289	—
	2	802	806	4	0.0051
	3	24522	24320	201	0.0082
	4	3039368	3016171	23197	0.0076
17	1	0	-0.2836	0.2836	—
	2	1928	1819	110	0.0568
	3	163474	148291	15183	0.0929
	4	33715741	30763101	2952640	0.0876

Tableau A.14 Comparaison des matrices de corrélation $CORR_t$ et $CORR_t^*$

Semaine	$CORR_t$	$CORR_t^*$	RMSE
10	$\begin{pmatrix} 1 & 0.3708 & 0.2175 \\ 0.3708 & 1 & 0.3721 \\ 0.2175 & 0.3721 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.3979 & 0.2594 \\ 0.3979 & 1 & 0.3696 \\ 0.2594 & 0.3696 & 1 \end{pmatrix}$	0.0235
11	$\begin{pmatrix} 1 & 0.2685 & 0.2013 \\ 0.2685 & 1 & 0.4774 \\ 0.2013 & 0.4774 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.2594 & 0.1782 \\ 0.2594 & 1 & 0.4997 \\ 0.1782 & 0.4997 & 1 \end{pmatrix}$	0.0157
12	$\begin{pmatrix} 1 & 0.5231 & 0.1444 \\ 0.5231 & 1 & 0.5175 \\ 0.1444 & 0.5175 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.4877 & 0.0819 \\ 0.4877 & 1 & 0.5492 \\ 0.0819 & 0.5492 & 1 \end{pmatrix}$	0.0370
13	$\begin{pmatrix} 1 & 0.4069 & 0.0861 \\ 0.4069 & 1 & 0.3366 \\ 0.0861 & 0.3366 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.4423 & 0.0862 \\ 0.4423 & 1 & 0.3394 \\ 0.0862 & 0.3394 & 1 \end{pmatrix}$	0.0167
14	$\begin{pmatrix} 1 & 0.6631 & 0.08 \\ 0.6631 & 1 & 0.2338 \\ 0.08 & 0.2338 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.6739 & 0.1022 \\ 0.6739 & 1 & 0.2795 \\ 0.1022 & 0.2795 & 1 \end{pmatrix}$	0.0245
15	$\begin{pmatrix} 1 & 0.2951 & 0.0732 \\ 0.2951 & 1 & 0.0011 \\ 0.0732 & 0.0011 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.2908 & 0.0864 \\ 0.2908 & 1 & -0.0068 \\ 0.0864 & -0.0068 & 1 \end{pmatrix}$	0.0075
16	$\begin{pmatrix} 1 & 0.4236 & 0.1728 \\ 0.4236 & 1 & 0.1157 \\ 0.1728 & 0.1157 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.4514 & 0.2044 \\ 0.4514 & 1 & 0.1444 \\ 0.2044 & 0.1444 & 1 \end{pmatrix}$	0.0240
17	$\begin{pmatrix} 1 & 0.6258 & 0.2399 \\ 0.6258 & 1 & 0.0302 \\ 0.2399 & 0.0302 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.5779 & 0.1828 \\ 0.5779 & 1 & -0.0462 \\ 0.1828 & -0.0462 & 1 \end{pmatrix}$	0.0503

ANNEXE B : THÉORIE

B.1 Heuristique de Høyland et al.

Les modifications apportées à l'algorithme noyau se situent aux étapes 4 et 5. On introduit un processus itératif qui permet d'approcher les résultats souhaités. Même si les résultats restent approximatifs, on peut néanmoins contrôler la qualité en imposant une erreur maximale.

A l'étape 4 de l'algorithme noyau(3.1), on doit générer des variables indépendantes qui ont des moments cibles spécifiés. Or, il existe toujours une légère corrélation entre ces variables. L'étape 4 modifiée est un processus itératif qui, à l'aide de la transformation cubique et la transformation matricielle, permet d'approcher l'indépendance à ϵ_x près.

Algorithme B.1 Étape 4 (de l'algorithme noyau) modifiée

1: Générer X_t^{s*} avec les moments cibles $E[(X_t^s)^k]$ indépendamment pour $s = 1, \ldots, S$.
   $\rightarrow$ La matrice de corrélation $R \approx I$.
2: $X = X_t^{s*}$
3: Calculer R =Matrice de corrélation de X_t^{s*}
4: **tant que** $dist(R, I) > \epsilon_x$;
5: Calculer L par $R = LL^T$
6: $X \leftarrow L^{-1}X$.
   $\rightarrow R = I$ mais $E[X^k] \neq E[(X_t^s)^k]$.
7: $X \leftarrow$ Transformation cubique de X afin d'obtenir les moments cibles $E[(X_t^s)^k]$.
8: Calculer R = Matrice de corrélation de X
9: $X_t^{s*} = X$

À l'étape 5 de l'algorithme noyau(3.1), les moments sont déformés quand on applique la transformation matricielle $\mathbf{Y}_t = L\mathbf{X}_t$. Il faut alors procéder de manière itérative afin d'approcher les moments cibles $E[(Y_t^s)^k]$ et la matrice de corrélation cible $CORR_t$.

Algorithme B.2 Étape 5 (de l'algorithme noyau) modifiée

1: $Y = LX_t^{s*}$ et R la matrice de corrélation de Y.
2: **tant que** $dist(R, CORR_t) > \epsilon_y$;
3: Calculer L_{temp} : $R = L_{temp}L_{temp}^T$
4: $Y \leftarrow = L * L_{temp}^{-1}Y$
$\rightarrow E(Y^k) \neq E[(Y_t^s)^k]$.
5: $Y \leftarrow$ transformation cubique de Y afin d'obtenir les moments cibles $E[(Y_t^s)^k]$.
6: Calculer R = matrice de corrélation de Y.
7: $Y_t^{s*} = Y$

ANNEXE C : LES ARBRES

C.1 Méthode des moments

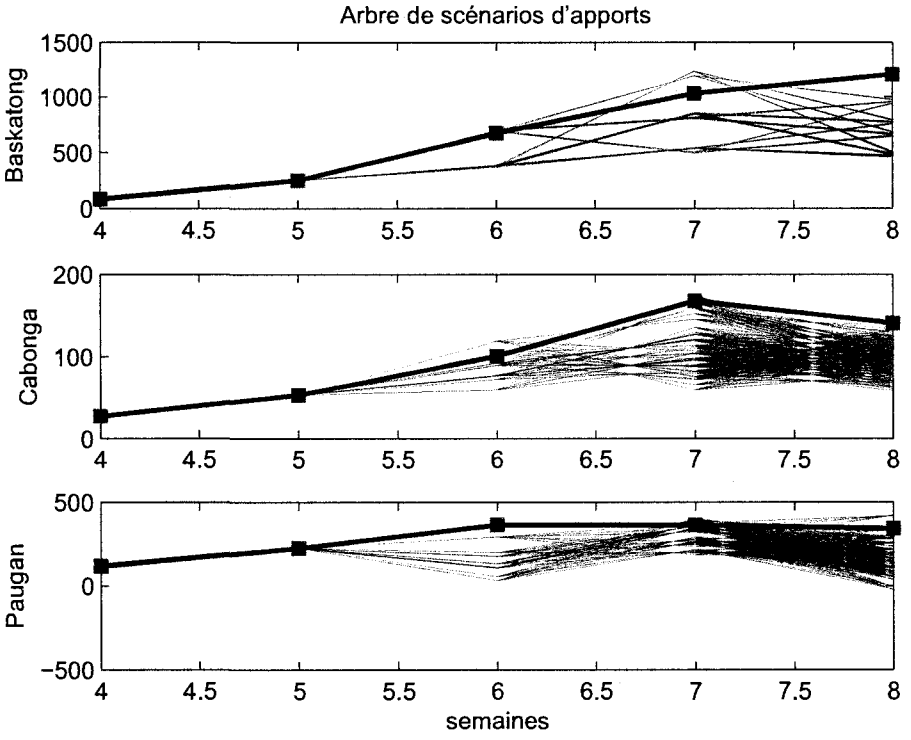


Figure C.1 Les prévisions d'apports pour les semaines 6, 7 et 8

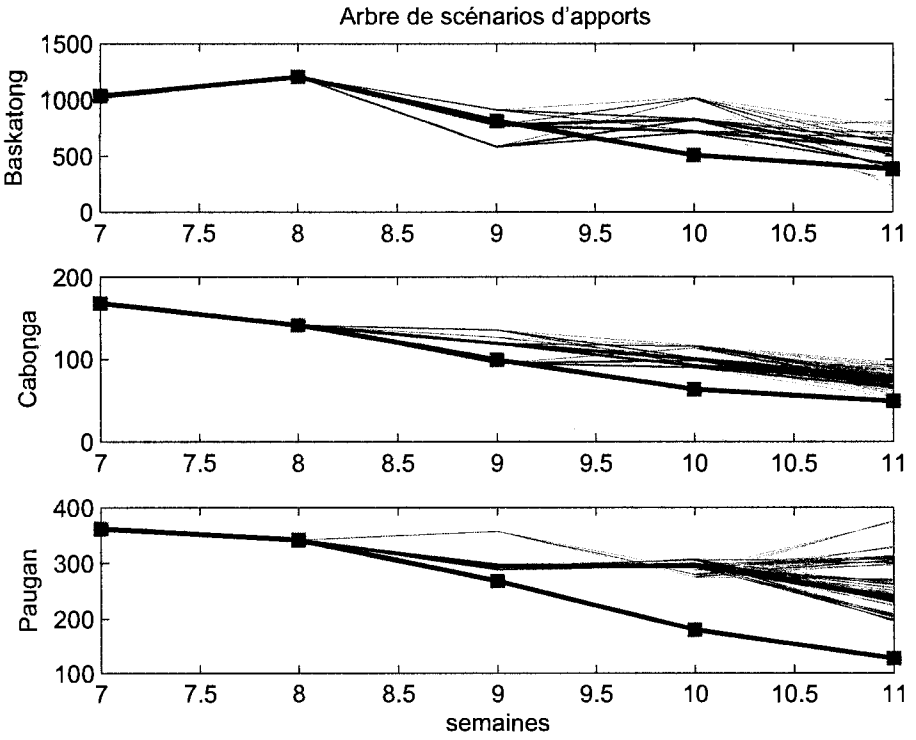


Figure C.2 Les prévisions d'apports pour les semaines 9, 10 et 11

C.2 Simulation séquentielle

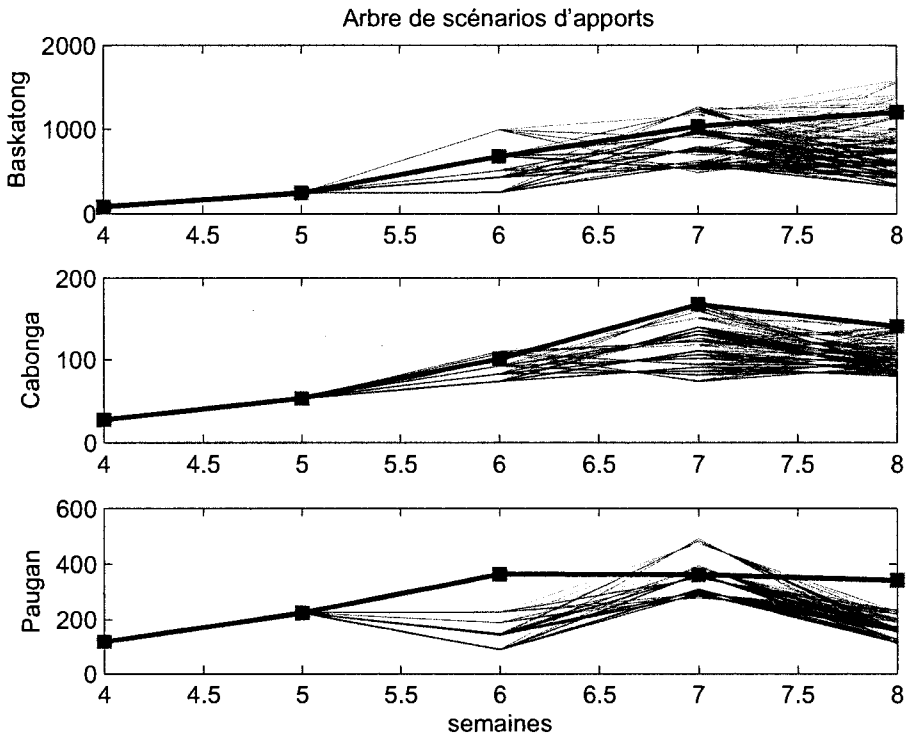


Figure C.3 Les prévisions d'apports pour les semaines 6, 7 et 8

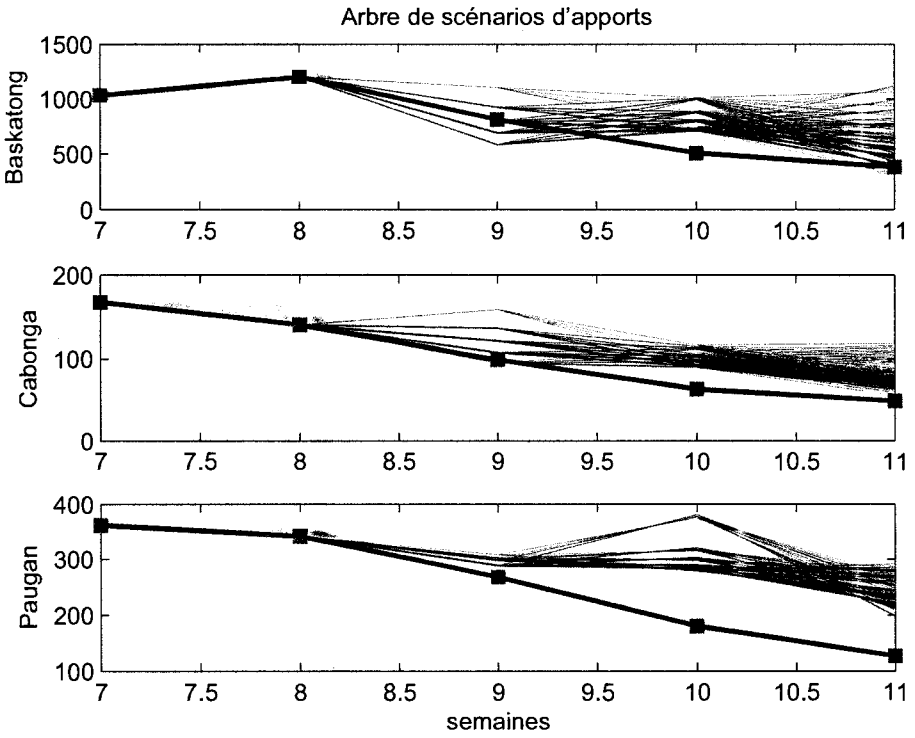


Figure C.4 Les prévisions d'apports pour les semaines 9, 10 et 11

C.3 Simulation parallèle

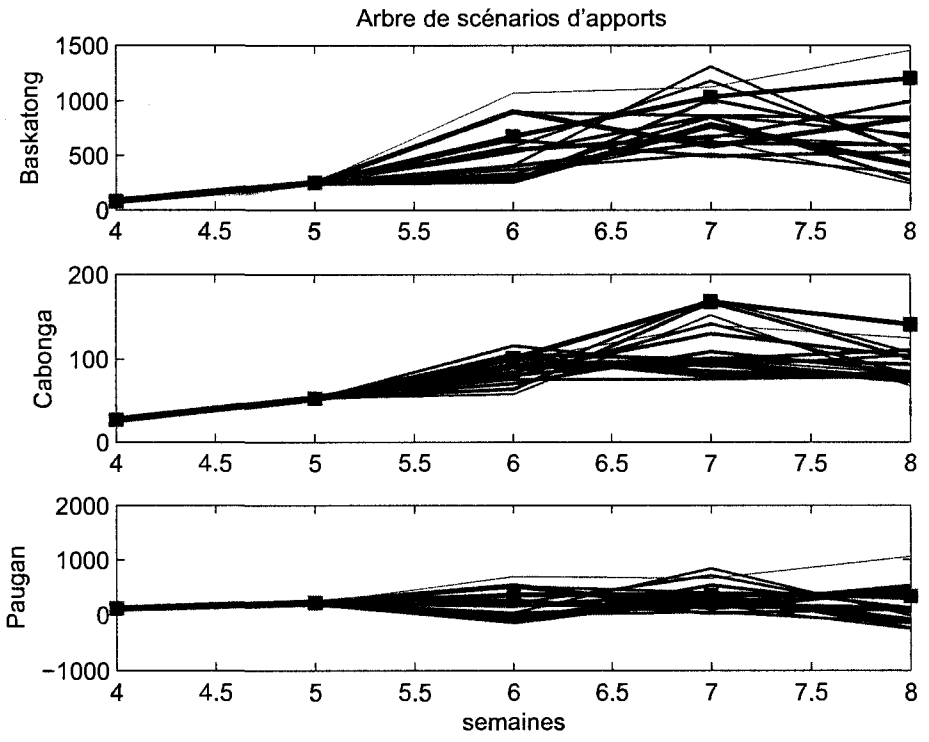


Figure C.5 Les prévisions d'apports pour les semaines 6, 7 et 8

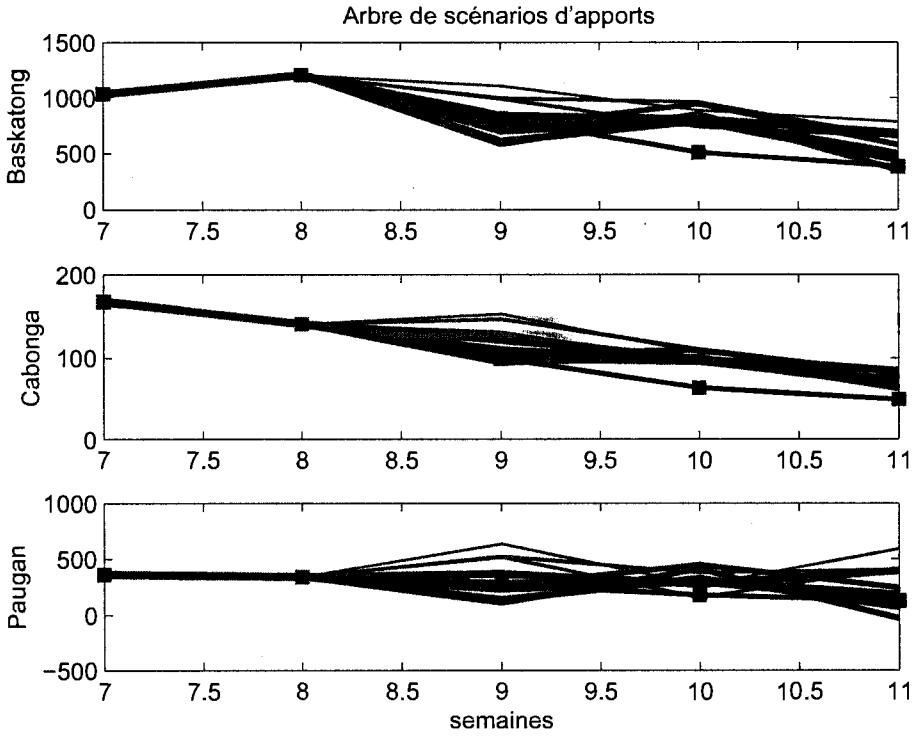


Figure C.6 Les prévisions d'apports pour les semaines 9, 10 et 11

C.4 Méthode de Høyland

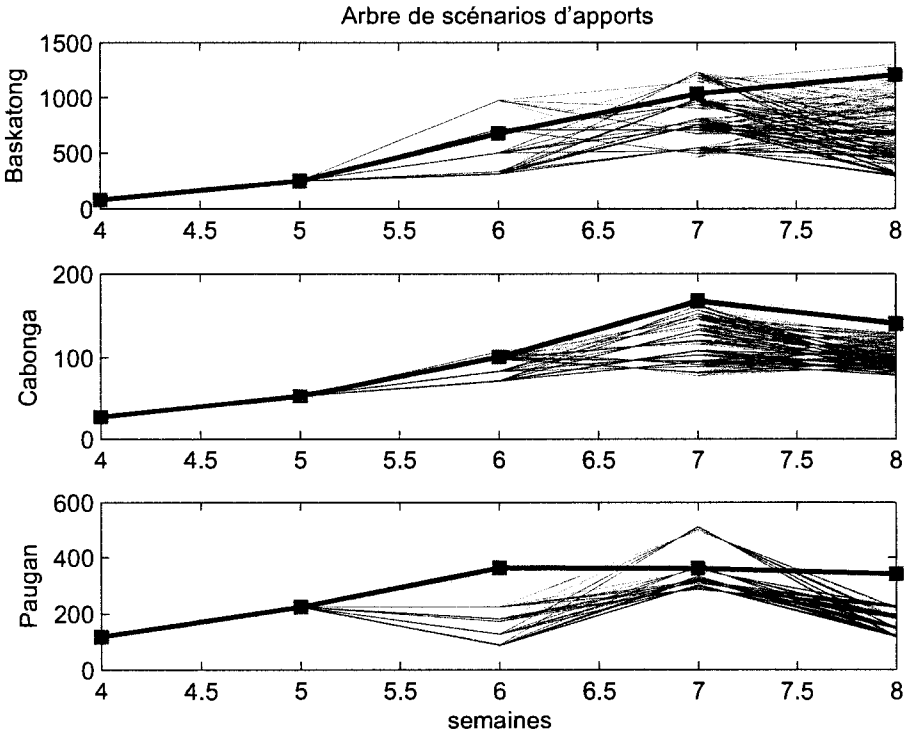


Figure C.7 Les prévisions d'apports pour les semaines 6, 7 et 8

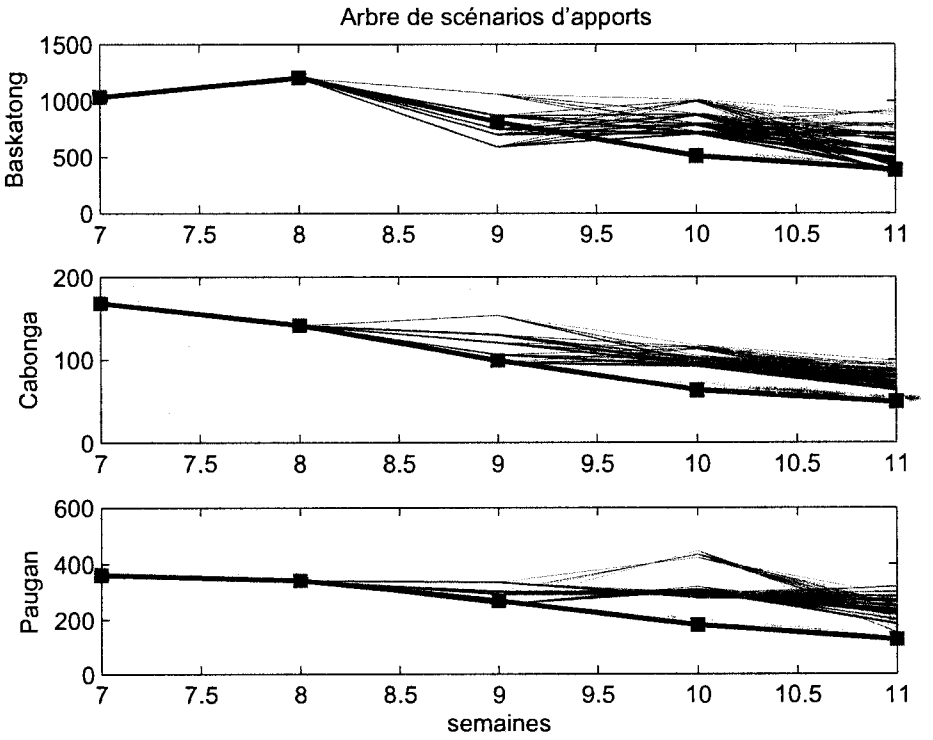


Figure C.8 Les prévisions d'apports pour les semaines 9, 10 et 11