

**ANALYSE STATISTIQUE DES SÉRIES
D'APPORTS ÉNERGÉTIQUES
RAPPORT FINAL DU PROJET C2**
Chaire CRSNG - Hydro-Québec
en Hydrologie statistique

PROJET C2

ANALYSE STATISTIQUE DES SÉRIES D'APPORTS ÉNERGÉTIQUES

par

**Luc Perreault
Mario Haché
Bernard Bobée**

Chaire en Hydrologie statistique Hydro-Québec / CRSNG
Institut National de la Recherche Scientifique, INRS-Eau
2800, rue Einstein, CP 7500, Sainte-Foy, Québec, G1V 4C7

Rapport de recherche No R-474

15 juillet 1996

ÉQUIPE DE RECHERCHE

CHARGÉ DE PROJET	INRS-Eau	LUC PERREAULT
CHARGÉ DE PROJET	Hydro-Québec	JEAN-LOUIS BISSON
COLLABORATEURS	INRS-Eau	MARIO HACHÉ
		MICHEL SLIVITZKY (Professeur émérite)
		PETER RASMUSSEN
		JACQUES BERNIER (Professeur invité)
	Hydro-Québec	PIERRE BRUNEAU
		DENIS TREMBLAY
		NARUT KANG
		LAURA FAGHERAZZI
	IREQ	ANDRÉ TURGEON
TITULAIRE DE LA CHAIRE	INRS-Eau	BERNARD BOBÉE

TABLE DES MATIÈRES

LISTE DES TABLEAUX	ix
LISTE DES FIGURES	xiii
SOMMAIRE	1
CHAPITRE 1 : Introduction.....	15
1.1 Situation actuelle.....	15
1.2 Problématique.....	16
1.3 Objectifs généraux	21
1.4 Description des objectifs spécifiques du projet.....	22
1.4.1. Planification du projet.....	22
1.4.2. Étude critique des données de base nécessaires au calcul des apports énergétiques	22
1.4.3. Analyse statistique des séries d'apports énergétiques	22
1.4.4. Modélisation de la série d'apports énergétiques annuels agrégés	23
CHAPITRE 2 : Étude critique des données de base	25
2.1 Problématique.....	25
2.2 Synthèse des reconstitutions et extensions de séries	27
2.2.1. Homogénéité temporelle	18
2.2.2. Homogénéité et concordance régionales.....	29
2.2.3. Reconstitution et extension de séries	30
2.2.3.1 Régression multiple (RéMuS).....	31
2.2.3.2 Rapport de superficies de bassins versants (RSBV).....	32
2.2.3.3 Désagrégation spatiale et temporelle (SPIGOT).....	33
2.2.4. Description des procédures de validation et synthèse.....	34
2.2.4.1 Procédure A	35
2.2.4.2 Procédure B	36
2.2.4.3 Procédure C	40

2.2.4.4 Procédure D	42
2.2.5. Apports mensuels utilisés par le service Prévisions	42
2.3 Analyse critique des méthodologies d'harmonisation des apports	45
2.3.1. Méthode du rapport des superficies de bassins versants	47
2.3.1.1 Réservoir OU4 (complexe Aux Outardes)	48
2.3.1.2 Réservoirs MA5, LSA, MAT ET GOU	57
2.3.1.3 Recommandations et alternatives possibles	64
2.3.2. Tests statistiques utilisés pour l'harmonisation des apports	65
2.3.2.1 Homogénéité temporelle: date de changement connue	65
2.3.2.2 Homogénéité temporelle: date de changement inconnue	69
2.3.2.3 Homogénéité spatiale	70
2.3.3. Application de la classification ascendante hiérarchique	74
2.4 Conclusions et recommandations	74
2.4.1 Conclusions et recommandations générales	74
2.4.2 Remarques spécifiques	77
 CHAPITRE 3 : Analyse statistique des apports En	 81
3.1 Calcul des apports énergétiques	81
3.2 Apports énergétiques mensuels	83
3.2.1. Statistiques de base	83
3.2.2. Corrélations mensuelles	95
3.3 Apports énergétiques annuels	100
3.3.1. Statistiques de base	103
3.3.2. Corrélations	105
3.3.3. Autocorrélations	108
3.4 Étude préliminaire de la stationnarité	112
3.4.1 Définition	112
3.4.2 Analyse de la stationnarité (date de changement connue)	113
3.4.3 Analyse de la stationnarité (date de changement inconnue)	115
3.5 Conclusions	120

LISTE DES TABLEAUX

Tableau 2.1. Sites traités lors de la validation de longues séries de débits observés.....	26
Tableau 2.2. Les complexes hydroélectriques classés selon les procédures employées	34
Tableau 2.3. Rivières de référence utilisées pour reconstituer les apports intermédiaires de chaque site (procédure B)	39
Tableau 2.4. Apports naturels de 1989 à 1994 utilisés par le service Prévisions.....	46
Tableau 2.5. Apports intermédiaires corrigés à OU3 et OU4.....	50
Tableau 2.6. Principales statistiques des résidus avant et après la mise en service de OU2	52
Tableau 2.7. Principales caractéristiques des sites MA5, LSA, MAT et GOU	58
Tableau 2.8. Principales observations tirées des figures 2.12 à 2.15.....	58
Tableau 2.9. Niveau de signification réel du test de Student.....	69
Tableau 2.10: Résultats du test de normalité de Shapiro-Wilk.....	73
Tableau 3.1. Statistiques de base des apports énergétiques (GWh) du complexe Bersimis.	84
Tableau 3.2. Statistiques de base des apports énergétiques (GWh) du complexe Ch. Falls.	85
Tableau 3.3. Statistiques de base des apports énergétiques (GWh) du complexe La Grande.	86
Tableau 3.4. Statistiques de base des apports énergétiques (GWh) du complexe Manic... ..	87
Tableau 3.5. Statistiques de base des apports énergétiques (GWh) du complexe Outardes	88
Tableau 3.6. Statistiques de base des apports énergétiques (GWh) du complexe Outaouais.....	89
Tableau 3.7. Statistiques de base des apports énergétiques (GWh) du complexe St-Maurice.	90
Tableau 3.8. Statistiques de base des apports énergétiques (GWh) du complexe St-Laurent.....	91
Tableau 3.9. Statistiques de base des apports énergétiques (GWh) de la série agrégée.....	92
Tableau 3.10. Corrélations des apports énergétiques mensuels pour Bersimis.....	95
Tableau 3.11. Corrélations des apports énergétiques mensuels pour Churchill Falls.....	96
Tableau 3.12. Corrélations des apports énergétiques mensuels pour La Grande	96
Tableau 3.13. Corrélations des apports énergétiques mensuels pour Manicouagan	97
Tableau 3.14. Corrélations des apports énergétiques mensuels pour Aux Outardes	97

Tableau 3.15. Corrélations des apports énergétiques mensuels pour Outaouais	98
Tableau 3.16. Corrélations des apports énergétiques mensuels pour St-Maurice.....	98
Tableau 3.17. Corrélations des apports énergétiques mensuels pour St-Laurent	99
Tableau 3.18. Corrélations des apports énergétiques mensuels: série agrégée.....	99
Tableau 3.19. Statistiques des apports énergétiques annuels des huit complexes.....	104
Tableau 3.20. Caractéristiques statistiques des apports énergétiques annuels agrégés	105
Tableau 3.21. Test de normalité de Shapiro-Wilk (apports énergétiques annuels).....	105
Tableau 3.22. Corrélations entre les apports énergétiques annuels des huit complexes...	105
Tableau 3.23. Autocorrélations des apports énergétiques annuels (1943-1993)	109
Tableau 3.24. Dates de ruptures méthodologiques (harmonisation des apports)	114
Tableau 3.25. Comparaison des moyennes des sous-échantillons.....	114
Tableau 3.26. Comparaison des moyennes des sous-échantillons (apports agrégés).....	115
Tableau 3.27. Année la plus probable d'un changement de moyenne selon la courbe cusum.....	119
Tableau 4.1. Résultats de la procédure bayésienne pour la série agrégée (1943-1993)	130
Tableau 4.2. Résultats de la procédure bayésienne pour chaque complexe.....	131
Tableau 4.3. Résultats de la procédure bayésienne pour les séries de précipitations	134
Tableau 4.4. Modèles considérés pour la simulation.....	141
Tableau 4.5. Statistiques sur les résidus des modèles.....	143
Tableau 4.6. Principales statistiques des séries simulées pour chaque modèle	146
Tableau 4.7. Caractéristiques de séquence d'apports agrégés déficitaires et excédentaires.....	149
Tableau 4.8. Prévision (GWh), écart-type, intervalle de confiance et probabilité de surplus et de déficit pour les modèles retenus (1994-1998)	153
Tableau 4.9. Déficit prédit correspondant à diverses probabilité au non-dépassement pour les modèles retenus (1994-1998)	154
Tableau 4.10. Prévision des écarts cumulés, écart-type, intervalle de confiance et probabilité de surplus et de déficit (1994-1998).....	156
Tableau 4.11. Déficit cumulés prédit correspondant à diverses probabilité au non-dépassement pour les modèles retenus (1994-1998).....	157
Tableau 4.12. Caractéristiques statistiques des apports énergétiques annuels agrégés (1943-1995)	159
Tableau 4.13. Résultats de la procédure bayésienne pour la série agrégée (1943-1995)	161

Tableau 4.14. Prévision, écart-type, intervalle de confiance et probabilité de surplus et de déficit pour les modèles retenus (1996-2000).....	161
Tableau 4.15. Déficit prédit correspondant à diverses probabilité au non-dépassement pour les modèles retenus (1996-2000)	162
Tableau 4.16. Prévision des écarts cumulés, écart-type, intervalle de confiance et probabilité de surplus et de déficit (1996-2000).....	162
Tableau 4.17. Déficit cumulés prédit correspondant à diverses probabilité au non-dépassement pour les modèles retenus (1996-2000).....	163

LISTE DES FIGURES

Figure 1.1. Localisation géographique des sept complexes hydroélectriques.....	17
Figure 1.2. Étapes menant au calcul des intrants des modèles de planification.....	20
Figure 2.1. Contenu des séries de référence d'apports naturels mensuels totaux pour la procédure A	37
Figure 2.2. Contenu des séries des débits des rivières de références pour la procédure B	41
Figure 2.3. Contenu de la série de référence d'apports naturels annuels pour la procédure C et apports intermédiaires résultant.....	43
Figure 2.4. Contenu de la série de référence d'apports naturels pour la procédure D et apports reconstitués résultant.....	44
Figure 2.5. Écarts entre les apports reconstitués (RSBV) et calculés par bilan hydrique (BILAN) à OU4.....	49
Figure 2.6. Écarts entre les apports reconstitués et calculés à OU4.....	50
Figure 2.7. Biais annuel brut et relatif de la méthode du rapport de bassins versants à OU4.....	51
Figure 2.8. Biais annuel brut de la reconstitution des apports à OU4 à partir de la série des apports totaux à OU3	52
Figure 2.9. Biais mensuel brut et relatif de la méthode du rapport de bassins versants à OU4	54
Figure 2.10. Rapports de moyennes mensuelles d'apports au site OU4 et à l'embouchure du complexe (OU4/App. Total)	55
Figure 2.11. Périodogramme : corrélations 2 mois consécutifs des résidus de la méthode RSBV	56
Figure 2.12. Biais annuel brut et relatif de la méthode du rapport de bassins versants: (A) Réservoir MA5, (B) Réservoir LSA, (C) Réservoir MAT, (D) Réservoir Gouin.....	59
Figure 2.13. Biais mensuel brut et relatif de la méthode du rapport de bassins versants: (A) Réservoir MA5, (B) Réservoir LSA, (C) Réservoir MAT, (D) Réservoir Gouin.....	60
Figure 2.14. Rapports de moyennes mensuelles d'apports (site d'intérêt/site de référence): (A) Réservoir MA5, (B) Réservoir LSA, (C) Réservoir MAT, (D) Réservoir Gouin.....	61
Figure 2.15. Corrélations 2 mois consécutifs des résidus de la méthode RSBV: (A) Réservoir MA5, (B) Réservoir LSA, (C) Réservoir MAT, (D) Réservoir Gouin.....	62

Figure 2.16. Apports totaux du mois d'avril au complexe Aux Outardes avant et après 1983 sur papier de probabilité normal	66
Figure 3.1. Répartition des apports énergétiques pour la période 1943-1993	82
Figure 3.2. Distribution des apports énergétiques mensuels - Complexe Bersimis	84
Figure 3.3. Distribution des apports énergétiques mensuels - Complexe Chur. Falls.....	85
Figure 3.4. Distribution des apports énergétiques mensuels - Complexe La Grande	86
Figure 3.5. Distribution des apports énergétiques mensuels - Complexe Manic.....	87
Figure 3.6. Distribution des apports énergétiques mensuels - Complexe Outardes	88
Figure 3.7. Distribution des apports énergétiques mensuels - Complexe Outaouais.....	89
Figure 3.8. Distribution des apports énergétiques mensuels - Complexe St-Maurice	90
Figure 3.9. Distribution des apports énergétiques mensuels - Complexe St-Laurent.....	91
Figure 3.10. Distribution des apports énergétiques mensuels - Série agrégée	92
Figure 3.11. Apports énergétiques moyens mensuels des huit complexes.....	93
Figure 3.12. Écart-type mensuel des apports énergétiques des huit complexes.....	94
Figure 3.13. Apports énergétiques annuels du complexe Bersimis	100
Figure 3.14. Apports énergétiques annuels du complexe Churchill Falls.....	100
Figure 3.15. Apports énergétiques annuels du complexe La Grande	101
Figure 3.16. Apports énergétiques annuels du complexe Manicouagan.....	101
Figure 3.17. Apports énergétiques annuels du complexe Aux Outardes	101
Figure 3.18. Apports énergétiques annuels du complexe Outaouais.....	102
Figure 3.19. Apports énergétiques annuels du complexe St-Maurice	102
Figure 3.20. Apports énergétiques annuels du complexe St-Laurent.....	102
Figure 3.21. Apports énergétiques annuels agrégés	103
Figure 3.22. Distribution des apports énergétiques annuels des huit complexes.....	104
Figure 3.23. Représentation des complexes dans l'espace des 3 premiers facteurs	107
Figure 3.24. Représentation des complexes dans l'espace des 2 premiers facteurs	108
Figure 3.25. Série chronologique des cusums centrés (Bersimis)	116
Figure 3.26. Série chronologique des cusums centrés (Churchill Falls)	116
Figure 3.27. Série chronologique des cusums centrés (La Grande)	117
Figure 3.28. Série chronologique des cusums centrés (Manicouagan).....	117
Figure 3.29. Série chronologique des cusums centrés (Aux Outardes).....	117
Figure 3.30. Série chronologique des cusums centrés (Outaouais).....	118
Figure 3.31. Série chronologique des cusums centrés (St-Maurice)	118

Figure 3.32. Série chronologique des cusums centrés (St-Laurent).....	118
Figure 3.33. Série chronologique des cusums centrés (Série agrégée).....	119
Figure 4.1a. δ_0 est beaucoup moins crédible que les valeurs de l'intervalle	128
Figure 4.1b. δ_0 est presque aussi crédible que les valeurs de l'intervalle.....	128
Figure 4.2. Résultats de la procédure bayésienne - Apports agrégés (1943-1993).....	130
Figure 4.3. Résultats de la procédure bayésienne - Complexe St-Laurent	133
Figure 4.4. Série simulée à partir d'un modèle stationnaire AR(1).....	135
Figure 4.5. Distribution des moyennes des séries simulées.....	147
Figure 4.6. Distribution des écart-types des séries simulées	147
Figure 4.7. Distribution des coefficients d'autocorrélation $\tilde{\rho}_1$ des séries simulées	147
Figure 4.8. Distribution des coefficients de Hurst des séries simulées	148
Figure 4.9. Écarts énergétiques annuels 1943-1993 (déficits et surplus).....	149
Figure 4.10. Résultats de la procédure bayésienne - Apports agrégés (1943-1995).....	160

SOMMAIRE

Ce rapport présente les travaux réalisés dans le cadre du projet C2 de la chaire en hydrologie statistique intitulé *Étude des séries historiques d'apports énergétiques*. Cinq grands thèmes ont été abordés: l'harmonisation des apports naturels mensuels, l'analyse statistique descriptive des apports énergétiques mensuels et annuels, le problème de non-stationnarité de la moyenne des apports énergétiques annuels, la modélisation de la série d'apports énergétiques agrégés annuels (simulation et prévision) et la distribution statistique des écarts énergétiques annuels par rapport à la moyenne (déficit et surplus énergétique).

Problématique générale

Dans un but de planification, Hydro-Québec présente périodiquement depuis 1979, un plan de développement de ses activités pour les 10 années suivantes. Depuis 1990, ce plan est réactualisé tous les trois ans. Cet exercice repose sur une analyse détaillée de l'évolution du contexte économique et énergétique. Il permet, en particulier, d'établir les politiques de développement à court, moyen et long termes qu'adopte l'entreprise. C'est pourquoi, il est important, du point de vue économique, de prévoir avec un maximum de précision la demande et l'offre hydroélectrique, puisque ces deux composantes sont déterminantes dans le choix des interventions à privilégier. On s'intéresse ici à la prévision de l'offre, c'est-à-dire du stock d'eau disponible pour produire de l'électricité aux centrales.

Afin de connaître et de prévoir avec précision l'état des réserves énergétiques, les apports naturels (en m³/s) sont calculés chaque jour et pour chacun des réservoirs du parc. Le cumul des apports naturels quotidiens à chaque site permet de connaître les apports naturels mensuels de chacun des réservoirs. Ces séries s'échelonnent sur une période variable selon les sites. Pour obtenir une plage commune d'au moins cinquante années d'apports naturels mensuels à chacun des sites, différentes interventions sont effectuées sur les données. Cette opération, réalisée périodiquement à Hydro-Québec, est appelée harmonisation des apports naturels. En effet, pour certains réservoirs, les composantes entrant dans le calcul des apports naturels (Haché *et al.*, 1996) n'ont pu être mesurées sur la période minimale considérée de cinquante années ou sont erronées. Des ensembles de données complets à chaque site sont alors obtenus à l'aide de différentes méthodes de reconstitution et d'extension de séries d'apports naturels (Bruneau, 1984, 1986a, b, c). Le pourcentage d'apports naturels reconstitués par l'une ou l'autre des approches est estimé à plus de 50%

des historiques de 1943 à 1993. Connaissant la capacité de production de chaque centrale, on peut ensuite transformer le stock d'eau emmagasinée et les apports naturels harmonisés en valeur énergétique exprimée en gigawatts-heure (GWh). Hydro-Québec dispose donc de séries d'apports énergétiques mensuels calculés ou reconstitués pour l'ensemble des centrales de son parc d'équipement sur au moins cinquante années.

Les données d'apports énergétiques constituent la base du système de gestion et de planification de l'énergie produite par Hydro-Québec. En particulier, certaines caractéristiques statistiques issues de ces séries d'apports énergétiques (moyenne, écart-type, autocorrélation) sont utilisées comme intrants dans les modèles de planification employés pour:

- l'optimisation de l'utilisation de l'énergie à long terme (horizon de 15 ans);
- l'optimisation de la gestion des réservoirs à moyen terme (horizon de 1 à 2 ans);
- le chargement optimal des centrales;
- l'analyse de la puissance de pointe;
- la gestion hydraulique des rivières;
- l'adéquation des évacuateurs;
- la mise en place de nouveaux équipements;
- le déclenchement, en temps opportun, de mesures préventives.

Une bonne connaissance des caractéristiques statistiques des séries d'apports énergétiques mensuels et annuels est donc nécessaire au bon fonctionnement du système de planification d'Hydro-Québec. En particulier, dans le contexte de faible hydraulité qui prévaut depuis quelques années, Hydro-Québec doit se doter d'outils lui permettant de gérer efficacement les ressources hydriques dont elle dispose. Toute l'information possible doit donc être extraite des séries d'apports énergétiques de façon à évaluer les probabilités d'occurrence de différents scénarios d'hydraulité et la marge de manoeuvre qu'Hydro-Québec a pour réagir et garantir la production d'énergie avec une certaine fiabilité.

Objectifs généraux

Le présent projet ne vise pas à reconsidérer les modèles de planification énergétique mais bien à analyser les séries d'apports énergétiques historiques. Les apports énergétiques mensuels et annuels des huit complexes hydroélectriques (Outaouais, St-Maurice, Aux

Outardes, Manicouagan, Bersimis, St-Laurent, La Grande et Churchill Falls) ainsi que leur agrégation (énergie totale) ont été étudiés.

Le premier objectif général du projet est de faire le point sur la construction des séries d'apports énergétiques (harmonisation des apports) et sur le calcul des principales caractéristiques statistiques servant d'intrants aux modèles de planification.

Le second objectif général est d'extraire toute l'information possible des séries d'apports énergétiques de façon à évaluer les probabilités d'occurrence de différents scénarios d'hydraulicité. Cela peut être réalisé d'abord à l'aide d'une analyse statistique descriptive des séries d'apports énergétiques annuels et mensuels et ensuite, en modélisant la série d'apports énergétiques annuels agrégés.

Le dernier objectif du projet est de répondre à certaines questions spécifiques d'Hydro-Québec. Un document précisant les points sur lesquels Hydro-Québec souhaite obtenir des recommandations de l'équipe de la chaire a été rédigé et remis à l'INRS-Eau le 31 août 1995. Les questions de recherche prioritaires soumises dans ce document concernent la stationnarité de la moyenne, la modélisation de la série d'apports énergétiques agrégés annuels (simulation et prévision) et la distribution statistique des écarts énergétiques annuels par rapport à la moyenne (déficit et surplus énergétique).

Principaux travaux réalisés

Comme on l'a mentionné précédemment, cinq grands thèmes ont été abordés dans le projet C2. Il s'agit de l'harmonisation des apports naturels mensuels, de l'analyse statistique descriptive des apports énergétiques mensuels et annuels, du problème non-stationnarité de la moyenne des apports énergétiques annuels, de la modélisation de la série d'apports énergétiques agrégés annuels (simulation et prévision) et de la distribution statistique des écarts énergétiques annuels par rapport à la moyenne (déficit et surplus énergétique).

Les travaux ont été réalisés dans le cadre de quatre sous-projets. Ils sont présentés brièvement dans ce qui suit.

Sous-projet 2.1: Planification du projet

Ce premier sous-projet consistait à identifier les besoins des différents utilisateurs de données d'apports énergétiques à Hydro-Québec ainsi que toutes les étapes de construction des séries d'apports énergétiques. Des rencontres avec les principaux intervenants en

planification énergétique à Hydro-Québec ont permis de tirer un portrait général de la situation actuelle et de définir les objectifs spécifiques du projet.

Sous-projet 2.2: Étude critique des données de base

Les interventions effectuées sur les séries d'apports naturels (harmonisation des apports) sont nombreuses et peuvent influencer le calcul des caractéristiques statistiques des données d'apports énergétiques utilisées comme intrants des modèles de planification. Ce sous-projet a donc été consacré à l'étude critique des données de base et des procédures employées lors de l'harmonisation des apports. Plus précisément, les travaux suivants ont été réalisés:

- Étude des rapports internes rédigés à Hydro-Québec concernant l'harmonisation des apports naturels mensuels (en particulier, Bruneau, 1984, 1986a, b, c);
- Élaboration de fiches synthèses de toutes les interventions effectuées sur les données d'apports naturels mensuels des principaux sites des complexes hydroélectriques (45 sites);
- Analyse critique des différentes procédures utilisées pour chacun des complexes hydroélectriques;
- Identification des problèmes d'ordre méthodologique;
- Recommandations d'ordre méthodologique et proposition de méthodes statistiques alternatives.

Sous-projet 2.3: Analyse statistique des séries d'apports énergétiques

Pour extraire toute l'information possible des séries d'apports énergétiques, ce sous-projet a été consacré à une analyse statistique descriptive complète des apports énergétiques mensuels et annuels des huit complexes hydroélectriques et de la série agrégée. En particuliers, les travaux suivants ont été effectués:

Apports énergétiques mensuels (1943-1993)

- Calcul des caractéristiques statistiques de base des neuf séries mensuelles;
- Examen de la variation temporelle des principales caractéristiques statistiques et de la distribution des apports énergétiques (examen graphique);
- Calcul des corrélations mensuelles;

Apports énergétiques annuels (1943-1993)

- Calcul des caractéristiques statistiques de base des neuf séries annuelles;
- Examen de la variation spatiale de la distribution des apports énergétiques annuels (examen graphique) et vérification de la normalité des données;
- Calcul des corrélations entre les séries d'apports énergétiques annuels des 8 complexes et analyse en composantes principales;
- Calcul des autocorrélations des apports énergétiques annuels;
- Examen de la contribution des séries d'apports énergétiques des complexes à l'autocorrélation significative de la série agrégée;
- Étude préliminaire de la stationnarité de la moyenne des séries d'apports énergétiques annuels.

Sous-projet 2.4: Modélisation de la série d'apports énergétiques annuels agrégés

Les résultats des analyses statistiques effectuées au sous-projet 2.3 ont permis à Hydro-Québec d'identifier certaines questions de recherche prioritaires (cf. document du 31 août 1995). Ces questions concernaient la stationnarité de la moyenne, la modélisation de la série d'apports énergétiques agrégés annuels (simulation et prévision) et la distribution statistique des écarts énergétiques annuels par rapport à la moyenne (déficit et surplus énergétique). Pour répondre aux besoins d'Hydro-Québec, les travaux suivants ont été réalisés:

- Analyse approfondie de la stationnarité de la moyenne des apports énergétiques annuels agrégés: application d'une procédure bayésienne de détection de changement de la moyenne à toutes les séries annuelles d'apports énergétiques ainsi qu'à un ensemble de 22 séries de précipitations annuelles;
- Modélisation des apports énergétiques annuels agrégés (1943-1993): comparaison de neuf modèles (analyse des résidus et simulations);
- Prévision des écarts énergétiques annuels par rapport à la moyenne à partir des quatre modèles théoriques retenus (estimation des modèles sur la période 1943-1993);
- Prévision des écarts énergétiques annuels cumulés sur différents horizons à partir des quatre modèles théoriques retenus (prévision des années 1994 à 1998);

- Détermination de la distribution statistique des prévisions des écarts énergétiques annuels et cumulés;
- Détermination des probabilités d'occurrence de divers événements (déficits, surplus, etc.);
- Mise à jour des prévisions: modèles estimés sur la période 1943-1995 et prévisions des années 1996 à 2000.

Principales conclusions et recommandations

Les conclusions obtenues ainsi que les principales recommandations soumises suite aux travaux réalisés dans le cadre des sous-projets 2.1, 2.2, 2.3 et 2.4 sont résumées dans ce qui suit.

Sous-projet 2.1 (Chapitre 1)

- Les séries d'apports énergétiques utilisées pour la planification sont constituées, en partie, de données reconstituées. Il est donc nécessaire, avant d'analyser les apports énergétiques, de faire le point sur l'ensemble des interventions effectuées sur la banque de données. Un sous-projet a donc été consacré à ce problème (sous-projet 2.2).

Sous-projet 2.2 (Chapitre 2)

- Le pourcentage d'apports naturels reconstitués par l'une ou l'autre des approches (harmonisation des apports) est estimé à plus de 50% des historiques de 1943 à 1993. Ces séries d'apports harmonisés ont été utilisées lors des analyses effectuées dans le cadre des sous-projets 2.3 et 2.4. Les résultats obtenus reposent donc sur l'hypothèse que les observations sont représentatives des apports énergétiques réels.
- Certains points communs méthodologiques ont permis d'identifier quatre procédures générales d'harmonisation des apports et ainsi de classer les complexes selon le type de procédure employée. Les complexes hydroélectriques d'un même groupe, et donc soumis au même type de procédure d'harmonisation des apports naturels, possèdent des caractéristiques semblables.
- Les données d'apports naturels employées par le service Prévisions correspondent aux apports harmonisés par le service Hydraulique jusqu'en 1988. Toutefois, de 1989 à nos jours, les apports naturels mensuels utilisés diffèrent. Le service Prévisions utilise depuis

1989 les apports calculés par bilan hydrique plutôt que ceux issus des méthodes de reconstitution du service Hydraulique.

- La reconstitution ou l'extension de séries d'apports mensuels à l'aide de la méthode du rapport des superficies de bassins versants RSBV est problématique: présence d'un biais annuel systématique, présence d'un biais saisonnier, périodicité et variance des données reconstituées imposée par la méthode et entièrement indépendante de la variance des apports historiques du site d'intérêt. Quelques approches alternatives sont recommandées dans le présent rapport, notamment l'emploi d'un rapport saisonnier des apports naturels (coefficient de transfert variable) ou de régressions multiples mois par mois.
- Dans certaines circonstances, l'utilisation des tests de Student, de Fisher et de la procédure de détection de changement de la moyenne de Lee et Heghenian (1977), lors de l'harmonisation des apports, n'est pas valide:
 - non-normalité des apports naturels;
 - biais des tests de Student et de Fisher lorsqu'ils sont appliqués suite à la procédure de Lee et Heghenian;
 - corrélation temporelle lors de la comparaison des caractéristiques statistiques mensuelles des séries d'apports de deux sites différents.

Des procédures statistiques mieux adaptées à ces circonstances sont recommandées. En particulier, des tests non paramétriques unidimensionnels et des tests paramétriques multidimensionnels sont proposés dans ce rapport.

- Les variables retenues pour l'analyse de classification hiérarchique (CAH) nous semblent insuffisantes pour bien caractériser les séries de données et discriminer les séries hétérogènes. Nous recommandons, en particulier, de considérer le coefficient d'autocorrélation d'ordre 1 lors de l'application de la CAH.

Sous-projet 2.3 (Chapitre 3)

Apports énergétiques mensuels (1943-1993)

- Les systèmes La Grande, Churchill Falls et Manicouagan cumulent près de 73% de la production énergétique totale. Les caractéristiques statistiques des apports énergétiques de ces sites influencent donc considérablement celles de la série agrégée.

- La distribution des apports varie d'un mois à l'autre pour la plupart des complexes (Bersimis, Churchill Falls, La Grande, Aux Outardes, Manicouagan, Outaouais, St-Maurice). En particulier, la variabilité des apports énergétiques est systématiquement plus grande en période de crue (avril à juillet). Par contre, le comportement de la distribution mensuelle des apports énergétiques du complexe St-Laurent est plus uniforme.
- Une forte corrélation est observée presque systématiquement entre les apports de deux mois consécutifs pour les complexes Churchill Falls, La Grande, Outaouais, St-Maurice, St-Laurent et la série des apports agrégés. Ces corrélations sont toutefois moins élevées pour les complexes Bersimis, Manicouagan et Aux Outardes.
- Toutes les corrélations mensuelles sont élevées pour le complexe St-Laurent. Les apports énergétiques mensuels sont donc très liés quel que soit le délai (1 à 12 mois).
- La matrice des corrélations mensuelles des apports énergétiques agrégés présente une structure assez semblable aux corrélations observées pour les complexes Churchill Falls et La Grande, les systèmes contribuant le plus aux apports totaux.

Apports énergétiques annuels (1943-1993)

- La dispersion des apports énergétiques varie en fonction de la dimension des complexes (plus le système est grand, plus la variabilité des apports est grande).
- Le test de normalité de Shapiro-Wilk a été appliqué aux neuf séries d'apports énergétiques annuels et confirme l'hypothèse de normalité. Seule la série d'apports annuels au complexe Bersimis n'est pas distribuée selon une loi normale au niveau de signification de 5%.
- Des corrélations généralement élevées et significatives sont obtenues pour des complexes voisins. En particulier, les corrélations entre les apports énergétiques annuels des complexes Manicouagan/Aux Outardes, Outaouais/St-Maurice et Aux Outardes/Bersimis sont respectivement 0.86, 0.82 et 0.81.
- Les apports énergétiques annuels du complexe St-Laurent ne sont pas corrélés avec ceux des autres complexes. Ces apports se comportent donc de façon tout à fait différente.
- L'analyse en composantes principales montre très clairement trois groupes de complexes: St-Laurent, [Outaouais/St-Maurice] et [Bersimis/Outardes/Manicouagan/Churchill Falls/La Grande].

- Parmi les complexes, seuls les apports énergétiques annuels du St-Laurent sont significativement autocorrélés. Les autocorrélations sont d'ailleurs significatives pour tous les délais considérés (une à cinq années).
- La série d'apports énergétiques annuels agrégés possède un coefficient d'autocorrélation d'ordre 1 significatif à 5% (0.356). La somme des apports énergétiques annuels de tous les complexes d'une année donnée est donc liée au cumul obtenu l'année précédente.
- Le coefficient d'autocorrélation d'ordre 1 de la série agrégée n'est pas influencée, comme on aurait pu le croire, par celui des apports du complexe St-Laurent (forte autocorrélation). La valeur obtenue à partir des observations (0.356) proviendrait plutôt de l'accumulation de faibles autocorrélations positives de certaines séries individuelles dont la contribution est plus importante (en particulier, Churchill Falls, La Grande et Manicouagan). Cela tend à démontrer que, même si les autocorrélations d'ordre 1 de ces séries ne sont pas significatives à 1% (dû à la taille limitée de l'échantillon), elles sont probablement réelles.
- Aucun changement de la moyenne, significatif à 5%, n'est induit par les interventions effectuées lors de l'harmonisation des apports naturels mensuels.

Sous-projet 2.4 (Chapitre 4)

- La procédure bayésienne de détection de non-stationnarité indique que la date la plus probable d'occurrence d'un changement de moyenne est l'année 1984 pour la série d'apports énergétiques annuels agrégés. On peut observer aussi une probabilité non négligeable, mais plus faible, que cette rupture survienne en 1963 ou en 1964. Cela pourrait signifier la présence de deux changements de moyenne dans la série. La série d'apports énergétiques agrégés annuels pourrait être non stationnaire et constituée de trois sous-séries de moyennes différentes (1943-1964, 1965-1983, 1984-1993).
- Le calcul d'intervalles de crédibilité bayésiens montrent, pour la série agrégée et la plupart des complexes (à l'exception de Manicouagan et du St-Laurent), que l'hypothèse de non-changement est crédible. Toutefois, les dates de rupture les plus probables obtenues permettent de constater une certaine cohérence régionale quant au moment d'un éventuel changement de la moyenne.
- La rupture observée en 1970 pour le complexe St-Laurent est nettement plus significative statistiquement que celle des autres séries.

- Les résultats de l'application de la procédure de détection du changement de moyenne obtenus pour les séries de précipitations montrent une probabilité de changement non négligeable entre 1963 et 1971 pour plusieurs de ces séries. Cela tend à montrer qu'un phénomène climatique serait effectivement survenu durant la décennie 1960-1970 et que cet événement aurait davantage influencé les apports énergétiques du St-Laurent que ceux des autres systèmes hydriques.
- Compte tenu de l'information restreinte dont on dispose ici (51 données d'apports énergétiques), on peut difficilement distinguer entre la non-stationnarité de la moyenne et la présence de déviations soutenues engendrées par l'autocorrélation (persistance). Malgré les résultats obtenus à partir des séries de précipitations, supportant la présence d'un changement de la moyenne, il est difficile de conclure "hors de tout doute raisonnable" à la non-stationnarité de la moyenne dans la série des apports énergétiques annuels agrégés. L'emploi de la moyenne générale comme intrant aux modèles de planification est donc recommandé. Toutefois, puisqu'elle est estimée à partir d'un échantillon restreint et qu'une non-stationnarité pourrait être détectée à l'aide d'un nombre de données plus élevé, il faut considérer cette valeur comme incertaine. Il serait approprié, notamment, d'effectuer une étude de sensibilité des modèles de planification en faisant varier la valeur de cet intrant.
- La série d'apports énergétiques annuels est donc considérée comme stationnaire et les modèles de prévision employés reposent sur cette hypothèse. Des modèles admettant une persistance accrue ont tout de même été considérés (modèles ARMA(1,1) dont les paramètres ont été estimés par la méthode de O'Connell) compte tenu de la présence possible d'un changement de moyenne.
- L'estimateur classique du coefficient d'autocorrélation d'ordre 1 est biaisé (sous-estimation) lorsque les observations sont représentées par un modèle AR(1). Une correction de biais est proposée et recommandée pour calibrer le modèle AR(1). Puisque ce modèle semble adéquat pour les apports énergétiques annuels agrégés, l'autocorrélation de cette série serait plutôt de 0.41.
- La procédure utilisée actuellement à Hydro-Québec pour déterminer la distribution des écarts énergétiques repose entièrement sur une simulation dont la valeur de départ est un écart nul. Telle qu'appliquée, la simulation ne donne aucune information concernant les écarts pouvant survenir dans le futur à une date spécifique. D'une part, les données simulées ne tiennent pas compte du dernier événement de la série, malgré le fait qu'un écart énergétique est fortement influencé par l'événement de l'année précédente. En

effet, la valeur initiale de la procédure est fixée à zéro. D'autre part, l'approche consiste à simuler une longue série que l'on partitionne ensuite pour déterminer les probabilités. Ceci peut induire un biais dans les résultats (Hipel et McLeod, 1994). Nous recommandons donc l'emploi de prévisions qui permettent d'estimer l'écart énergétique susceptible de se produire, avec la plus grande probabilité, à une période donnée dans le futur. L'incertitude d'échantillonnage (variance et intervalle de confiance) peut être évaluée analytiquement. Cette approche permet aussi de calculer aisément la plupart des événements ou probabilités qui intéressent Hydro-Québec.

- Nous recommandons l'emploi du modèle AR(1) pour des fins de prévision et de simulation. Ce modèle, qui donne des résultats très comparables au modèle ARMA(1,1), a l'avantage d'être très simple d'utilisation. Toutefois, nous ne rejetons pas tous les modèles de O'Connell qui préservent le coefficient de Hurst et qui admettent une gamme de scénarios plus vaste. Leurs caractéristiques statistiques sont plus variables et supposent une plus grande persistance du phénomène. Ils permettent donc de générer des scénarios d'apports ou d'écarts énergétiques plus extrêmes qui peuvent être ensuite soumis comme intrant à des modèles de planification. Nous croyons, par contre, que l'emploi de tels modèles nécessitent une étude approfondie du phénomène de Hurst.
- Les propriétés théoriques des prévisions issues des modèles considérés sont bien connues, ce qui permet de déterminer analytiquement les probabilités d'événements spécifiques plus complexes (cumul des écarts, par exemple). Toutefois, si on s'intéresse particulièrement aux durées des séquences consécutives de déficits ou de surplus, il n'existe pas à notre connaissance d'équations analytiques directes. On peut alors avoir recours à une simulation dont la valeur initiale doit correspondre à l'écart observé lors de la dernière année.

Retombées pratiques

Les principales retombées du projet C2, présentement identifiables, sont données dans ce qui suit.

Sous-projets 2.1 et 2.2

- Une revue et une synthèse de toutes les reconstitutions effectuées lors de l'harmonisation des apports naturels a été réalisée. Le rapport contient des fiches synthèses (figures 2.1 à 2.4) de toutes les interventions effectuées sur les données permettant ainsi, d'un seul

coup d'oeil, de savoir d'où proviennent les apports énergétiques mensuels de chacun des principaux sites du réseau.

- Cette synthèse a permis de mettre en commun les besoins des divers services d'Hydro-Québec.
- L'analyse critique des méthodes de reconstitution et d'extension de données a permis de déceler quelques lacunes d'ordre méthodologique concernant en particulier :
 - la méthode de transfert par le rapport des superficies de bassins versants (biais).
 - les problèmes particuliers aux complexes Aux Outardes et La Grande.
 - l'application du test de Student suite à la procédure de Lee et Heghenian (niveau de signification).
 - le problème lié à l'hypothèse de normalité des tests de Student et de Fisher.
 - le problème lors de la comparaison des apports mensuels de deux séries par les tests de Student et de Fisher (corrélations entre les mois).

Les recommandations émises et les méthodes alternatives proposées permettront d'améliorer la procédure systématique d'harmonisation des apports et de compter sur une base de données d'apports énergétiques de meilleure qualité. Cela est très important pour la planification de l'exploitation et pour celle de l'équipement (aidant, par exemple, à mieux calculer le dimensionnement des ouvrages de production).

Sous-projets 2.3

- La mise à jour des principales caractéristiques statistiques des apports énergétiques mensuels et annuels de tous les complexes qui a été faite (statistiques de base, corrélations, autocorrélations, analyse en composantes principales, etc.) permet une meilleure connaissance du système. De plus, cette analyse est nécessaire pour effectuer une modélisation efficace des séries d'apports énergétiques.
- Cette étude a permis de mettre en évidence la présence d'autocorrélations significatives pour les séries d'apports énergétiques annuels (en particulier, les apports agrégés). Cette information doit être prise en compte lors de la modélisation.
- Une analyse préliminaire de la stationnarité de la moyenne des apports énergétiques annuels a été effectuée. L'hypothèse de stationnarité des apports énergétiques n'avait pas été étudiée de manière approfondie auparavant même si les spécialistes d'Hydro-Québec soupçonnaient la présence d'un changement de moyenne. Cette étude a permis:

- de définir clairement la non-stationnarité, de mieux comprendre le problème et de montrer que la communauté scientifique (hydrologie, météorologie, climatologie, etc.) se penche sur ce problème.
- de montrer que ce problème est complexe (distinction difficile à effectuer entre autocorrélation et changement de moyenne), que les outils de détection sont limités et qu'il faut appuyer nos conclusions à l'aide d'analyses complémentaires (précipitations, information régionale, etc.).
- de mettre en évidence et de quantifier l'incertitude sur la moyenne (principal intrant).
- de proposer une méthode de prise en compte de cette incertitude dans la planification de l'exploitation.

Sous-projets 2.4

- Un intervalle de confiance bayésien (intervalle de crédibilité) est proposé pour vérifier si le changement de moyenne est significatif. Cette procédure a été appliquée aux séries annuelles d'apports énergétiques et à un ensemble de séries annuelles de précipitations. Compte tenu de la taille limitée des échantillons, il est difficile de conclure, "hors de tout doute raisonnable", à la non-stationnarité de la moyenne.
- Une étude comparative de divers modèles confirme la validité du modèle AR(1) pour représenter les apports énergétiques annuels agrégés. Toutefois, des modèles admettant une plus grande persistance ont aussi été considérés. Ceux-ci ne doivent pas être rejetés puisqu'ils produisent des scénarios plus extrêmes qui peuvent être ensuite utilisés comme intrant dans les modèles de planification. Cela peut aider la prise de décision.
- De nombreuses discussions ont permis de faire la distinction entre la simulation (génération de scénarios) et la prévision. Ces deux outils sont complémentaires.
- L'importance d'effectuer des prévisions (tenir compte du caractère autorégressif des apports et de la dernière donnée observée) et d'utiliser une correction de biais de l'estimateur de l'autocorrélation d'ordre 1 a été démontrée.
- Des formules analytiques pour le calcul de prévisions d'écart énergétique et de leur incertitude (variance, intervalle de confiance) ont été développées.
- Une méthodologie simple a été proposée pour le calcul analytique des probabilités d'occurrence de divers événements susceptibles de se produire dans les années à venir: surplus et déficits énergétiques, écarts énergétiques cumulés. Les propriétés théoriques des prévisions issues des modèles considérés sont bien connues, ce qui a permis de

déterminer analytiquement les probabilités d'événements spécifiques plus complexes (cumul des écarts, par exemple). Toutefois, si on s'intéresse particulièrement aux durées des séquences consécutives de déficits ou de surplus, il n'existe pas à notre connaissance d'équations analytiques directes. On peut alors avoir recours à une simulation dont la valeur initiale doit correspondre à l'écart observé la dernière année. Une modélisation à partir de modèles de processus aléatoires devrait aussi être envisagée.

- Des recommandations ont été proposées concernant les questions de recherche prioritaires soumises par Hydro-Québec (document du 31 août 1995).

Retombées générales

La réalisation de ce projet a:

- permis de réunir les principaux intervenants en planification énergétique pour réfléchir sur le problème du risque associé à la possible présence de non-stationnarité.
- montré la nécessité de rechercher un modèle de risque pour permettre la prise de décision.
- permis d'envisager la possibilité de développer un estimateur de la moyenne qui soit robuste à la présence de non-stationnarité.
- montré la nécessité de mettre à jour les modèles de prévision à chaque année.

1 INTRODUCTION

1.1 Situation actuelle

Hydro-Québec est une entreprise de service public qui produit, transporte et distribue presque toute l'électricité consommée au Québec. De plus, Hydro-Québec vend de l'énergie à d'autres réseaux électriques du Canada et du nord-est des États-Unis. C'est également l'une des seules entreprises productrices d'électricité au monde à exploiter un réseau où l'énergie d'origine hydraulique représente plus de 95 % de l'énergie totale produite. L'eau est emmagasinée dans de larges réservoirs à remplissage annuel et multi-annuel. Une portion de l'électricité est aussi produite au fil de l'eau. Le réseau de transport et de distribution d'Hydro-Québec est l'un des plus vastes en Amérique du Nord.

Dans un but de planification, Hydro-Québec présente périodiquement depuis 1979, un plan de développement de ses activités pour les 10 années à venir. Au début, le plan de développement décennal était ré-actualisé une fois l'an. Depuis 1990, ce plan est réactualisé tous les trois ans. Cet exercice repose sur une analyse détaillée de l'évolution du contexte économique et énergétique. Il permet, en particulier, d'établir les politiques de développement à court, moyen et long termes qu'adopte l'entreprise par l'entremise de la planification de sa production hydro-électrique, de la planification des équipements de production et des stratégies de développement des marchés externes.

Jusqu'à tout récemment Hydro-Québec pouvait négocier d'intéressants contrats d'exportation d'énergie, avec ses voisins américains et d'autres provinces canadiennes, en raison de surplus d'énergie essentiellement attribuables à des conditions d'hydraulicité favorables. Depuis 1985, Hydro-Québec a connu plusieurs années de précipitations déficitaires, donc de faible hydraulicité. C'est la mise en place de nouveaux équipements (La Grande phase 2), combiné à une diminution de la demande, qui ont permis jusqu'à présent de faire face à cette période de déficit cumulé durant la dernière décennie.

Hydro-Québec doit effectuer une gestion rigoureuse de ses ressources et prendre différentes mesures afin d'assurer le maintien de l'équilibre entre l'offre et la demande en énergie pour les prochaines années, et ainsi permettre à sa clientèle de bénéficier d'un service adéquat. C'est pourquoi des mesures telles que:

- le devancement ou le report de projets de suréquipement de centrales hydro-électriques et d'aménagement de nouveaux complexes;
- le rachat de contrats d'exportation;
- la mise en route d'équipements de production thermique,

sont prises afin de permettre à Hydro-Québec d'accroître sa marge de manoeuvre.

La mise en place de telles mesures peut s'avérer très onéreuse si elles ne sont pas appliquées en temps opportun. C'est pourquoi il est important du point de vue économique de prévoir avec un maximum de précision la demande et l'offre hydro-électrique, puisque ces deux composantes sont déterminantes dans le choix des interventions à privilégier. On s'intéresse, dans cette étude, à la prévision de l'offre, c'est-à-dire du stock d'eau disponible pour produire de l'énergie hydroélectrique aux centrales.

1.2 Problématique

Le potentiel hydroélectrique du Québec, actuellement aménagé, repose sur les réserves hydrauliques emmagasinées dans les réservoirs de 7 grands complexes : Outaouais, St-Maurice, Aux Outardes, Manicouagan, Bersimis, St-Laurent et La Grande. Un huitième grand complexe, Churchill-Falls, n'est pas géré par Hydro-Québec mais, puisque la majorité de sa production est dirigée vers le réseau d'Hydro-Québec, ses règles de gestion sont conformes aux besoins du réseau et sa production est incluse dans les simulations de gestion du réseau d'Hydro-Québec. La figure 1.1 donne la localisation géographique des huit grands complexes hydroélectriques.

Afin de connaître et de prévoir avec précision l'état des réserves énergétiques, on calcule, chaque jour et pour chacun des réservoirs du réseau, les apports naturels exprimés en m^3/s . Les apports naturels sont calculés par bilan hydrique et il est important de les connaître avec précision. Le projet C1 de la Chaire en hydrologie statistique est d'ailleurs consacré à ce problème (Haché *et al.* 1996). Le cumul mensuel des apports naturels quotidiens à chaque site permet de connaître les apports naturels mensuels de l'ensemble des sites. Connaissant la capacité de production de chaque centrale, on peut ensuite transformer le stock d'eau emmagasinée et les apports naturels en valeur énergétique exprimée en gigawatts-heure (GWh).

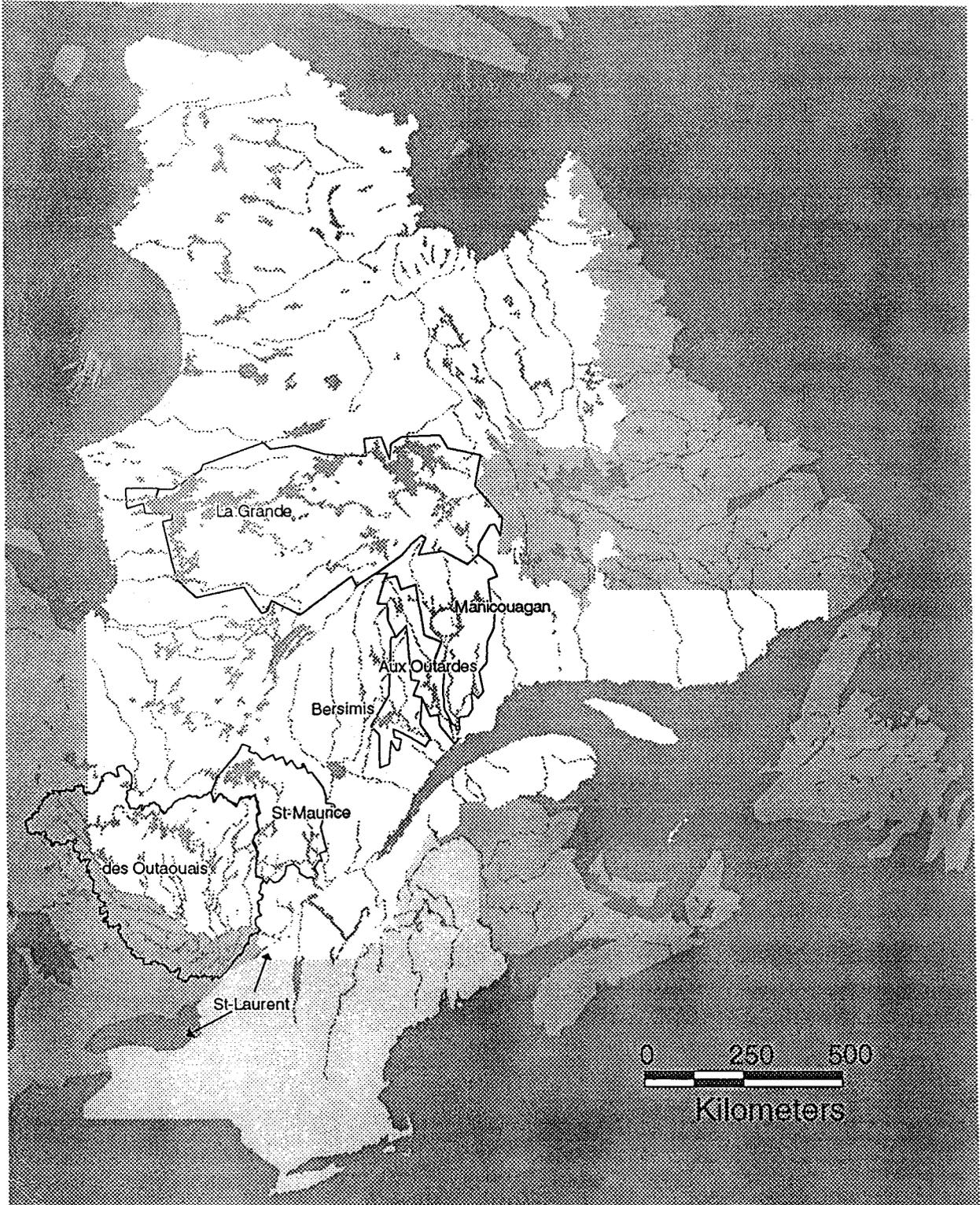


Figure 1.1. Localisation géographique des complexes hydroélectriques.

L'apport énergétique mensuel E_n d'une centrale donnée est donc proportionnel à l'apport naturel mensuel A_n calculé au site. Pour chaque centrale d'un complexe hydroélectrique, un facteur de production exprimé en KW/(m³/s) est d'abord calculé. Les facteurs de production d'un même complexe sont ensuite cumulés de l'aval vers l'amont. Ainsi, la centrale située la plus en amont d'un complexe possède le facteur de production le plus élevé. Pour un mois donné, l'apport naturel mensuel A_n (en m³/s) de chaque centrale est multiplié par son facteur de production pour obtenir l'énergie mensuelle E_n disponible au site (en GWh). Hydro-Québec dispose de séries d'apports énergétiques mensuels calculés pour l'ensemble des centrales de son parc d'équipements. Ces séries s'échelonnent sur une période variable selon les sites. Les données d'apports énergétiques sont la base du système de gestion et de planification de l'énergie produite par Hydro-Québec.

L'énergie mensuelle totale produite par un complexe hydroélectrique est alors calculée en cumulant les apports énergétiques mensuels de chaque site du complexe de l'amont vers l'aval. Ensuite, l'énergie annuelle totale du complexe est obtenue en sommant les apports énergétiques mensuels sur toute l'année. Enfin, une série annuelle agrégée, représentant la production totale de l'ensemble du parc d'équipement d'Hydro-Québec, est construite en sommant les apports énergétiques annuels totaux de chaque complexe. Ainsi, quatre types de séries d'apports énergétiques sont obtenues:

1. Apports énergétiques mensuels de chaque centrale;
2. Apports énergétiques mensuels des huit complexes hydroélectriques (cumul spatial);
3. Apports énergétiques annuels des huit complexes hydroélectriques (cumul temporel);
4. Apports énergétiques annuels agrégés (cumul spatial).

Ces différentes étapes de calcul et d'agrégation menant à la construction de ces quatre types de séries sont précédées d'une opération préliminaire réalisée sur les apports naturels mensuels A_n aux sites. Il s'agit d'une reconstitution et d'une extension de séries permettant d'obtenir une plage commune d'au moins cinquante années de valeurs d'apports à chaque site. En effet, pour certains réservoirs, les composantes entrant dans le calcul des apports naturels A_n (Haché *et al.*, 1996) n'ont pu être mesurées sur la période minimale considérée de cinquante années ou sont tout simplement erronées. Des ensembles de données complets à chaque site ont alors été obtenus à l'aide de différentes méthodes de reconstitution et d'extension de séries d'apports naturels (Bruneau, 1984, 1986a, b, c). Le pourcentage

d'apports naturels A_n reconstitués par l'une ou l'autre des approches est estimé à plus de 50% des historiques de 1943 à 1993¹.

L'analyse des quatre types de séries d'apports énergétiques est un outil d'aide à la décision pour la gestion optimale du réseau d'Hydro-Québec. En particulier, certaines caractéristiques statistiques issues de ces séries d'apports énergétiques (moyennes, écarts-types, autocorrélation) sont utilisées comme intrants dans les modèles de planification. Ces modèles sont employés pour :

- l'optimisation de l'utilisation de l'énergie à long terme (horizon de 15 ans);
- l'optimisation de la gestion des réservoirs à moyen terme (horizon de 1 à 2 ans);
- le chargement optimal des centrales;
- l'analyse de la puissance de pointe;
- la gestion hydraulique des systèmes hydriques;
- l'adéquation des évacuateurs;
- la mise en place de nouveaux équipements.

Le schéma de la figure 1.2 illustre, à partir de l'acquisition des données aux sites, les différentes étapes menant au calcul des intrants des modèles de planification énergétique.

Une bonne connaissance des caractéristiques statistiques (intrants des modèles de planification) des différentes séries d'apports énergétiques est donc nécessaire au bon fonctionnement du système de planification d'Hydro-Québec. Or, les manipulations effectuées sur les données (reconstitutions, extensions, etc.) soulèvent quelques questions sur la validité de cette information de base. En effet, différentes irrégularités non expliquées ont été observées dans les caractéristiques statistiques des séries d'apports énergétiques. Une étude approfondie de la construction des ces séries ainsi qu'une analyse statistique descriptive doivent être effectuées. De plus, afin de déterminer plus précisément les risques de défaillance énergétique associées à différents scénarios d'exploitation et de bien planifier le parc d'équipements de production hydroélectrique, une modélisation statistique de ces

¹ Depuis le début du projet C2, nous avons considéré la série qui s'échelonne de 1943 à 1993. Dans la section 4.4, les données des années 1994 et 1995 ont été prises en compte pour les prévisions.

séries doit être réalisée. Cette modélisation permettrait, en particulier, de déterminer les probabilités d'occurrence associées à des volumes de surplus ou de déficits énergétiques.

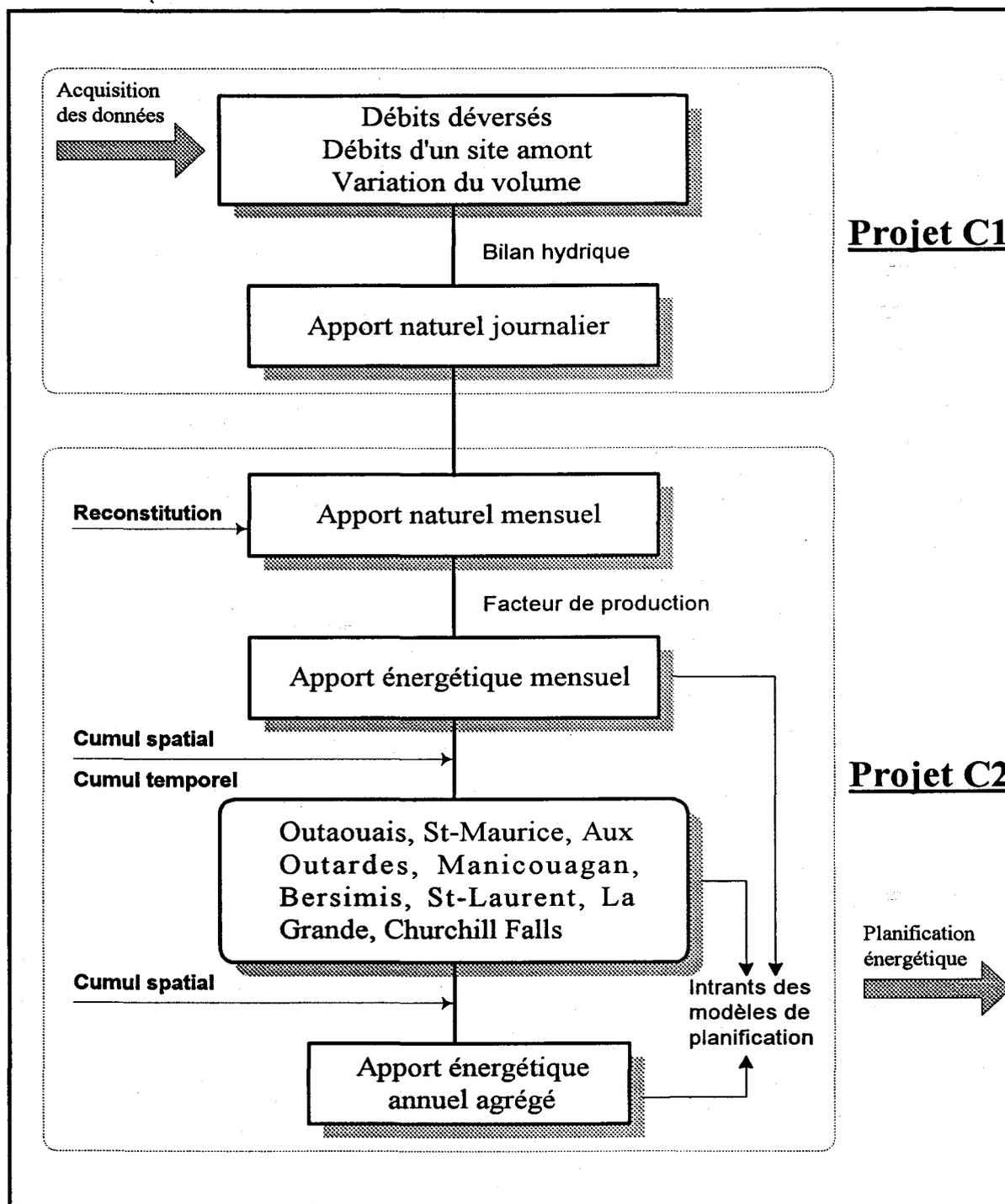


Figure 1.2. Étapes menant au calcul des intrants des modèles de planification.

1.3 Objectifs généraux du projet

Les différents modèles de planification de la production à Hydro-Québec reposent principalement sur des caractéristiques statistiques issues des données d'apports énergétiques. Ces données ont été remaniées de façon à obtenir divers types de séries s'échelonnant sur une plage commune de cinquante années. Comme il a été mentionné précédemment, ces interventions soulèvent quelques questions de recherche quant à la validité de cette information de base. En particulier :

- peut-on expliquer théoriquement le fait que certaines caractéristiques statistiques soient non significatives pour des séries d'apports énergétiques E_n calculés aux sites mais deviennent significatives suite à une agrégation temporelle et/ou spatiale ?
- les données d'apports naturels A_n reconstituées influencent-elles la valeur des intrants des modèles utilisés pour la planification énergétique à court, moyen et long terme ?
- doit-on nécessairement, compte-tenu de la possible présence de non-stationnarité, utiliser les séries d'observations s'échelonnant sur cinquante années pour les modèles de planification énergétique ?

Le présent projet ne vise pas à reconsidérer les modèles de planification énergétique mais bien à étudier les série d'apports E_n . Le premier objectif général du projet est donc de faire le point sur la construction des séries d'apports énergétiques et sur le calcul des principales caractéristiques statistiques servant d'intrants aux modèles de planification.

Dans le contexte de faible hydraulicité qui prévaut depuis plusieurs années, Hydro-Québec doit se doter d'outils lui permettant de gérer efficacement les ressources hydriques dont elle dispose. Pour ce faire, toute l'information possible doit être extraite des séries d'apports énergétiques de façon à évaluer les probabilités d'occurrence de différents scénarios d'hydraulicité qui conditionnent le déclenchement de mesures correctives. Cela peut être réalisé d'abord à l'aide d'une analyse statistique descriptive des séries d'apports énergétiques annuels et mensuels des complexes et ensuite, en modélisant la série d'apports énergétiques annuels agrégés. Ceci constitue le second objectif général du présent projet.

1.4 Description des objectifs spécifiques du projet

Le projet C2 de la chaire en hydrologie statistique intitulé "Étude des séries historiques d'apports énergétiques annuels" est divisé en quatre sous-projets. Les sous-projets sont décrits sommairement dans les sections qui suivent.

1.4.1 Planification du projet (Chapitre 1)

Ce premier sous-projet consiste à identifier les besoins des différents utilisateurs de données d'apports énergétiques à Hydro-Québec ainsi que toutes les étapes de construction des séries d'apports énergétiques. Les objectifs spécifiques du projet ont été aussi déterminés.

1.4.2 Étude critique des données de base nécessaires au calcul des apports énergétiques (Chapitre 2)

Les manipulations effectuées sur les séries d'apports naturels sont nombreuses et peuvent influencer le calcul des caractéristiques statistiques des données d'apports énergétiques utilisées comme intrants des modèles de planification. Ce sous-projet est donc entièrement consacré à l'étude critique de ces données de base. La synthèse de toutes les reconstitutions et extensions effectuées sur les séries d'apports naturels mensuels, afin d'obtenir une plage commune de cinquante années d'apports énergétiques, est réalisée. Une analyse des différentes approches utilisées pour chacun des sept complexes hydroélectriques est ensuite effectuée afin d'identifier d'éventuels problèmes méthodologiques. Enfin, des recommandations d'ordre méthodologique sont données à l'issue du sous-projet.

1.4.3 Analyse statistique des séries d'apports énergétiques (Chapitre 3)

Dans le but d'extraire toute l'information possible des séries d'apports énergétiques, ce sous-projet consiste à effectuer une analyse statistique descriptive complète des apports totaux mensuels et annuels des sept complexes hydroélectriques considérés. Cette analyse comprend en particulier le calcul des principales caractéristiques statistiques des séries ainsi que la détermination des corrélations entre les différents complexes. Quelques tests statistiques sont aussi appliqués afin de vérifier certaines hypothèses sur lesquelles reposent l'utilisation de ces séries dans les modèles de planification (tests de normalité).

1.4.4 Modélisation de la série d'apports énergétiques annuels agrégés (Chapitre 4)

Pour établir les politiques de gestion qu'adoptera l'entreprise, il est important, du point de vue économique, de prévoir avec un maximum de précision la demande et l'offre hydroélectrique, puisque ces deux composantes sont déterminantes dans le choix des interventions à privilégier. Afin de déterminer les risques de défaillance énergétique associés à différents scénarios d'exploitation et de bien planifier le parc d'équipements de production hydroélectrique, Hydro-Québec utilise différents modèles de simulation. Ces modèles reposent, entre autres, sur les caractéristiques statistiques de la série d'apports énergétiques annuels agrégés (moyenne, écarts-type, autocorrélation) et sont basés sur l'hypothèse que les observations sont stationnaires.

Les résultats des analyses statistiques effectuées dans le sous-projet 2.3, en particulier ceux concernant la présence éventuelle d'un changement de moyenne dans la série des apports énergétiques annuels agrégés, ont permis de définir plus précisément les objectifs du sous-projet 2.4. Un document précisant les points sur lesquels Hydro-Québec souhaite obtenir des recommandations de l'équipe de la chaire a été rédigé et remis à l'INRS-Eau le 31 août 1995. Les questions de recherche prioritaires soumises dans ce document concernaient la stationnarité de la moyenne, la modélisation de la série d'apports énergétiques annuels (simulation et prévision) et la distribution statistique des écarts énergétiques annuels par rapport à la moyenne (déficit et surplus énergétique).

Pour répondre aux besoins d'Hydro-Québec, le sous-projet 2.4 est divisé en quatre étapes :

- Analyse approfondie de la stationnarité de la moyenne des apports énergétiques annuels agrégés
- Modélisation des écarts énergétiques annuels agrégés par rapport à la moyenne générale et prévisions
- Distribution statistique des écarts énergétiques annuels agrégés
- Sensibilité des prévisions au changement de moyenne

2 ÉTUDE CRITIQUE DES DONNÉES DE BASE

Ce chapitre présente les travaux réalisés dans le cadre du sous-projet 2 intitulé "Étude critique des données de base nécessaire au calcul des apports énergétiques".

2.1 Problématique

Pour obtenir au moins cinquante années d'apports énergétiques mensuels de bonne qualité pour les principaux sites du réseau, Hydro-Québec a procédé à des reconstitutions et des extensions de séries d'apports naturels mensuels. Outre l'acquisition d'une plage commune de cinquante années de données pour tous les sites, les deux principaux critères de qualité considérés sont l'homogénéité temporelle de la moyenne et de la variance ainsi qu'un comportement similaire à celui des débits de rivières voisines. Les reconstitutions et les extensions ont donc été effectuées dans un souci de conservation, d'une part, des caractéristiques statistiques (moyennes et variances), et d'autre part, de la concordance régionale. De nombreuses manipulations ainsi que plusieurs tests statistiques ont donc été effectués sur les données d'apports naturels mensuels des sites des sept complexes hydroélectriques. Ce projet, intitulé "Validation de longues séries de débits observés", a été réalisé par le Service Hydraulique d'Hydro-Québec entre 1984 et 1986 et est bien documenté dans Bruneau (1984; 1986a, b, c).

Depuis 1986, des mises à jour périodiques des données d'apports naturels ont été effectuées par le Service Hydraulique afin d'inclure les observations nouvellement acquises, tout en respectant les méthodologies utilisées en 1986 et les critères considérés. Les données d'apports naturels mensuels ainsi remaniées (validation, extension, reconstitution, etc.) et s'échelonnant sur une période d'au moins cinquante années, sont appelées "apports naturels harmonisés". Elles constituent l'ensemble de l'information de base employée pour construire les différents types de séries d'apports énergétiques (cf. section 1.2). Les caractéristiques statistiques issues des séries d'apports énergétiques résultantes sont celles qu'Hydro-Québec utilise comme intrants dans les modèles de planification énergétique.

Le tableau 2.1 donne, pour chaque complexe hydroélectrique, les sites traités dans le cadre du projet "Validation de longues séries de débits observés". Les abréviations de chaque site, données au tableau 2.1, sont utilisées tout au long du présent rapport. Puisque les mesures de débit et de production à Churchill Falls ne relèvent pas d'Hydro-Québec, elles ne sont pas incluses dans la présente étude critique.

Tableau 2.1. Sites traités lors de la validation de longues séries de débits observés

COMPLEXE	SOUS-BASSIN	SITES	ABRÉVIATION
Outaouais	Gatineau	Cabonga	CAB
		Baskatong	BAS
		Maniwaki	MANI
	Outaouais Supérieur	Paugan	PAU
		Chelsea	CHE
		Rapides-Farmers	RAF
		Dozois	DOZ
		Dozois/Rapides-Sept	D/R7
		Rapides-Sept/Lac des Quinzes	R7/L15
	Outaouais Inférieur	Kipawa	KIP
		Témiscamingue	TEM
		Témiscamingue/Chenaux	TEM/CHE
		Chenaux/Chats falls Chats Falls/Carillon	CHE/CHF CHF/CAR
St-Maurice		Gouin	GOU
		Manouane	MANO
		Matawin	MAT
		Rapide-Blanc	RAB
		Trenche	TRE
		Beaumont	BEA
		LaTuque	LAT
		Grand-Mère	GRM
		Shawinigan	SHA
		La Gabelle	LAG
Aux Outardes		Outardes 4	OU4
		Outardes 3	OU3
		Outardes 2	OU2
Manicouagan		Manic 5	MA5
		Lac Ste-Anne	LSA
		Manic 3	MA3
		Manic 2	MA2
		Manic 1	MA1
Bersimis		Bersimis 1	BE1
		Bersimis 2	BE2
St-Laurent		Cornwall	COR
		Lac St-François (inter)	LSFI
		Lac St-François (total)	LSFT
La Grande		Caniapiscau	CAN
		Laforge 2	LA2
		Laforge 1	LA1
		La Grande 4	LG4
		La Grande 3	LG3
		La Grande 2	LG2
		La Grande 1 Eastmain 1	LG1 EM1

Étant donnée l'importance des travaux réalisés pour harmoniser les données de base nécessaires au calcul des apports énergétiques, le premier objectif de ce sous-projet était de faire le point sur l'ensemble des interventions effectuées sur les apports naturels mensuels. Une étude des principaux documents relatifs à ces travaux d'harmonisation des apports (Bruneau, 1986a, b, c) a d'abord permis de regrouper les complexes hydroélectriques en quatre groupes selon les procédures employées. Ensuite, pour chaque groupe de complexes, une synthèse de toutes les reconstitutions et extensions de séries a été effectuée. La section 2.2 est consacrée à la présentation de cette synthèse.

Les caractéristiques statistiques issues des séries d'apports énergétiques dépendent directement de celles des séries d'apports naturels harmonisés et de la manière dont elles ont été construites. C'est pourquoi, le second objectif du sous-projet 2 était l'étude critique des méthodes de reconstitution et d'extension employées, en vue de détecter, s'il y a lieu, les problèmes méthodologiques et d'effectuer des recommandations d'ordre général pour les travaux futurs d'harmonisation. Cette étude critique repose d'une part sur l'étude de rapports internes d'Hydro-Québec, et d'autre part, sur des analyses statistiques descriptives et graphiques réalisées à l'aide des données de certains sites témoins. Les résultats de ces travaux, ainsi que quelques recommandations, sont présentés à la section 2.3.

2.2 Synthèse des reconstitutions et extensions de séries

La procédure utilisée par le Service Hydraulique pour obtenir, en chaque site du tableau 2.1, une série d'apports naturels mensuels valides s'échelonnant sur au moins cinquante années varie d'un complexe hydroélectrique à l'autre. Certains points communs méthodologiques permettent toutefois d'identifier quatre procédures générales et ainsi de classer les sept complexes selon le type de méthodes employées. Les quatre méthodologies identifiées sont notées A, B, C et D et font appel à différents outils statistiques de validation, de reconstitution et d'extension de séries. Avant de décrire les étapes de chacune des quatre procédures (cf. section 2.2.4), les principaux outils statistiques utilisés sont présentés brièvement. Les détails mathématiques ne sont pas donnés dans les sections qui suivent puisque ces techniques statistiques sont connues et bien documentées dans les ouvrages de base en statistique. Les principales références sont données dans le texte.

Les travaux de synthèse nous ont permis d'observer que le Service Prévisions d'Hydro-Québec a utilisé de 1989 à 1994 des données d'apports naturels mensuels différentes des

apports harmonisés par le Service Hydraulique. La méthodologie employée par le Service Prévision est décrite à la section 2.2.5.

2.2.1 Homogénéité temporelle

De façon générale, pour les sept complexes hydroélectriques, les séries d'apports naturels mensuels disponibles aux différents sites ont toutes été validées selon un critère d'homogénéité temporelle. Si nécessaire, suite à la vérification de ce critère, des reconstitutions, partielles ou complètes, ainsi que des extensions de séries ont été effectuées de façon à obtenir des séries valides d'au moins cinquante années d'apports mensuels.

Une série est considérée homogène dans le temps si ses caractéristiques statistiques sont invariantes sur toute la période disponible. Les deux caractéristiques statistiques considérées par Hydro-Québec sont la moyenne et la variance des données. La méthode la plus utilisée dans les documents d'Hydro-Québec (Bruneau, 1984, 1986a, b, c) pour vérifier l'homogénéité temporelle de la moyenne consiste, dans un premier temps, à appliquer la procédure bayésienne proposée par Lee et Heghinian (1977) (cette procédure est présentée en détail dans la section 4.2.1 du présent rapport). Cette approche permet de déterminer, sous l'hypothèse qu'un changement de moyenne soit effectivement présent dans la série, la date t où ce changement possède la plus grande probabilité de se produire. Ayant identifié le moment le plus probable d'un éventuel changement de moyenne, la série est ensuite divisée en deux sous-ensembles (avant et après la date critique t) et un test de Student est appliqué aux deux sous-échantillons. Le test de Student (Bickel et Doksum, 1977) permet de comparer les moyennes de deux échantillons et donc de vérifier si le changement de moyenne à la date critique t est bel et bien significatif.

Cette approche est appliquée quand aucune information a priori n'est disponible concernant la date d'un changement éventuel de moyenne. Toutefois, dans certains cas, cette information est connue. Par exemple, la mise en service d'un nouveau site peut influencer le comportement de la série des valeurs d'apports naturels du bassin total et engendrer une hétérogénéité temporelle. Le test de Student est alors appliqué directement sans utiliser la procédure bayésienne de Lee et Heghinian (1977). Dans ces circonstances, c'est-à-dire lorsque la date d'un éventuel changement dans les caractéristiques statistiques est connue, la variance ainsi que la distribution des deux sous-échantillons ont aussi été comparées. Les tests statistiques utilisés sont respectivement le test de Fisher (Bickel et Doksum, 1977) et le test de Kolmogorov-Smirnov (D'Agostino et Stephens, 1986).

Pour chaque ensemble d'apports naturels mensuels disponibles (apports intermédiaires aux sites et apports du bassin total si nécessaire), ces méthodes de détection d'hétérogénéité ont été appliquées à des séries de pas de temps annuels : apports minimums annuels, apports maximums annuels, apports annuels moyens et apports mensuels considérés mois par mois (série annuelle pour chaque mois). Il est important de souligner que les tests de Student et de Fisher, ainsi que la procédure de Lee et Heghinian, reposent principalement sur deux hypothèses mathématiques :

1. les observations sont indépendantes;
2. les observations sont distribuées selon une loi normale.

Le test de Kolmogorov-Smirnov ne repose toutefois que sur l'hypothèse d'indépendance des observations. En effet, ce test est non paramétrique et la loi de sa statistique est tout à fait indépendante de la distribution des observations.

2.2.2 Homogénéité et concordance régionales

La détection d'un changement temporel significatif des caractéristiques statistiques d'une série d'apports naturels n'est pas nécessairement suivie du rejet des données de la série traitée et de leur reconstitution. En effet, pour appuyer la conclusion des tests et pour s'assurer d'une homogénéité spatiale, les débits de rivières voisines (si disponibles) sont aussi soumis au même type d'analyse de l'homogénéité temporelle. Si la même tendance est détectée dans les séries de débits, la série d'apports naturels est alors retenue malgré son hétérogénéité. On suppose que le changement ne provient pas d'erreurs de mesures ou d'une modification de la méthode d'acquisition des données mais plutôt de variations de certains facteurs physiques à l'échelle régionale (par exemple, fluctuations climatiques normales ou changements climatiques).

La concordance régionale d'une série d'apports avec une série de débits mesurés en rivières naturelles est aussi examinée en comparant directement les caractéristiques statistiques de ces deux types de données. Pour ce faire, les tests de Student, de Fisher et de Kolmogorov-Smirnov sont appliqués aux séries de pas de temps annuel : apports minimums annuels, apports maximums annuels, apports annuels moyens et apports mensuels mois par mois. Les observations ont été préalablement standardisées en les divisant soit par la superficie de bassin versant, soit par le module annuel.

Enfin, pour certains complexes (Aux Outardes, par exemple), Hydro-Québec pouvait disposer de deux séries différentes d'apports naturels à un même site. Ces séries étant acceptables du point de vue de l'homogénéité temporelle, le choix d'un ensemble de données était donc basé sur le critère de concordance régionale. Dans ce cas, en plus d'effectuer les tests de comparaison de leurs caractéristiques statistiques avec celles de données mesurées à des rivières naturelles voisines (tests de Student, Fisher et Kolmogorov-Smirnov), une analyse de classification ascendante hiérarchique (CAH) a été réalisée. La CAH (Jambu, 1978) est une technique permettant de regrouper les éléments semblables d'un tableau de données. Ces regroupements se font par étapes en calculant, à partir de variables mesurées pour chaque élément, une distance entre ceux-ci. La distance utilisée pour évaluer les écarts entre les éléments est la distance Euclidienne qui est définie comme étant la somme des différences au carré. Un arbre hiérarchique est ensuite produit qui permet d'interpréter toutes les étapes de cette procédure de regroupement et de former des groupes distincts d'éléments homogènes selon les variables considérées.

La CAH appliquée par Hydro-Québec utilise comme éléments les deux séries d'apports naturels ainsi que les séries de débits mesurés de plusieurs rivières voisines (Bruneau, 1986b, c). Les variables considérées sont les moyennes mensuelles mois par mois, les moyennes annuelles, les maximums annuels et les minimums annuels. La CAH a donc permis d'étudier régionalement les similarités entre les caractéristiques statistiques et de vérifier si l'une ou l'autre des deux séries d'apports naturels au même site se démarque par rapport à celles de sites environnants.

2.2.3 Reconstitution et extension de séries

Après avoir effectué les vérifications d'homogénéité temporelle et spatiale, les séries non-valides et/ou incomplètes sont soumises, en partie ou entièrement, à différents modèles de reconstitution. De plus, si cela est nécessaire, on fait une extension de série afin d'obtenir au moins cinquante années d'apports naturels mensuels. Les modèles statistiques utilisés sont les mêmes que l'on procède à une reconstitution ou à une extension.

Trois approches sont principalement utilisées pour reconstituer¹ les données d'une série d'apports naturels :

¹ Nous confondons, dans la suite, reconstitution et extension puisque les mêmes modèles sont utilisés.

- le modèle de régression multiple avec conservation de la moyenne et de la variance de la série originale (HEC-4, REMUL, *RéMuS*²);
- la méthode du rapport de superficies de bassins versants (RSBV);
- le modèle de désagrégation spatiale et temporelle (SPIGOT).

2.2.3.1 Régression multiple (*RéMuS*)

Dans bien des domaines, la régression linéaire multiple est l'une des méthodes statistiques les plus utilisées pour étudier la relation entre une variable dépendante et plusieurs variables indépendantes (ou variables explicatives). En théorie statistique, la régression multiple a fait l'objet de nombreux articles et d'ouvrages scientifiques très complets (en particulier, Neter *et al.*, 1985 et Draper et Smith, 1966). Considérons une variable dépendante Y et p variables explicatives supposées indépendantes X_1, X_2, \dots, X_p . Le modèle de régression multiple est de la forme suivante :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon \quad (2.1)$$

où ε est une variable aléatoire distribuée selon une loi normale.

Tel qu'utilisée par Hydro-Québec dans le projet "Validation de longues séries de débits observés", la régression est effectuée mois par mois et permet de reconstituer la variable Y (apport au site d'intérêt) à partir des variables X (débit des rivières voisines) tout en conservant, au sens statistique, la moyenne et la variance de la série originale des données observées de la variable Y . Pour un mois donné l'ensemble des données sur lesquelles le modèle est ajusté comprend donc les apports mensuels disponibles au site d'intérêt (y_1, y_2, \dots, y_n) et les débits de p rivières voisines ($x_{1j}, x_{2j}, \dots, x_{pj}$, $j = 1, \dots, p$) calculés pour n années concomitantes. Suite à l'ajustement du modèle de régression (estimation des paramètres), l'équation générale permettant de reconstituer l'apport non valide ou manquant à l'aide des p mesures disponibles aux autres sites s'exprime comme suit :

² REMUL est une version légèrement modifiée par Hydro-Québec du logiciel HEC-4 (Beard, 1971); *RéMuS* est une version considérablement modifiée et améliorée disponible depuis peu (Perron *et al.*, 1996; Perreault *et al.*, 1994a) et qui n'a donc pas été utilisée dans les études d'harmonisation de 1984 et 1986 (Bruneau, 1986a, b, c). Nous utiliserons le terme *RéMuS* indifféremment pour désigner l'un ou l'autre des logiciels.

$$\tilde{Y}_i = \hat{\beta}_0 + \sum_{j=1}^p \hat{\beta}_j x_{ij} + \delta_j \quad (2.2)$$

où:

- \tilde{Y}_i est l'estimation de l'apport de l'année i au site d'intérêt;
- x_{ij} est le débit mensuel de l'année i mesuré au site voisin j ;
- $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p$ sont les paramètres estimés par la régression multiple obtenus par la méthode des moindres carrés;
- δ_i est un nombre aléatoire provenant d'une loi normale et permettant de conserver la variance de la série originale y_1, y_2, \dots, y_n .

Pour certains sites à reconstituer, les variables explicatives utilisées dans le modèle de régression ne sont pas des débits en rivières mais plutôt des précipitations. Une description détaillée des aspects théoriques de la reconstitution des débits par régression multiple est donnée dans Perreault *et al.* (1994a). Ce document est lié au développement du logiciel *RéMuS* utilisé actuellement à Hydro-Québec pour faire ce genre d'opérations sur les séries de temps.

Notons que la régression multiple ne peut être employée que pour une reconstitution partielle d'une série d'apports. En effet, cette méthode nécessite, pour l'estimation des paramètres, de disposer d'un ensemble de données valides sur une certaine période, à la fois au site à reconstituer (y_1, y_2, \dots, y_n) et aux rivières de référence utilisées comme variables explicatives ($x_{1j}, x_{2j}, \dots, x_{nj}$, $j = 1, \dots, p$). En raison de cette contrainte, pour effectuer une reconstitution complète d'une série, Hydro-Québec a dû se tourner vers d'autres méthodes, en particulier l'estimation par la méthode des rapports de superficies de bassins versants.

2.2.3.2 Rapport de superficies de bassins versants (RSBV)

La méthode de transfert de données par rapport de superficies de bassins versants s'apparente à celle de la régression multiple en ce sens qu'elle repose aussi sur une relation linéaire entre les apports de divers sites. Toutefois, cette relation ne s'appuie pas sur des techniques statistiques mais plutôt sur des prémisses hydrologiques et physiques : des sites voisins soumis à des conditions climatiques et physiographiques équivalentes devraient posséder le même débit spécifique. Le débit spécifique à un site est obtenu en divisant le débit par la superficie du bassin versant correspondant.

Généralement, dans ce type d'approche, le bassin versant total d'un complexe hydroélectrique est d'abord divisé en sous-bassins correspondant à chacun des sites. Pour un sous-bassin donné Y dont les données ne sont pas valides, une recherche est ensuite effectuée pour trouver un sous-bassin voisin X (appartenant au même complexe de préférence) possédant des données valides sur la période d'intérêt. Les apports du sous-bassin Y sont alors calculés à partir des apports spécifiques du sous-bassin ayant des données valides (site de référence). Plus précisément, l'apport naturel \hat{Q}_Y estimé par cette méthode s'exprime de la façon suivante :

$$\hat{Q}_Y = \frac{A_Y}{A_X} Q_X \quad (2.3)$$

où A_X et A_Y désignent respectivement les superficies des sous-bassins versants X et Y et Q_X l'apport du sous-bassin X (site de référence). Cette méthode peut être facilement généralisée de façon à utiliser plus d'un sous-bassin voisin (UNESCO, 1981). L'apport estimé devient alors une moyenne pondérée des apports aux sites voisins. Il est important de noter que la relation (2.3) est applicable lorsque l'ensemble des sous-bassins utilisés possède des régimes hydrologiques et climatiques semblables.

Dans le projet "Validation de longues séries de débits mensuels observés", Hydro-Québec a abondamment utilisé cette approche puisque la plupart des séries reconstituées l'ont été entièrement et que la régression multiple ne pouvait s'appliquer. De plus, pour quelques complexes (Outaouais et St-Maurice), l'équation (2.3) a été appliquée à partir de sites de référence situés dans des bassins versants voisins. Cette approche donne de bons résultats en autant que les régimes hydrologiques des différents bassins versants ne diffèrent pas significativement.

2.2.3.3 Désagrégation spatiale et temporelle (SPIGOT)

Le modèle de génération stochastique SPIGOT ("Synthetic Streamflow Generation Software Package"), proposé par Grygier et Stedinger (1990), permet la désagrégation temporelle et spatiale d'une série globale en conservant les principales caractéristiques statistiques des observations (moyennes, variances, corrélations et autocorrélations).

Cette méthode a été appliquée au complexe hydroélectrique de la rivière La Grande où les corrélations mois par mois entre les débits mensuels des rivières mesurés sur une longue période (rivières de référence) et les apports aux sites n'étaient pas significatives (Bruneau, 1986c). La reconstitution sur une base mensuelle ne pouvait pas être effectuée. Hydro-Québec s'est donc tournée vers une reconstitution de la série des apports annuels du bassin

total. À ce pas de temps, la série des apports totaux s'avère très bien corrélée avec celle des débits annuels des rivières de référence. C'est pourquoi, après avoir obtenu une série globale d'apports annuels du bassin total reconstitués à partir de modèles de régression et de la méthode du rapport de superficies de bassins, le modèle de désagrégation **SPIGOT** a été utilisé pour obtenir des données mensuelles en chaque site.

Les apports annuels totaux sont d'abord distribués sur les douze mois de chacune des années. Ces derniers sont ensuite désagrégés dans l'espace. Le modèle **SPIGOT** effectue ces opérations en conservant les caractéristiques statistiques des séries d'apports observés aux sites.

2.2.4 Description des procédures de validation et synthèse des reconstitutions

Comme on l'a mentionné précédemment, la procédure utilisée par le service Hydraulique pour obtenir, en chaque site du tableau 2.1, une série d'apports naturels mensuels harmonisés, c'est-à-dire des données valides et s'échelonnant sur au moins cinquante années, varie d'un complexe hydroélectrique à l'autre. Certains points communs méthodologiques ont toutefois permis d'identifier quatre procédures générales et ainsi de classer les sept complexes selon le type de procédure employée. Les quatre méthodologies A, B, C et D, décrites en détail dans cette section, font appel aux outils statistiques présentés brièvement aux sections 2.2.1, 2.2.2 et 2.2.3. Le tableau 2.2 donne les regroupements de complexes considérés.

Tableau 2.2. Les complexes hydroélectriques classés selon les procédures employées.

PROCÉDURES	COMPLEXES HYDROÉLECTRIQUES
A	Aux Outardes, Manicouagan, Bersimis
B	Outaouais, St-Maurice
C	La Grande
D	St-Laurent

On remarque que les complexes hydroélectriques d'un même groupe, et donc soumis au même type de procédure d'harmonisation des apports naturels, possèdent des caractéristiques semblables. Comme on peut l'observer sur la figure 1.1, les groupes obtenus correspondent à quatre régions géographiques bien distinctes du Québec : nord-est (procédure A), sud-est (procédure B), nord-ouest (procédure C) et centre-sud (procédure D).

D). Les bassins versants des complexes appartenant à un même groupe sont donc soumis à un climat ainsi qu'à des caractéristiques géomorphologiques relativement semblables. Outre les complexes La Grande et St-Laurent qui diffèrent des autres systèmes principalement par leur localisation géographique et qui forment chacun un groupe, les complexes soumis aux procédures A et B se distinguent particulièrement par le nombre de sites et la complexité de leur répartition sur le bassin versant. Les complexes Aux Outardes, Manicouagan et Bersimis possèdent un nombre restreint de sites (3, 5 et 2 sites respectivement, cf. tableau 2.1) répartis, à partir d'un réservoir de tête, du nord vers le sud au fil d'un long cours d'eau principal. Les bassins versants des complexes Outaouais et St-Maurice ont une forme plus circulaire et sont divisés en nombreux sous-bassins (14 et 10 sites respectivement, cf. tableau 2.1) répartis assez uniformément le long de différents cours d'eau. Cette structure est plus complexe.

Les caractéristiques physiques et géographiques des bassins versants ont donc une influence sur le type de méthodologie employée pour obtenir des séries d'apports naturels harmonisés en chaque site. Les procédures A, B, C et D sont présentées dans ce qui suit.

2.2.4.1 Procédure A

L'idée générale de la procédure A est d'obtenir d'abord une longue série de référence d'apports naturels mensuels valides à l'embouchure du système hydrique (apports du bassin versant total). Cette série est ensuite répartie aux différents sites du complexe en respectant la contribution de chacun des sous-bassins à l'apport du bassin versant total. Les complexes hydroélectriques de la Côte-Nord (Aux Outardes, Manicouagan et Bersimis) se prêtent bien à ce type d'approche. En effet, leurs sites sont situés du nord au sud le long d'un même cours d'eau pour lequel, en général, une longue série de débits naturels mesurés en aval est disponible avant la mise en service des différents ouvrages.

La procédure A est divisée en quatre grandes étapes :

1. Examen des données disponibles à l'embouchure du bassin versant (apports naturels mensuels du bassin versant total). Plusieurs séries d'observations en aval du complexe peuvent être disponibles (séries de banques de données différentes). Ces séries sont souvent constituées de différents types de données selon la période (débits naturels, apports calculés, etc.).
2. Validation des données et choix d'une série valide (si nécessaire) sur la période disponible. Les tests d'homogénéité temporelle et spatiale sont effectués (procédure

de Lee et Heghinian, tests de Student, de Fisher, et Kolmogorov-Smirnov, cf. sections 2.2.1 et 2.2.2).

3. Reconstitution et/ou extension des apports totaux (si nécessaire) par régression multiple avec les débits des rivières voisines (section 2.2.3). Dans certaines situations, la reconstitution peut être faite à partir des apports totaux d'un complexe voisin.
4. Calcul des apports intermédiaires à chaque site du système en distribuant les apports totaux validés (série de référence) aux sites à l'aide de la méthode du rapport de superficies de bassins versants (section 2.2.3).

Les apports intermédiaires obtenus par la procédure A dépendent donc de la méthode de transfert par rapport de superficies de bassins versants et des différentes interventions effectuées sur la série de référence. La figure 2.1 illustre, sous forme de diagramme des concomitances, le contenu de la série d'apports de référence utilisée pour chaque complexe. Notons que pour le système hydrique Bersimis, la série de référence ne correspond pas aux apports naturels calculés en aval du complexe mais plutôt à ceux du réservoir de tête BE1. Les apports de cette série ont été reconstitués avant 1956 à partir de régressions multiples avec les apports totaux au complexe Aux Outardes (série N9519003). Après 1956, cette série contient les apports calculés directement par bilan hydrique au site BE1. La série de référence du système hydrique Aux Outardes (N9519003) est composée de débits naturels en aval du système (Chute aux Outardes) transformés avant 1937, de débits naturels bruts entre 1937 et 1968 et du cumul des apports intermédiaires (bilan hydrique) après la mise en service des sites. La transformation des débits naturels a été effectuée suite à la détection d'un changement significatif de la variance de la série avant et après 1937. Enfin, la série des apports totaux à Manicouagan est constituée des débits naturels mesurés en aval entre 1946 et 1964, et de débits estimés par régression multiple mois par mois avec les données du complexe Aux Outardes (série N9519003) avant et après cette période.

2.2.4.2 Procédure B

Plutôt que de construire une série d'apports totaux valides à l'embouchure du complexe et de la répartir en apports intermédiaires aux différents sites, la procédure B consiste à transférer directement à chaque site les débits naturels de rivières voisines. Ces rivières de référence, dont les débits naturels ont été validés préalablement, ne sont pas nécessairement situées dans le bassin versant du complexe.

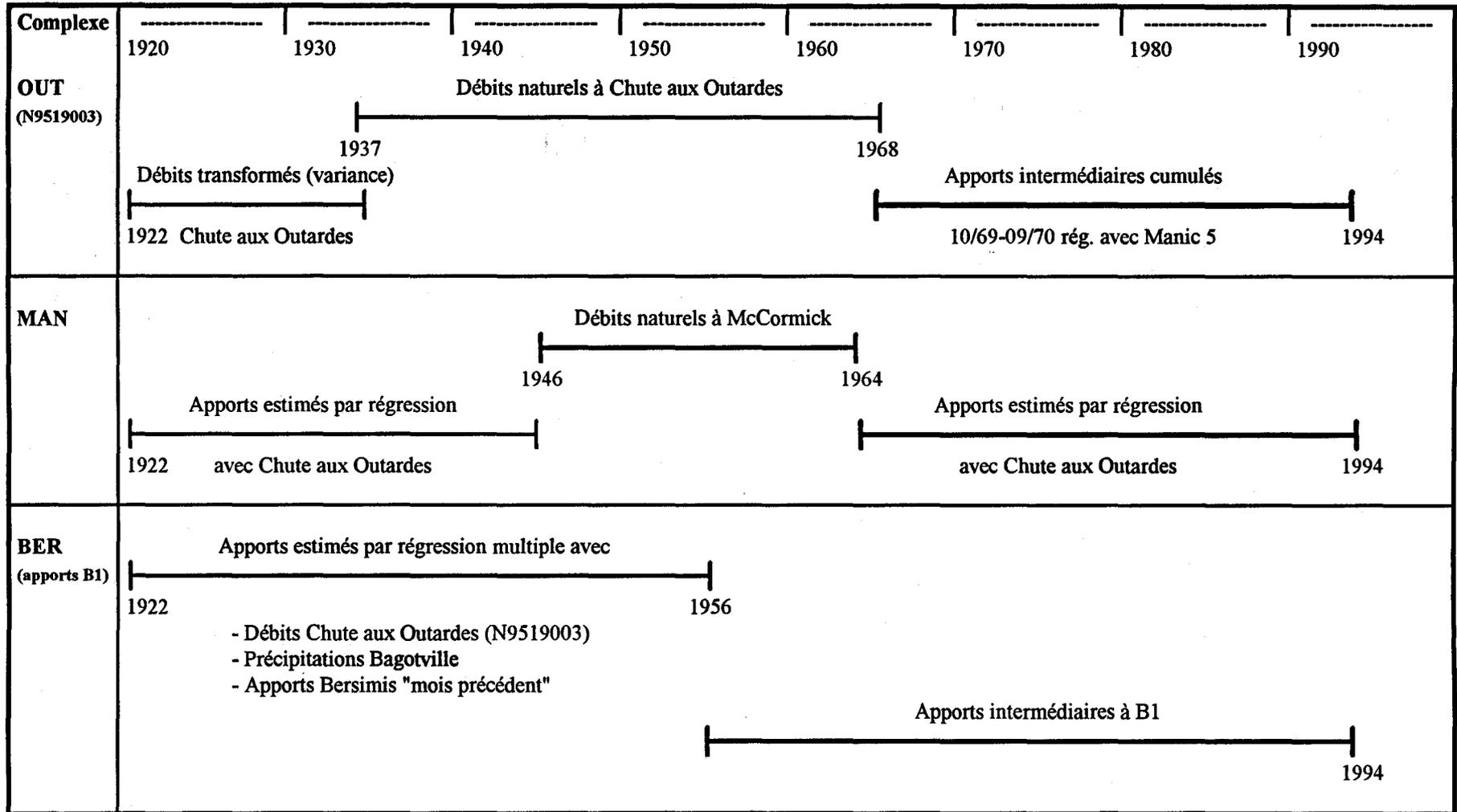


Figure 2.1. Contenu des séries de référence d'apports naturels mensuels totaux pour la procédure A.

Pour un site donné, la ou les rivières de références sont choisies en fonction de leur corrélation mois par mois avec les données d'apports naturels disponibles au site. La série utilisée pour la reconstitution peut donc être différente d'un mois à l'autre. Cette procédure est adaptée aux systèmes hydriques possédant de nombreux sous-bassins et dont la répartition des sites est plus complexe que celle des systèmes hydriques de la Côte-Nord. Elle est appliquée lorsque les séries observées aux sites présentent une qualité trop douteuse pour être utilisées telles quelles, mais suffisante pour permettre de déterminer quelle station de référence est la plus représentative.

La procédure B a été appliquée aux complexes hydroélectriques de l'Outaouais et du St-Maurice. Elle est divisée en quatre étapes :

1. Examen des données disponibles aux sites et aux rivières voisines susceptibles d'être utilisées comme ensemble de données de référence. Les séries dont les données d'apports aux sites semblent de mauvaise qualité sont rejetées (on fait référence ici à un problème de mesures erronées traité dans le cadre du projet C1 de la Chaire en hydrologie statistique).
2. Validation des données de débits naturels des rivières voisines de référence. Les tests d'homogénéité temporelle et spatiale sont effectués (procédure de Lee et Heghinian, tests de Student, de Fisher, et Kolmogorov-Smirnov, cf. sections 2.2.1 et 2.2.2).
3. Reconstitution des débits des rivières de référence (si nécessaire) par régression multiple ou par la méthode de transfert par rapports des bassins versants.
4. Calcul des apports intermédiaires à chaque site du système en transférant les débits validés de rivières de référence par la méthode du rapport de superficies de bassins versants (section 2.2.3). Les rivières de référence sont choisies à partir des corrélations mois par mois avec les données d'apports intermédiaires disponibles au site.

Les apports intermédiaires issus de la procédure B dépendent, en grande partie, de la méthode de transfert par rapport de superficies de bassins versants. Le tableau 2.3 indique (cellule ombragée du tableau) pour chaque site des complexes de l'Outaouais et du St-Maurice la ou les rivières de référence utilisées pour obtenir une série d'apports valides d'au moins cinquante années (étape 4).

Tableau 2.3. Rivières de référence utilisées pour reconstituer les apports intermédiaires de chaque site (procédure B)

		RIVIÈRES DE RÉFÉRENCE							
		VERMILLON	MATAWIN	MÉGISCANE	COULONGE	HARRICANA	KINOJEVIS	MISSISSIPI	ROUGE
ST-MAURICE									
GOU									
MANO									
MAT									
RAB									
TRE									
BEA									
LAT									
GRM									
SHA									
LAG									
GATINEAU									
CAB	Pas de transfert par rapports de bassins versants, mais plutôt les apports de la base mensuelle de la banque de données								
BAS	Pas de transfert par rapports de bassins versants, mais plutôt les apports mensuels moyens de la base journalière de la banque de données								
MANi									
PAU									
CHE									
RAF									
OUT. SUPÉ.									
DOZ									
D/R7									
R7/L15									
KIP									
TEM									
OUT. INFÉ.									
CHE/TEM									
CHF/CHE									
CAR/CHF									

Rappelons que le complexe de l'Outaouais est divisé en trois sous-bassins : Gatineau, Outaouais Supérieur et Outaouais Inférieur. On remarque que les données des réservoirs Baskatong et Cabonga n'ont pas été reconstituées. En effet, les séries disponibles à ces sites contenaient des données d'apports de bonne qualité (étape 1).

La figure 2.2 illustre, sous forme de diagramme des concomitances, le contenu des séries de référence et les interventions effectuées pour valider ces débits naturels (étape 2). On remarque que les reconstitutions de données effectuées sur les débits des rivières de référence sont peu nombreuses (étape 3).

2.2.4.3 Procédure C

La procédure C est appliquée lorsqu'une reconstitution sur une base mensuelle ne peut être effectuée. Elle consiste à construire, dans un premier temps, une série d'apports totaux annuels valides à l'embouchure du complexe que l'on désagrège ensuite dans le temps et dans l'espace. La procédure C a été appliquée au complexe de la rivière La Grande pour lequel aucune donnée d'apport intermédiaire fiable aux sites n'est disponible. Les seules données de qualité disponibles sont des débits mesurés à des rivières situées dans le bassin versant mais s'échelonnant sur une courte période de 20 années concomitantes. Ces séries courtes ont été étendues à partir de longues séries de débits mesurés pour des rivières voisines situées à l'extérieur du bassin versant. Cette extension a été effectuée à un pas de temps annuel puisque les corrélations mois par mois étaient non significatives.

Plus précisément, la procédure C est constituée des cinq étapes suivantes :

1. Validation des séries de données disponibles à l'intérieur du bassin versant : séries courtes (20 années de débits mesurés en rivières naturelles). Les tests d'homogénéité temporelle et spatiale sont effectués (Lee et Heghinian, Student, Fisher, et Kolmogorov-Smirnov) ainsi qu'une analyse de regroupement (CAH).
2. Reconstitution sur une base annuelle des apports intermédiaires à chaque site par la méthode des rapports de superficies de bassins versants avec les rivières validées à l'étape 1. On dispose alors d'apports annuels aux sites sur une courte période de 20 années.
3. Construction de la série d'apports totaux annuels à l'embouchure du système hydrique en cumulant les apports intermédiaires annuels reconstitués de chaque site.

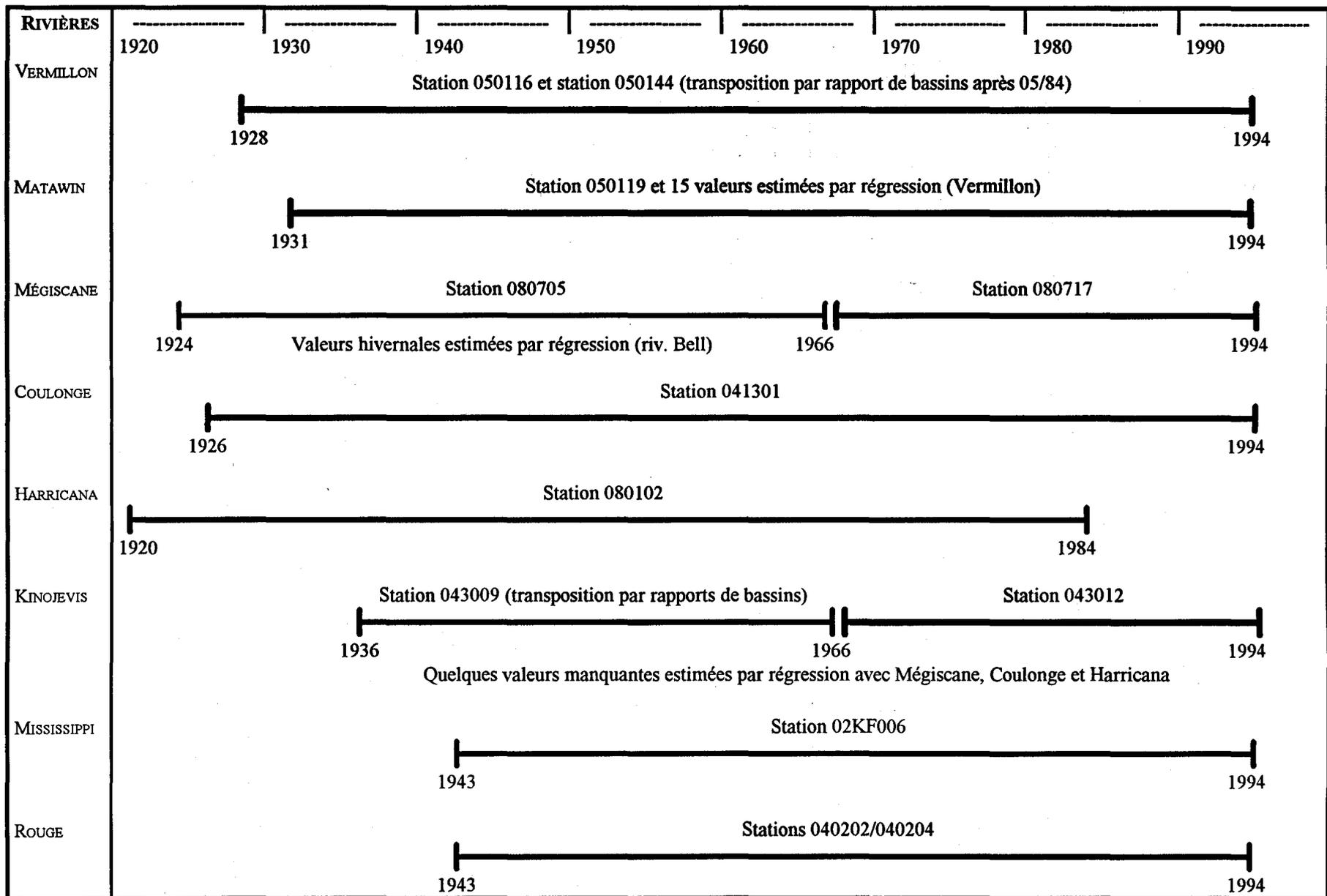


Figure 2.2. Contenu des séries de débits des rivières de référence pour la procédure B

4. Extension de la série des apports annuels cumulés par régression multiple avec les débits annuels des longues séries des rivières voisines (à l'extérieur du bassin versant) ainsi qu'avec des données de précipitations.
5. Calcul des apports intermédiaires à chaque site du système : répartition spatiale et temporelle de la longue série des apports totaux annuels. Cette opération est effectuée à l'aide du modèle de génération stochastique SPIGOT (section 2.2.3) qui permet de désagréger dans le temps et dans l'espace une série globale en conservant les principales caractéristiques statistiques des observations (moyennes, variances, corrélations et autocorrélations).

La figure 2.3 illustre, sous forme de diagramme des concomitances, le contenu de la série d'apports totaux annuels complète désagrégée pour obtenir les apports mensuels à chaque site. De 1961 à 1981, les apports annuels totaux sont obtenus en cumulant les apports intermédiaires reconstitués (étapes 2 et 3). À l'extérieur de cette période, les apports totaux sont déduits de modèles de régression multiple (étape 4).

2.2.4.4 Procédure D

La procédure D s'apparente à la procédure A en ce sens qu'elle cherche à obtenir une série d'apports totaux à l'embouchure du complexe. Cette méthode a été appliquée au complexe hydroélectrique du St-Laurent où, pour les études de planification énergétique, Hydro-Québec s'intéresse essentiellement aux apports naturels à la sortie du Lac-St-François. Cette série de données est obtenue en sommant les apports naturels calculés à Cornwall et au réservoir du lac St-François. La série à Cornwall contient des apports naturels corrigés selon le plan 1958-D de gestion des Grands Lacs (International St-Laurence Board of Control, 1963). Les données au réservoir du lac St-François ont été obtenues en grande partie à l'aide d'un transfert des débits mesurés sur la rivière St-Régis à Brasher Center (État de New-York, USGS 04269000) en utilisant la méthode du rapport des superficies de bassins versants. La figure 2.4 donne le contenu des différentes séries d'apports utilisées dans cette procédure.

2.2.5 Apports mensuels utilisés par le Service Prévisions

Les données d'apports naturels employées par le Service Prévisions correspondent aux apports harmonisés par le service Hydraulique jusqu'en 1988 (cf. figures 2.1 à 2.4). Toutefois, de 1989 à nos jours, les apports naturels mensuels utilisés diffèrent. Le service

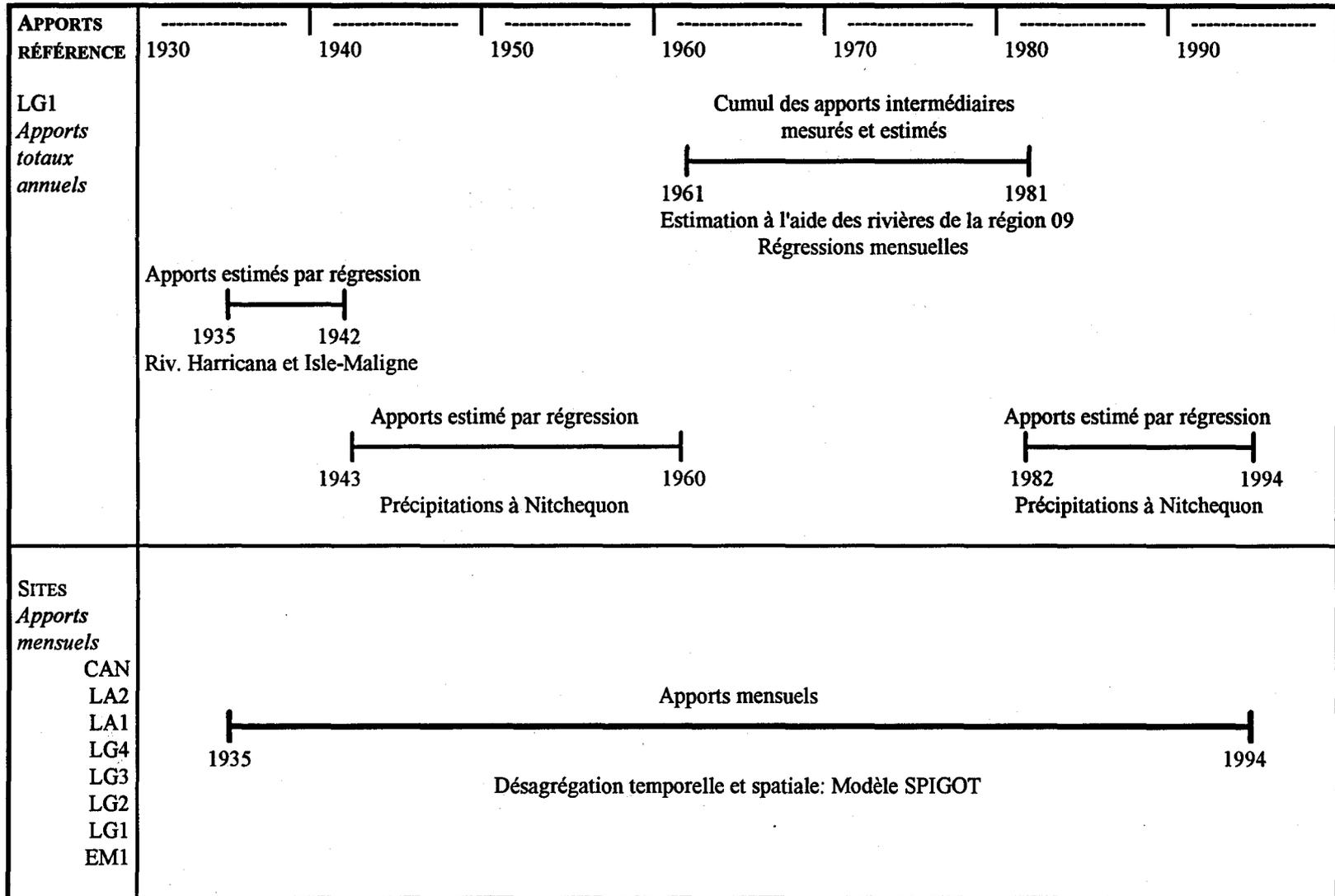


Figure 2.3. Contenu de la série de référence d'apports naturels annuels pour la procédure C et apports intermédiaires résultant.

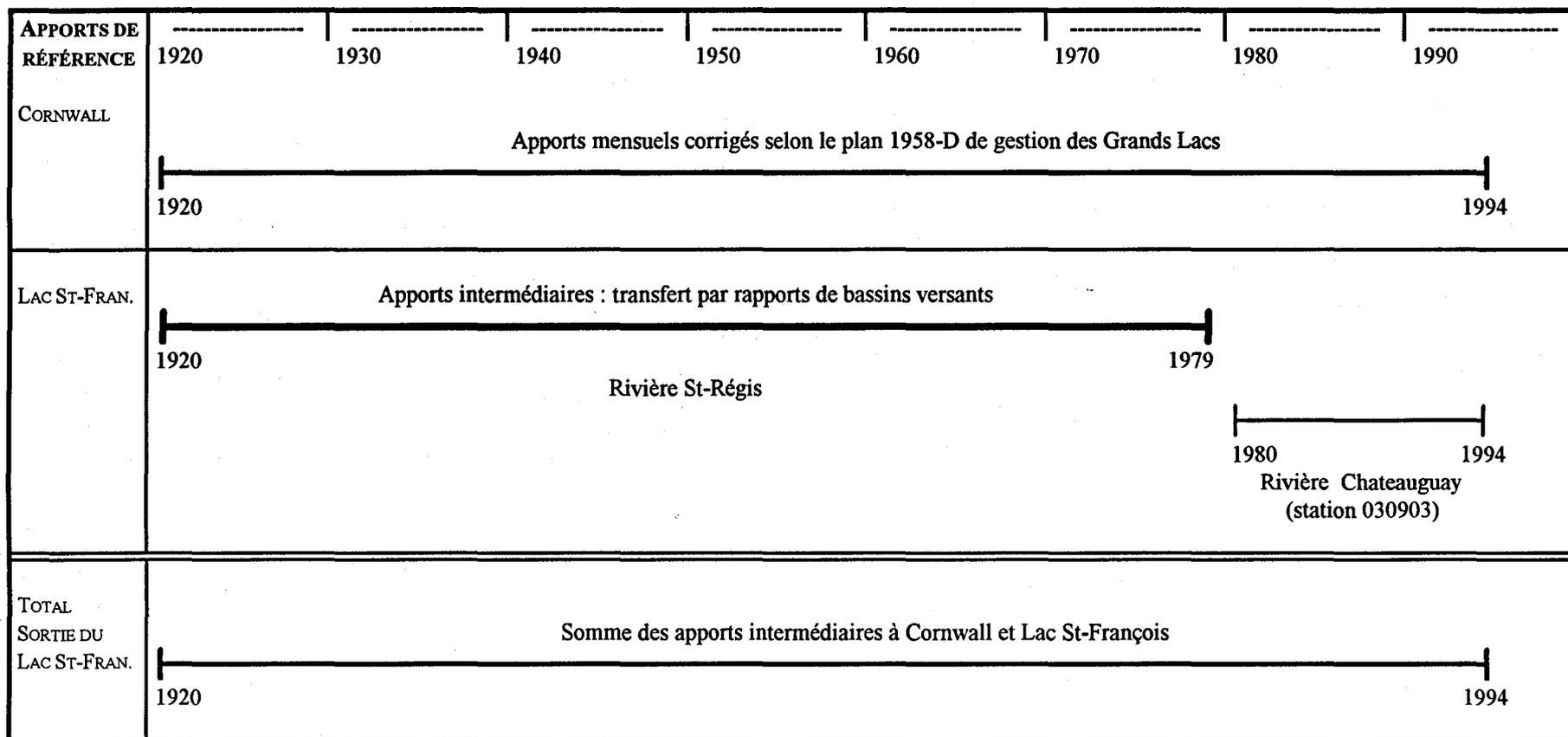


Figure 2.4. Contenu des séries de référence d'apports naturels pour la procédure D et apports reconstitués résultant.

Prévisions publie chaque mois le bilan énergétique des apports naturels et des stocks utilisés. Ces données sont comparées à la production mesurée aux centrales avant publication. Ainsi, pour éviter que les apports mensuels ne concordent pas avec la production, ce service utilise depuis 1989 les apports calculés par bilan hydrique plutôt que ceux issus des méthodes de reconstitution du service Hydraulique. On tente donc d'utiliser au maximum les mesures enregistrées aux sites. Cependant, pour certains sites particuliers, les apports ainsi obtenus peuvent être négatifs. Deux situations peuvent alors se produire :

1. les apports mensuels négatifs au site engendrent un volume annuel négatif (pour un petit réservoir, les apports mensuels peuvent être systématiquement négatifs);
2. les apports mensuels négatifs n'influencent pas significativement le volume annuel (par exemple, lorsqu'un ou deux apports mensuels seulement sont négatifs pendant l'année).

Lorsque la première situation survient, les apports mensuels sont tous reconstitués au site problématique à partir d'un site voisin du même système hydrique (ou, si disponible, d'une rivière naturelle) par la méthode du rapport de bassins versants. Si la deuxième situation se produit, une proportion des apports du même site est transférée d'un mois à l'autre tout en conservant le volume annuel. Cette nouvelle répartition des apports mensuels est obtenue en général à l'aide d'un modèle de simulation déterministe.

Le tableau 2.3 donne pour chaque site la méthode de calcul des apports naturels mensuels de 1989 à 1994 employée par le service Prévisions.

2.3 Analyse critique des méthodologies d'harmonisation des apports

Les caractéristiques statistiques issues des séries d'apports énergétiques dépendent directement de celles des séries d'apports naturels harmonisés et de la manière dont elles ont été construites. C'est pourquoi, l'un des objectifs du sous-projet 2 était l'étude critique des procédures de reconstitution et d'extension employées afin d'identifier, s'il y a lieu, les problèmes d'ordre méthodologique. Une analyse exhaustive et systématique de toutes les interventions effectuées sur les données n'a pu être réalisée étant donnée la quantité importante d'observations à traiter. L'étude de rapports internes d'Hydro-Québec ainsi que la synthèse des travaux d'harmonisation des apports naturels mensuels (section 2.2) ont

toutefois permis d'identifier quelques aspects méthodologiques problématiques communs à toutes les procédures. Les problèmes observés sont liés :

- à la méthode de transfert de données par rapport de superficies de bassins versants (RSBV);
- aux tests paramétriques de Student et de Fisher;
- à l'analyse de regroupement CAH.

Tableau 2.4. Calcul des apports naturels de 1989 à 1994 utilisés par le service Prévisions.

COMPLEXE	SITE	MÉTHODE DE CALCUL
GATINEAU SUPÉRIEUR	CAB, BAS MANI, PAU	Calculé par bilan hydrique Redistribution de la somme des apports calculés par bilan hydrique à chacun de ces deux sites
	CHE, RAF	Transfert par rapports de bassin versants (Paugan)
	DOZ D/R7	Calculé par bilan hydrique Redistribution de la somme des apports calculés par bilan hydrique à chacun des sites
	R7/L15	Redistribution de la somme des apports calculés par bilan hydrique à chacun des sites
INFÉRIEUR	KIP, TEM	Calculé par bilan hydrique
	CHE	Apport total moins débit déversé à Témiscamingue
	CHF CAR	Apport total moins débit déversé à Chenaux Apport total moins débit déversé à Carillon
ST-MAURICE	GOU, RAB, MAT MANO TRE, BEA, LAT GRM, SHA, LAG	Calculé par bilan hydrique Somme des apports inter. calculés aux sous-bassins Redistribution de la somme des apports calculés par bilan hydrique à chacun de ces trois sites Redistribution de la somme des apports calculés par bilan hydrique à chacun de ces trois sites
OUTARDES	OU4 OU3 et OU2	Calculé par bilan hydrique Transfert par rapports de bassins versants (O4)
MANICOUAGAN	MA5, MA3, MA2, LSA MA1	Calculé par bilan hydrique Transfert par rapports de bassins versants (M2)
BERSIMIS	BE1, BE2	Même méthode que le Service Hydraulique
ST-FRANÇOIS	COR, LSFI, LSFT	Même méthode que le Service Hydraulique
LA GRANDE	Tous	Apport calculé par bilan hydrique et répartition temporelle de l'eau si l'apport est négatif

La méthode de reconstitution par le rapport des superficies des bassins versants nous apparaît la plus problématique. Quelques analyses descriptives et graphiques ont été réalisées à l'aide des données de certains sites témoins afin d'étudier plus en profondeur cette méthode. Les résultats de ces analyses sont présentés à la section 2.3.1. Les sections 2.3.2 et 2.3.3 sont consacrées respectivement aux tests paramétriques et à la CAH.

2.3.1 Méthode du rapport des superficies de bassins versants

Dans le projet d'harmonisation des apports, Hydro-Québec a employé abondamment la méthode du transfert de données par rapport des superficies des bassins versants (cf. RSBV, section 2.2.4). Cette approche, utilisée pour reconstituer les apports mensuels à un site, repose sur la relation linéaire entre ces apports et ceux d'un site voisin³ (éq. 2.3). Ce transfert de données est valide si les deux sites (site d'intérêt Y et site de référence X) sont soumis à des conditions climatiques et physiographiques équivalentes quelle que soit la période de l'année. En effet, le paramètre de transfert $k = A_Y/A_X$ est constant et implique donc que la relation linéaire entre les données des deux sites est invariante dans le temps. Or, les sites utilisés dans l'équation (2.3) sont souvent assez éloignés l'un de l'autre et les conditions météorologiques peuvent alors différer considérablement, en particulier en période de crue. Il y a de fortes raisons de croire que cette approche, utilisée avec un pas de temps mensuel, engendre des biais saisonniers (biais à court-terme) pour la reconstitution d'un ensemble de données à partir d'un site éloigné. Par exemple, le transfert des apports totaux à l'embouchure des complexes Aux Outardes et Manicouagan aux réservoirs de tête OU4 et MA5 (procédure A, section 2.2.4.1) peut être problématique. L'évaluation du paramètre de transfert peut aussi être la cause d'une sur-estimation ou d'une sous-estimation systématique des apports mensuels au site d'intérêt. En effet, les superficies de bassin sont calculées approximativement et peuvent être la source d'erreurs engendrant un biais à long-terme. Enfin, la variance des données reconstituées \hat{Q}_Y est entièrement déterminée et imposée par la méthode RSBV. Elle correspond à $k^2 Var\{Q_X\}$ et est entièrement indépendante de la variance des apports historiques Q_Y du site d'intérêt Y . À notre connaissance, il n'y a aucune raison pratique ou théorique d'obtenir systématiquement une telle variance.

³ Plus d'un site voisin (site de référence) peuvent être utilisés (cf. Section 2.2.3.3).

Cinq sites témoins, qui ont fait l'objet d'une reconstitution de données par la méthode du rapport des superficies de bassins versants, ont été étudiés : OU4, MA5, LSA, MAT et GOU. Ces réservoirs ont été choisis principalement parce qu'Hydro-Québec dispose de données d'apports intermédiaires calculés de bonne qualité pour ces sites. Pour un réservoir donné Y , les analyses ont été effectuées en grande partie sur les résidus mensuels ε_Y de la méthode. Ces écarts sont définis de la façon suivante :

$$\varepsilon_Y = \hat{Q}_Y - Q_Y = \frac{A_Y}{A_X} Q_X - Q_Y \quad (2.4)$$

où A_X et A_Y désignent respectivement les superficies des bassins versants X et Y , \hat{Q}_Y la valeur prédite (éq. 2.3), Q_X l'apport du site de référence X et Q_Y l'apport calculé du réservoir d'intérêt Y . Le résidu ε_Y est l'erreur de prédiction du modèle linéaire (2.3) et on peut l'utiliser afin d'étudier la qualité de la reconstitution. Comme dans tout modèle linéaire (régression, etc.), la qualité de la reconstitution est conditionnelle au fait que les résidus soient distribués aléatoirement selon une loi normale de moyenne nulle. Tout écart par rapport à cette condition (moyenne non-nulle, cycles et tendances dans les écarts, etc.) implique que le modèle est inadéquat.

La section qui suit est entièrement consacrée aux principaux résultats des analyses effectuées sur les données du réservoir OU4, qui a été étudié de manière plus approfondie. Les conclusions étant similaires pour l'ensemble des sites soumis à une reconstitution par RSBV, nous avons regroupé dans une même section (2.3.1.2) les résultats relatifs aux réservoirs MA5, LSA, MAT et GOU.

2.3.1.1 Réservoir OU4 (complexe Aux Outardes)

Les apports naturels mensuels du réservoir OU4 sont obtenus en transférant les données d'apports harmonisés de la série N9519003 (apports totaux à la centrale OU2 sur la rivière Aux Outardes, cf. figure 2.1). La superficie du bassin versant OU4 est de 17 489 km², ce qui constitue plus de 89% ($k=0.89752$) du bassin versant total du complexe Aux Outardes. Les apports naturels journaliers de ce réservoir sont calculés par l'équation du bilan hydrique (Haché *et al.*, 1996) depuis sa mise en service en 1968 et sont, à un pas de temps mensuel, de bonne qualité selon les spécialistes du service Prévisions d'Hydro-Québec. Ainsi, l'efficacité de la reconstitution au site OU4 peut être étudiée en comparant les valeurs issues du transfert par rapport de superficies de bassins \hat{Q}_Y et celles calculées par bilan hydrique Q_Y (c'est-à-dire en analysant les résidus du modèle ε_Y , éq. (2.4)). Les écarts mensuels de 1971

à 1989 sont présentés à la figure 2.5. Pour ces années, la série de référence N9519003 contient les apports mensuels totaux, c'est-à-dire la somme des apports intermédiaires des sites OU4, OU3 et OU2 (cf. section 2.2.4.1). Pour la période considérée, le module annuel des apports naturels calculés par bilan hydrique à OU4 est de 354,15 m³/s.

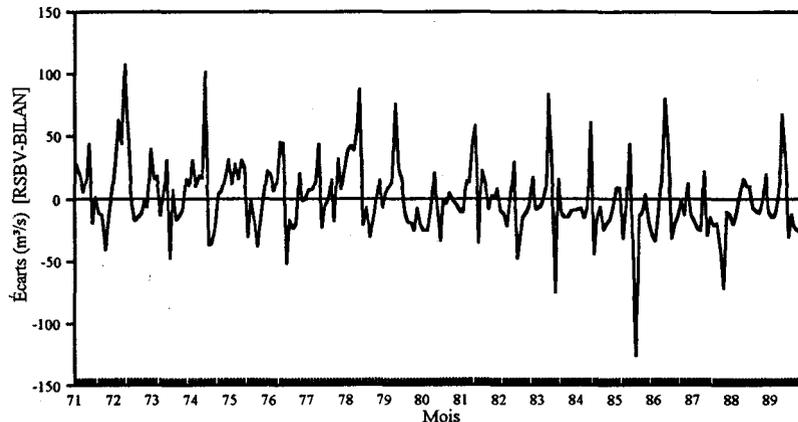


Figure 2.5. Écarts entre les apports reconstitués (RSBV) et calculés par bilan hydrique (BILAN) à OU4.

Les écarts de prédictions varient entre -125,44 et 107,10 m³/s. L'erreur moyenne sur toute la période est toutefois presque nulle (0,75 m³/s) mais l'examen de la figure 2.5 permet de constater que les écarts ne sont pas aléatoires. Premièrement, les résidus semblent changer de signe après 1980 ce qui indiquerait une coupure dans la série qui n'est probablement pas due à des variations purement aléatoires. Deuxièmement, des cycles annuels sont clairement présents de sorte que les écarts sont périodiques et autocorrélés.

Avant de présenter les résultats de l'étude critique concernant la reconstitution des apports intermédiaires de 1971 à 1989 par RSBV, nous discutons brièvement d'un problème particulier constaté lors de l'examen des écarts obtenus d'octobre 1969 à septembre 1970 et en avril 1972. En effet, comme le montre la figure 2.6, les écarts de prévisions sont très grands par rapports à ceux présentés à la figure 2.5. En particulier, l'écart observé en mai 1970 est quatre fois supérieur à l'écart maximum de prévision calculé entre 1971 et 1989.

Les apports de la série de référence correspondant à ces dates ont été calculés à l'aide d'une régression linéaire avec les apports intermédiaires de MA5 (cf. figure 2.1); cette reconstitution est clairement problématique. Un examen plus détaillé de la banque de données a permis l'identification de quelques données d'apports intermédiaires erronées à OU3 et OU4. Cela aurait justifié, lors de l'harmonisation des apports, l'emploi de la

régression plutôt que du cumul des apports intermédiaires pour calculer les données de référence. Le tableau 2.5 présente les apports intermédiaires erronés et les nouvelles valeurs après correction. On constate que celles-ci sont beaucoup plus adéquates. Nous recommandons à Hydro-Québec leur utilisation pour le calcul des apports totaux du complexe Aux Outardes à la place des valeurs obtenues par régression.

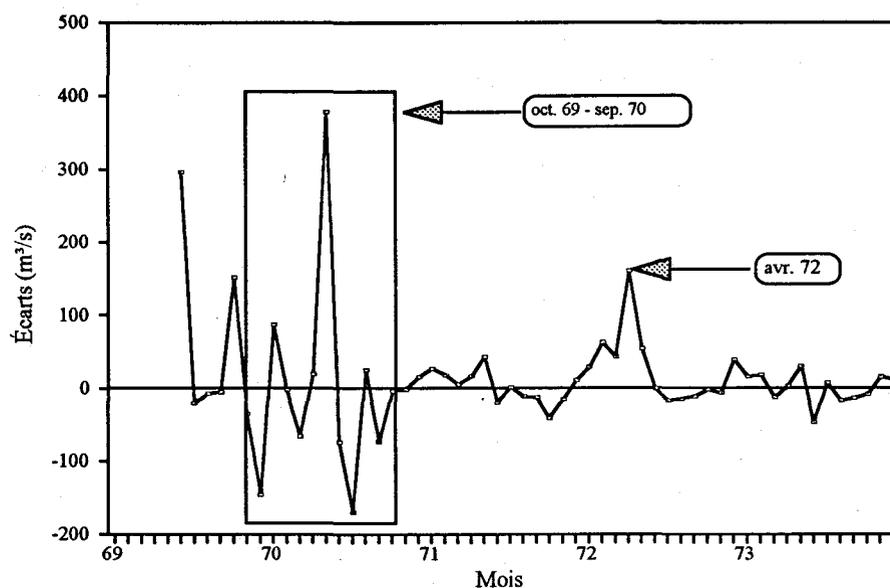


Figure 2.6. Écarts entre les apports reconstitués et calculés par bilan hydrique à OU4.

Tableau 2.5. Apports intermédiaires corrigés à OU3 et OU4.

Site	Date	Valeur erronée	Valeur corrigée
OU3	10/69	-301,1	11,3
	05/70	-3048,6	44,5
	06/70	965,0	35,0
	07/70	-1848,2	13,0
	08/70	-5057,1	18,0
OU4	04/72	-53,76	Donnée manquante

La tendance à long terme observée à la figure 2.5 est induite par un changement (en moyenne) du signe des résidus à la fin des années 1970 et indique la présence d'un biais annuel variable dans le temps : les apports reconstitués sur-estiment les apports naturels avant 1980 pour ensuite les sous-estimer. La figure 2.7 permet de mieux observer ce biais annuel.

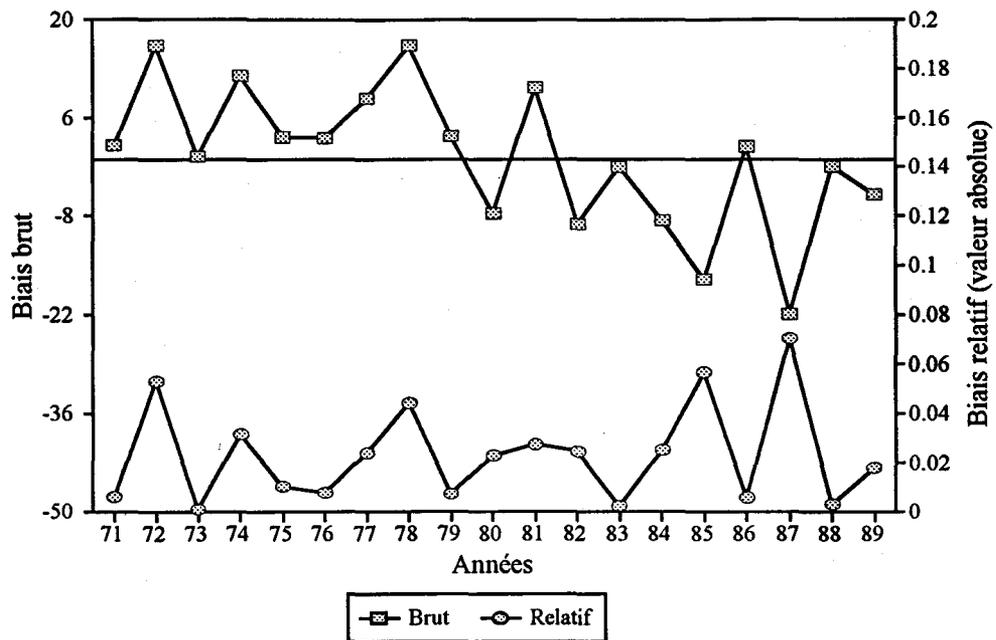


Figure 2.7. Biais annuel brut et relatif de la méthode du rapport de bassins versants à OU4.

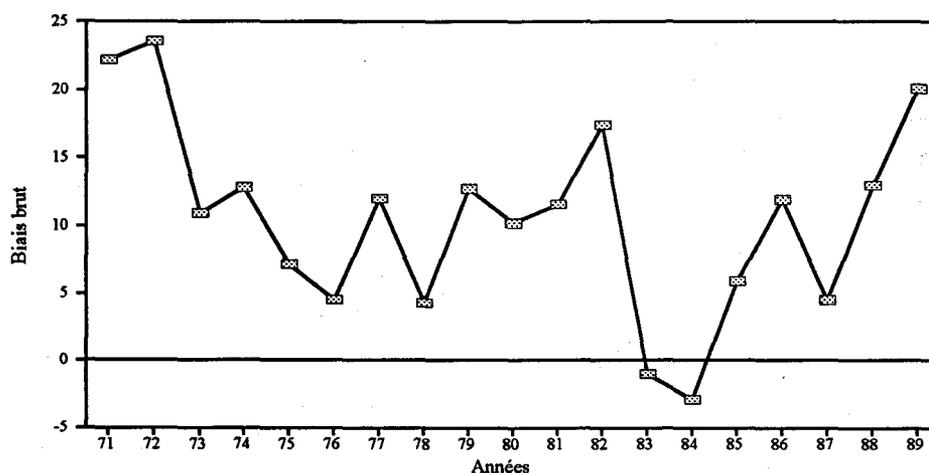
Deux courbes y sont tracées: les moyennes annuelles des écarts bruts (biais brut) afin d'illustrer la coupure dans la série et les écarts relatifs en valeur absolue (biais relatif: division par la moyenne annuelle) pour tenir compte de la différence entre les moyennes de chaque année. Le biais relatif donne une meilleure idée de l'erreur réelle de prévision. Le biais relatif annuel varie, en valeur absolue, de 0.1% en 1973 ($0,43 \text{ m}^3/\text{s}$) à 7% ($-21,84 \text{ m}^3/\text{s}$) en 1987. Le changement de signe du biais annuel est clairement illustré par la figure 2.7. Le paramètre de transfert k de l'équation (2.3) est donc trop élevé avant 1980 et trop faible après cette année. Ce type de variation temporelle du biais annuel de la méthode du rapport des superficies de bassins versants est quelque peu étonnant à première vue. Cela pourrait toutefois s'expliquer par un changement climatique ou par une transformation dans l'aménagement du complexe Aux Outardes. Cette dernière hypothèse est retenue parce que la coupure dans la série des écarts coïncide assez bien avec la mise en service du site OU2 en 1978. Le tableau 2.5 donne les principales statistiques descriptives des deux sous-échantillons avant et après la mise en service du réservoir OU2.

Tableau 2.6. Principales statistiques des résidus avant et après la mise en service de OU2.

Statistiques	Écarts de 1971-1977	Écarts de 1978-1989
Taille d'échantillon	84	144
Moyenne	7,65	-3,29
Médiane	7,39	-8,22
Minimum	-51,57	-125,44
Maximum	107,10	86,91
Écart-type	28,12	28,68

On remarque en particulier que, tout en possédant des écart-types presque identiques, les deux sous-échantillons ont des mesures de tendances centrales (moyenne et médiane) qui diffèrent considérablement. Le test non paramétrique de Wilcoxon (section 2.3.2) appliqué aux données confirme que la différence entre les moyennes des écarts (biais annuels) est significative à un niveau de signification de 1%. On peut émettre l'hypothèse que le changement dans l'aménagement du complexe Aux Outardes a influencé significativement la reconstitution des apports intermédiaires à OU4 après 1978.

Pour vérifier cette hypothèse, les apports intermédiaires à OU4 ont été reconstitués par la méthode du rapport des superficies de bassins versants en considérant, cette fois-ci, le sous-bassin formé des réservoirs OU3 et OU4. Ainsi, les apports reconstitués du site OU4 ne font pas intervenir les données observées au site OU2. La figure 2.8 présente les écarts moyens annuels obtenus (biais annuels bruts).

**Figure 2.8.** Biais annuel brut de la reconstitution des apports à OU4 à partir de la série des apports totaux à OU3.

On remarque, lorsque le site OU2 n'est pas considéré, la présence d'un biais annuel positif et relativement stable dans le temps. Ce biais est de 10 m³/s en moyenne et correspond approximativement au biais moyen observé avant la mise en service du réservoir OU2 en 1978 (7.65 m³/s, cf. tableau 2.6). L'aménagement du site OU2 a donc un effet déterminant sur le comportement à long terme des données reconstituées à OU4. À première vue, cette conclusion peut paraître étonnante étant donnée la petite taille du bassin versant OU2 (superficie de 1241 km²) par rapport à la superficie du bassin total. Toutefois, les débits de sorties à ce site contribuent de façon importante au calcul des apports totaux de la série de référence (apport à l'embouchure du complexe). En effet, étant donné que les niveaux des réservoirs OU3 et OU2 sont maintenus constants à l'échelle mensuelle, l'apport total de la série de référence, c'est-à-dire le cumul des apports intermédiaires, n'est fonction que de la variation du volume au réservoir OU4 et du débit de sortie au site OU2.

Ce type de variation du biais annuel (changement de signe), attribuable à un changement dans l'aménagement du complexe ou à une modification de la méthode de calcul des apports de la série de référence, est une particularité du complexe Aux Outardes. On ne peut pas, a priori, généraliser cette observation à l'ensemble des sites soumis à une reconstitution par rapport de superficies de bassins versants. Toutefois, lorsque l'effet de ce site OU2 est enlevé, un biais annuel stable et de signe constant dans le temps est toujours présent (figure 2.8). Un biais à long terme de signe invariable est directement lié aux hypothèses de la méthode de reconstitution et peut être probablement généralisé à l'ensemble des sites reconstitués dans des conditions similaires.

Un biais stable dans le temps s'explique en grande partie par le fait que la méthode de reconstitution ne tient pas compte de la variabilité spatiale des conditions météorologiques. En effet, la méthode du rapport des superficies de bassins versants repose sur l'hypothèse que les sites sont soumis simultanément aux mêmes conditions climatiques. Or, les conditions climatiques diffèrent considérablement au réservoir OU4 situé au nord (site d'intérêt) et à l'embouchure du complexe Aux Outardes (site de référence). En particulier, la crue printanière survient plus tôt à l'embouchure du complexe qu'au réservoir OU4. Ce décalage temporel des hydrogrammes printaniers aux deux sites (approximativement une semaine) engendre nécessairement un biais dans les apports intermédiaires mensuels reconstitués à OU4. On peut s'attendre à ce que l'écart printanier contribue considérablement au biais annuel.

La figure 2.9 montre les biais mensuels bruts et relatifs (moyennes des écarts mois par mois) calculés sur la période complète de 1971 à 1989.

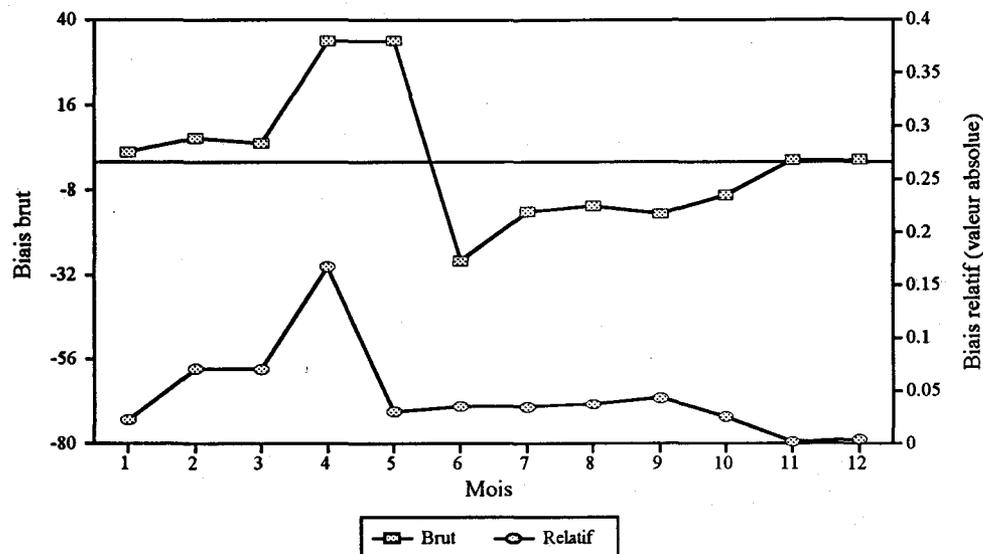


Figure 2.9. Biais mensuel brut et relatif de la méthode du rapport de bassins versants à OU4.

Le biais relatif maximal est observé en avril (17%) et coïncide avec le début de la crue printanière au sud du complexe. Ce biais mensuel, quoiqu'amointri par la prise en compte des données observées après la mise en service du site OU2, explique en grande partie la présence d'un biais annuel positif stable avant 1978. Cette sur-estimation des apports intermédiaires du mois d'avril à OU4 est suivie d'une sous-estimation de moindre importance en juin. Celle-ci correspond assez bien avec la décrue à l'embouchure du bassin versant.

Le paramètre de transfert k du modèle (2.3) est constant quelle que soit la période de l'année. Cela suppose que la relation linéaire entre les apports des deux sites est la même de saison en saison. Or, cette hypothèse du modèle de reconstitution n'est pas vérifiée dans le cas des sites étudiés ici. Les biais mensuels (figure 2.9) montrent bien l'influence qu'a le décalage temporel des hydrogrammes printaniers aux deux sites sur la qualité de la reconstitution. Le paramètre de transfert k doit traduire au mieux la relation qui existe entre les deux séries d'apports naturels quelle que soit la période de l'année de façon à tenir compte de la variabilité spatiale des conditions hydrométéorologiques entre les sites. Les rapports des moyennes mensuelles, présentés à la figure 2.10, montrent bien la nécessité d'utiliser un paramètre de transfert variable dans le temps.

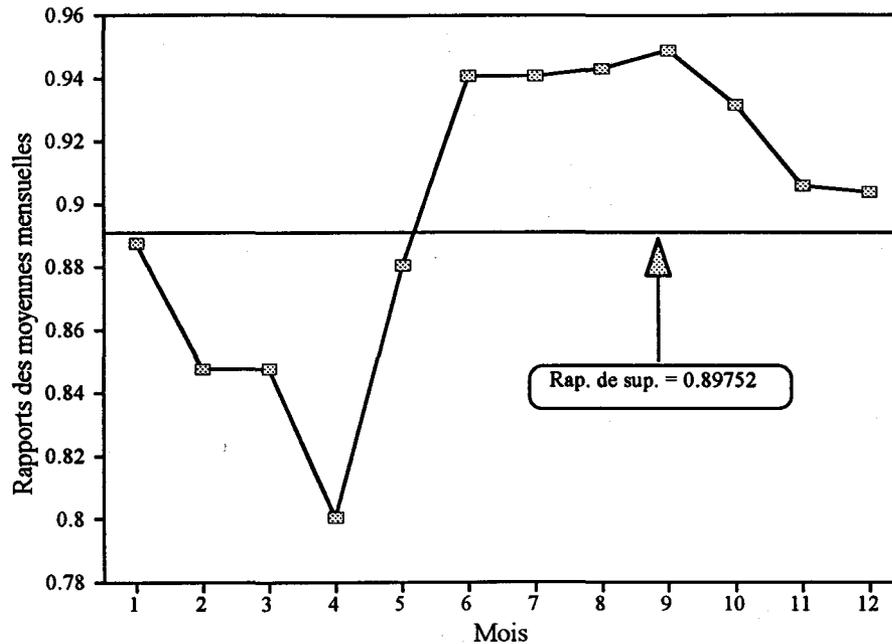


Figure 2.10. Rapports de moyennes mensuelles d'apports au site OU4 et à l'embouchure du complexe (OU4/App. Total).

En effet, on déduit de cette figure que la relation linéaire (hypothèse de la méthode RSBV) entre les apports des deux sites varie dans le temps. Les rapports de moyennes mensuelles sont de plus conformes aux observations tirées de la figure 2.9. En particulier, un paramètre de transfert k plus faible doit être utilisé au début de la crue (mois d'avril) et plus élevé à la décrue (mois de juin) du site de référence.

En plus des biais dans les valeurs reconstituées, une certaine périodicité dans la série des résidus mensuels a été observée à la figure 2.5. Cette périodicité indique que les résidus sont autocorrélés et donc non-aléatoires. Pour étudier de manière plus approfondie ce comportement saisonnier des résidus et pour vérifier si la périodicité est significative, la corrélation des données de deux mois consécutifs a été calculée (autocorrélation d'ordre 1). Les résultats sont présentés sous forme de diagramme en bâtons à la figure 2.11. Chaque bâton correspond à la corrélation entre les résidus de deux mois consécutifs calculées sur la période 1971-1989. Les corrélations supérieures à 0.4 sont significatives à 5%.

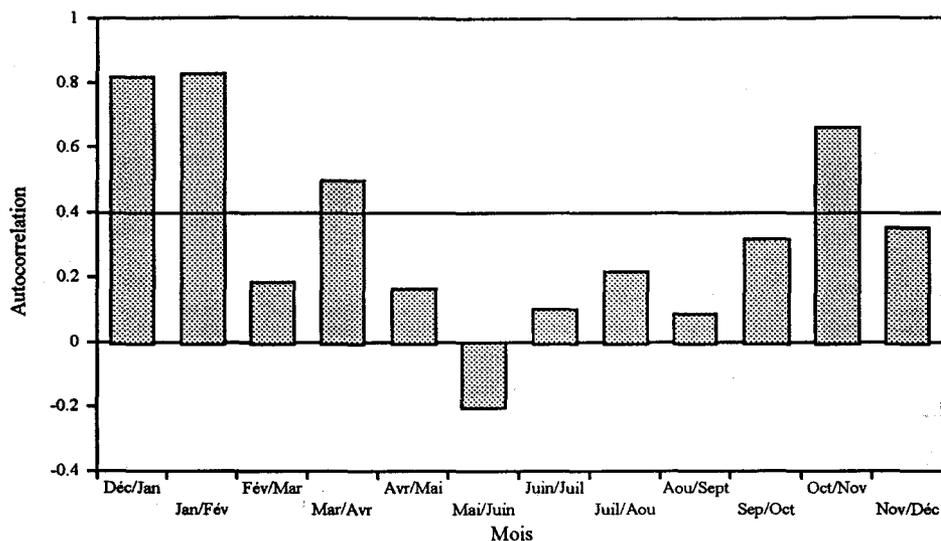


Figure 2.11. Périodogramme: corrélations 2 mois consécutifs des résidus de la méthode RSBV.

Même si la taille d'échantillon est restreinte (19 années), la corrélation des résidus de deux mois consécutifs est significativement différente de zéro au niveau de 5% pour quatre paires: (décembre, janvier), (janvier, février), (mars, avril), (octobre, novembre). Ceci montre, par exemple, que les résidus du mois de janvier dépendent fortement de ceux du mois de décembre.

En résumé, les analyses réalisées et présentées dans cette section ont d'abord permis d'identifier deux problèmes particuliers au complexe Aux Outardes liés à la reconstitution de la série de référence N9519003 (apports totaux) qui sont dus :

1. à la reconstitution des apports mensuels d'octobre 1969 à septembre 1970 et d'avril 1972 à l'aide d'un modèle de régression linéaire avec les apports intermédiaires à MA5;
2. au cumul des apports intermédiaires suite à la mise en service du réservoir OU2 en 1978.

Ces problèmes ont une influence directe sur la qualité de la reconstitution des apports à OU4. La régression linéaire avec MA5 engendre des écarts de prévision très élevés. Ce problème peut être évité en corrigeant les valeurs erronées qui ont motivé cette intervention (tableau 2.4). Les valeurs de référence obtenues ensuite en cumulant les apports intermédiaires corrigés permettent de diminuer substantiellement les écarts. Le calcul des apports de référence par le cumul des apports intermédiaires est toutefois problématique après la mise en service du réservoir OU2. En effet, les débits de sorties à ce site, qui

n'intervenaient pas dans le calcul avant 1978, influencent grandement les apports totaux et produisent un changement de signe du biais des apports reconstitués à OU4. Deux solutions peuvent être envisagées pour corriger cette situation :

- valider les débits de sortie à OU2 et utiliser toujours le cumul des apports comme série de référence;
- employer plutôt les apports intermédiaires à OU4 comme valeurs de référence à partir de la date de mise en service de ce réservoir.

Outre ces deux aspects problématiques propres au complexe Aux Outardes, des problèmes plus généraux ont été mis en évidence:

1. présence d'un biais annuel systématique. Dans le cas d'OU4, ce biais est positif, du moins avant la mise en service du site OU2 ou lorsque ce réservoir n'est pas considéré dans le cumul (figures 2.7 et 2.8);
2. présence d'une saisonnalité dans les résidus (biais saisonnier). Ce biais est particulièrement important en période de crue (figure 2.9);
3. périodicité des résidus (figure 2.11).

Ces trois problèmes sont directement liés à la méthode du rapport des superficies de bassins et sont probablement présents dans le cas des autres sites soumis à ce type de procédure. Toutefois, on peut s'attendre à ce que les biais annuel et saisonnier ne soient pas nécessairement de même signe et de même amplitude. La périodicité peut aussi être différente. En effet, ces problèmes sont dus en particulier aux différences dans les conditions hydrométéorologiques entre les sites intervenant dans le modèle (2.3).

Les résultats des principales analyses réalisées sur les écarts de prévision des apports reconstitués par RSBV pour les autres sites considérés sont présentés dans la section qui suit.

2.3.1.2 Réservoirs MA5, LSA, MAT et GOU

Les apports intermédiaires mensuels des sites MA5, LSA, MAT et GOU sont obtenus, au moins sur une portion des cinquante années, à partir de la méthode du rapport de superficies de bassins versants (figures 2.1 et 2.2, tableau 2.2). Pour MA5 et LSA, les données de la série de référence sont les apports totaux du complexe Manicouagan, tandis que pour les

sites MAT et GOU, la série de référence contient des débits mesurés en rivière naturelle. Le tableau 2.7 donne les principales caractéristiques de ces sites.

Tableau 2.7. Principales caractéristiques des sites MA5, LSA, MAT et GOU

Complexe	Sites	Superficie de bassin (km ²)	Période étudiée	Module annuel (m ³ /s)	Série de référence	Paramètre de transfert
Manicouagan	Manic 5	28 823	1967-1989	618.53	Apports totaux	0.6506
	Lac Ste-Anne	8 036	1967-1989	186.52	Apports totaux	0.1759
St-Maurice	Matawin	4 053	1963-1987	74.15	Rivière Matawin	2.922
	Gouin	8 779	1963-1987	176.48	Rivière Mégiscane	1.056

Les figures 2.12 à 2.15 donnent, pour les quatre sites, les biais annuels, les biais mensuels, les rapports de moyennes mensuelles d'apports (site d'intérêt divisé par site de référence) et les corrélations 2 mois consécutifs respectivement. Le tableau 2.8 résume les principales observations que l'on peut tirer de ces graphiques.

Tableau 2.8. Principales observations tirées des figures 2.12 à 2.15.

	SITES			
	MA5	LSA	MAT	GOU
Situation géograph. p/r au site de référence (approx.)	162 km au nord	100km au nord-est	20 km au nord-est	? km au sud-est
Biais ann. brut (figure 2.12)	> 0 en général	< 0 en général	< 0 avant 1978 > 0 après 1978	< 0 en général s'accroît (1978 ...)
Biais ann. relatif (figure 2.12)	[0%, 28%] variable	[0%, 24%] peu variable	[0%, 22%] variable	[0%, 20%] variable
Biais mens. brut (figure 2.13)	< 0, 02-05, 09 > 0, 01, 06-07, 10-12	< 0, 02-06, 09-11 > 0, 01, 07-08, 12	< 0, 03-07 > 0, 01-02 et 08-12	< 0, 01-05 > 0, 06-12
Biais mens. relatif (figure 2.13)	Max: 06 (32%)	Max: 03 (40%)	Max: 01, 12 (40%)	Max: 03, 04 (40%)
Rapp. des modules (figure 2.14)	> rap. de superf. : 02-05, 09 < rap. de superf. : 01, 06-08, 10-12	> rap. de superf. : 02-06, 09-11 < rap. de superf. : 01, 07-08, 12	> rap. de superf. : 03-07 < rap. de superf. : 01-02, 08-12	> rap. de superf. : 01-05, 08 < rap. de superf. : 06-07, 09-12
Périodicité (figure 2.15)	Corr. signif. à 5%: 02/01, 03/02, 08/07, 09/08, 10/09	Corr. signif. à 5%: 10/09	Corr. signif. à 5%: 01/12, 02/01, 03/02, 06/05, 07/06, 08/07, 09/08, 11/10	Corr. signif. à 5%: 05/04, 08/07, 10/09

* Le symbole 0X indique le mois (01: janvier, 02: février, ...).

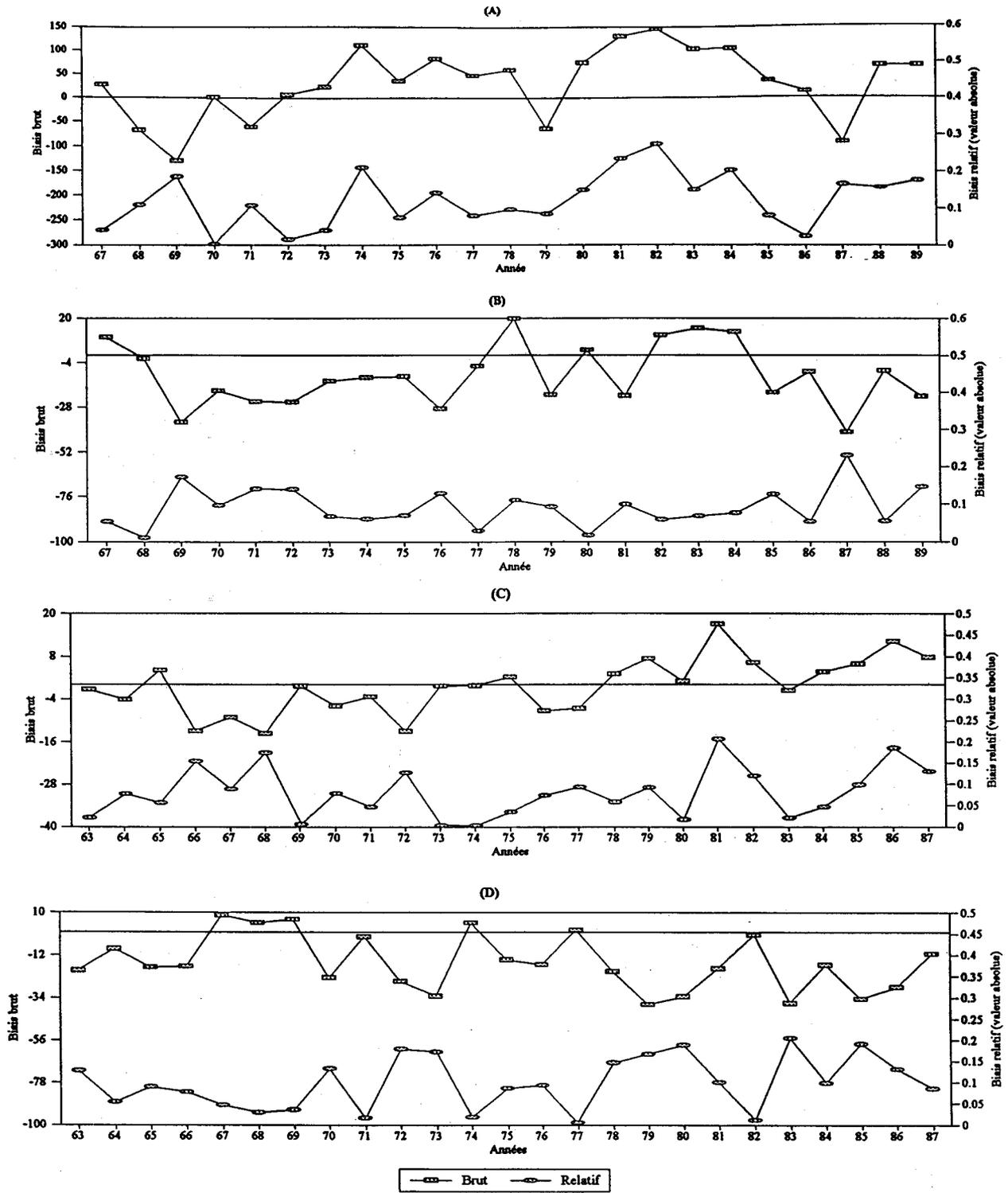


Figure 2.12. Biais annuel brut et relatif de la méthode du rapport de bassins versants : (A) Réservoir MA5, (B) Réservoir LSA, (C) Réservoir MAT, (D) Réservoir Gouin

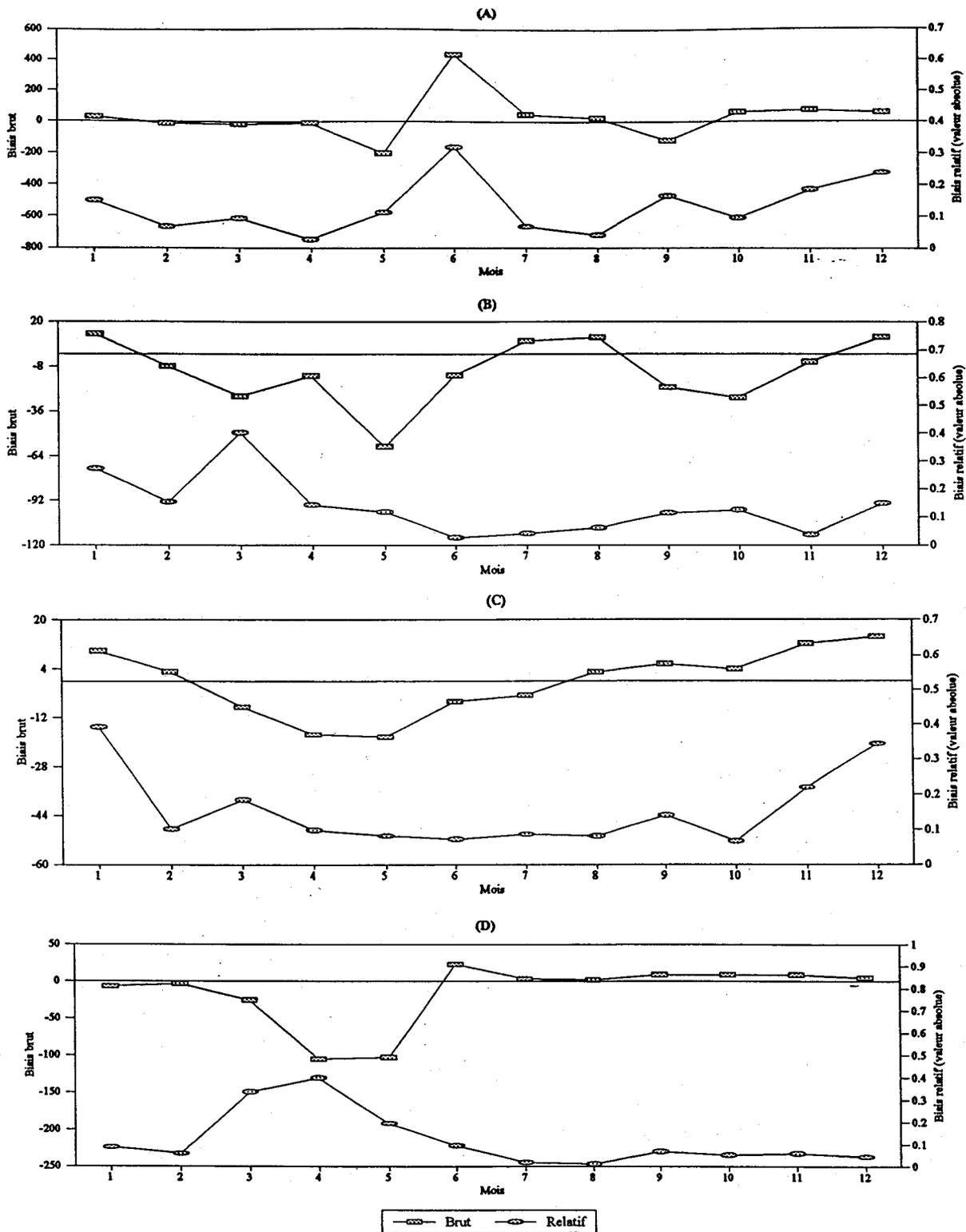


Figure 2.13. Biais mensuel brut et relatif de la méthode du rapport de bassins versants :
 (A) Réservoir MA5, (B) Réservoir LSA, (C) Réservoir MAT, (D) Réservoir Guoin

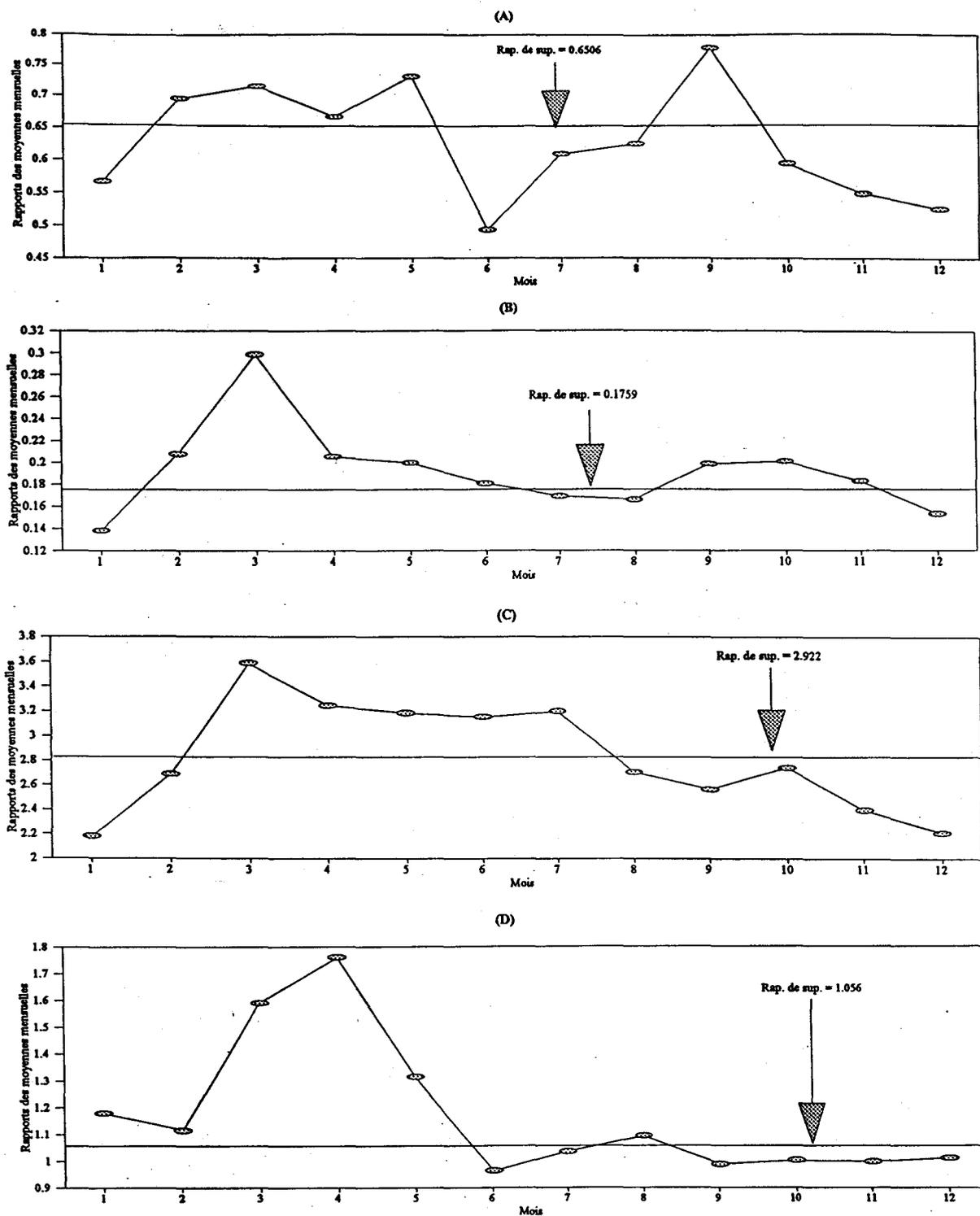


Figure 2.14. Rapports de moyennes mensuelles d'apports (site d'intérêt/site de référence) :
 (A) Réservoir MA5, (B) Réservoir LSA, (C) Réservoir MAT, (D) Réservoir Gouin

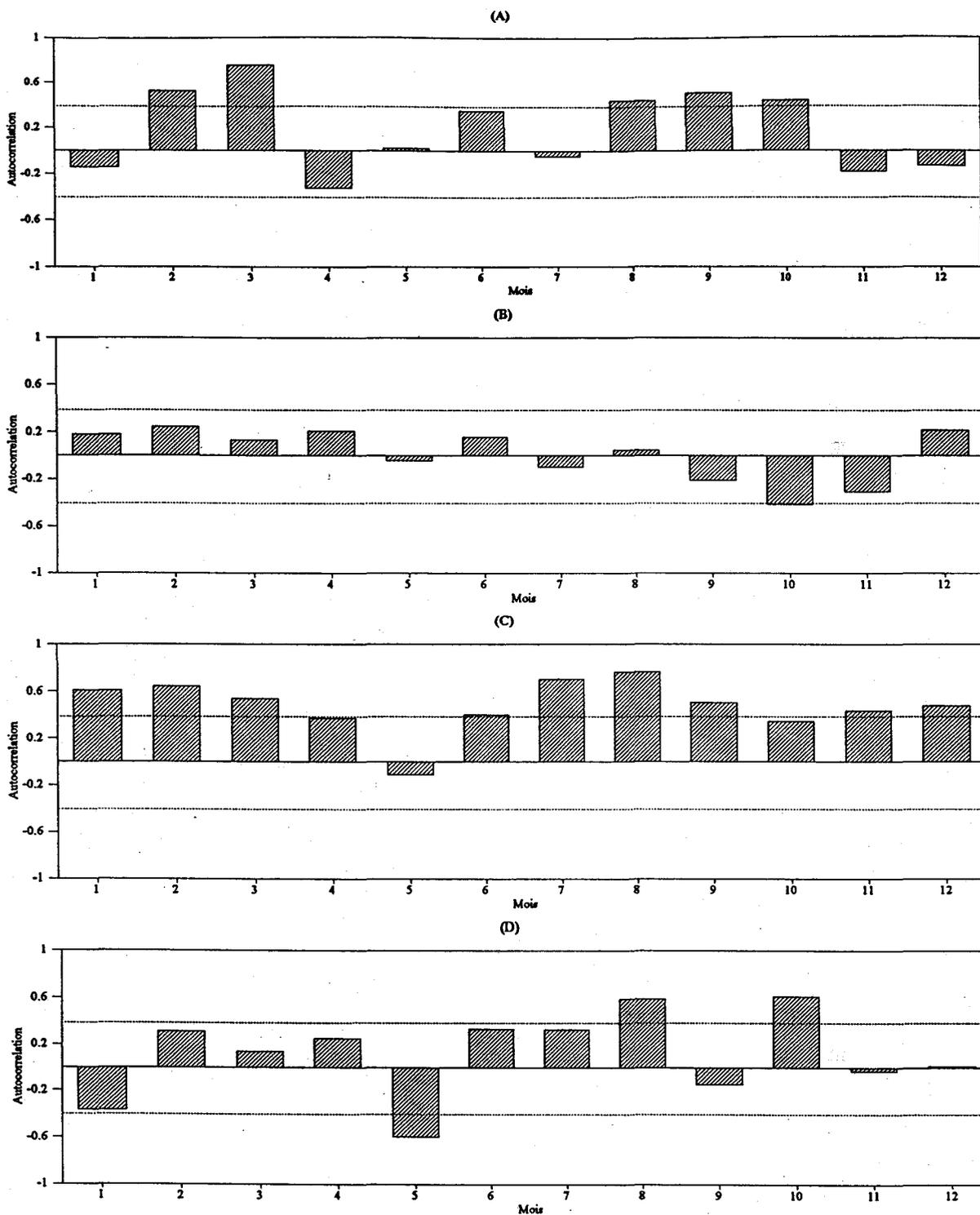


Figure 2.15. Corrélations 2 mois consécutifs des résidus de la méthode RSBV: (A) Réservoir MA5, (B) Réservoir LSA, (C) Réservoir MAT, (D) Réservoir Gouin; 1 = déc.-janv., 2 = janv.-fév., ...

Les figures 2.12 et 2.13 montrent, pour l'ensemble des sites considérés, que les estimations des apports par la méthode RSBV sont biaisées annuellement et mensuellement. Pour les sites MA5, LSA et GOU, le biais annuel est relativement stable d'une année à l'autre. Toutefois, la figure 2.12(C) illustre clairement, pour le site MAT, la présence d'une tendance dans le biais annuel qui s'apparente à ce qui a été observé au site OU4 (section 2.3.1.1). En effet, le biais change de signe en 1978. Ce site se distingue aussi en ce qui a trait au biais mensuel relatif où le maximum est observé en hiver (décembre et janvier) alors que pour les autres sites le maximum survient plutôt au printemps. Enfin, les corrélations deux mois consécutifs présentées à la figure 2.15 montrent la présence d'une périodicité dans les résidus de la méthode pour les sites MA5, MAT et GOU. Remarquons que les corrélations significatives sont observées généralement en saison hors-crue et qu'elles sont positives. Cependant, le site LSA ne possède qu'une seule corrélation significativement différente de zéro et celle-ci est négative (septembre/octobre).

Les résultats de l'analyse des biais annuels et mensuels ainsi que de la périodicité des résidus pour les sites MA5, LSA, MAT et GOU confirment ceux obtenus pour le site OU4 et illustrent bien les problèmes liés à l'estimation par la méthode RSBV:

1. présence d'un biais annuel systématique négatif ou positif. Ce biais peut être dû à une mauvaise estimation des superficies de bassins versants. Dans le cas du site MAT, ce biais change de signe;
2. présence d'un biais saisonnier. Ce biais est probablement dû au fait que la quantité de précipitation est différente au site d'intérêt et au site de référence. Le décalage temporel (en crue particulièrement) peut aussi expliquer ce problème.;
3. périodicité des résidus.

Outre ces trois problèmes, rappelons que la variance des données reconstituées \hat{Q}_Y est entièrement déterminée et imposée par la méthode RSBV. Elle correspond à $k^2 Var\{Q_X\}$ et est entièrement indépendante de la variance des apports historiques Q_Y du site d'intérêt Y . Ceci constitue un problème supplémentaire puisque, comme il a été mentionné à la section plus haut, il n'y a aucune raison pratique ou théorique d'obtenir systématiquement une telle variance.

Quoique certaines similarités puissent être observées quant au comportement des biais et des corrélations de deux mois consécutifs, aucune structure systématique commun n'a été identifiée pour les sites étudiés. Cela n'est pas vraiment étonnant. En effet, si la variation

spatiale des paramètres hydrométéorologiques entre les séries de référence (origine du transfert) et les sites d'intérêt (à reconstituer) diffère, le biais de la méthode de transfert, par exemple, n'aura pas la même forme d'un site à l'autre.

2.3.1.3 Recommandations et alternatives possibles

Compte tenu des contraintes propres à chaque réservoir (disponibilité de données, etc.), il est difficile d'identifier une approche alternative applicable à l'ensemble des sites qui ont été soumis à une reconstitution par rapport de superficies de bassins versants. Toutefois, la présente étude critique de l'harmonisation des apports nous a permis d'identifier différentes avenues possibles qui peuvent corriger certains problèmes inhérents à la méthode RSBV. Quelques recommandations sont proposées dans ce qui suit :

- Si des données d'apports naturels de bonne qualité sont disponibles pour une certaine période au site d'intérêt, on peut envisager l'utilisation d'un rapport d'apports saisonnier ou mensuel comme paramètre de transfert. La détermination des apports intermédiaires serait alors effectuée à l'aide d'un paramètre de transfert variable avec les mois ou les saisons. Cette approche diminuerait probablement le biais annuel et certainement la périodicité des résidus. Cependant, la variance des estimations serait toujours entièrement déterminée et indépendante de la variance des apports historiques Q_Y du site d'intérêt Y .
- Dans la même situation, on peut envisager d'utiliser des régressions linéaires mois par mois afin de déterminer les relations mensuelles entre les deux sites. La détermination des apports intermédiaires serait alors effectuée à l'aide d'un paramètre de transfert variable avec les mois (pente de la régression). Les prévisions peuvent être effectuées en conservant la variance des apports de la série historique éliminant ainsi le problème mentionné plus haut.
- Pour éviter le transfert par la méthode RSBV de l'embouchure d'un complexe aux différents sites (Aux Outardes, Manicouagan et Bersimis), il serait possible de répartir les apports totaux (série de référence) à l'aide du modèle de désagrégation SPIGOT. Ce modèle, a l'avantage de conserver les principales caractéristiques des séries d'apports intermédiaires historiques. Ici encore, cette approche repose sur la disponibilité de données de bonne qualité au site.
- Si le modèle SPIGOT n'est pas considéré pour la répartition des apports totaux aux sites d'un complexe, nous recommandons d'utiliser comme série de référence (si possible) les

apports intermédiaires du réservoir de tête qui, en général, sont de bonne qualité. Cela évitera une sur-paramétrisation dû au fait que les apports totaux sont, pour certaines périodes, le résultat d'une première modélisation (cf. complexe Manicouagan). Des estimations à la chaîne peuvent entraîner une augmentation de la variance des prévisions.

2.3.2 Tests statistiques utilisés pour l'harmonisation des apports

Les tests de Student et de Fisher ont été utilisés, lors de l'harmonisation des apports naturels, pour comparer respectivement les moyennes et les variances dans trois situations (cf. sections 2.2.1 et 2.2.2) :

- pour comparer deux sous-échantillons d'une même série annuelle d'apports mensuels d'un mois donné, d'apports maximums annuels, d'apports minimums annuels ou d'apports moyens annuels lorsque la date du changement est connue a priori (homogénéité temporelle, section 2.2.1);
- pour comparer deux sous-échantillons d'une même série annuelle d'apports mensuels d'un mois donné, d'apports maximums annuels, d'apports minimums annuels ou d'apports moyens annuels lorsque la date de changement est inconnue (homogénéité temporelle, section 2.2.1). Les tests sont alors effectués suite à l'application préalable de la procédure de Lee et Heghinian (1977);
- pour comparer deux séries d'apports à des sites différents (homogénéité spatiale ou sélection entre deux séries, section 2.2.2). Dans ce cas, les tests de Student et de Fisher sont appliqués systématiquement aux échantillons annuels d'apports mensuels standardisés pour chaque mois (24 tests statistiques) ainsi qu'aux séries d'apports maximums annuels, minimums annuels et moyens annuels standardisés.

2.3.2.1 Homogénéité temporelle: date de changement connue

Les recommandations liées à la première situation concernent particulièrement les hypothèses de ces tests statistiques. En effet, les tests de Student et de Fisher sont paramétriques et reposent principalement sur deux hypothèses mathématiques:

1. les observations sont indépendantes;
2. les observations sont distribuées selon une loi normale.

Il est raisonnable de supposer que les données soient indépendantes puisque les tests de Student et Fisher sont effectués sur des séries annuelles (mois par mois, par exemple). Toutefois, l'hypothèse de normalité peut, dans certains cas, être problématique. En effet,

rien ne permet a priori d'affirmer que les apports mensuels, maximums, minimums et moyens sont distribués systématiquement selon une loi normale.

À titre d'exemple, considérons la situation hypothétique où l'appareil de mesure aurait été déplacé ou modifié au cours de l'année 1983 au complexe Aux Outardes. Les caractéristiques statistiques de la série annuelle des apports totaux risquent d'être différentes avant et après 1983. Pour vérifier si ce changement de moyenne est significatif et si la variance est influencée, les tests de Student et de Fisher seraient utilisés mois par mois (séries annuelles), lors de l'harmonisation des apports naturels, afin de comparer ces caractéristiques statistiques avant et après la date du changement hypothétique de la moyenne. Or, la figure 2.16, où sont tracées les valeurs des deux sous-échantillons sur un papier de probabilité normal (moi d'avril), indiquent bien que ces observations ne sont pas distribuées selon une loi normale (la répartition des points sur le papier normal n'est pas linéaire).

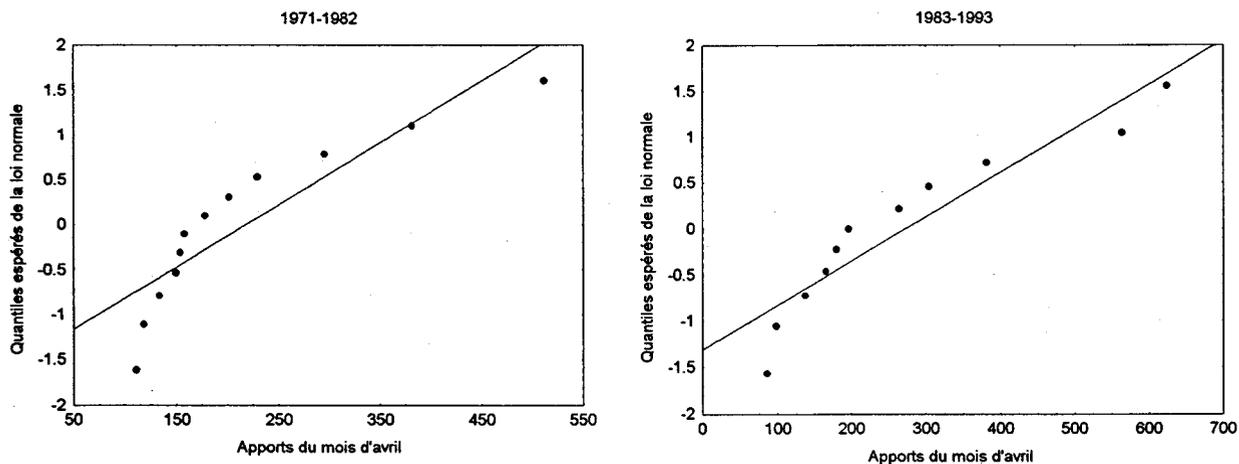


Figure 2.16. Apports totaux du mois d'avril au complexe Aux Outardes avant et après 1983 tracés sur un papier de probabilité normal.

Lorsque les observations ne suivent pas une loi normale, la puissance des tests de Student et de Fisher diminue et ils ont alors tendance à conduire à l'acceptation de l'hypothèse nulle (égalité des moyennes ou des variances) alors qu'elle devrait être rejetée (erreur de type 2). Pour pallier ce problème et pour éviter d'effectuer systématiquement un test de normalité avant l'application des tests de Student et de Fisher (tests qui sont les plus puissants lorsque les observations sont normales), il est recommandé de valider la conclusion de ces tests en utilisant en parallèle des tests qui ne sont pas sensibles à l'hypothèse de normalité. Les deux tests statistiques proposés ici sont le test de Wilcoxon (1945) et le test de Levene (1960).

Ces tests non paramétriques ne reposent pas sur l'hypothèse de normalité des observations. Ils sont donc plus puissants que les tests paramétriques (Student et Fisher) lorsque les observations ne suivent pas une loi normale.

Test de Wilcoxon (égalité des moyennes)

Considérons deux échantillons (ou sous-échantillons) à comparer de taille n_1 et n_2 respectivement : $y_{11}, y_{12}, \dots, y_{1n_1}$ et $y_{21}, y_{22}, \dots, y_{2n_2}$. Supposons de plus qu'ils proviennent respectivement de deux populations de moyenne et de variance (μ_1, σ_1^2) et (μ_2, σ_2^2) . Le test de Wilcoxon, tout comme celui de Student, examine l'hypothèse

$$H_0 : \mu_1 = \mu_2 \quad \text{contre} \quad H_1 : \mu_1 \neq \mu_2$$

La statistique utilisée, qui est asymptotiquement distribuée selon une loi normale centrée-réduite, est donnée par :

$$W = \frac{V - \frac{n_2(n+1)}{2} + \frac{1}{2}}{\sqrt{\text{Var}\{V\}}} \quad (2.5)$$

où :

- $V = \sum_{i=1}^{n_1+n_2} R_i s(R_i)$
- R_i est le rang correspondant à l'observation i de l'échantillon combiné de taille $n_1 + n_2 = n$ classé en ordre croissant
- $s(R_i) = \begin{cases} 0 & \text{si } R_i \text{ correspond à une donnée de l'échantillon des } y_2 \\ 1 & \text{si } R_i \text{ correspond à une donnée de l'échantillon des } y_1 \end{cases}$
- $\text{Var}\{V\} = \frac{n_1 n_2 (n+1)}{12} - \frac{n_1 n_2 \sum_{k=1}^h d_k^3 - d_k}{12n(n-1)}$, en supposant que nous avons dans l'échantillon combiné h groupes distincts contenant des observations égales, et que le nombre d'observations égales dans chacun de ces groupes soient respectivement d_1, d_2, \dots, d_h . Si toutes les observations sont distinctes, on a $h = n$, $d_k^3 - d_k = 0$ et donc :

$$\text{Var}\{V\} = \frac{n_1 n_2 (n+1)}{12}$$

La règle de décision pour effectuer ce test à un niveau de signification donné α est la suivante :

- Si $|W| > z_{\alpha/2}$, on rejette H_0
- Sinon, on ne rejette pas H_0

où $z_{\alpha/2}$ est le quantile de probabilité au dépassement $\alpha/2$ de la loi normale centrée-réduite. Un test tout à fait équivalent au test de Wilcoxon est celui de Mann-Whitney. Lehmann (1975, Chap. 1) donne les détails théoriques concernant ces deux tests.

Test de Levene (égalité des variances)

Le test de Levene (1960) examine l'hypothèse d'égalité des variances des deux échantillons, c'est-à-dire :

$$H_0 : \sigma_1^2 = \sigma_2^2 \quad \text{contre} \quad H_1 : \sigma_1^2 \neq \sigma_2^2$$

Conover *et al.* (1981) ont montré, à l'aide de simulations, que ce test est le plus efficace parmi 50 procédures considérées pour comparer les variances. Pour effectuer ce test, on calcule d'abord pour chaque échantillon (ou sous-échantillon) de taille n_i les écarts en valeurs absolues des observations par rapport à la médiane :

$$EC_{ij} = |y_{ij} - Med_i|, \quad i = 1, 2 \quad \text{et} \quad j = 1, \dots, n_i$$

La statistique du test de Levene (1960) est alors donnée par :

$$L = \frac{(n_1 + n_2 - 2) \left[n_1 (\overline{EC}_1 - \overline{EC})^2 + n_2 (\overline{EC}_2 - \overline{EC})^2 \right]}{\sum_{j=1}^{n_1} (EC_{1j} - \overline{EC}_1)^2 + \sum_{j=1}^{n_2} (EC_{2j} - \overline{EC}_2)^2} \quad (2.6)$$

où \overline{EC}_1 , \overline{EC}_2 et \overline{EC} sont respectivement les moyennes des écarts pour l'échantillon chacun des deux échantillons (ou sous-échantillons) et pour l'échantillon combiné.

La règle de décision pour effectuer ce test à un niveau de signification donné α est la suivante :

- Si $L \leq F_{1, n_1 + n_2 - 2}(1 - \alpha)$, on accepte H_0
- Sinon, on rejette H_0

où $F_{1, n_1 + n_2 - 2}(1 - \alpha)$ est le quantile de probabilité au non-dépassement $(1 - \alpha)$ de la loi de Fisher à 1 et $(n_1 + n_2 - 2)$ degrés de liberté.

2.3.2.2 Homogénéité temporelle: date de changement inconnue

Lorsque la date de changement de la moyenne est inconnue a priori, Hydro-Québec utilise la procédure bayésienne de Lee et Heghenian (1977) pour déterminer la date la plus probable d'une éventuelle coupure dans la série. Une fois la date identifiée, les tests de Student et de Fisher sont appliqués pour vérifier respectivement si les moyennes et les variances des deux sous-échantillons sont significativement différentes (cf. section 2.2.1).

Cette procédure en deux étapes est problématique puisque les mêmes observations sont utilisées deux fois. En effet, les données sont utilisées une première fois pour déterminer la date de coupure et une seconde fois pour appliquer les tests statistiques. Ces deux étapes n'étant pas indépendantes, le niveau de signification réel des tests de Student et de Fisher (erreur de Type I, la probabilité de rejeter à tort l'hypothèse nulle) est différent de celui que l'on se fixe a priori et est en fait conditionnel au résultat de la procédure de Lee et Heghenian (1977).

Pour étudier ce problème, nous avons déterminé le niveau de signification réel du test de Student suite à l'application de la méthode bayésienne. Pour cela, 1000 séries de taille 10, 20, 30, 40 et 50 observations ont été simulées. Les résultats obtenus sont présentés dans le tableau 2.9 pour des niveaux de signification nominaux de 5% et 1%.

Tableau 2.9. Niveau de signification réel du test de Student.

Niveau nominal	TAILLE D'ÉCHANTILLON				
	10	20	30	40	50
0.05 (5%)	0.281	0.373	0.439	0.452	0.482
0.01 (1%)	0.074	0.099	0.127	0.129	0.143

Les niveaux de signification réels sont systématiquement supérieurs aux niveaux nominaux. Par exemple, pour une série de 50 observations (ordre de grandeur des séries considérées ici), la probabilité de rejeter à tort l'égalité des moyennes est de 48% alors qu'a priori l'erreur de Type I choisie était de 5%. Les tests statistiques appliqués suite à la détection de la date de coupure par la procédure de Lee et Heghenian ne sont pas valides pour déterminer si le changement est statistiquement significatif. Tel qu'employée actuellement à Hydro-Québec, cette approche détecte des changements de moyenne qui n'en sont probablement pas, compte tenu de l'information restreinte disponible (taille d'échantillon limitée).

Pour pallier ce problème, nous recommandons, lorsque la date de changement est inconnue, l'application d'un test bayésien construit à partir de la procédure de détection de Lee et

Heghenian. Bernier (1994), qui a généralisé l'approche de Lee et Heghenian (1977), donne dans son article quelques pistes à ce sujet. Toutefois, une étude plus approfondie du test bayésien qu'il propose est nécessaire; nous reviendrons sur ce problème au chapitre 4 lors de l'analyse de la stationnarité des séries d'apports énergétiques annuels.

2.3.2.3 Homogénéité spatiale

Les recommandations qui suivent concernent la troisième situation où sont comparées systématiquement, mois par mois, les moyennes et les variances de deux séries d'apports standardisés (division par la superficie de bassin versant ou par le module annuel) calculés à des sites différents. Cette approche nécessite douze comparaisons des moyennes et des variances de deux échantillons à partir des tests de Student et de Fisher. Les douze tests (Student ou Fisher) sont appliqués de façon indépendante sans tenir compte du lien statistique entre les apports de différents mois. Or, en particulier, il est commun d'observer une corrélation significative entre les apports de deux mois consécutifs. Ce type de situation se prête donc plutôt à une procédure multidimensionnelle qui tient compte des corrélations entre les mois et qui permet une comparaison simultanée des 12 caractéristiques statistiques (moyennes ou variances) ainsi que des covariances entre les mois. Une telle approche a le double avantage de diminuer le nombre de tests à effectuer et de comparer, en plus des moyennes et des variances mensuelles, les corrélations mois par mois des deux séries.

Deux tests multidimensionnels sont présentés dans ce qui suit. Les équations et les règles de décisions données sont applicables seulement au cas traité dans l'harmonisation des apports, c'est-à-dire à la comparaison de deux populations de lois normales multivariées de dimension 12 (12 mois à comparer) et mesurées toutes les deux sur n années. Les formules générales sont présentées dans Srivastava et Carter (1983).

Pour comparer simultanément les 12 moyennes mensuelles de deux échantillons, il est recommandé d'utiliser le test d'Hotelling (Srivastava et Carter, 1983) qui est en fait une généralisation du test unidimensionnel de Student. Considérons les deux échantillons multidimensionnels X et Y contenant les 12 séries annuelles (n années) d'apports mensuels:

$$\mathbf{X}_{12 \times n} = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1,n} \\ x_{21} & x_{22} & x_{23} & \cdots & x_{2,n} \\ \vdots & \vdots & \vdots & & \vdots \\ x_{12,1} & x_{12,2} & x_{12,3} & \cdots & x_{12,n} \end{bmatrix} = [\mathbf{x}_1 \quad \mathbf{x}_2 \quad \cdots \quad \mathbf{x}_n]$$

$$\mathbf{Y}_{12 \times n} = \begin{bmatrix} y_{11} & y_{12} & y_{13} & \cdots & y_{1,n} \\ y_{21} & y_{22} & y_{23} & \cdots & y_{2,n} \\ \vdots & \vdots & \vdots & & \vdots \\ y_{12,1} & y_{12,2} & y_{12,3} & \cdots & y_{12,n} \end{bmatrix} = [\mathbf{y}_1 \quad \mathbf{y}_2 \quad \cdots \quad \mathbf{y}_n]$$

Supposons que chacun des n vecteurs aléatoires soient distribués selon une loi normale multidimensionnelle de dimension 12, de vecteurs-moyennes $\boldsymbol{\mu}_1$ et $\boldsymbol{\mu}_2$ respectivement pour les matrices \mathbf{X} et \mathbf{Y} , et de même matrice variances-covariances $\boldsymbol{\Sigma}$. L'hypothèse d'une même matrice des variances-covariances généralise ici à l'hypothèse d'égalité des variances du test unidimensionnel de Student.

Le test d'Hotelling examine l'hypothèse

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \quad \text{contre} \quad H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$$

et compare donc deux vecteurs contenant les apports moyens des 12 mois de l'année pour les deux sites d'intérêt. Pour effectuer ce test, il suffit d'abord de calculer les estimations des paramètres $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ et $\boldsymbol{\Sigma}$, qui sont données respectivement par les vecteurs des moyennes mensuelles :

$$\hat{\boldsymbol{\mu}}_1 = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_{12}) \quad \text{et} \quad \hat{\boldsymbol{\mu}}_2 = (\bar{y}_1, \bar{y}_2, \dots, \bar{y}_{12})$$

et par la matrice des variances mensuelles et des covariances entre les mois :

$$\hat{\boldsymbol{\Sigma}} = [(n-1)(\mathbf{S}_1 + \mathbf{S}_2)] / \nu, \quad \nu = 2n - 2$$

où

$$(n-1)\mathbf{S}_1 = \sum_{i=1}^n (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_1)(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_1)^T \quad \text{et} \quad (n-1)\mathbf{S}_2 = \sum_{i=1}^n (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_2)(\mathbf{y}_i - \hat{\boldsymbol{\mu}}_2)^T$$

La statistique du test d'Hotelling, distribuée à une constante près selon une loi de Fisher, est donnée par :

$$T = \frac{n}{2} (\hat{\boldsymbol{\mu}}_1 - \hat{\boldsymbol{\mu}}_2)^T \hat{\boldsymbol{\Sigma}}^{-1} (\hat{\boldsymbol{\mu}}_1 - \hat{\boldsymbol{\mu}}_2) \quad (2.7)$$

où $\boldsymbol{\Sigma}^{-1}$ désigne la matrice inverse de $\boldsymbol{\Sigma}$.

La règle de décision pour effectuer ce test à un niveau de signification donné α est la suivante :

- Si $(\nu - 12 + 1/(12\nu))T > F_{12, \nu-11, \alpha}$, on rejette H_0
- Sinon, on ne rejette pas H_0

où $F_{12, \nu-11, \alpha}$ est le quantile de probabilité au dépassement α de la loi de Fisher à 12 et $\nu - 11$ degrés de liberté.

Pour comparer les variances mensuelles et les covariances mois par mois des deux séries, nous recommandons un test utilisé fréquemment en analyse de variance multidimensionnelle (MANOVA) et documenté dans Srivastava et Carter (1983, pages 333 à 335). Ce test, construit à partir du rapport des vraisemblances maximales, vérifie les hypothèses :

$$H_0 : \Sigma_1 = \Sigma_2 \quad \text{contre} \quad H_1 : \Sigma_1 \neq \Sigma_2$$

où Σ_1 et Σ_2 sont des matrices symétriques ayant comme éléments les variances et covariances des deux ensembles de données (12 séries annuelles d'apports mensuels des deux sites).

La statistique générale qui examine ces hypothèses est distribuée asymptotiquement selon une loi de khi-deux et est donnée par :

$$\lambda = \frac{(n_1 + n_2)^{p(n_1 + n_2)/2} |n_1 S_1|^{n_1/2} |n_2 S_2|^{n_2/2}}{n_1^{pn_1/2} n_2^{pn_2/2} |n_1 S_1 + n_2 S_2|^{(n_1 + n_2)/2}} \quad (2.8)$$

où la notation $|A|$ signifie le déterminant de la matrice A , n_1 et n_2 correspondent au nombre d'observations de chaque ensemble de données (nombre d'années) et p désigne le nombre de séries (ici, $p = 12$). La règle de décision pour effectuer ce test à un niveau de signification donné α est la suivante :

- Si $-2(1 - 2k) \ln \lambda > \chi_{0.5, p(p+1), \alpha}^2$, on rejette H_0
- Sinon, on ne rejette pas H_0

où $\chi_{\nu, \alpha}^2$ est le quantile de probabilité au dépassement α de la loi de khi-deux à ν degrés de liberté et k est donné par :

$$k = (n_1^2 + n_1 n_2 + n_2^2)(2p^2 + 3p - 1) / [12(p+1)(n_1 + n_2)n_1 n_2]$$

Les deux tests multidimensionnels (équations (2.7) et (2.8)) reposent sur l'hypothèse que les 12 séries annuelles d'apports mensuels sont distribuées selon une loi normale multivariée. Évidemment, cette hypothèse peut être contestée. Par exemple, lors de l'harmonisation des apports des sites du complexe Aux Outardes (Bruneau, 1986b), les spécialistes d'Hydro-Québec ont été amenés à choisir entre deux séries d'apports totaux pour la période 1968-1986 (apports cumulés des sites ou apports transférés par RSBV à partir du site OU4). Pour comparer les moyennes et les variances des deux échantillons, les tests de Student et de Fisher ont été appliqués mois par mois. Or, le test de normalité de Shapiro et Wilk (1965) (Perreault *et al.*, 1994b) conclu à la non-normalité pour quelques ensembles de données comme le montre le tableau 2.10.

Tableau 2.10. Résultats du test de normalité de Shapiro-Wilk.

Mois	APPORTS CUMULÉS		APPORTS TRANSFÉRÉS DE OU4	
	Statistique W	Prob. au dép.	Statistique W	Prob. au dép.
Janvier	0.9429	0.209	0.8907	<u>0.014</u>
Février	0.9128	<u>0.045</u>	0.9722	0.735
Mars	0.9748	0.792	0.9765	0.827
Avril	0.8368	<u>0.001</u>	0.9222	<u>0.073</u>
Mai	0.9866	0.976	0.9816	0.920
Juin	0.9158	<u>0.053</u>	0.9239	<u>0.080</u>
Juillet	0.9807	0.905	0.9704	0.692
Août	0.9558	0.386	0.9405	0.186
Septembre	0.9717	0.722	0.9753	0.802
Octobre	0.9682	0.641	0.9668	0.610
Novembre	0.9269	<u>0.093</u>	0.9295	0.106
Décembre	0.9813	0.916	0.8772	<u>0.007</u>

0.xxx : Significatif à 5% 0.xxx : Significatif à 10%

Nous recommandons tout de même, dans des situations similaires, les tests paramétriques multidimensionnels puisque les procédures non paramétriques analogues (tests de rang multidimensionnels) sont fort complexes et leur efficacité n'est pas bien démontrée dans la littérature. Cela est dû principalement au fait que la notion de rang multivarié est difficilement applicable et interprétable. De plus, il nous apparaît important, dans cette situation, de tenir davantage compte de la structure de corrélations temporelles que de la non-normalité des observations.

2.3.3 Application de la classification ascendante hiérarchique (CAH)

Pour certains complexes hydroélectriques, une analyse de classification ascendante hiérarchique (CAH) a été réalisée afin d'étudier l'homogénéité spatiale d'un certain nombre de séries d'apports mensuels (cf. section 2.2.2). Telle qu'utilisée par Hydro-Québec lors de

l'harmonisation des apports, la CAH a comme objectif d'étudier les similarités entre les séries d'apports mensuels de sites voisins. La CAH permet en effet de regrouper les éléments semblables d'un tableau de données en calculant, à partir de variables mesurées pour chaque élément, une distance entre ceux-ci. Un arbre hiérarchique est ensuite produit qui permet d'interpréter toutes les étapes de cette procédure de regroupement et de former des groupes distincts d'éléments homogènes selon les variables considérées.

Cette analyse dépend fortement du choix des variables caractérisant le comportement de chaque série d'observations. Les variables retenues sont les moyennes mensuelles et annuelles, les maximums annuels et les minimums annuels. On peut se demander si ces statistiques sont suffisantes pour bien caractériser les séries de données et discriminer les séries hétérogènes. A priori, il serait intéressant d'ajouter à l'analyse de CAH certaines caractéristiques statistiques. Dans un premier temps, il serait souhaitable de considérer l'autocorrélation d'ordre 1 car ce paramètre explique une bonne partie du comportement d'une série d'apports mensuels.

Il est recommandé toutefois qu'une étude en profondeur soit réalisée afin de déterminer les caractéristiques statistiques pertinentes pour analyser les similarités entre les séries d'apports mensuels de sites voisins.

2.4 Conclusions et recommandations

Cette section résume les principales conclusions et recommandations d'ordre méthodologique suite à l'analyse critique des données de base (sous-projet 2.2). Quelques remarques spécifiques y sont aussi données.

2.4.1 Conclusions et recommandations générales

- *Transfert par rapport de superficies de bassins versants (section 2.3.1)*

Les résultats de l'analyse des biais annuels et mensuels ainsi que de la périodicité des résidus pour les sites OU4, MA5, LSA, MAT et GOU illustrent bien les problèmes liés à l'estimation par la méthode RSBV (section 2.3.1) :

1. présence d'un biais annuel systématique négatif ou positif;
2. présence d'un biais saisonnier;
3. périodicité des résidus.

Outre ces trois problèmes, la variance des données reconstituées \hat{Q}_Y est imposée par la méthode RSBV et est entièrement indépendante de la variance des apports historiques Q_Y du site d'intérêt Y . Ceci constitue un problème supplémentaire puisque qu'il n'y a aucune raison pratique ou théorique d'obtenir systématiquement une telle variance.

L'étude critique de l'harmonisation des apports nous a permis d'identifier différentes avenues possibles qui peuvent corriger certains problèmes associés à l'utilisation de la méthode RSBV:

- Si des données d'apports naturels de bonne qualité sont disponibles pour une certaine période au site d'intérêt, on peut envisager l'utilisation d'un rapport d'apports saisonnier ou mensuel comme paramètre de transfert.
- Dans la même situation, on peut envisager d'utiliser des régressions linéaires mois par mois afin de déterminer les relations mensuelles entre les deux sites.
- Pour éviter le transfert par la méthode RSBV de l'embouchure d'un complexe aux différents sites (Aux Outardes, Manicouagan et Bersimis), il serait possible de répartir les apports totaux (série de référence) à l'aide d'un modèle de désagrégation (SPIGOT, par exemple). Ce modèle a l'avantage de conserver les principales caractéristiques des séries d'apports intermédiaires historiques.
- Si le modèle SPIGOT n'est pas considéré pour la répartition des apports totaux aux sites d'un complexe, nous recommandons d'utiliser comme série de référence (si possible) les apports naturels du réservoir de tête qui, en général, semblent de bonne qualité.
- *Tests statistiques (section 2.3.2)*

À la section 2.3.2, quelques problèmes d'ordre méthodologique ont été identifiés concernant l'application du test de Student, du test de Fisher et de la procédure de détection de changement de la moyenne de Lee et Heghenian (1977):

1. non-normalité des apports naturels aux sites;
2. biais des tests de Student et de Fisher lorsqu'ils sont appliqués suite à la procédure de Lee et Heghenian;
3. corrélation temporelle lors de la comparaison des caractéristiques statistiques mensuelles des séries d'apports de deux sites différents.

Pour pallier le problème de non-normalité et pour éviter d'effectuer systématiquement un test de normalité avant l'application des tests de Student et de Fisher, il est recommandé de valider la conclusion de ces tests en utilisant en parallèle le test de Wilcoxon (1945) et le test de Levene (1960). Ces tests sont non paramétriques et ne reposent pas sur la normalité des observations. Ils sont donc plus puissants que les tests paramétriques lorsque les observations ne suivent pas une loi normale (cf. section 2.3.2.1).

Concernant le second problème traité à la section 2.3.2.2, nous déconseillons l'emploi de tests classiques de comparaison des moyennes ou des variances suite à l'application de la procédure bayésienne de Lee et Heghenian (1977). En effet, la probabilité de rejeter à tort l'hypothèse nulle (par exemple, l'égalité des moyennes) est beaucoup plus élevée que celle que l'on se fixe a priori (cf. tableau 2.9). Cela est dû au fait que les mêmes données sont utilisées deux fois (une première fois pour déterminer la date de coupure et une seconde fois pour appliquer les tests statistiques). Nous recommandons donc l'application d'un test bayésien construit à même la procédure de détection de Lee et Heghenian. Cette approche permettrait de tenir compte de la double utilisation des observations. Bernier (1994) donne dans son article quelques pistes à ce sujet. Toutefois, une étude plus approfondie du test bayésien qu'il propose doit être réalisée avant sa mise en oeuvre.

Enfin, la comparaison des moyennes et des variances mensuelles des séries d'apports de deux sites différents est effectuée en appliquant douze tests (Student et Fischer pour chaque mois) de façon indépendante, sans tenir compte du lien statistique entre les apports de différents mois. Or, il est commun d'observer une corrélation significative entre les apports de deux mois consécutifs. Ce type de situation se prête donc plutôt à une procédure multidimensionnelle qui tient compte des corrélations entre les mois et qui permet une comparaison simultanée des 12 caractéristiques statistiques (moyennes ou variances) ainsi que des covariances entre les mois. Une telle approche a le double avantage de diminuer le nombre de tests à effectuer et de comparer, en plus des moyennes et des variances mensuelles, les corrélations mois par mois des deux séries.

Deux tests multidimensionnels, présentés à la section 2.3.2.3, sont recommandés:

- le test d'Hotelling (Srivastava et Carter, 1983) qui permet de comparer simultanément les moyennes mensuelles de deux séries;

- un test utilisé fréquemment en analyse de variance multidimensionnelle (MANOVA) et documenté dans Srivastava et Carter (1983) qui compare les variances mensuelles et les covariances mois par mois de deux séries.

- *Classification ascendante hiérarchique CAH (section 2.3.3)*

Telle qu'utilisée par Hydro-Québec lors de l'harmonisation des apports, la CAH a comme objectif d'étudier les similarités entre les séries d'apports mensuels de sites voisins. Cette analyse dépend fortement du choix des variables caractérisant le comportement de chaque série d'observations. Les variables retenues par Hydro-Québec sont les moyennes mensuelles et annuelles, les maximums annuels et les minimums annuels. On peut se demander si ces statistiques sont suffisantes pour bien caractériser les séries de données et discriminer les séries hétérogènes. Il serait intéressant d'ajouter à l'analyse de CAH certaines caractéristiques statistiques. Dans un premier temps, il serait souhaitable de considérer l'autocorrélation d'ordre 1 car ce paramètre explique une bonne partie du comportement d'une série d'apports mensuels.

Il est recommandé toutefois qu'une étude en profondeur soit réalisée afin de déterminer les caractéristiques statistiques les plus pertinentes pour analyser les similarités entre les séries d'apports mensuels de sites voisins.

2.4.2 Remarques spécifiques

- *Complexe Aux Outardes (section 2.3.1.1)*

L'étude critique de la méthode de transfert par superficies de bassins versants repose, en particulier, sur les apports naturels reconstitués du site témoin Outardes 4. Les analyses présentées à la section 2.3.1.1 ont permis d'identifier deux problèmes spécifiques au complexe Aux Outardes liés à la reconstitution de la série de référence N9519003 (apports totaux):

1. la reconstitution des apports mensuels d'octobre 1969 à septembre 1970 et d'avril 1972 à l'aide d'un modèle de régression linéaire avec les apports intermédiaires à MA5;
2. le cumul des apports intermédiaires suite à la mise en service du réservoir OU2 en 1978.

Ces problèmes ont une influence directe sur la qualité de la reconstitution des apports à OU4. La régression linéaire avec MA5 engendre des écarts de prévision très élevés. Ce problème peut être évité en corrigeant les valeurs erronées qui ont motivé cette intervention (cf. tableau 2.4). Les valeurs de référence obtenues ensuite en cumulant les apports intermédiaires corrigés permettent de diminuer substantiellement les écarts. Le calcul des apports de référence par le cumul des apports intermédiaires est toutefois problématique après la mise en service du réservoir OU2. En effet, les débits de sortie à ce site, qui n'intervenaient pas dans le calcul avant 1978, influencent grandement les apports totaux et produisent un changement de signe du biais des apports reconstitués à OU4. Deux solutions peuvent être envisagées pour corriger cette situation :

- valider les débits de sortie à OU2 et utiliser toujours le cumul des apports comme série de référence;
 - employer plutôt les apports intermédiaires à OU4 comme valeurs de référence à partir de la date de mise en service de ce réservoir.
- *Complexe La Grande*

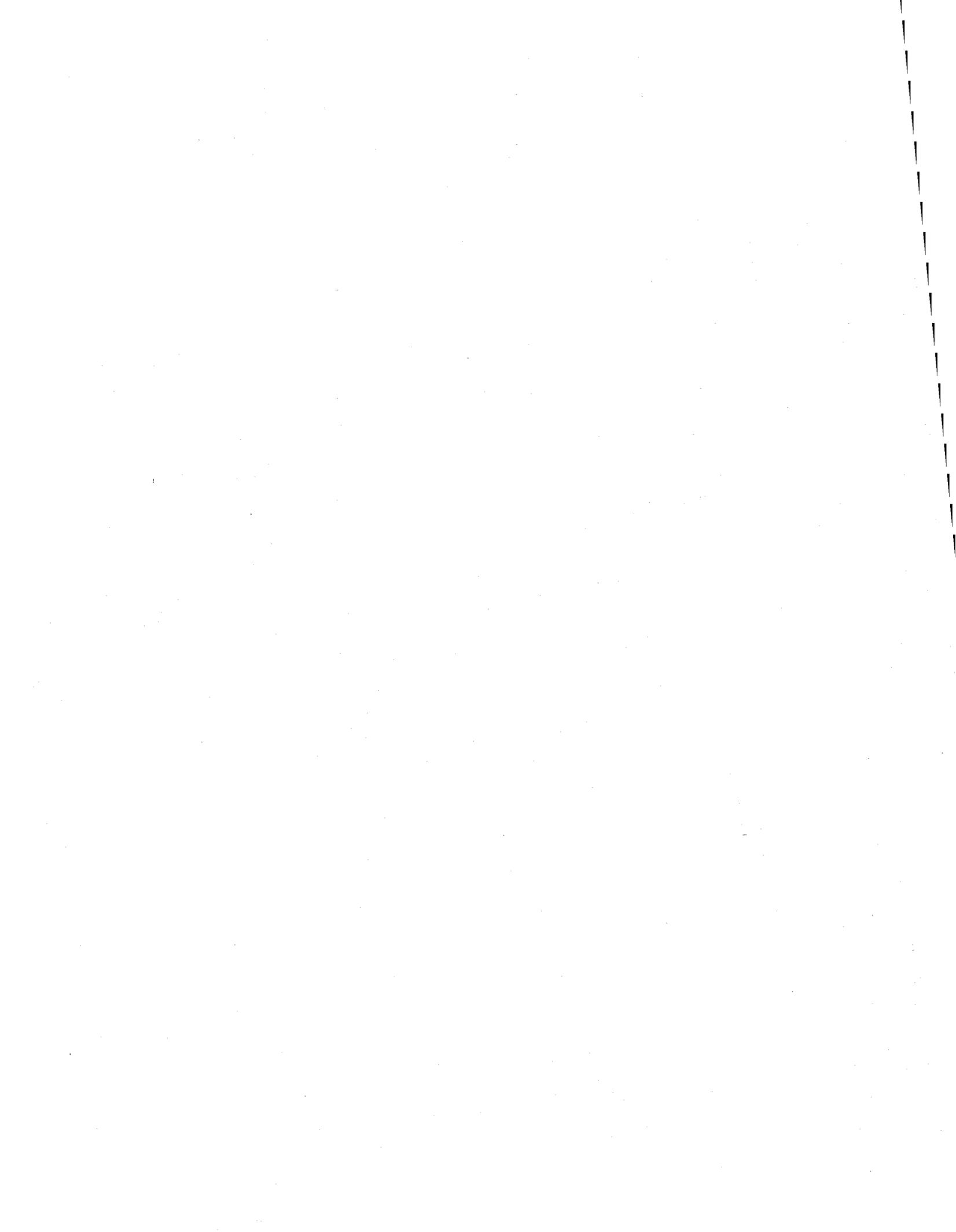
Le système hydrique de La Grande est le complexe qui contribue le plus aux apports totaux (cf. chapitre 3). Les caractéristiques statistiques des apports naturels de ce complexe influencent donc fortement celles des apports énergétiques annuels agrégés. Or, l'objectif principal du sous-projet 2.4 est la modélisation de cette série (cf. chapitre 4). Ainsi, une attention particulière doit être portée aux données de base de ce complexe hydroélectrique.

L'approche d'harmonisation des apports pour le complexe La Grande nous apparaît adéquate compte tenu de toutes les contraintes inhérentes au peu d'observations disponibles aux sites du complexe ainsi qu'aux rivières situées dans le bassin versant (cf. figure 2.3). Toutefois, quelques remarques et recommandations sont émises dans ce qui suit.

- Une première reconstitution des apports intermédiaires aux sites est effectuée par la méthode RSBV à partir de rivières voisines (pas de temps annuel, 1961-1980). Ce transfert est sûrement moins problématique au pas de temps annuel qu'au pas de temps mensuel. Toutefois, cette approche peut quand même engendrer un biais dû à la variabilité spatiale des précipitations. Dans Bruneau (1986c), il est mentionné qu'une correction a été appliquée au paramètre de transfert (par la SEBJ) et que ce nouveau facteur est acceptable. Maintenant que des données sont disponibles aux sites, nous recommandons qu'une étude graphique sur les résidus, semblable à celle

réalisée pour le complexe Aux Outardes (section 2.3.1.1), soit effectuée pour examiner ce problème. Cela nécessite toutefois que des débits en rivière soient disponibles pour la même période de mesure.

- Une fois que 20 années communes d'apports naturels annuels sont disponibles à chaque site, on les cumule et on effectue une extension de cette nouvelle série d'apports totaux à partir de régressions. Cette série est ensuite désagrégée dans le temps et dans l'espace à l'aide du logiciel SPIGOT. Il serait intéressant de comparer cette approche à celle qui consisterait à effectuer une extension directement aux sites par une régression multidimensionnelle qui conserverait automatiquement les corrélations entre les sites. Une désagrégation temporelle serait toutefois nécessaire pour obtenir ensuite des apports mensuels. La régression multidimensionnelle est disponible dans le logiciel *RéMuS* (Perron *et al.*, 1996). Les aspects théoriques de cette approche sont présentés dans Perreault *et al.* (1994a)
- Le logiciel SPIGOT permet de conserver explicitement toutes les caractéristiques statistiques que l'on désire. Toutefois, on se limite en général à la conservation d'un nombre restreint de caractéristiques à cause des nombreux paramètres à estimer. Dans ce cas, l'ordre dans lequel la désagrégation est effectuée peut avoir une certaine importance. Si on désagrège d'abord aux sites pour ensuite répartir ces apports annuels en données mensuelles, la corrélation entre les apports annuels des différents sites sera conservée explicitement mais pas celle entre les apports mensuels. Si on désagrège d'abord les apports annuels en apports mensuels pour ensuite les répartir spatialement, la corrélation mois par mois (d'ordre choisi a priori) des apports totaux sera conservée explicitement mais pas celle entre sites. Il est donc important de bien identifier les caractéristiques que l'on désire conserver avant d'appliquer le modèle SPIGOT.



3 ANALYSE STATISTIQUE DES APPORTS E_n

Dans le but d'extraire toute l'information possible des séries d'apports énergétiques, le sous-projet 3, intitulé "Analyse statistique des séries d'apports énergétiques", consistait à effectuer une analyse descriptive complète des apports énergétiques (E_n) totaux mensuels et annuels des huit complexes hydroélectriques du parc d'Hydro-Québec. Cette analyse comprend en particulier le calcul des principales caractéristiques statistiques des séries ainsi que la détermination des corrélations entre les différents complexes. Quelques procédures statistiques sont aussi appliquées afin de vérifier certaines hypothèses sur lesquelles reposent l'utilisation de ces séries dans les modèles de planification (normalité, stationnarité de la moyenne, etc.).

Mentionnons qu'une analyse statistique similaire, mais moins complète, a déjà été réalisée par Nguyen (1993). D'autre part, une étude concernant les apports naturels, axée sur les représentations graphiques, a aussi été réalisée par Technik-Eaucan Inc (1995).

3.1 Calcul des apports énergétiques

Suite à l'harmonisation des apports (cf. chapitre 2) qui permet d'obtenir une plage commune d'au moins cinquante années d'apports naturels mensuels à chaque site, ces valeurs sont converties en apports énergétiques. En effet, connaissant la capacité de production de chaque centrale, on peut transformer le stock d'eau emmagasinée et les apports naturels (m^3/s) en valeur énergétique exprimée en gigawatts-heure (GWh).

L'apport énergétique mensuel E_n , d'une centrale donnée est donc proportionnel à l'apport naturel mensuel A_n calculé au site. Pour chaque centrale d'un complexe hydroélectrique, un facteur de production, exprimé en $kW/(m^3/s)$, est d'abord calculé à partir de la production enregistrée à la centrale au cours des quatre dernières années. Les facteurs de production d'un même complexe sont ensuite cumulés de l'aval vers l'amont. Ainsi, la centrale la plus en amont d'un complexe possède le facteur le plus élevé. Pour un mois donné, l'apport naturel mensuel A_n (en m^3/s) de chaque centrale est multiplié par son facteur de production pour obtenir l'énergie mensuelle E_n au site (en GWh). Hydro-Québec dispose de séries d'apports énergétiques mensuels calculés pour l'ensemble des centrales de son parc d'équipements sur au moins cinquante années.

L'énergie mensuelle totale produite par un complexe hydroélectrique est alors calculée en cumulant les apports énergétiques de chaque site du complexe de l'amont vers l'aval. Ensuite, l'énergie annuelle totale du complexe est obtenue en sommant les apports énergétiques mensuels sur toute l'année. Enfin, une série annuelle agrégée, représentant la production totale de l'ensemble du parc d'équipement d'Hydro-Québec, est construite en sommant les apports énergétiques annuels totaux de chaque complexe. Quatre types de séries d'apports énergétiques sont obtenues:

1. Apports énergétiques mensuels de chaque centrale;
2. Apports énergétiques mensuels des huit complexes hydroélectriques (cumul spatial);
3. Apports énergétiques annuels des huit complexes hydroélectriques (cumul temporel);
4. Apports énergétiques annuels agrégés (cumul spatial).

Comme il a été mentionné précédemment, l'objectif du présent sous-projet est d'effectuer une analyse statistique descriptive des données d'apports énergétiques pour les huit principaux complexes hydroélectriques du Québec. La banque de données fournie contient, pour chaque complexe, les apports énergétiques historiques au pas de temps mensuel et annuel pour la période s'échelonnant de 1943 à 1993¹. La figure 3.1 donne, pour cette période, la répartition des apports énergétiques entre les huit complexes hydroélectriques (parc du 1^{er} janvier 1995).

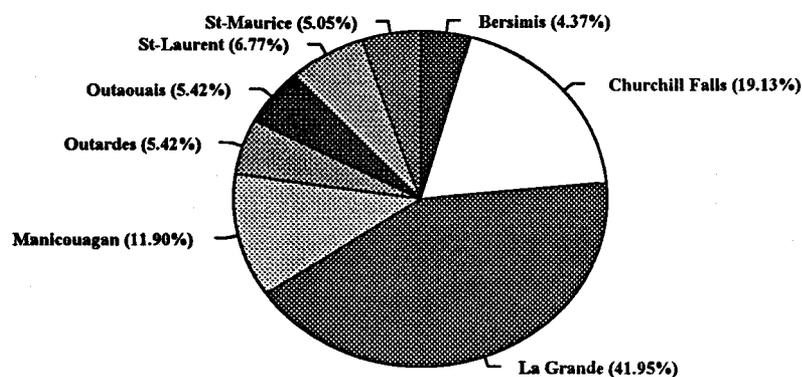


Figure 3.1. Répartition des apports énergétiques pour la période 1943-1993.

¹ Depuis le début du projet C2, nous avons considéré la série qui s'échelonne de 1943 à 1993. Dans la section 4.4, les années 1994 et 1995 ont été prises en compte pour les prévisions.

On constate que les systèmes La Grande, Churchill Falls et Manicouagan cumulent près de 73% de la production énergétique totale. Les caractéristiques statistiques des apports énergétiques de ces sites influencent donc considérablement celles de la série agrégée.

Les résultats présentés dans les sections qui suivent concernent les apports énergétiques totaux aux complexes et les apports agrégés.

3.2 Apports énergétiques mensuels

Les statistiques de base permettant de caractériser les séries d'apports énergétiques mensuels sont présentées dans cette section. Ces statistiques permettent de tracer un portrait général des données d'apports mensuels de chaque complexe hydroélectrique.

3.2.1 Statistiques de base

Les tableaux 3.1 à 3.9 donnent les principales caractéristiques statistiques des apports énergétiques mensuels des huit complexes et de la série agrégée. Les figures 3.2 à 3.10 illustrent la distribution mensuelle des apports de chacune des séries. Ces graphiques sont des diagrammes en boîte (boxplot) représentant schématiquement la forme de la distribution des observations. Cette illustration, au même titre qu'un histogramme, est très utile pour comparer visuellement les caractéristiques statistiques de plusieurs séries. Ici, elle est particulièrement utile pour étudier la variabilité temporelle des apports à un site donné ou pour comparer leur évolution dans le temps pour différents sites. Une brève description de ce type de représentation graphique est donnée en annexe A.

Les diagrammes en boîte montrent que la distribution des apports varie considérablement d'un mois à l'autre pour la plupart des complexes (Bersimis, Churchill Falls, La Grande, Aux Outardes, Manicouagan, Outaouais, St-Maurice). En particulier, la variabilité des apports énergétiques est systématiquement plus grande en période de crue (avril à juillet). Par contre, le comportement de la distribution mensuelle des apports énergétiques du complexe St-Laurent est plus uniforme (figure 3.9). En effet, les moyennes et les écart-types mensuels de ses apports énergétiques sont beaucoup plus stables d'un mois à l'autre. Les figures 3.11 et 3.12, où ces caractéristiques statistiques sont tracées, illustrent bien ce comportement. Toutefois, cela n'influence pas les apports énergétiques mensuels de la série agrégée (cf. tableau 3.9 et figure 3.10). Ceux-ci se comportent plutôt comme ceux du complexe La Grande (figure 3.4) dont la contribution est beaucoup plus importante (42% pour La Grande et 7% pour St-Laurent).

Tableau 3.1. Statistiques de base des apports énergétiques (GWh) du complexe Bersimis.

	JAN	FÉV	MAR	AVR	MAI	JUIN
MOYENNE	223.92	172.40	182.67	405.30	2017.72	1360.91
MÉDIANE	220.7	169.5	175.3	317.5	1947.4	1328.4
ÉCART-TYPE	48.23	48.06	58.18	249.74	521.85	433.43
CV	0.215	0.279	0.318	0.616	0.258	0.318
Cs	0.446	0.144	1.529	1.213	-0.057	1.588
MINIMUM	135.7	71.8	96.2	93.1	807.3	695
MAXIMUM	335.8	271.8	387.7	1113.0	3354.2	3164.9
TOTAL	11420.2	8792.7	9316.3	20670.4	102904.1	69406.7
	JUIL	AOÛ	SEP	OCT	NOV	DÉC
MOYENNE	809.81	684.02	638.13	723.70	483.98	326.46
MÉDIANE	797.2	629.7	671.0	695.7	454.0	333.5
ÉCART-TYPE	215.36	229.91	196.34	224.73	172.27	80.42
CV	0.265	0.336	0.307	0.310	0.355	0.246
Cs	0.525	0.752	0.099	0.468	2.292	0.886
MINIMUM	439.7	315.7	274.7	365.3	240.9	170.6
MAXIMUM	1483.6	1271.1	1086.7	1204.4	1242.1	622.0
TOTAL	41300.4	34885.3	32544.8	36909.1	24683	16649.9

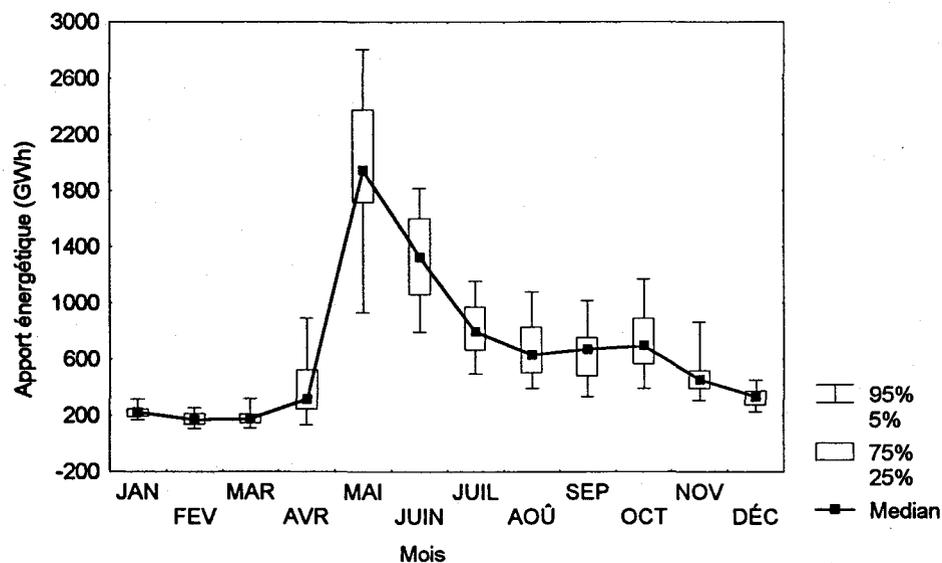
**Figure 3.2.** Distribution des apports énergétiques mensuels - Complexe Bersimis.

Tableau 3.2. Statistiques de base des apports énergétiques (GWh) du complexe Ch. Falls.

	JAN	FÉV	MAR	AVR	MAI	JUIN
MOYENNE	1128.93	838.94	836.84	1048.85	4740.86	9002.98
MÉDIANE	1119.6	889.7	852.5	895.3	4326.6	9410.1
ÉCART-TYPE	369.17	267.86	275.85	488.85	2267.90	2145.85
CV	0.327	0.319	0.329	0.466	0.478	0.238
CS	-0.148	-0.181	-0.239	2.316	1.327	-0.389
MINIMUM	170.9	275.8	23.5	477.6	1205.1	3475.8
MAXIMUM	1846	1377.9	1489.2	3178.0	12635.9	12968.6
TOTAL	57575.6	42786.4	42679.2	53491.4	241783.9	459152.2

	JUIL	AOÛ	SEP	OCT	NOV	DÉC
MOYENNE	4742.19	3102.11	3028.31	3045.80	2103.48	1538.61
MÉDIANE	4670.6	3098.1	2880.3	3061.8	2007.7	1489.2
ÉCART-TYPE	1208.75	798.13	937.60	757.38	616.96	397.69
CV	0.254	0.257	0.309	0.248	0.293	0.258
CS	0.246	-0.044	0.470	0.023	0.525	0.403
MINIMUM	1722.1	1390.9	1358.5	1512.7	971.8	726.4
MAXIMUM	7749.5	4598.0	5264.3	4700.6	3653.6	2602.4
TOTAL	241852.1	158207.9	154444.2	155336.1	107277.7	78469.6

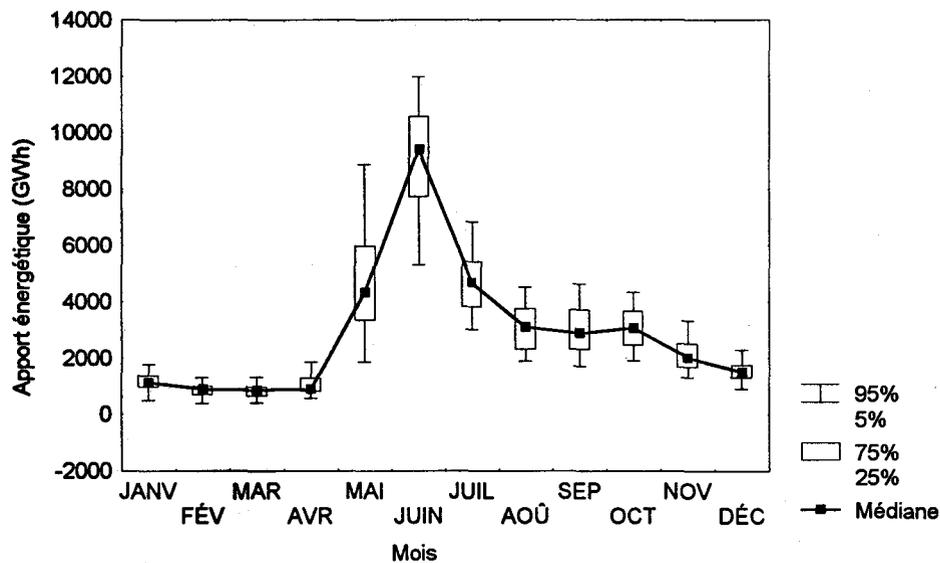


Figure 3.3. Distribution des apports énergétiques mensuels - Complexe Churchill Falls.

Tableau 3.3. Statistiques de base des apports énergétiques (GWh) du complexe La Grande.

	JAN	FÉV	MAR	AVR	MAI	JUIN
MOYENNE	2871.28	1917.80	1756.33	1875.38	9506.15	14792.88
MÉDIANE	2803.8	1907.4	1741.0	1699.7	8933.9	15220.6
ÉCART-TYPE	379.84	251.61	222.20	696.66	3047.74	4302.75
Cv	0.132	0.131	0.126	0.371	0.320	0.299
Cs	0.206	-0.500	1.179	2.549	0.298	-0.005
MINIMUM	2166.2	1150.5	1368.4	1059.2	3949.1	7157.4
MAXIMUM	3705.5	2364.4	2551.3	4898.7	16190.5	22783.2
TOTAL	146435.7	97807.8	89573.3	95644.4	484813.8	754437.2
	JUIL	AOÛ	SEP	OCT	NOV	DÉC
MOYENNE	9581.97	7584.75	7329.30	8685.94	6775.09	4445.50
MÉDIANE	10118.5	6691.8	7020.6	7992.4	6359.6	4368.3
ÉCART-TYPE	2731.09	3518.54	2402.59	2204.54	1646.79	929.91
Cv	0.285	0.463	0.327	0.253	0.243	0.209
Cs	0.000	1.072	0.766	0.752	0.597	0.256
MINIMUM	4087.7	2568.3	3365.2	4994.1	3634.3	2811.7
MAXIMUM	15125.9	16874.5	14184	14479.1	11599.3	6232.5
TOTAL	488680.5	386822.5	373794.5	442983.2	345529.6	226720.6

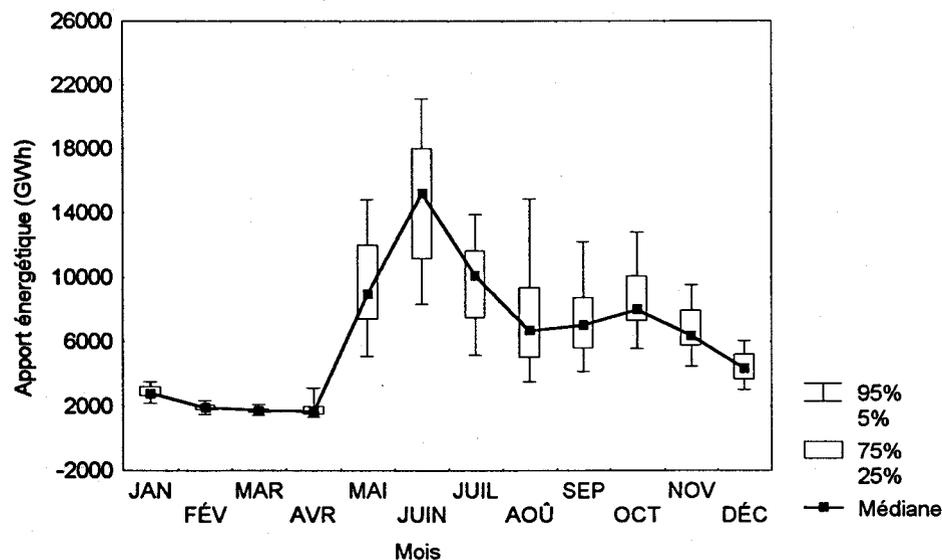
**Figure 3.4.** Distribution des apports énergétiques mensuels - Complexe La Grande.

Tableau 3.4. Statistiques de base des apports énergétiques (GWh) du complexe Manic..

	JAN	FÉV	MAR	AVR	MAI	JUIN
MOYENNE	601.03	424.67	415.29	748.53	4246.70	4912.62
MÉDIANE	583.2	423.0	405.6	555.3	4227.0	4985.8
ÉCART-TYPE	165.19	116.73	111.12	537.46	1506.68	1544.73
CV	0.274	0.274	0.267	0.718	0.354	0.314
Cs	1.109	0.255	0.341	2.189	0.159	0.027
MINIMUM	307.6	106.6	176.1	236.7	984.9	2011.9
MAXIMUM	1161.5	772.2	676.5	3095.2	7898.9	8447.7
TOTAL	30652.8	21658.5	21180.0	38175.2	216581.8	250543.4
	JUIL	AOÛ	SEP	OCT	NOV	DÉC
MOYENNE	2481.02	1976.52	1828.15	2059.06	1284.80	887.99
MÉDIANE	2393.5	1981.9	1851.1	2061.2	1244.6	834.8
ÉCART-TYPE	833.43	534.40	576.61	662.66	355.48	327.99
CV	0.335	0.270	0.315	0.321	0.276	0.369
Cs	0.764	0.002	-0.123	1.101	1.284	1.577
MINIMUM	848.0	910.9	665.2	795.9	760.3	465.3
MAXIMUM	5014.5	3056.8	2899.5	4275.2	2582.9	2004.3
TOTAL	126531.8	100802.7	93235.9	105012.2	65525	45287.5

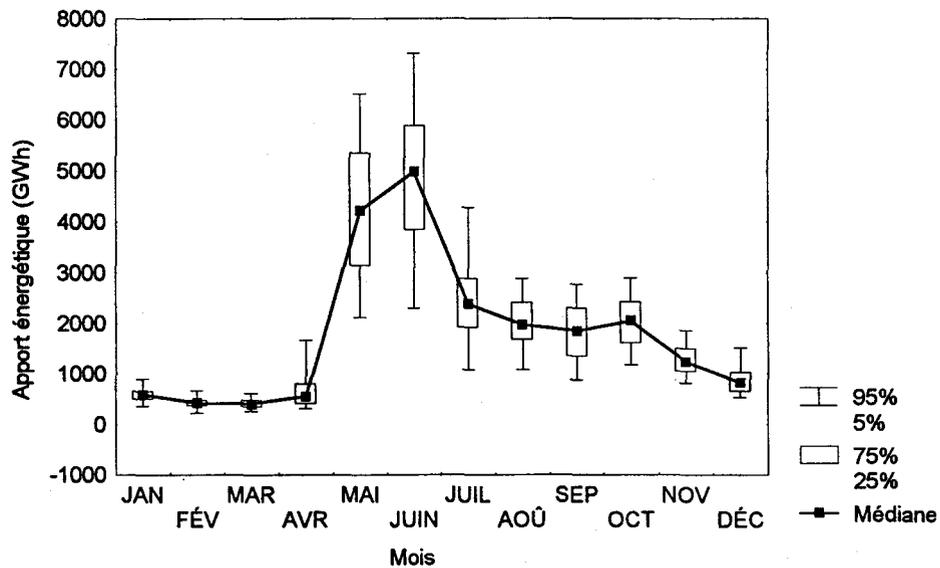


Figure 3.5. Distribution des apports énergétiques mensuels - Complexe Manicouagan.

Tableau 3.5. Statistiques de base des apports énergétiques (GWh) du complexe Outardes.

	JAN	FÉV	MAR	AVR	MAI	JUIN
MOYENNE	303.03	218.45	201.90	429.28	2358.66	1876.94
MÉDIANE	283.5	211.9	198.5	310.9	2322.9	1783.2
ÉCART-TYPE	94.00	79.45	86.81	285.53	639.06	635.47
CV	0.310	0.363	0.429	0.665	0.271	0.338
Cs	0.965	0.331	1.407	1.435	0.170	0.861
MINIMUM	124.3	37.0	53.0	140.8	806.1	869.6
MAXIMUM	550.0	400.4	560.7	1248.7	3936.8	3797.3
TOTAL	15455.0	11141.4	10297.2	21893.6	120291.9	95724.1
	JUIL	AOÛ	SEP	OCT	NOV	DÉC
MOYENNE	1020.01	848.44	796.05	876.18	600.77	426.19
MÉDIANE	1001.7	824.4	809.4	843.8	565.5	406.5
ÉCART-TYPE	289.28	248.30	246.72	277.92	184.38	120.52
CV	0.283	0.292	0.309	0.317	0.306	0.282
Cs	0.591	0.084	-0.218	0.691	2.541	1.757
MINIMUM	496.5	300.1	288.2	393.1	339.5	226.5
MAXIMUM	1859.2	1454.7	1361.8	1780.5	1467.9	903.1
TOTAL	52020.8	43270.5	40599	44685.6	30639.5	21736

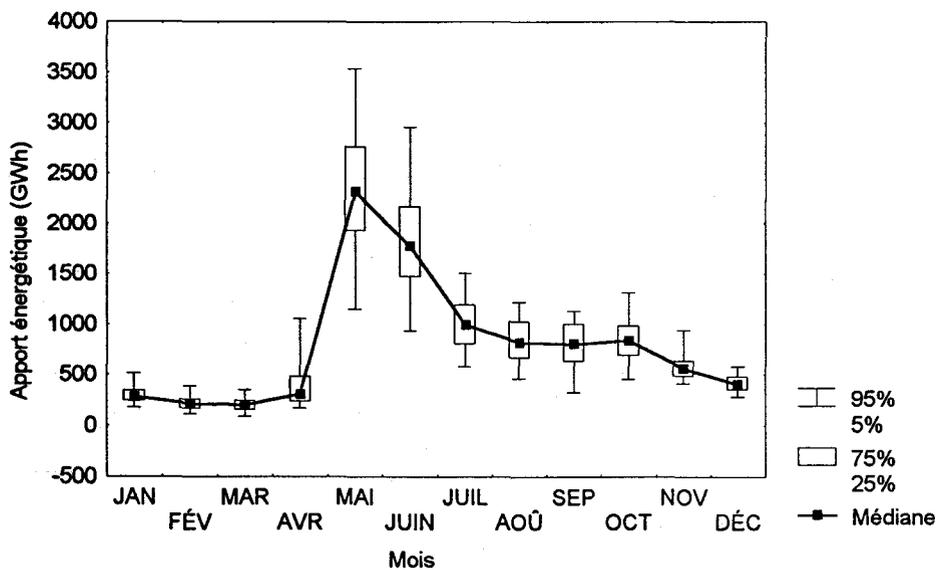
**Figure 3.6.** Distribution des apports énergétiques mensuels - Complexe Aux Outardes.

Tableau 3.6. Statistiques de base des apports énergétiques (GWh) du complexe Outaouais.

	JAN	FÉV	MAR	AVR	MAI	JUIN
MOYENNE	408.60	308.91	430.49	1739.86	2037.64	1093.65
MÉDIANE	387	286.3	371.6	1681.9	1978.4	987.3
ÉCART-TYPE	106.54	92.74	196.04	583.36	669.00	411.88
CV	0.260	0.300	0.455	0.335	0.328	0.376
Cs	0.941	1.851	1.750	0.343	0.926	2.001
MINIMUM	211.4	178.2	199.8	697.9	941.0	539.8
MAXIMUM	716.2	718.5	1096.4	3055.6	4018.9	2946.7
TOTAL	20838.9	15754.4	21955.4	88733.3	103919.9	55776.5

	JUIL	AoÛ	SEP	OCT	Nov	DÉC
MOYENNE	715.21	535.56	548.84	754.50	780.57	607.27
MÉDIANE	657.8	505.2	506.6	658.4	703.0	566.0
ÉCART-TYPE	282.50	206.59	222.01	313.66	312.49	217.39
CV	0.394	0.385	0.404	0.415	0.400	0.357
Cs	0.990	0.637	1.313	0.838	0.732	1.104
MINIMUM	199.4	237.1	240.9	236.1	298.1	245.7
MAXIMUM	1618.4	1066.1	1328.6	1628.0	1669.2	1349.9
TOTAL	36475.8	27313.8	27991.3	38480.0	39809.1	30971.0

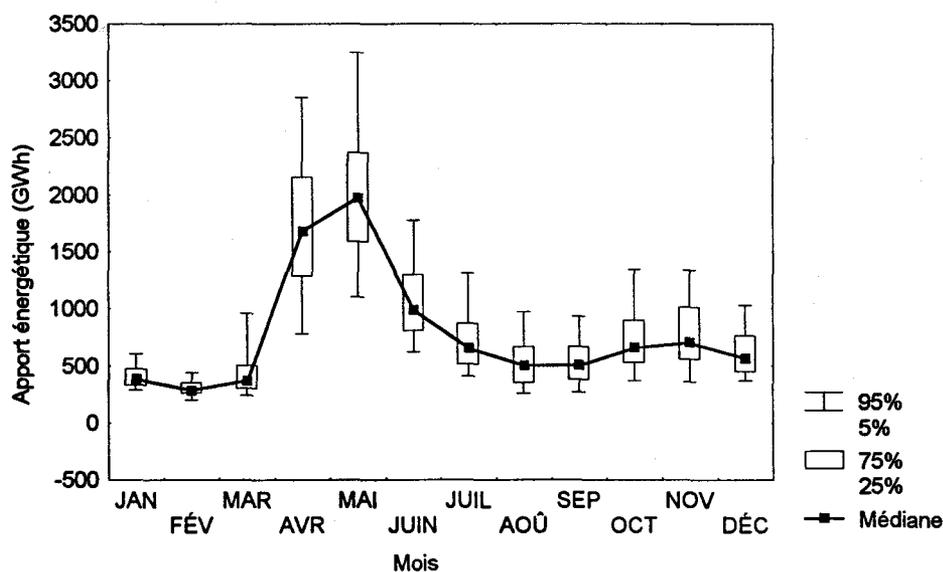


Figure 3.7. Distribution des apports énergétiques mensuels - Complexe Outaouais.

Tableau 3.7. Statistiques de base des apports énergétiques (GWh) du complexe St-Maur..

	JAN	FÉV	MAR	AVR	MAI	JUIN
MOYENNE	357.08	252.84	288.66	1125.86	2183.20	1102.74
MÉDIANE	355.0	255.0	265.6	1089.8	2179.2	1017.4
ÉCART-TYPE	86.16	68.88	93.52	547.14	688.07	410.56
Cv	0.241	0.272	0.324	0.486	0.315	0.372
Cs	0.923	0.455	2.609	0.957	1.014	1.702
MINIMUM	205.9	57.3	162.8	295.7	982.6	494.3
MAXIMUM	639.4	495.6	724.6	2698.2	4580.1	2559.6
TOTAL	18211.5	12895.2	14721.8	57418.9	111343.4	56240.2

	JUIL	AOÛ	SEP	OCT	NOV	DÉC
MOYENNE	731.69	598.04	573.52	776.22	742.10	558.05
MÉDIANE	685.2	576.4	586.8	709.3	672.1	495.7
ÉCART-TYPE	230.54	218.40	227.44	319.74	276.67	173.74
Cv	0.315	0.365	0.396	0.412	0.373	0.311
Cs	0.515	0.828	0.647	0.754	0.824	0.688
MINIMUM	249.4	233.5	254.6	237.1	288.0	296.0
MAXIMUM	1447.0	1195.4	1147.6	1514.1	1513.2	1042.3
TOTAL	37316.3	30500.3	29250.0	39587.7	37847.4	28460.8

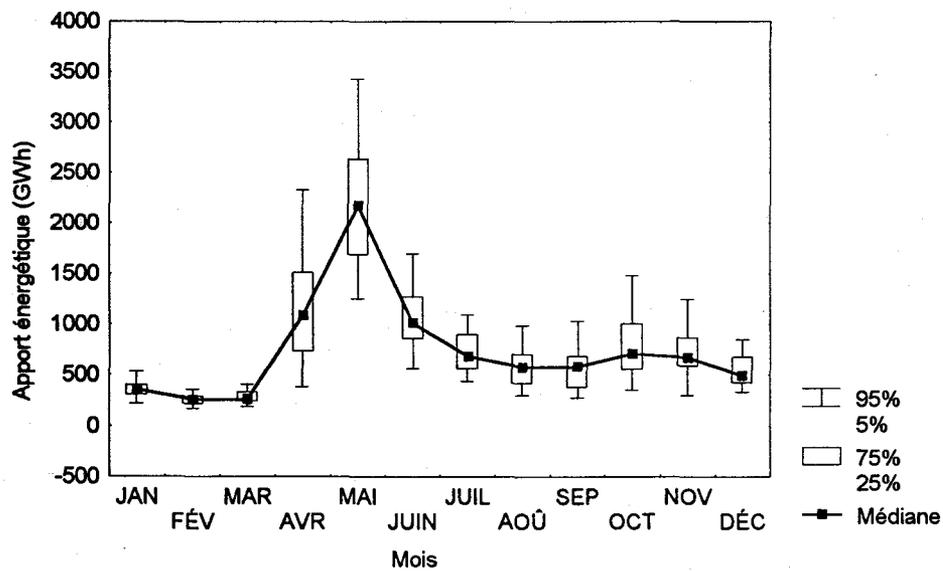
**Figure 3.8.** Distribution des apports énergétiques mensuels - Complexe St-Maurice.

Tableau 3.8. Statistiques de base des apports énergétiques (GWh) du complexe St-Laur..

	JAN	FÉV	MAR	AVR	MAI	JUIN
MOYENNE	925.65	898.14	1041.23	1061.36	1116.21	1082.69
MÉDIANE	902.3	913.2	1038.2	1086.7	1152.8	1119.5
ÉCART-TYPE	76.96	95.15	136.20	150.22	170.17	160.69
CV	0.083	0.105	0.131	0.141	0.152	0.148
Cs	1.046	-0.061	-0.118	-0.672	-0.601	-0.317
MINIMUM	754.0	673.5	734.3	721.9	722.3	738.0
MAXIMUM	1210.6	1080.7	1295.2	1306.9	1391.3	1387.5
TOTAL	47208.6	45805.4	53102.8	54129.4	56927.1	55217.5
	JUIL	AOÛ	SEP	OCT	NOV	DÉC
MOYENNE	1105.98	1093.39	1045.87	1059.90	1009.87	1004.21
MÉDIANE	1116.0	1084.7	1050.8	1053.8	1010.9	1006.3
ÉCART-TYPE	150.31	132.25	117.51	122.55	125.03	120.73
CV	0.135	0.121	0.112	0.115	0.124	0.120
Cs	-0.024	-0.094	-0.187	0.013	0.077	0.555
MINIMUM	800.2	819.7	787.8	820.0	775.5	778.8
MAXIMUM	1425.8	1345.4	1285.9	1340.0	1340.1	1381.2
TOTAL	56405.1	55763.0	53339.8	54055.4	51503.7	51214.9

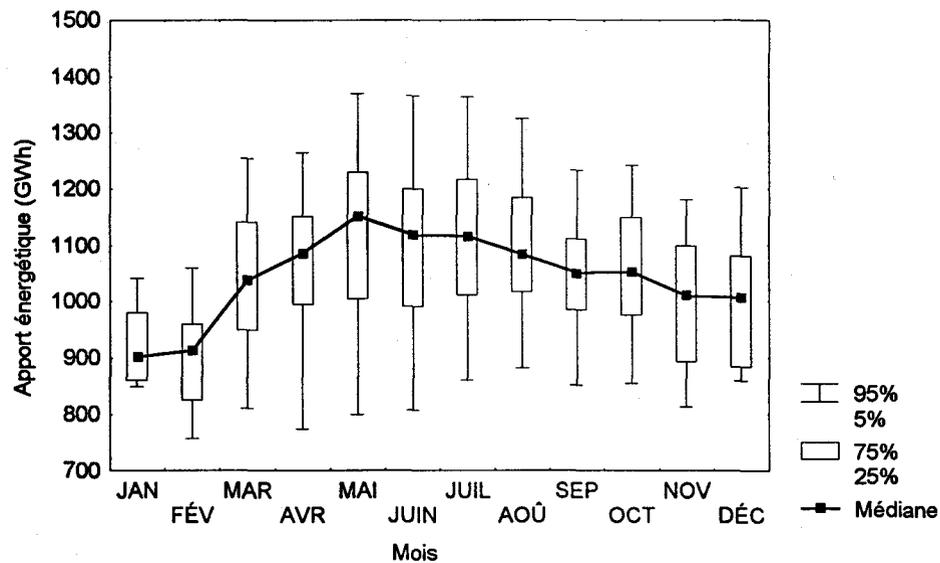
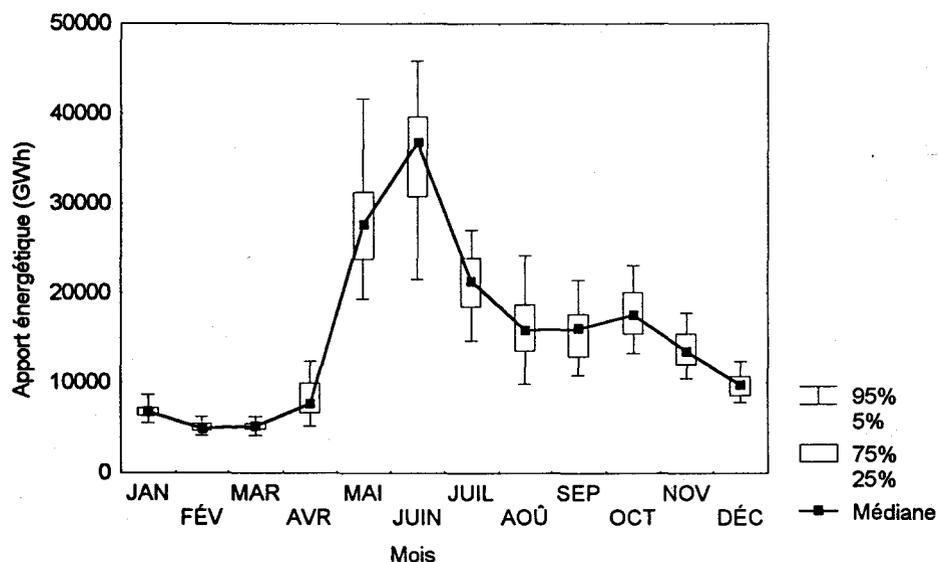


Figure 3.9. Distribution des apports énergétiques mensuels - Complexe St-Laurent.

Tableau 3.9. Statistiques de base des apports énergétiques (GWh) de la série agrégée.

	JAN	FÉV	MAR	AVR	MAI	JUIN
MOYENNE	6819.55	5032.17	5153.45	8434.42	28207.13	35225.41
MÉDIANE	6797.7	4915.4	5168.0	7693.0	27663.7	36757.5
ÉCART-TYPE	800.70	598.46	603.29	2566.63	5798.61	7116.27
CV	0.117	0.119	0.117	0.304	0.205	0.202
Cs	0.707	0.409	0.069	1.198	0.799	-0.354
MINIMUM	5347.5	3816.9	3752.2	4898.6	18454.3	19205.5
MAXIMUM	9143.6	6472.1	6742.6	17706.3	42124.5	48250.5
TOTAL	347797.4	256640.9	262826.1	430155.2	1438563.8	1796496.4
	JUIL	AOÛ	SEP	OCT	Nov	DÉC
MOYENNE	21187.86	16422.83	15788.21	17981.34	13780.68	9794.29
MÉDIANE	21378.6	15911.5	16091.0	17596.5	13532.8	9831.0
ÉCART-TYPE	3985.94	4212.06	3270.81	3175.12	2203.93	1418.32
CV	0.188	0.256	0.207	0.176	0.159	0.145
Cs	-0.530	0.494	0.530	0.306	0.049	0.182
MINIMUM	9569.4	8005.0	9981.1	11208.2	8173.5	6388.7
MAXIMUM	28296.4	27422.4	24308.4	26413.2	18095.3	12819.8
TOTAL	1080581.3	837564.3	805198.7	917048.6	702815	499508.8

**Figure 3.10.** Distribution des apports énergétiques mensuels - Série agrégée.

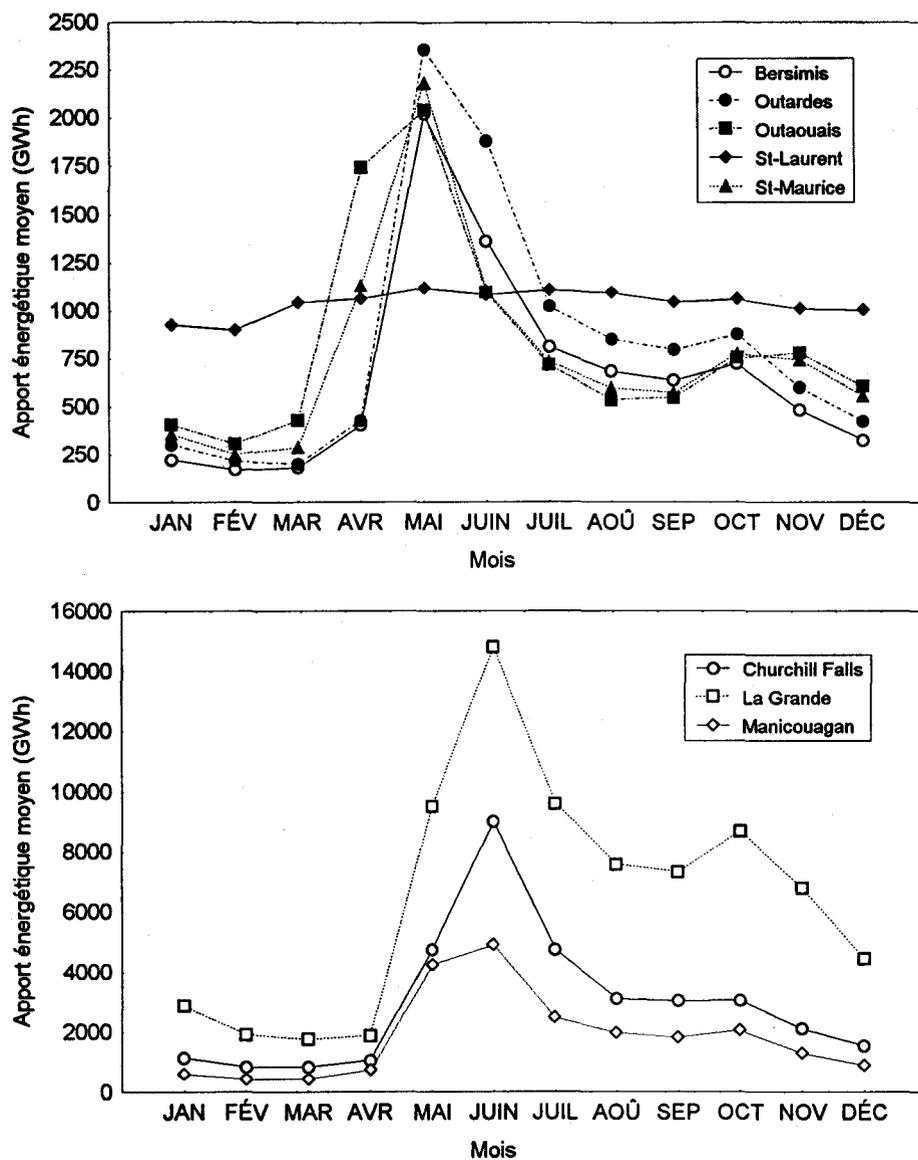


Figure 3.11. Apports énergétiques moyens mensuels des huit complexes.

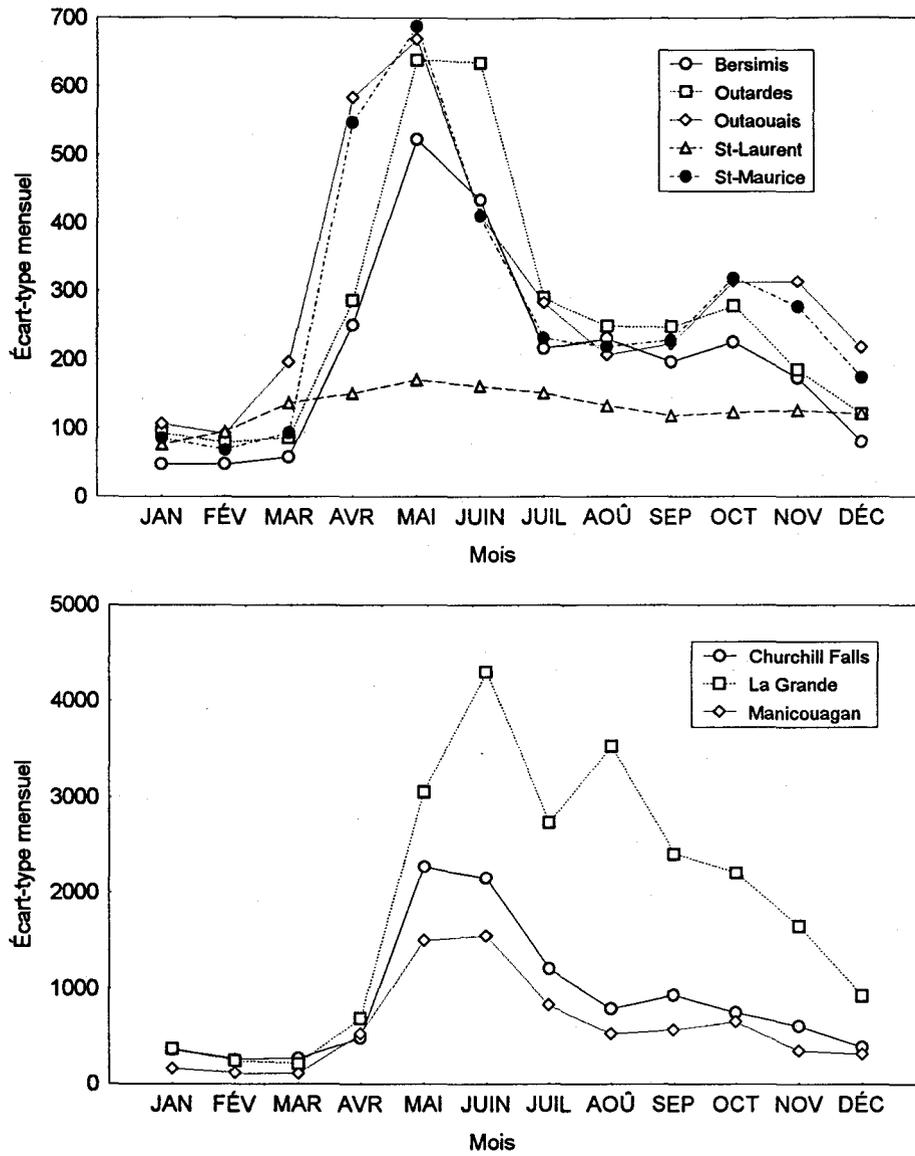


Figure 3.12. Écart-type mensuel des apports énergétiques (GWh) des huit complexes.

3.2.2 Corrélations mensuelles

Les tableaux 3.10 à 3.18 donnent les matrices de corrélations entre les apports énergétiques mensuels des neuf séries considérées. Les corrélations ombragées sont significativement différentes de zéro à un niveau de signification de 1%. On remarque en particulier:

- qu'une forte corrélation est observée presque systématiquement entre les apports de deux mois consécutifs pour les complexes Churchill Falls (tableau 3.11), La Grande (tableau 3.12), Outaouais (tableau 3.15), St-Maurice (tableau 3.16), St-Laurent (tableau 3.17) et la série des apports agrégés (tableau 3.18). Ces corrélations sont toutefois moins élevées pour les complexes Bersimis (tableau 3.10), Manicouagan (tableau 3.13) et Aux Outardes (tableau 3.14).
- que toutes les corrélations mensuelles sont élevées pour le complexe St-Laurent. Les apports énergétiques mensuels sont donc très liés quel que soit le délai (1 à 12 mois).
- que la matrice de corrélations des apports énergétiques agrégés présente une structure assez semblable aux corrélations observées pour les complexes Churchill Falls et La Grande, les systèmes contribuant le plus à ces apports (figure 3.1).

Tableau 3.10. Corrélations des apports énergétiques mensuels pour Bersimis.

	JAN	FÉV	MAR	AVR	MAI	JUIN	JUIL	AOÛ	SEP	OCT	NOV	DÉC
JAN	1.000											
FÉV	0.180	1.000										
MAR	0.198	0.475	1.000									
AVR	0.064	0.296	0.241	1.000								
MAI	0.069	0.292	0.389	0.288	1.000							
JUIN	0.037	0.078	0.136	-0.239	-0.048	1.000						
JUIL	-0.186	-0.137	0.041	-0.221	-0.195	0.336	1.000					
AOÛ	0.086	0.074	0.072	0.019	0.100	0.003	0.096	1.000				
SEP	0.149	0.125	-0.068	0.249	0.038	-0.050	0.066	0.376	1.000			
OCT	0.132	0.028	0.087	0.016	-0.013	-0.270	-0.227	0.087	0.268	1.000		
NOV	0.008	-0.131	0.054	0.001	-0.067	-0.029	-0.034	0.008	0.036	0.261	1.000	
DÉC	-0.003	-0.130	0.018	0.107	0.024	0.073	0.045	-0.036	0.044	0.105	0.633	1.000

Tableau 3.11. Corrélations des apports énergétiques mensuels pour Churchill Falls.

	JAN	FÉV	MAR	AVR	MAI	JUIN	JUIL	AoÛ	SEP	OCT	NOV	DÉC
JAN	1.000											
FÉV	0.465	1.000										
MAR	0.440	0.628	1.000									
AVR	0.276	0.141	0.274	1.000								
MAI	0.065	-0.053	0.173	0.614	1.000							
JUIN	0.385	0.426	0.278	-0.300	-0.264	1.000						
JUIL	0.330	0.220	0.166	-0.062	0.011	0.692	1.000					
AoÛ	0.296	0.170	0.036	0.077	0.017	0.403	0.550	1.000				
SEP	0.142	0.047	0.037	0.065	0.073	0.268	0.258	0.742	1.000			
OCT	-0.123	-0.078	0.126	0.130	0.235	-0.114	0.050	0.353	0.470	1.000		
NOV	0.059	-0.018	0.122	0.364	0.289	-0.183	-0.040	0.259	0.260	0.614	1.000	
DÉC	-0.052	0.205	0.270	0.278	0.259	-0.060	-0.051	0.058	0.175	0.478	0.565	1.000

Tableau 3.12. Corrélations des apports énergétiques mensuels pour La Grande.

	JAN	FÉV	MAR	AVR	MAI	JUIN	JUIL	AoÛ	SEP	OCT	NOV	DÉC
JAN	1.000											
FÉV	0.840	1.000										
MAR	0.524	0.674	1.000									
AVR	-0.086	0.111	0.343	1.000								
MAI	-0.312	-0.232	0.023	0.542	1.000							
JUIN	0.062	-0.035	-0.227	-0.390	-0.265	1.000						
JUIL	-0.049	-0.010	-0.227	-0.216	-0.283	0.647	1.000					
AoÛ	-0.162	-0.076	-0.282	-0.172	-0.202	0.260	0.707	1.000				
SEP	-0.293	-0.210	-0.142	-0.062	-0.025	0.131	0.580	0.661	1.000			
OCT	0.069	-0.063	-0.207	-0.207	-0.190	0.030	0.183	0.254	0.195	1.000		
NOV	0.162	0.101	-0.061	-0.228	-0.370	-0.012	0.132	0.168	0.043	0.810	1.000	
DÉC	0.154	0.098	-0.109	-0.204	-0.358	-0.066	0.021	-0.034	-0.057	0.609	0.798	1.000

Tableau 3.13. Corrélations des apports énergétiques mensuels pour Manicouagan.

	JAN	FÉV	MAR	AVR	MAI	JUIN	JUIL	AOÛ	SEP	OCT	NOV	DÉC
JAN	1.000											
FÉV	0.830	1.000										
MAR	0.731	0.808	1.000									
AVR	0.047	-0.097	-0.060	1.000								
MAI	0.008	-0.044	-0.056	0.355	1.000							
JUIN	0.267	0.191	0.204	-0.206	-0.246	1.000						
JUIL	-0.156	-0.101	-0.046	-0.146	-0.388	0.657	1.000					
AOÛ	-0.028	-0.049	0.052	0.061	-0.281	0.108	0.189	1.000				
SEP	-0.131	-0.106	-0.040	0.307	0.015	-0.034	0.017	0.276	1.000			
OCT	0.234	0.124	-0.025	-0.047	-0.173	0.082	-0.086	0.199	0.159	1.000		
NOV	0.195	0.133	-0.025	-0.225	-0.306	0.226	0.224	0.192	-0.181	0.397	1.000	
DÉC	0.138	0.097	0.039	0.089	-0.132	0.134	0.233	0.347	-0.122	0.263	0.569	1.000

Tableau 3.14. Corrélations des apports énergétiques mensuels pour Aux Outardes.

	JAN	FÉV	MAR	AVR	MAI	JUIN	JUIL	AOÛ	SEP	OCT	NOV	DÉC
JAN	1.000											
FÉV	0.736	1.000										
MAR	0.553	0.558	1.000									
AVR	0.073	0.055	0.298	1.000								
MAI	0.020	0.208	0.116	0.321	1.000							
JUIN	0.164	0.248	0.324	-0.209	-0.054	1.000						
JUIL	-0.140	-0.085	0.139	-0.170	-0.193	0.518	1.000					
AOÛ	-0.125	-0.138	-0.129	-0.084	-0.081	0.102	0.313	1.000				
SEP	-0.042	-0.132	-0.080	0.169	-0.083	0.017	0.150	0.502	1.000			
OCT	0.312	0.217	-0.025	-0.011	-0.079	-0.180	-0.177	0.254	0.309	1.000		
NOV	0.090	-0.013	0.066	-0.103	-0.236	0.061	0.079	0.190	-0.022	0.429	1.000	
DÉC	-0.067	0.002	-0.011	0.026	0.001	0.056	0.270	0.195	-0.123	0.235	0.599	1.000

Tableau 3.15. Corrélations des apports énergétiques mensuels pour Outaouais.

	JAN	FÉV	MAR	AVR	MAI	JUIN	JUIL	AoÛ	SEP	OCT	NOV	DÉC
JAN	1.000											
FÉV	0.478	1.000										
MAR	0.092	0.540	1.000									
AVR	0.098	0.357	0.577	1.000								
MAI	0.081	0.097	-0.177	-0.230	1.000							
JUIN	0.137	0.223	-0.073	-0.173	0.712	1.000						
JUIL	0.089	0.246	-0.117	-0.100	0.383	0.603	1.000					
AoÛ	0.025	0.075	-0.163	-0.108	0.250	0.302	0.449	1.000				
SEP	-0.120	-0.123	-0.238	-0.267	0.085	0.133	0.347	0.646	1.000			
OCT	0.179	0.008	-0.047	0.064	-0.066	-0.088	0.010	0.414	0.513	1.000		
NOV	0.212	0.054	-0.147	0.032	-0.020	-0.006	0.084	0.319	0.285	0.689	1.000	
DÉC	0.153	0.026	0.034	0.060	-0.148	-0.052	0.106	0.333	0.230	0.470	0.673	1.000

Tableau 3.16. Corrélations des apports énergétiques mensuels pour St-Maurice.

	JAN	FÉV	MAR	AVR	MAI	JUIN	JUIL	AoÛ	SEP	OCT	NOV	DÉC
JAN	1.000											
FÉV	0.655	1.000										
MAR	0.221	0.717	1.000									
AVR	0.055	0.275	0.658	1.000								
MAI	0.078	0.183	0.135	-0.128	1.000							
JUIN	0.169	0.249	0.055	-0.279	0.676	1.000						
JUIL	0.151	0.209	-0.069	-0.210	0.097	0.435	1.000					
AoÛ	0.051	0.074	0.034	-0.087	-0.051	0.192	0.370	1.000				
SEP	0.012	-0.011	-0.098	-0.173	-0.052	0.068	0.285	0.593	1.000			
OCT	0.115	-0.128	-0.129	-0.017	-0.155	-0.076	-0.055	0.455	0.513	1.000		
NOV	0.090	-0.049	-0.113	-0.109	-0.037	-0.040	0.046	0.256	0.320	0.726	1.000	
DÉC	0.015	-0.003	-0.047	-0.019	-0.207	0.016	0.287	0.349	0.232	0.436	0.649	1.000

Tableau 3.17. Corrélations des apports énergétiques mensuels pour St-Laurent.

	JAN	FÉV	MAR	AVR	MAI	JUIN	JUIL	AOÛ	SEP	OCT	NOV	DÉC
JAN	1.000											
FÉV	0.817	1.000										
MAR	0.803	0.908	1.000									
AVR	0.722	0.815	0.917	1.000								
MAI	0.691	0.788	0.867	0.952	1.000							
JUIN	0.627	0.738	0.797	0.875	0.956	1.000						
JUIL	0.583	0.712	0.746	0.813	0.890	0.962	1.000					
AOÛ	0.571	0.691	0.714	0.776	0.843	0.907	0.971	1.000				
SEP	0.556	0.660	0.696	0.766	0.818	0.852	0.911	0.967	1.000			
OCT	0.515	0.527	0.650	0.735	0.751	0.745	0.790	0.853	0.925	1.000		
NOV	0.439	0.435	0.559	0.654	0.654	0.632	0.662	0.729	0.818	0.942	1.000	
DÉC	0.423	0.422	0.517	0.616	0.584	0.544	0.563	0.636	0.726	0.830	0.892	1.000

Tableau 3.18. Corrélations des apports énergétiques mensuels: série agrégée.

	JAN	FÉV	MAR	AVR	MAI	JUIN	JUIL	AOÛ	SEP	OCT	NOV	DÉC
JAN	1.000											
FÉV	0.781	1.000										
MAR	0.486	0.687	1.000									
AVR	0.142	0.103	0.474	1.000								
MAI	-0.178	-0.062	0.250	0.467	1.000							
JUIN	0.399	0.398	0.211	-0.290	-0.234	1.000						
JUIL	0.382	0.288	0.097	-0.087	-0.215	0.726	1.000					
AOÛ	0.212	0.089	-0.068	0.064	0.015	0.343	0.682	1.000				
SEP	0.083	-0.136	-0.080	0.201	0.065	0.147	0.554	0.676	1.000			
OCT	0.229	0.059	0.135	0.096	0.166	0.126	0.197	0.269	0.378	1.000		
NOV	0.197	0.084	0.115	0.020	0.083	0.166	0.206	0.227	0.266	0.747	1.000	
DÉC	0.120	0.099	0.141	0.048	-0.055	0.190	0.244	0.161	0.264	0.547	0.730	1.000

3.3 Apports énergétiques annuels

Cette section est consacrée aux caractéristiques statistiques des apports énergétiques annuels de 1943 à 1993. Les figures 3.13 à 3.21 présentent les courbes chronologiques des neuf séries d'apports énergétiques annuels étudiées.

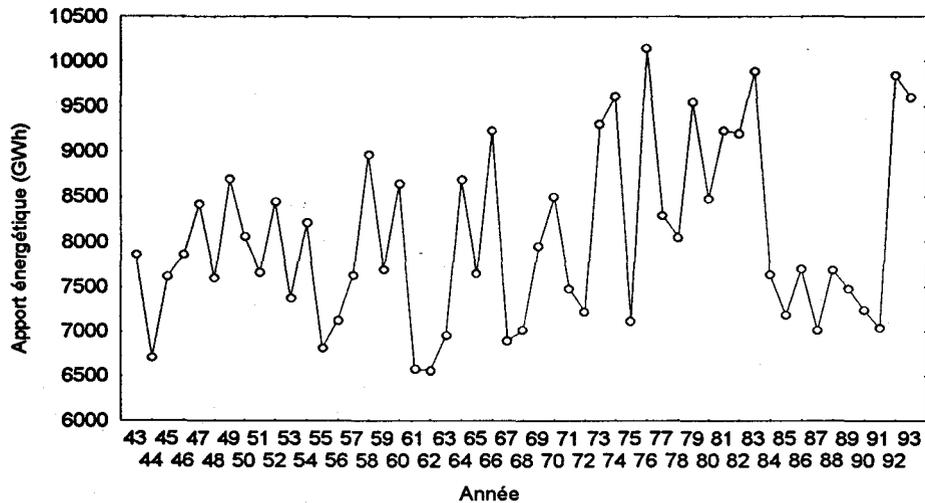


Figure 3.13. Apports énergétiques annuels du complexe Bersimis.

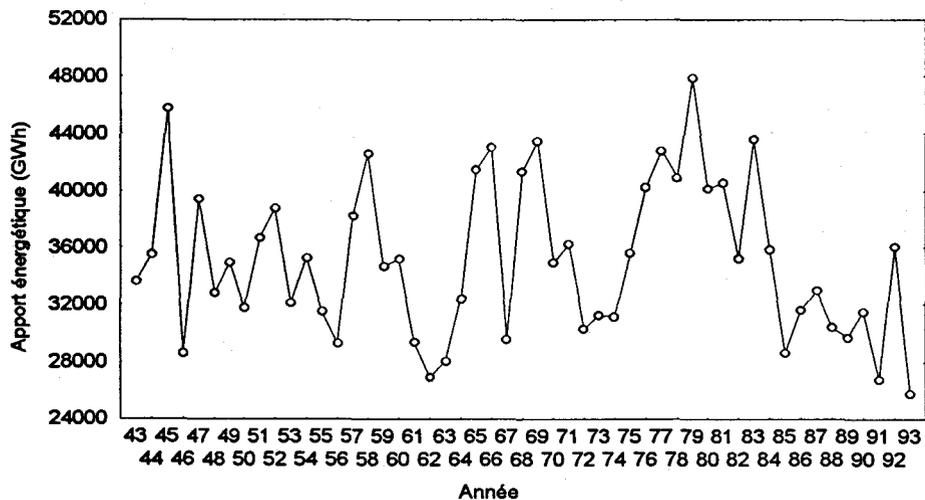


Figure 3.14. Apports énergétiques annuels du complexe Churchill Falls.

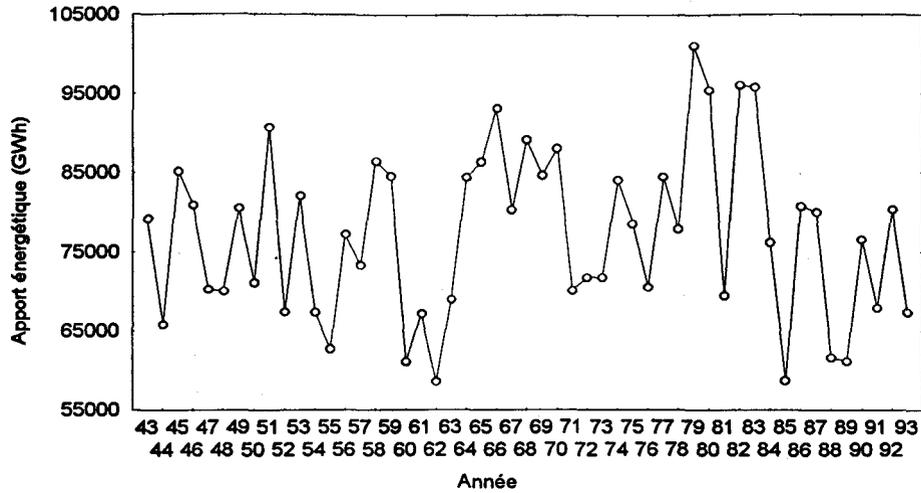


Figure 3.15. Apports énergétiques annuels du complexe La Grande.

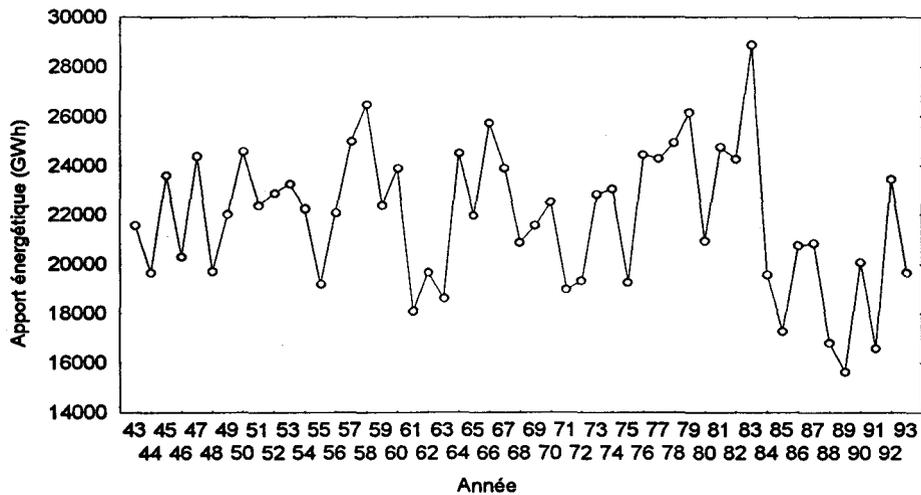


Figure 3.16. Apports énergétiques annuels du complexe Manicouagan.

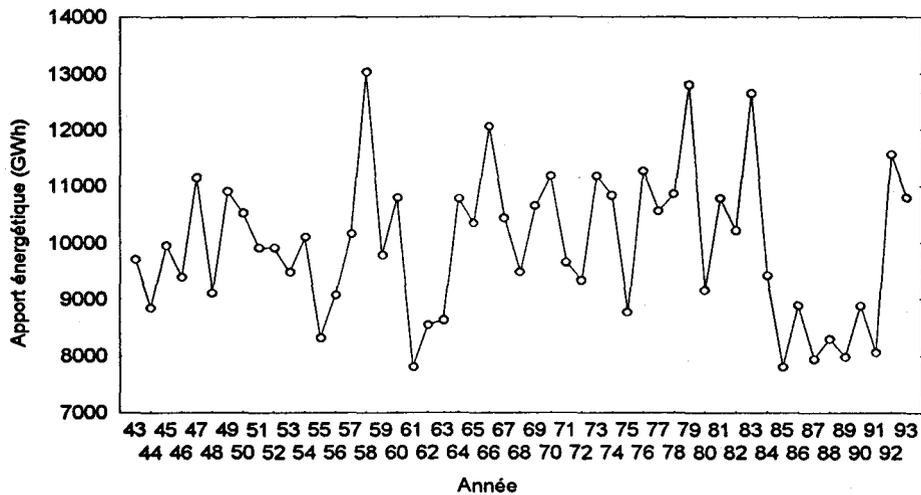


Figure 3.17. Apports énergétiques annuels du complexe Aux Outardes.

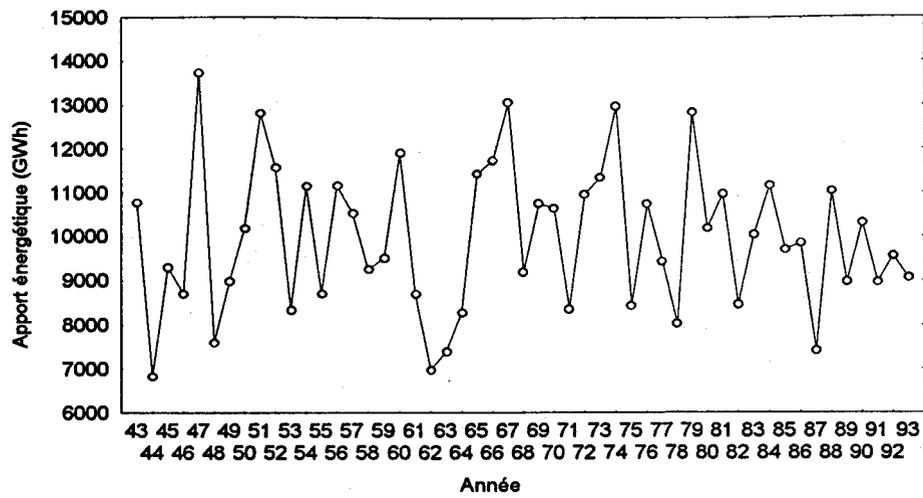


Figure 3.18. Apports énergétiques annuels du complexe Outaouais.

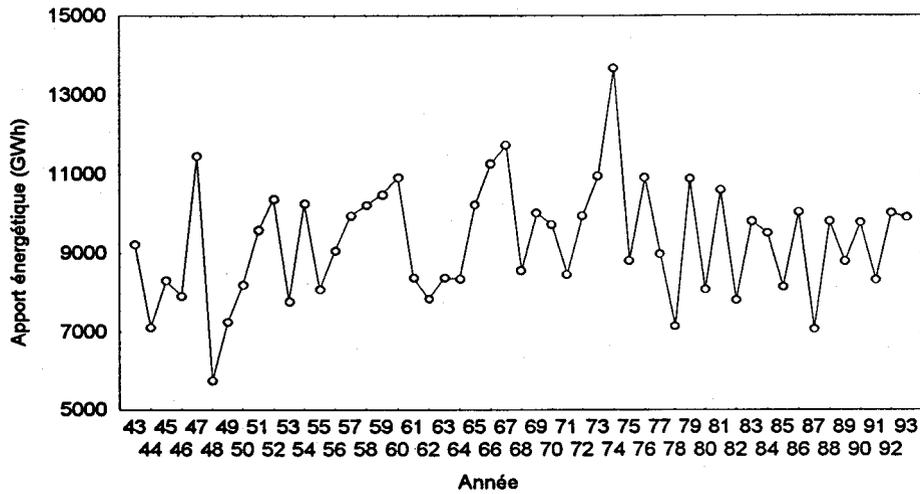


Figure 3.19. Apports énergétiques annuels du complexe St-Maurice.

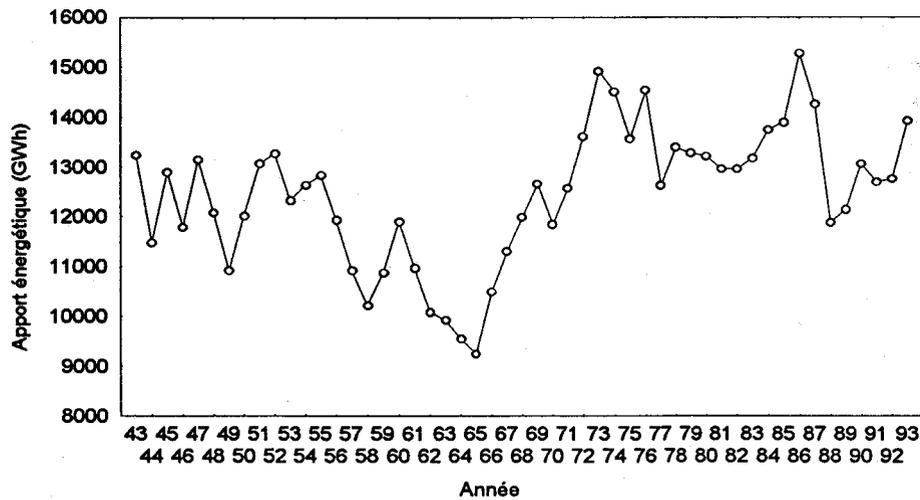


Figure 3.20. Apports énergétiques annuels du complexe St-Laurent.

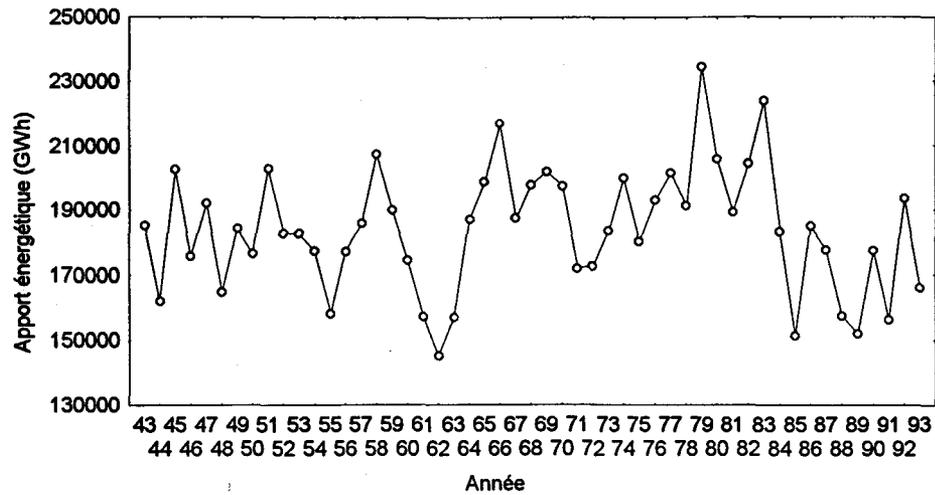


Figure 3.21. Apports énergétiques annuels agrégés.

3.3.1 Statistiques de base

Le tableau 3.19 donne les principales caractéristiques statistiques des apports énergétiques des huit complexes. Leur distribution est représentée par des diagrammes en boîte à la figure 3.22. En examinant cette figure, on peut constater que:

- la dispersion des apports énergétiques varie en fonction de la dimension des complexes (plus le système est grand, plus la variabilité des apports est grande).
- leur distribution semble symétrique. On observe d'ailleurs de faibles valeurs du coefficient d'asymétrie (cf. tableau 3.19).

Les statistiques de base des apports énergétiques annuels agrégés sont présentées au tableau 3.20. On remarquera, entre autres, la faible valeur du coefficient d'asymétrie C_s qui suggère que les apports annuels agrégés sont distribués selon une loi normale. Ce résultat n'est pas étonnant puisqu'en vertu du théorème de la limite centrale une somme de variables aléatoires aura tendance à se distribuer selon une loi normale. Le test de normalité de Shapiro-Wilk (cf. section 2.3.2) a été appliqué aux neuf séries d'apports énergétiques annuels et confirme d'ailleurs cette hypothèse (cf. tableau 3.21). Seule la série d'apports annuels au complexe Bersimis n'est pas distribuée selon une loi normale au niveau de signification de 5%.

Tableau 3.19. Statistiques de base des apports énergétiques annuels des huit complexes.

	BERSIMIS	CHUR. FALLS	LA GRANDE	MANIC.
MOYENNE	8029.05	35157.94	77121.38	21866.32
INT. CONF. (95%)	[7758 ; 8300]	[33629 ; 36686]	[74149 ; 80093]	[21078 ; 22654]
MÉDIANE	7702.7	34945.2	77882.1	22074.1
ÉCART-TYPE	963.20	5434.05	10566.52	2800.56
Cv	0.120	0.154	0.137	0.128
Cs	0.519	0.373	0.185	-0.059
MINIMUM	651.4	25790.3	58546.0	15636.6
MAXIMUM	10149.6	47884.3	101050.6	28882.1
TOTAL	409481.6	1793055.3	3933190.4	1115182.5

	OUTARDES	OUTAOUAIS	ST-Maurice	ST-LAURENT
MOYENNE	9955.88	9960.94	9289.90	12444.51
INT. CONF. (95%)	[9592 ; 10319]	[9494 ; 10427]	[8880 ; 9699]	[12058 ; 12830]
MÉDIANE	9908.3	9843.5	9515.2	12659.3
ÉCART-TYPE	1292.92	1658.49	1456.95	1371.92
Cv	0.129	0.166	0.156	0.110
Cs	0.325	0.222	0.243	-0.354
MINIMUM	7804.9	6829.6	5754.4	9259.1
MAXIMUM	13033.8	13731.0	13674.2	15273.2
TOTAL	507750.1	508008.3	473785.2	634670.1

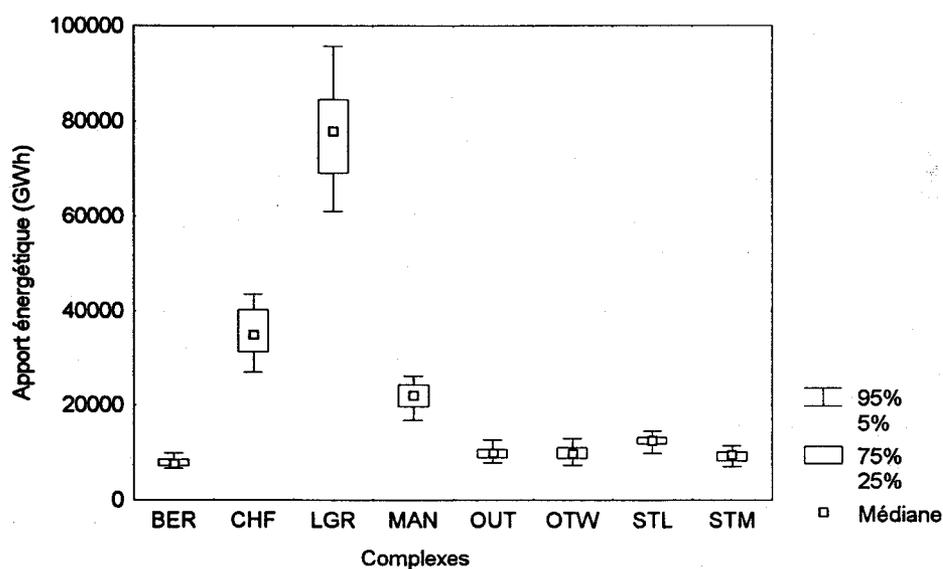
**Figure 3.22.** Distribution des apports énergétiques annuels des huit complexes.

Tableau 3.20. Caractéristiques statistiques des apports énergétiques annuels agrégés.

MOYENNE	INT. CONF (95%)	MÉDIANE	ÉCART-TYPE	CV	Cs	MINIMUM	MAXIMUM	TOTAL
183 822	[178 456 ; 189 188]	184 238	19 079	0.104	0.156	145 155	234 438	9 374 943

Tableau 3.21. Test de normalité de Shapiro-Wilk (apports énergétiques annuels).

	BER	CHF	LGR	MAN	OUT	OTW	STM	STL	AGREG
W	0.9359*	0.9578	0.9662	0.9848	0.9637	0.9729	0.9826	0.9727	0.9793
Prob.	0.0119	0.1209	0.2690	0.8801	0.2110	0.4605	0.8153	0.4538	0.6971

* Significatif au seuil de 5%.

3.3.2 Corrélations

Une des façons de caractériser les apports énergétiques annuels du réseau d'Hydro-Québec est d'étudier les corrélations entre les séries d'apports des différents complexes. Le tableau 3.22 présente la matrice de corrélations pour les huit systèmes étudiés. Les corrélations ombragées sont significatives au niveau de signification de 1%.

Tableau 3.22. Corrélations entre les apports énergétiques annuels des huit complexes.

	BER	CHF	LGR	MAN	OUT	OTW	STM	STL
BER	1.000							
CHF	0.421	1.000						
LGR	0.376	0.582	1.000					
MAN	0.645	0.649	0.575	1.000				
OUT	0.812	0.622	0.528	0.860	1.000			
OTW	0.373	0.274	0.206	0.365	0.435	1.000		
STM	0.485	0.201	0.139	0.374	0.506	0.822	1.000	
STL	0.261	0.012	-0.014	-0.041	-0.035	0.228	0.176	1.000

L'examen de la matrice des corrélations permet, principalement, de constater que:

- des corrélations généralement élevées sont obtenues pour des complexes voisins (figure 1.1). En particulier, les corrélations entre les apports énergétiques annuels des complexes Manicouagan/Aux Outardes, Outardes/Bersimis et Outaouais/St-Maurice sont respectivement 0.86, 0.81 et 0.82.

- les apports énergétiques annuels du complexe St-Laurent ne sont pas corrélés avec ceux des autres complexes. Ces apports se comportent donc de façon tout à fait différente.

Pour mieux interpréter les différents liens existants entre les apports énergétiques des différents complexes, une analyse en composantes principales a été réalisée. Cette technique est très utile puisqu'elle permet de dégager l'essentiel de l'information contenue dans la matrice des corrélations et d'obtenir une représentation graphique des liens entre les complexes. On retrouve une description détaillée des aspects théoriques de l'analyse en composantes principales dans Lebart, Morineau et Tabard (1977). Nous donnons ici une brève interprétation géométrique de la méthode.

Considérons le tableau des apports énergétiques annuels dont les colonnes représentent les complexes et les lignes correspondent aux années d'observations. Pour comprendre les opérations pratiquées par ce type de technique, il est utile de représenter géométriquement les années et les complexes du tableau de données par des points d'un espace à p ou n dimensions (ici, $p=8$ complexes et $n=51$ années), points dont les coordonnées sont précisément les éléments du tableau. Les 51 années d'observations peuvent donc être considérées comme 51 points d'un espace à 8 dimensions, alors que les 8 complexes sont des points d'un espace à 51 dimensions.

Si les valeurs prises par deux complexes particuliers sont très voisines pour l'ensemble des années d'observations, ces systèmes seront représentés par deux points très proches dans l'espace des années (51 dimensions). Cela peut vouloir dire que leurs apports énergétiques sont liés par une relation particulière et que leurs caractéristiques statistiques sont semblables. Toutefois, l'interprétation de ces proximités dans les deux espaces est très lourde à faire (8 et 51 dimensions). L'analyse en composantes principales consiste à représenter graphiquement les proximités entre les points au terme d'une série d'ajustements: afin de simplifier le problème, on cherche, dans l'espace des années (à 51 dimensions) des sous-espaces (facteurs) de faibles dimensions (une, deux ou trois) s'ajustant le mieux possible aux nuages de points-complexes, c'est-à-dire de façon à ce que les proximités, mesurées dans les sous-espaces, reflètent autant que possible les proximités réelles.

Pour ce faire, on détermine des axes factoriels, expliquant chacun un pourcentage décroissant de la variance totale, sur lesquels on projette les points-complexes. Par exemple, en ne retenant que les 3 premiers axes expliquant le plus de variabilité, nous obtenons une représentation graphique des points-complexes sur un sous-espace à trois dimensions.

L'interprétation est donc plus facile, cependant la qualité de la représentation dépend de la variance expliquée cumulée des trois axes factoriels.

La Figure 3.23 présente les résultats de l'analyse en composantes principales. On y donne la représentation des points-complexes dans l'espace des trois premiers facteurs. Le pourcentage de variance totale expliquée par les trois axes factoriels (représentation à 3 dimensions au lieu de 51) est de 80.7%.

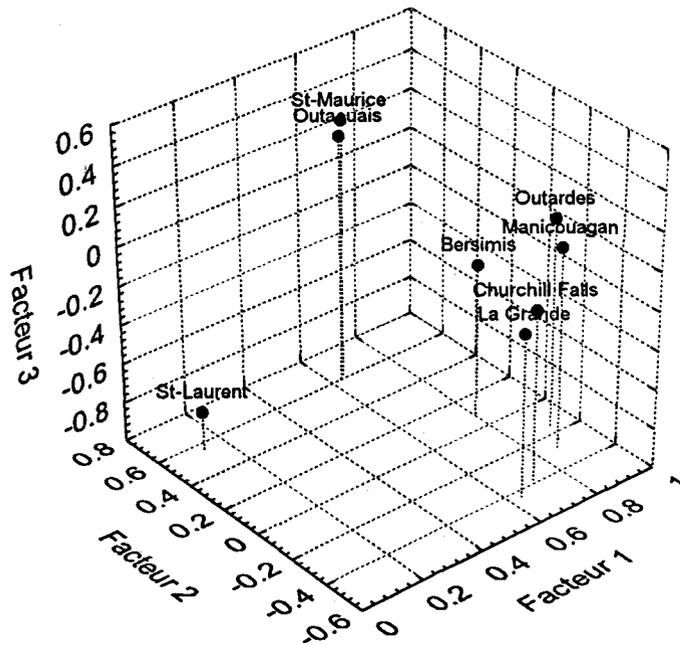


Figure 3.23. Représentation des complexes dans l'espace des 3 premiers facteurs.

La figure 3.23 montre très clairement trois groupes de complexes: St-Laurent, Outaouais/St-Maurice et Bersimis/Outardes/Manicouagan/Churchill Falls/La Grande. Les distances entre les complexes dans l'espace des trois premiers facteurs traduisent bien les corrélations obtenues (cf. tableau 3.22). Ainsi, par exemple, les apports énergétiques du St-Laurent ont un comportement tout à fait différent des autres systèmes. En approfondissant l'analyse, on peut aussi remarquer que le groupe formé des complexes Bersimis, Outardes, Manicouagan, Churchill Falls et La Grande pourrait être divisé. En particulier, si on examine la représentation des points-complexes dans l'espace des 2 premiers facteurs (cf. figure 3.24), on remarque que le système Bersimis se distingue quelque peu.

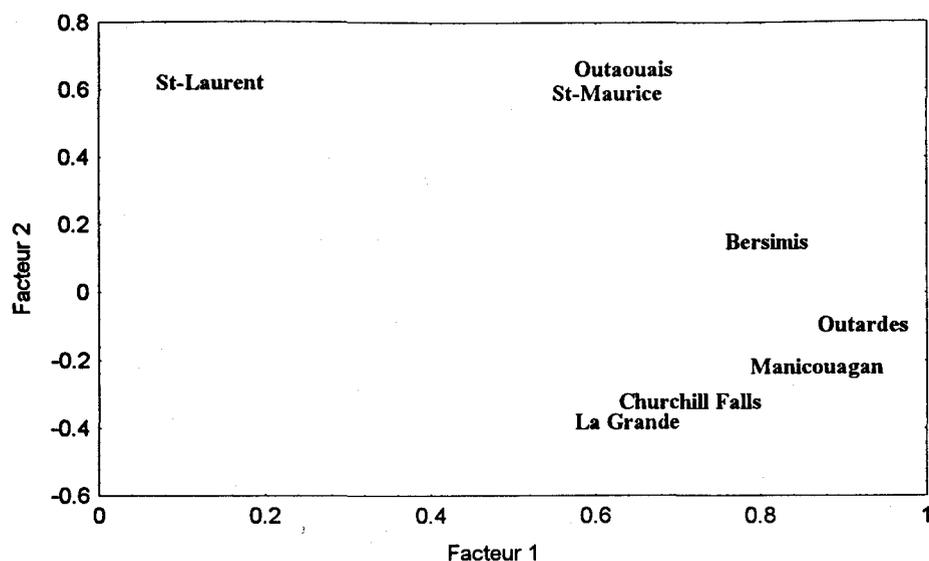


Figure 3.24. Représentation des complexes dans l'espace des 2 premiers facteurs.

3.3.3 Autocorrélations

Une des caractéristiques statistiques les plus importantes lors de la modélisation de séries chronologiques est l'autocorrélation. Les autocorrélations sont des mesures qui indiquent comment une série chronologique est liée à elle-même dans le temps. Plus précisément, une autocorrélation mesure l'importance de la corrélation entre les valeurs de la série séparées par un nombre de périodes fixe dans le temps. L'intervalle de temps qui sépare les valeurs est appelé l'ordre de l'autocorrélation. Par exemple, l'autocorrélation d'ordre 1 évalue la façon dont les observations successives (intervalle de temps d'une période) sont corrélées entre elles. L'autocorrélation d'ordre 2 mesure comment les valeurs de la série, distantes de deux périodes, sont corrélées entre elles. De manière générale, l'autocorrélation d'ordre k est notée ρ_k .

Plusieurs estimateurs de l'autocorrélation ρ_k ont été proposés dans la littérature et leurs propriétés ont été étudiées, en particulier, par Jenkins et Watts (1968). Dans cette étude, il est recommandé d'utiliser l'estimateur $\hat{\rho}_k$ dont l'expression mathématique est donnée par:

$$\hat{\rho}_k = \frac{\sum_{t=1}^{n-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^n (x_t - \bar{x})^2} \quad (3.1)$$

où x_1, x_2, \dots, x_n est une série chronologique de taille n et de moyenne \bar{x} .

Le tableau 3.23 présente les autocorrélations d'ordre 1 à 5 des apports énergétiques annuels des huit complexes et de la série agrégée. Les valeurs ombragées sont statistiquement différentes de zéro au niveau de signification 1%.

Tableau 3.23. Autocorrélations des apports énergétiques annuels (1943-1993).

COMPLEXES									
ORDRE	BER	CHF	LGR	MAN	OUT	OTW	STM	STL	AGREG
1	.163	.284	.194	.244	.195	-.070	-.020	.748	.356
2	.083	.240	.062	.203	.161	-.001	.187	.586	.203
3	.055	.096	.066	.017	-.077	-.271	-.120	.479	.014
4	.013	-.021	.036	.096	-.001	.028	-.082	.417	.072
5	-.058	-.121	-.033	-.086	-.116	.092	.131	.347	-.123

Parmi les complexes, seuls les apports énergétiques annuels du St-Laurent sont significativement autocorrélés. Les autocorrélations sont d'ailleurs significatives pour tous les délais considérés (une à cinq années). Ainsi, pour le complexe St-Laurent, la probabilité d'occurrence d'un apport énergétique annuel donné est fortement influencée par les apports observés les années précédentes. On dira que cette série admet une forte persistance. Toutefois, mentionnons qu'une autocorrélation élevée peut être induite par la non-stationnarité d'une des caractéristiques statistiques de la série (saut de la moyenne, par exemple). Ce sujet est abordé dans la section 3.4 ainsi qu'au chapitre 4.

D'autre part, on constate que la série d'apports énergétiques annuels agrégés possède un coefficient d'autocorrélation d'ordre 1 significatif. La somme des apports énergétiques annuels de tous les complexes d'une année donnée est donc liée au cumul obtenu l'année précédente. Cette information est très importante puisqu'elle peut être utilisée lors de la modélisation de cette série (chapitre 4).

Puisque les apports annuels agrégés sont constitués de la somme des apports de chaque complexe, l'autocorrélation d'ordre 1 (0.356) provient en quelque sorte des caractéristiques statistiques de chacune des séries individuelles. En particulier, on peut se demander si celle-ci est engendrée principalement par la forte autocorrélation d'ordre 1 des apports énergétiques annuels du complexe St-Laurent. Pour examiner cette hypothèse et tenter d'expliquer la provenance de cette autocorrélation, l'exercice théorique suivant a été réalisé.

Considérons d'abord que les apports énergétiques annuels des 7 complexes non autocorrélés significativement sont représentés par des variables aléatoires normales x_1, x_2, \dots, x_7 de moyennes $\mu_i, i = 1, \dots, 7$, et de variances $\sigma_i^2, i = 1, \dots, 7$, respectivement. Cette hypothèse est acceptable puisque le test de Shapiro-Wilk n'a permis de rejeter la normalité, au niveau de signification de 5%, que pour le complexe Bersimis (cf. tableau 3.21). Supposons, de plus, que l'apport énergétique annuel du complexe St-Laurent est représenté par un modèle qui relie les apports successifs, soit un modèle autorégressif d'ordre 1. L'apport énergétique y_t de l'année t au complexe St-Laurent s'exprime alors de la façon suivante:

$$y_t = \mu_y + \rho_y(y_{t-1} - \mu_y) + \varepsilon_t \quad (3.2)$$

où μ_y désigne la moyenne de la série, ε_t un terme d'erreur aléatoire distribué selon une loi normale de moyenne nulle et de variance σ_ε^2 , et ρ_y l'autocorrélation d'ordre 1. En supposant que la série soit stationnaire (cf. chapitre 4), ce modèle semble adéquat suite à une analyse approfondie des résidus.

L'apport énergétique agrégé z_t de l'année t peut donc s'écrire comme suit:

$$\begin{aligned} z_t &= \sum_{i=1}^7 x_{i,t} + y_t \\ &= \sum_{i=1}^7 x_{i,t} + \mu_y + \rho_y(y_{t-1} - \mu_y) + \varepsilon_t \end{aligned} \quad (3.3)$$

Or, la somme de 7 variables aléatoires normales corrélées entre elles (cf. tableau 3.22) est distribuée selon une loi normale de moyenne et de variance données respectivement par:

$$\mu = \sum_{i=1}^7 \mu_i \quad (3.4)$$

$$\sigma^2 = \sum_{i=1}^7 \sigma_i^2 + 2 \sum_{i \neq j}^7 \sigma_i \sigma_j \rho_{ij} \quad (3.5)$$

où ρ_{ij} désigne la corrélation entre les variables x_i et x_j .

Enfin, en faisant l'hypothèse que les apports du St-Laurent sont indépendants de ceux des autres complexes (hypothèse raisonnable puisque les corrélations sont très faibles, c.f. tableau 3.22), on peut montrer (cf. annexe B) que l'autocorrélation d'ordre 1 de la série des z_t (série agrégée) est donnée par l'expression suivante:

$$\begin{aligned}\rho_z &= \left(\frac{\sigma_y^2}{\sigma_y^2 + \sigma^2} \right) \rho_y^3 \\ &= R \rho_y^3\end{aligned}\tag{3.6}$$

où σ_y^2 est la variance de la série des y_t (St-Laurent).

Dans ces conditions, l'autocorrélation d'ordre 1 des apports énergétiques annuels agrégés est donc fonction de l'autocorrélation d'ordre 1 des apports du St-Laurent ainsi que du rapport R . Ce rapport, dont les valeurs appartiennent à l'intervalle $[0, 1]$, représente la contribution de la variance des apports énergétiques du St-Laurent à la variance des apports énergétiques agrégés. On constate, pour ρ_y fixé, que l'autocorrélation est une fonction linéaire et croissante de R , ce qui signifie que plus la variance des apports du St-Laurent est grande par rapport à celle des apports agrégés, plus l'autocorrélation des apports agrégés est élevée. À la limite, si la variance de la somme des apports des 7 complexes non autocorrélés est très faible par rapport à celle des apports du complexe St-Laurent, R tend vers 1 et ρ_z vers sa valeur maximale ρ_y^3 . À l'inverse, si la contribution de la variance des apports énergétiques du St-Laurent est négligeable, R tend vers 0 et ρ_z est nul.

Le rapport R est ici égal à 0.0052 et l'autocorrélation d'ordre 1 ρ_z déduite de l'équation (3.6) est alors de 0.002. Or, l'autocorrélation des apports énergétiques agrégés est égale à 0.356 (cf. tableau 3.23). Ainsi, le coefficient d'autocorrélation d'ordre 1 de la série agrégée n'est pas influencée, comme on aurait pu le croire, par celui des apports du complexe St-Laurent. La valeur obtenue à partir des observations (0.356) proviendrait plutôt de l'accumulation de faibles autocorrélations positives de certaines séries individuelles dont la contribution est plus importante (en particulier, Churchill Falls, La Grande et Manicouagan). Cela tend à démontrer, même si les autocorrélations d'ordre 1 de ces séries ne sont pas significatives à 1% (dû à la taille limitée de l'échantillon), qu'elles sont probablement réelles.

Un exercice théorique semblable pourrait être réalisé en représentant plutôt les apports énergétiques de tous les complexes par des modèles autorégressifs d'ordre 1 (éq. 3.2). Toutefois, le problème devient beaucoup plus complexe analytiquement. En effet, l'autocorrélation des apports agrégés est alors fonction, non seulement des autocorrélations d'ordre 1 de chacune des séries, mais aussi des autocorrélations croisées d'ordre 1.

3.4 Étude préliminaire de la stationnarité

Les différents modèles de planification de la production à Hydro-Québec reposent, en particulier, sur les caractéristiques statistiques issues des séries d'apports énergétiques annuels (moyenne, écart-type et autocorrélation). Ces données ont été modifiées de façon à obtenir une plage commune de cinquante années d'observations (cf. chapitre 2). Comme il a été mentionné précédemment, ces interventions soulèvent quelques questions quant à la validité des caractéristiques statistiques. En particulier, doit-on nécessairement, compte tenu de la possible présence de non-stationnarité engendrée par ces interventions, utiliser les séries d'observations s'échelonnant sur cinquante années pour le calcul de ces statistiques? D'autre part, l'objectif principal du sous-projet 4 est de construire un modèle permettant de représenter les apports énergétiques annuels agrégés et de prédire les écarts énergétiques des années à venir. Or, les modèles statistiques classiques de séries chronologiques en hydrologie reposent en particulier sur l'hypothèse de stationnarité de la moyenne.

Une attention particulière a donc été portée à la stationnarité. Quelques résultats d'analyses préliminaires sont présentés dans cette section. Nous nous intéressons principalement à la stationnarité de la moyenne des apports énergétiques annuels.

3.4.1 Définition

En pratique, une série chronologique est dite *stationnaire* si les observations s'éloignent assez uniformément autour d'un niveau fixé (aléatoirement). Les données peuvent toutefois s'écarter de ce niveau mais elles auront tendance à se redistribuer nécessairement autour de cette valeur. Cette déviation peut se maintenir sur une période de temps plus ou moins longue selon la corrélation temporelle entre les observations. Ce phénomène est appelé ici *persistance* et peut être mesuré à l'aide du coefficient d'autocorrélation (éq. 3.1). Plus le coefficient d'autocorrélation est élevé, plus cette déviation par rapport au niveau moyen risque de s'échelonner sur une longue période. Le niveau considéré, autour duquel les données d'une série stationnaire gravitent, est habituellement la moyenne générale de la série.

De plus en plus d'hydrologues sont maintenant soucieux du fait que l'hypothèse de stationnarité de la moyenne est susceptible d'influencer considérablement les résultats d'un modèle de série chronologique. En effet, si cette hypothèse n'est pas respectée, les prévisions issues d'un modèle classique peuvent s'avérer erronées et par conséquent influencer grandement la prise de décision.

Un changement de moyenne dans une série d'apports énergétiques annuels peut survenir dans trois situations distinctes :

- le changement de moyenne est engendré par la modification des données de base (reconstitution et extension de séries, cf. chapitre 2);
- le changement de moyenne est dû à une modification des appareils de mesures ou de l'aménagement du bassin;
- le changement de moyenne est lié à un phénomène physique (changement climatique, changement physiographique du bassin, etc.).

Les deux premières situations sont plus faciles à traiter puisque la date du changement éventuel est connue. Les techniques statistiques usuelles de détection du changement de moyenne sont alors applicables (par exemple, les tests statistiques de comparaison des moyennes avant et après la date d'intérêt). Des tests statistiques ont donc été utilisés pour comparer les moyennes avant et après certaines dates correspondant à des changements méthodologiques de reconstitution et d'extension de séries. Les principaux résultats sont présentés dans la section suivante. La troisième situation (date de changement inconnue a priori) est traitée à la section 3.4.3 ainsi qu'au chapitre 4.

3.4.2 Analyse de la stationnarité (date de changement connue a priori)

Les changements méthodologiques lors de l'harmonisation des apports naturels sont susceptibles d'engendrer une non-stationnarité de la moyenne dans les séries d'apports énergétiques annuels. Les figures 2.1 à 2.4 et le tableau 2.3 donnent le contenu des séries d'apports de chacun des complexes et indiquent les dates correspondant à des changements méthodologiques de reconstitution et d'extension de données. Ces dates de rupture méthodologique sont présentées au tableau 3.24. Le complexe Churchill Falls n'ayant pas été traité lors de l'harmonisation des apports naturels, aucune date de changement méthodologique n'est connue a priori et il n'est pas pris en compte ici.

Le test paramétrique de Student, permettant de comparer les moyennes des sous-échantillons avant et après la date de rupture, a été appliqué aux séries d'apports énergétiques des complexes Bersimis, Manicouagan, Aux Outardes, Outaouais et St-Maurice. Le complexe La Grande ayant deux dates de rupture, ce test statistique n'a pu être effectué. Dans ce cas, le test de Fisher, qui permet de comparer plus de deux moyennes, a été utilisé (analyse de variance à un facteur, cf. Montgomery, 1984). D'autre part, le test de

Student n'a pas été appliqué aux apports énergétiques du complexe St-Laurent puisque ceux-ci sont très autocorrélés (cf. tableau 3.23) et que la statistique du test repose sur l'hypothèse d'indépendance des observations. Les résultats sont présentés au tableau 3.25. Les moyennes des sous-échantillons, la valeur de la statistique du test ainsi que le niveau de signification observé (valeur entre parenthèses) y sont donnés. Les tests de Student et de Fisher supposent la normalité des observations. Cette hypothèse n'est pas problématique (cf. tableau 3.21) mais les tests non paramétriques correspondants ont quand même été appliqués par mesure de prudence. Il s'agit des tests de Wilcoxon (équivalent non paramétrique du test de Student, cf. section 2.3.2) et de Kruskal-Wallis (équivalent non paramétrique du test de Fisher, cf. Lehmann, 1975). Les conclusions obtenues sont systématiquement les mêmes.

Tableau 3.24. Dates de ruptures méthodologiques lors de l'harmonisation des apports.

COMPLEXE	DATE	COMMENTAIRES	FIGURE
BERSIMIS	1956	Débits naturels avant et apports calculés par bilan hydrique après	2.1
LA GRANDE	1960	Changement de série de référence	2.3
	1982	Changement de série de référence	2.3
MANICOUAGAN	1964	Débits naturels avant et apports calculés par bilan hydrique après	2.1
AUX OUTARDES	1968	Débits naturels avant et apports calculés par bilan hydrique après	2.1
OUTAOUAIS	1966	Rupture dans la série de référence (rivière Mégiscane)	2.2
ST-MAURICE	1966	Rupture dans la série de référence (rivière Kinojevis)	2.2
ST-LAURENT	1979	Rivière de référence St-Régis avant et Chateauguay après	2.4

Tableau 3.25. Comparaison des moyennes des sous-échantillons.

Complexe hydroélectrique	MOYENNES DES SOUS-ÉCHANTILLONS			Résultat du test	
	1	2	3	(Student ou Fisher)	
BERSIMIS	7 746.86	8 135.83	-	-1.296	(0.2011)
LA GRANDE	75 305.93	80 504.10	73 326.66	2.204*	(0.1214)
MANICOUAGAN	22 098.60	21 690.12	-	0.512	(0.6109)
AUX OUTARDES	9 931.02	9 981.75	-	-0.139	(0.8902)
OUTAOUAIS	9 816.04	10 089.76	-	-0.584	(0.5616)
ST-MAURICE	9 013.42	9 535.67	-	-1.286	(0.2045)

* Test de Fisher (comparaison de trois moyennes).

Aucun changement de la moyenne, significatif à 5%, n'a été détecté dans les séries d'apports énergétiques annuels aux dates de rupture méthodologique. En effet, les niveaux de

signification observés sont tous supérieurs à 5%. D'ailleurs, un des critères de validation des reconstitutions et des extensions de séries effectuées par Hydro-Québec était la conservation de l'homogénéité temporelle (cf. chapitre 2).

La contribution du complexe La Grande aux apports énergétiques annuels totaux est très importante (plus de 41%, cf. figure 3.1). Ainsi, les manipulations de données effectuées sur cette série (cf. figure 2.3) peuvent influencer considérablement la stationnarité de la moyenne de la série des apports agrégés. C'est pourquoi, les moyennes des apports énergétiques annuels agrégés des sous-échantillons formés à partir des dates de rupture méthodologique identifiées pour le complexe La Grande (1960 et 1982) ont été comparées. Les résultats sont présentés au tableau 3.26. Les moyennes ne sont pas statistiquement différentes à un niveau de signification de 5%.

Tableau 3.26. Comparaison des moyennes des sous-échantillons (apports agrégés).

Complexe	MOYENNES DES SOUS-ÉCHANTILLONS			Test de Fisher
	1943-1960	1961-1981	1982-1993	
SÉRIE AGRÉGÉE	182 203.8	189 619.9	174 876.2	2.420 (0.0997)

3.4.3 Analyse de la stationnarité (date de changement inconnue)

Aucun changement de moyenne significatif ne semble être engendré par les interventions effectuées sur les données lors de l'harmonisation des apports naturels. Toutefois, une non-stationnarité de la moyenne peut être causée par d'autres phénomènes (changement climatique, etc.). Dans cette situation, la date de changement est inconnue et les méthodes usuelles de comparaison des moyennes ne peuvent être employées.

Diverses méthodes sont utilisées pour la détection de changement de moyennes lorsque la date est inconnue a priori. En particulier, la procédure graphique des sommes cumulées (cusum) est souvent employée. La technique graphique du cusum consiste à porter en ordonnées le cumul dans l'ordre chronologique des valeurs de la série et en abscisse les périodes. On détecte usuellement les sauts de moyennes pouvant survenir à certaines périodes par les discontinuités de cette courbe "cusum" (changement brusque du signe de la pente).

Considérons une série chronologique x_1, x_2, \dots, x_n . Les sommes centrées cumulées S_i consécutives (cusum centré) sont définies par:

$$S_i = \sum_{j=1}^i (x_j - \bar{x}) \quad (3.7)$$

où \bar{x} est la moyenne générale de la série. Cette quantité est intéressante puisqu'elle est liée aux différences entre les moyennes avant et après la date i . De plus, elle permet de détecter d'éventuels changements de moyennes qu'il aurait été difficile d'identifier en examinant seulement la courbe chronologique des données originales.

Les figures 3.25 à 3.33 présentent les courbes "cusum" des apports énergétiques annuels pour chacun des huit complexes hydroélectriques et pour la série agrégée.

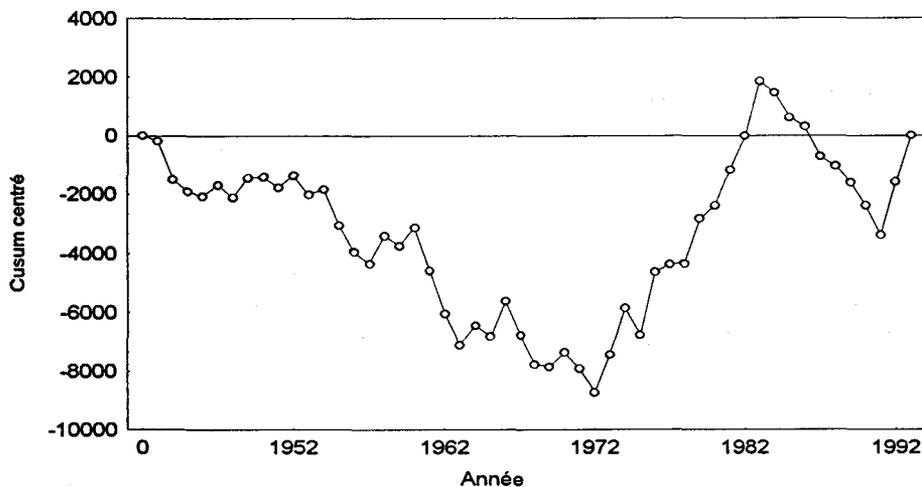


Figure 3.25. Série chronologique des cusums centrés (Bersimis).

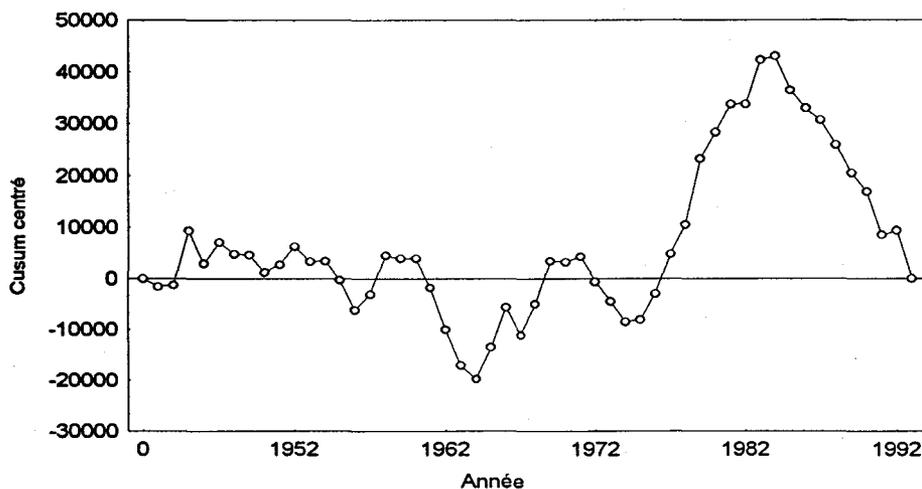


Figure 3.26. Série chronologique des cusums centrés (Churchill Falls).

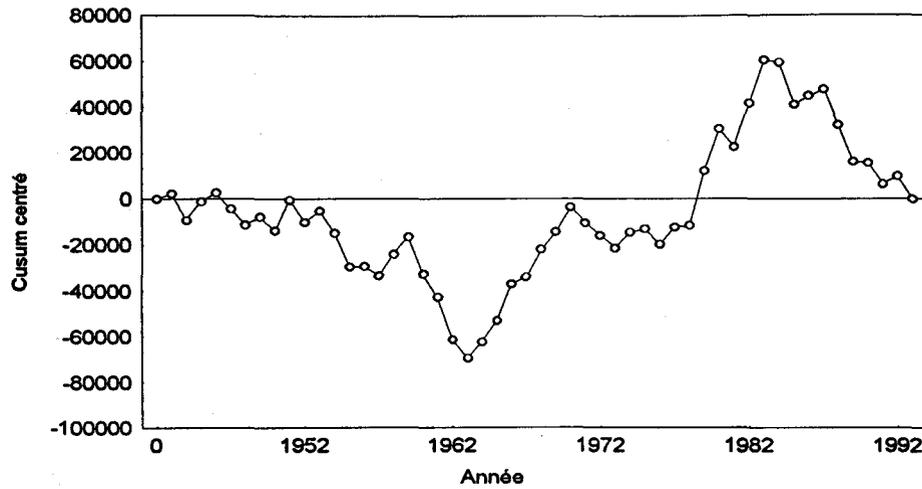


Figure 3.27. Série chronologique des cusums centrés (La Grande).

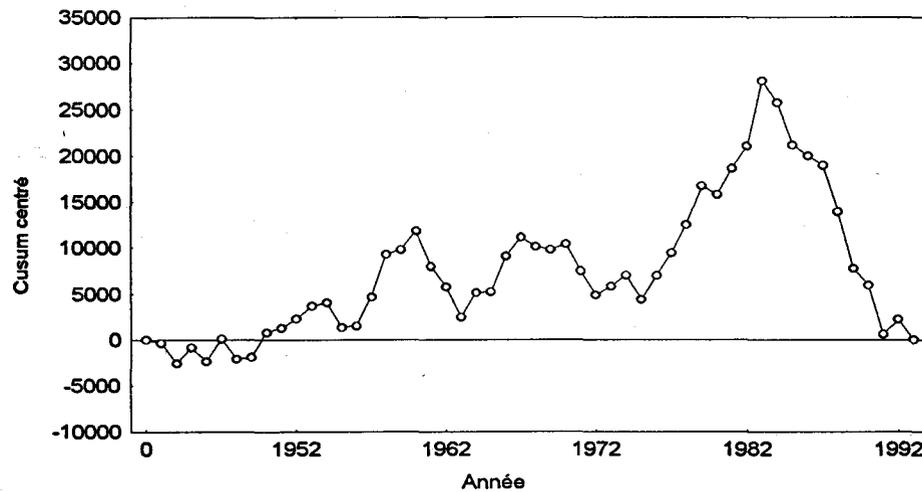


Figure 3.28. Série chronologique des cusums centrés (Manicouagan).

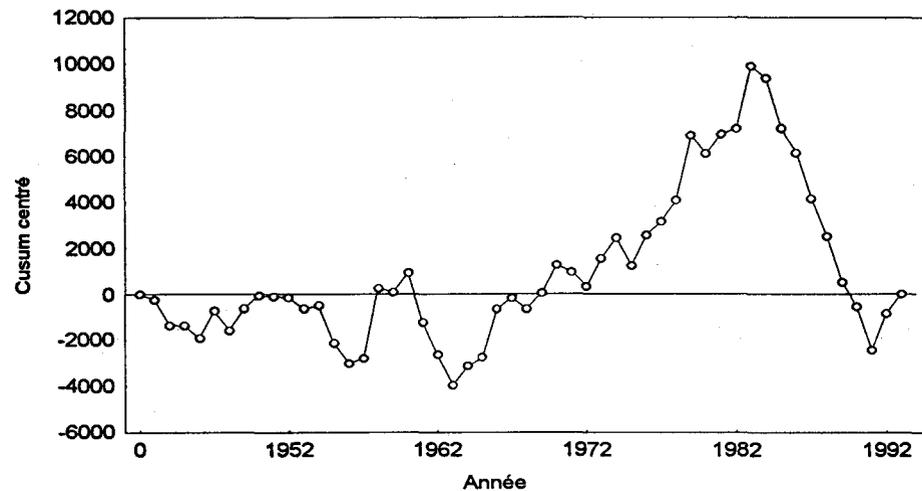


Figure 3.29. Série chronologique des cusums centrés (Aux Outardes).

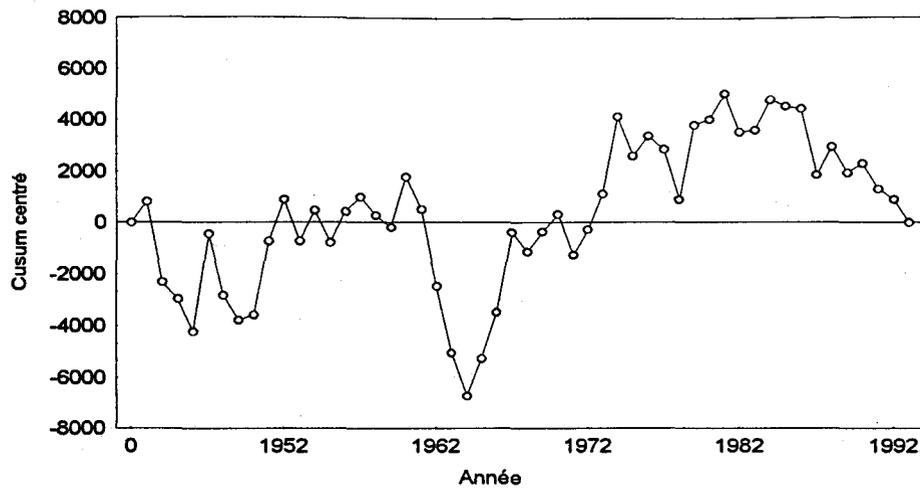


Figure 3.30. Série chronologique des cusums centrés (Outaouais).

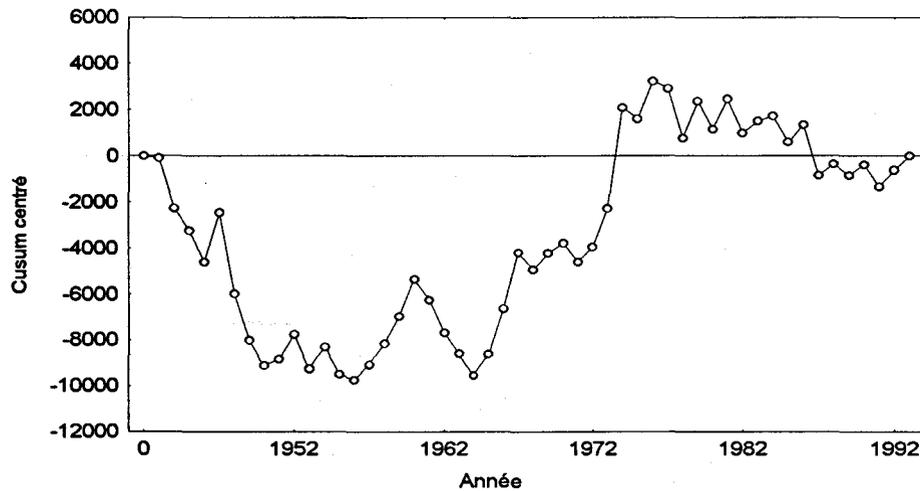


Figure 3.31. Série chronologique des cusums centrés (St-Maurice).

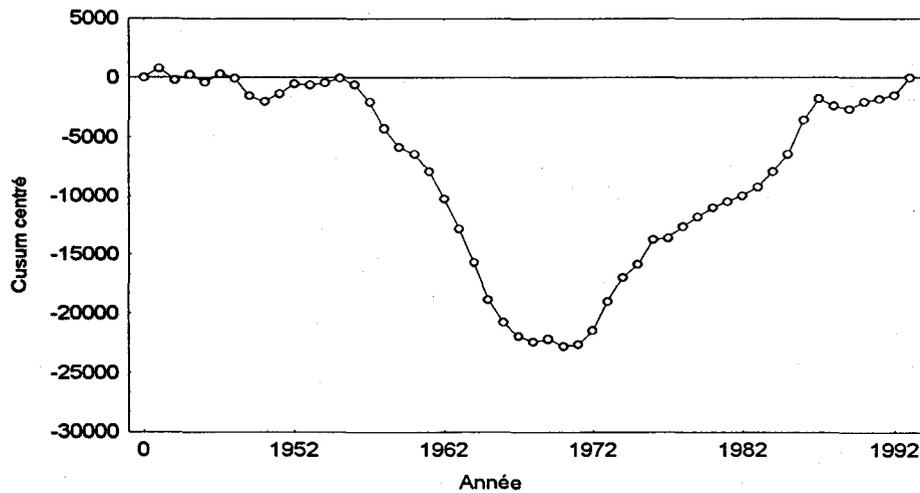


Figure 3.32. Série chronologique des cusums centrés (St-Laurent).

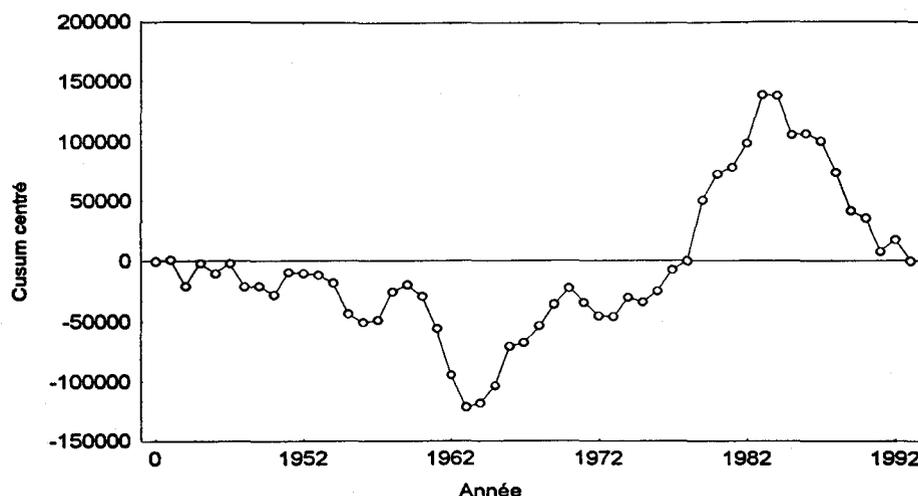


Figure 3.33. Série chronologique des cusums centrés (Série agrégée).

L'examen des courbes "cusum" permet d'identifier les dates les plus probables de changement de la moyenne. Le tableau 3.27 donne pour chacune des séries les périodes détectées. Mentionnons que les cassures de la courbe cusum pour les complexes Outaouais et St-Maurice sont beaucoup moins claires que celles observées pour les autres séries.

Tableau 3.27. Année la plus probable d'un changement de moyenne selon la courbe cusum.

Complexes	BER	CHF	LGR	MAN	OUT	OTW	STM	STL	Agrégée
Année	1972	1983	1963	1983	1983	1963	1963	1972	1963
	1983		1983						1983

Le problème de cette méthode est l'absence de tests statistiques valables pour préciser la signification des cassures apparentes (Bernier, 1977). En effet, la courbe "cusum" d'une série tout à fait stationnaire présente nécessairement de telles ruptures par sa nature même et il reste à les distinguer de celles imputables aux changements de moyennes. La technique du cusum demeure quand même utile pour identifier, de façon préliminaire, certaines dates où des changements de moyennes ont pu survenir.

Compte tenu de la complexité du problème de non-stationnarité, une analyse plus approfondie est nécessaire avant de conclure à un changement de moyenne significatif pour une série chronologique donnée. Une section du chapitre 4 est donc consacrée à ce problème. On s'attardera à la stationnarité de la série des apports énergétiques annuels agrégés, que l'on souhaite modéliser.

3.5 Conclusions

Pour extraire toute l'information possible des séries d'apports énergétiques, une analyse statistique descriptive des apports énergétiques mensuels et annuels des huit complexes hydroélectriques et de la série agrégée a été effectuée. Les principales observations tirées des différentes analyses sont résumées dans ce qui suit.

Apports énergétiques mensuels (1943-1993)

- Les systèmes La Grande, Churchill Falls et Manicouagan cumulent près de 73% de la production énergétique totale. Les caractéristiques statistiques des apports énergétiques de ces sites influencent donc considérablement celles de la série agrégée.
- La distribution des apports varie d'un mois à l'autre pour la plupart des complexes (Bersimis, Churchill Falls, La Grande, Aux Outardes, Manicouagan, Outaouais, St-Maurice). En particulier, la variabilité des apports énergétiques est systématiquement plus grande en période de crue (avril à juillet). Par contre, le comportement de la distribution mensuelle des apports énergétiques du complexe St-Laurent est plus uniforme.
- Une forte corrélation est observée presque systématiquement entre les apports de deux mois consécutifs pour les complexes Churchill Falls, La Grande, Outaouais, St-Maurice, St-Laurent et la série des apports agrégés. Ces corrélations sont toutefois moins élevées pour les complexes Bersimis, Manicouagan et Aux Outardes.
- Toutes les corrélations mensuelles sont élevées pour le complexe St-Laurent. Les apports énergétiques mensuels sont donc très liés quel que soit le décalage (1 à 12 mois).
- La matrice des corrélations mensuelles des apports énergétiques agrégés présente une structure assez semblable à celles des corrélations observées pour les complexes Churchill Falls et La Grande, les systèmes contribuant le plus aux apports totaux.

Apports énergétiques annuels (1943-1993)

- La dispersion des apports énergétiques varie en fonction de la dimension des complexes (plus le système est grand, plus la variabilité des apports est grande).
- Le test de normalité de Shapiro-Wilk a été appliqué aux neuf séries d'apports énergétiques annuels et confirme l'hypothèse de normalité. C'est seulement pour la série d'apports annuels au complexe Bersimis que l'hypothèse d'une distribution normale n'est pas acceptée au niveau de signification de 5%.

- Des corrélations généralement élevées sont obtenues pour des complexes voisins. En particulier, les corrélations entre les apports énergétiques annuels des complexes Manicouagan/Aux Outardes, Outardes/Bersimis et Outaouais/St-Maurice sont respectivement 0.86, 0.81 et 0.82.
- Les apports énergétiques annuels du complexe St-Laurent ne sont pas corrélés avec ceux des autres complexes. Ces apports se comportent donc de façon tout à fait différente.
- L'analyse en composantes principales montre très clairement trois groupes de complexes: St-Laurent, [Outaouais/St-Maurice] et [Bersimis/Outardes/Manicouagan/Churchill Falls/La Grande].
- Parmi les complexes, seuls les apports énergétiques annuels du St-Laurent sont significativement autocorrélés. Les autocorrélations sont d'ailleurs significatives pour tous les décalages considérés (une à cinq années).
- La série d'apports énergétiques annuels agrégés possède un coefficient d'autocorrélation d'ordre 1 significatif à 5% (0.356). La somme des apports énergétiques annuels de tous les complexes d'une année donnée est donc liée au cumul obtenu l'année précédente.
- Le coefficient d'autocorrélation d'ordre 1 de la série agrégée n'est pas influencée, comme on aurait pu le croire, par celui des apports du complexe St-Laurent (forte autocorrélation). La valeur obtenue à partir des observations (0.356) proviendrait plutôt de l'accumulation de faibles autocorrélations positives de certaines séries individuelles dont la contribution est plus importante (en particulier, Churchill Falls, La Grande et Manicouagan). Cela tend à démontrer, même si les autocorrélations d'ordre 1 de ces séries (tableau 3.2.3) ne sont pas significatives à 1% (dû à la taille limitée de l'échantillon), qu'elles sont probablement réelles.
- Aucun changement de la moyenne, significatif à 5%, n'est induit par les interventions effectuées lors de l'harmonisation des apports naturels mensuels (aux dates de rupture méthodologique).



4 MODÉLISATION DE LA SÉRIE AGRÉGÉE

4.1 Introduction

Comme on l'a mentionné au chapitre 1, Hydro-Québec présente périodiquement depuis 1979, un plan de développement de ses activités pour les 10 années à venir. Depuis 1990, ce plan est réactualisé tous les trois ans. Cet exercice repose sur une analyse détaillée de l'évolution du contexte économique et énergétique. Il permet, en particulier, d'établir les politiques de développement à court, moyen et long termes qu'adoptera l'entreprise. C'est pourquoi, il est important, du point de vue économique, de prévoir avec un maximum de précision la demande et l'offre hydroélectrique, puisque ces deux composantes sont déterminantes dans le choix des interventions à privilégier.

Afin de déterminer les risques de défaillance énergétique associés à différents scénarios d'exploitation et de bien planifier le parc d'équipements de production hydroélectrique, Hydro-Québec utilise différents modèles de simulation. Ces modèles reposent, entre autres, sur les caractéristiques statistiques de la série d'apports énergétiques¹ annuels agrégés (moyenne, écarts-type, autocorrélation) et sont basés sur l'hypothèse que les observations sont stationnaires.

Les résultats des analyses statistiques effectuées au chapitre 3, en particulier ceux concernant la présence éventuelle d'un changement de moyenne dans la série des apports énergétiques annuels agrégés, ont permis de définir plus précisément les objectifs du sous-projet 2.4. Un document précisant les points sur lesquels Hydro-Québec souhaite obtenir des recommandations de l'équipe de la chaire a été rédigé et remis à l'INRS-Eau le 31 août 1995. Les questions de recherche prioritaires soumises dans ce document concernent la stationnarité de la moyenne, la modélisation de la série d'apports énergétiques agrégés annuels (simulation et prévision) et la distribution statistique des écarts énergétiques annuels par rapport à la moyenne (déficits et surplus énergétiques).

Pour répondre aux besoins d'Hydro-Québec, le sous-projet 2.4 a été divisé en quatre étapes:

- Analyse de la stationnarité de la moyenne des apports énergétiques annuels agrégés

¹ Depuis le début du projet C2, nous avons considéré la série qui s'échelonne de 1943 à 1993.

- Modélisation des écarts énergétiques annuels agrégés par rapport à la moyenne générale et prévisions
- Distribution statistique des écarts énergétiques annuels agrégés

Ce chapitre présente donc les travaux réalisés dans le cadre de ce sous-projet. Des recommandations sont données à la fin de chacune des sections.

4.2 Stationnarité de la moyenne (série agrégée)

Certains résultats de l'analyse statistique des séries d'apports énergétiques annuels ont amené les membres du projet C2 à envisager sérieusement la présence, dans la série agrégée, de changements de moyennes (section 3.4.3). Puisque la moyenne est un des principaux intrants des modèles de planification utilisés à Hydro-Québec et que la stationnarité de cette statistique est une hypothèse de base en modélisation, la première étape du sous-projet 2.4 a été consacrée à l'étude plus approfondie de ce phénomène. Ces travaux avaient comme objectif d'étudier certaines questions de recherche soumises par Hydro-Québec (document du 31 août 1995). En particulier :

- (a) Quelle moyenne doit être utilisée pour les simulations stochastiques ?
- (b) Combien d'années sont nécessaires pour détecter un éventuel changement de moyenne ?
- (c) Ce phénomène a-t-il été étudié et détecté ailleurs dans le monde ?
- (d) Les changements de moyennes constituent-ils un cycle et dans cette éventualité la durée de ce cycle est-elle prévisible ?

Comme on l'a mentionné à la section 3.4.3, la vérification de la stationnarité des apports énergétiques annuels lorsque la date de changement de la moyenne est inconnue a priori ne peut être effectuée à partir des méthodes usuelles (test de comparaison de moyennes, etc.). Une procédure bayésienne, applicable dans ces circonstances, a donc été considérée. Cette approche repose sur la procédure de Lee et Heghenian (1977), employée actuellement à Hydro-Québec lors de l'harmonisation des apports (2.2.1 et 2.3.2.2), et sur la généralisation qu'en a fait Bernier (1994).

Avant d'appliquer cette approche bayésienne aux apports énergétiques annuels agrégés (section 4.2.2), les principaux aspects théoriques de cette méthode sont décrits dans la section qui suit.

4.2.1 Procédure bayésienne de détection de changement de la moyenne

Considérons une série chronologique $X_1, X_2, \dots, X_\tau, X_{\tau+1}, \dots, X_n$ de variables aléatoires. Supposons que la première (X_1, X_2, \dots, X_τ) et la seconde partie ($X_{\tau+1}, \dots, X_n$) de cette séquence de variables aléatoires possèdent des moyennes différentes, de valeurs respectives μ et $\mu + \delta$. Un modèle traduisant le comportement de cette série est donné par:

$$X_i = \begin{cases} \mu + \varepsilon_i, & i = 1, 2, \dots, \tau \\ \mu + \delta + \varepsilon_i, & i = \tau + 1, \dots, n \end{cases} \quad (4.1)$$

où $\varepsilon_i, i = 1, \dots, n$, sont des variables aléatoires indépendantes et distribuées selon une loi normale de moyenne nulle et de variance σ^2 . Les paramètres τ, μ, δ et σ sont inconnus. Le paramètre τ représente la date de rupture dans la série et δ l'amplitude du changement de moyenne.

Contrairement aux méthodes statistiques classiques, les méthodes bayésiennes ne considèrent pas les paramètres des modèles comme des valeurs fixes mais bien comme des variables aléatoires représentées par une distribution statistique. Cette approche permet d'intégrer dans l'analyse, à travers ces distributions dites "a priori", les informations provenant de connaissances subjectives du phénomène étudié. La formule de Bayes est utilisée ensuite pour déterminer, à partir des distributions a priori et du modèle considéré (modèle 4.1, par exemple), les lois a posteriori permettant d'effectuer de l'inférence statistique sur les paramètres (Box et Tiao, 1973; Berger, 1985).

Lee et Heghenian (1977) considèrent les distributions a priori suivantes pour représenter les paramètres τ, μ, δ et σ du modèle (4.1):

- $\pi_1(\tau) = 1/(n-1), \tau = 1, \dots, n-1$, une loi uniforme
- $\pi_2(\delta)$, une loi normale de moyenne nulle et de variance σ_δ^2
- $\pi_3(\mu)$, une loi normale de moyenne nulle et de variance σ_μ^2
- $\pi_4(\sigma) \propto \sigma^{-1}$

On remarque, en particulier, que la même probabilité a priori est affectée à chacune des valeurs de τ . Cela traduit le fait qu'aucune information subjective a priori ne permet de favoriser une période particulière d'occurrence du changement de moyenne. Lee et Heghenian supposent aussi que les paramètres τ, μ, δ et σ sont indépendants et que

$\sigma^2/n \ll \sigma_\mu^2$ et $\sigma^2 \ll \sigma_\delta^2$, ce qui signifie que l'information connue a priori sur la moyenne et sur l'amplitude de son changement est de beaucoup inférieure à l'information fournie par une réalisation x_1, x_2, \dots, x_n des n variables aléatoires X_i . Mentionnons que les lois a priori utilisées par Bernier (1994) ne sont pas les mêmes puisqu'il emploie la notion de famille de distributions conjuguées qui facilite la dérivation de la loi a posteriori. Les résultats concernant le modèle (4.1) sont toutefois équivalents.

Étant données ces hypothèses, la loi a posteriori de la date de rupture τ (Lee et Heghenian, 1977) est donnée par:

$$f(\tau|x_1, \dots, x_n) \propto \left[\frac{n}{\tau(n-\tau)} \right]^{1/2} [R(\tau)]^{-(n-2)/2}, \quad \tau \in \{1, 2, \dots, n-1\} \quad (4.2)$$

où,

$$R(\tau) = \left[\sum_{i=1}^{\tau} (x_i - \bar{x}_\tau)^2 + \sum_{i=\tau+1}^n (x_i - \bar{x}_{n-\tau})^2 \right] \left[\sum_{i=1}^n (x_i - \bar{x}_n)^2 \right]^{-1} \quad (4.3)$$

\bar{x}_τ , $\bar{x}_{n-\tau}$ et \bar{x}_n étant respectivement les moyennes des τ premières observations, des $n - \tau$ dernières observations et de la série complète.

La loi a posteriori $f(\tau|x_1, \dots, x_n)$ est donc une loi discrète donnant, pour chaque date de la série, la probabilité que la rupture se réalise à cette période étant donné l'échantillon observé x_1, \dots, x_n . L'examen de cette distribution permet donc d'identifier le moment où le changement de moyenne possède la plus grande probabilité d'occurrence. Mentionnons, que si la date τ de probabilité a posteriori maximale se situe au début ou à la fin de la série ($\tau = 1, 2, 3$, ou $n-3, n-2, n-1$), il faut être prudent lors de l'interprétation. D'une part, cela peut très bien indiquer qu'il n'y a pas de rupture dans la série, et d'autre part, le décideur à Hydro-Québec est intéressé par un changement de moyenne possible qui a une certaine "permanence" et non à un changement qui risque d'être fortuit s'il est limité à un trop petite période.

Pour tirer des conclusions concernant l'amplitude du changement de moyenne, il est nécessaire de déterminer la loi a posteriori de δ . On peut montrer que la distribution conditionnelle a posteriori de ce paramètre $f(\delta|\tau, x_1, \dots, x_n)$ est une loi de Student non centrée à $n - 2$ degrés de liberté (Lee et Heghenian, 1977; Bernier, 1994). Ce résultat est utilisé par Hydro-Québec (Bruneau, 1986b; Bruneau, 1988) qui compare, à l'aide du test classique de Student, les moyennes avant et après la date de rupture la plus probable identifiée à partir de la loi a posteriori (4.2). Tel que mentionné à la section 2.3.2.2, cette

approche est problématique. En effet, cette distribution est une loi de Student si et seulement si τ est fixé (loi conditionnelle). Or, la date la plus probable de rupture correspond à celle pour laquelle la différence entre les moyennes avant et après est maximale. En appliquant le test de Student de cette manière, on se trouve donc dans les conditions où le découpage en deux de la série n'est pas indépendant des données. On se place en fait dans les conditions les plus favorables à un résultat significatif. Il n'est donc pas étonnant que le niveau de signification réel du test de Student (erreur de Type I, la probabilité de rejeter à tort l'hypothèse nulle) soit supérieur à celui que l'on se fixe a priori (cf. tableau 2.9). D'autre part, l'emploi d'une procédure statistique classique (test de Student) alors que l'approche est bayésienne nous apparaît fallacieuse.

Pour évaluer l'importance du changement de moyenne δ , nous proposons d'utiliser la loi a posteriori inconditionnelle de δ et d'en déduire ce qu'on appelle en théorie bayésienne un intervalle de crédibilité. La loi a posteriori marginale de δ est obtenue en sommant la loi jointe a posteriori du couple (τ, δ) sur l'ensemble des valeurs de τ . Cette loi inconditionnelle correspond à une moyenne pondérée de lois de Student non centrées (Lee et Heghenian, 1977) et peut s'écrire comme suit:

$$f(\delta|x_1, \dots, x_n) = \sum_{\tau=1}^{n-1} f(\delta|\tau, x_1, \dots, x_n) \cdot f(\tau|x_1, \dots, x_n), \quad -\infty < \delta < +\infty \quad (4.4)$$

Cette distribution est tout à fait indépendante de la date τ et peut donc être employée pour faire de l'inférence sur l'amplitude du changement de moyenne sans supposer a priori que la date de rupture est connue. L'intervalle de crédibilité symétrique $[a, b]$ de niveau de $100(1-\alpha)\%$ pour δ est tel que:

$$\text{Prob}\{a \leq \delta \leq b\} = \int_a^b f(\delta|x_1, \dots, x_n) d\delta = 1 - \alpha \quad (4.5)$$

où $\text{Prob}\{\delta < a\} = \text{Prob}\{\delta > b\} = \alpha/2$.

Lorsque l'intervalle de crédibilité résume bien la distribution $f(\delta|x_1, \dots, x_n)$, et particulièrement pour de petites valeurs de α (0.05 ou 0.01), une valeur spécifique δ_0 du paramètre sera dite "non crédible" si $\delta_0 \notin [a, b]$. Dans la présente étude, nous avons donc examiné si une valeur δ_0 de l'amplitude du changement, en l'occurrence $\delta_0 = 0$ (aucun changement), est crédible ou non. Il est toutefois important de noter, tel qu'illustré à la figure 4.1, que la notion de non-crédibilité lorsque $\delta_0 \notin [a, b]$ dépend fortement de la forme

de la distribution a posteriori $f(\delta|x_1, \dots, x_n)$. L'examen graphique de la loi a posteriori est donc nécessaire pour valider la conclusion de l'intervalle de crédibilité.

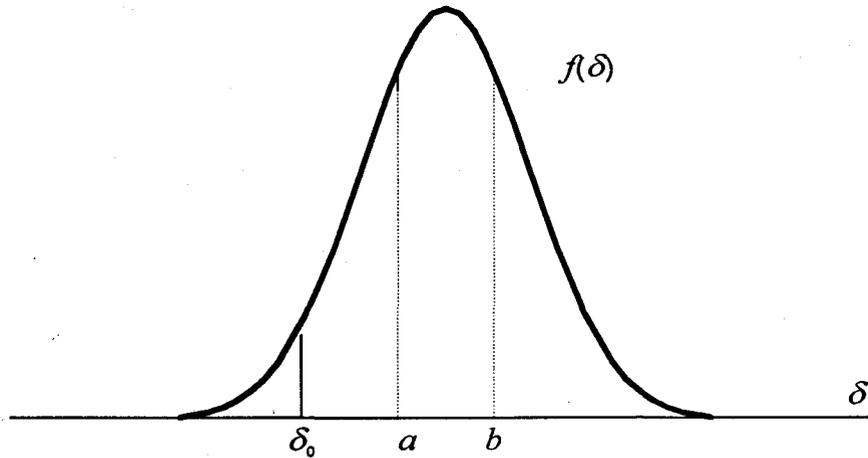


Figure 4.1a δ_0 est beaucoup moins crédible que les valeurs de l'intervalle

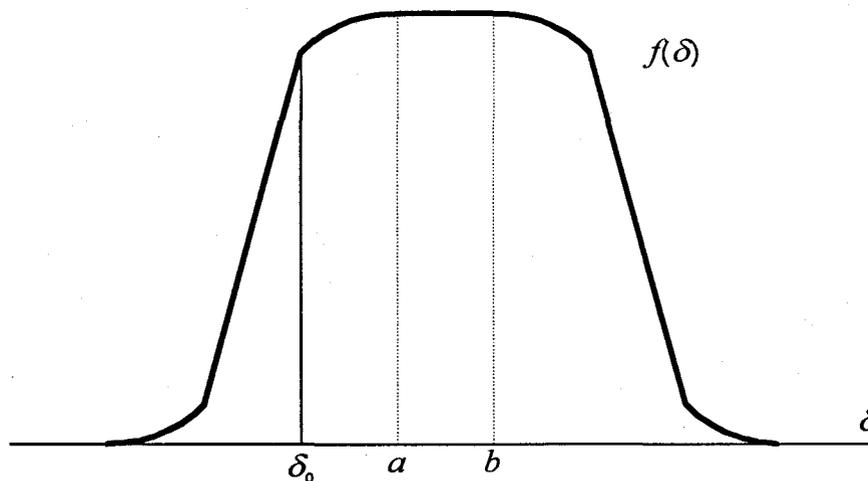


Figure 4.1b δ_0 est presque aussi crédible que les valeurs de l'intervalle

Mentionnons enfin que l'utilisation de l'intervalle de crédibilité ne permet pas de justifier une décision telle que "rejeter l'hypothèse $\delta = \delta_0$ ". Dans ce cas, il faut plutôt procéder à un test d'hypothèses bayésien en spécifiant une hypothèse alternative ainsi qu'une fonction de perte.

Bernier (1994) a proposé un test bayésien permettant de vérifier si le changement de moyenne est significatif. La statistique qu'il propose repose sur le fait que, sous l'hypothèse d'aucune rupture, la loi a posteriori $f(\tau|x_1, \dots, x_n)$ est uniforme. De plus, tel que mentionné plus haut, le fait que la date τ de probabilité a posteriori maximale se situe à l'une des extrémités de la série n'a pas beaucoup d'intérêt étant donné le petit nombre d'observations

permettant de calculer l'une des deux moyennes. Ainsi, en pratique, l'hypothèse nulle de non-changement peut s'exprimer de la manière suivante:

$$H_0: \tau \leq n_0 \text{ ou } \tau \geq n_1 \quad (4.6)$$

où n_0 et n_1 sont des valeurs respectivement près de 1 et de n convenablement choisies. La statistique R utilisée correspond au rapport entre la probabilité qu'il y ait changement de moyenne ($\tau \in [n_0, n_1]$) et la probabilité qu'aucune rupture ne soit présente dans la série ($\tau \notin [n_0, n_1]$). Cette statistique s'écrit comme suit:

$$R = \frac{\sum_{\tau \in [n_0, n_1]} f(\tau | x_1, \dots, x_n)}{\sum_{\tau \notin [n_0, n_1]} f(\tau | x_1, \dots, x_n)} \quad (4.7)$$

On rejette donc l'hypothèse nulle (aucun changement) si la statistique R est supérieure à une valeur critique donnée. En statistique classique, cette valeur critique correspondrait au quantile de la loi de R de probabilité au dépassement égale au niveau de signification fixé a priori (généralement 5%). En statistique bayésienne la valeur critique doit être déterminée à partir d'une fonction de perte. Bernier (1994) propose d'utiliser le rapport c_{nc}/c_c où c_{nc} et c_c correspondent, respectivement, au coût associé au rejet à tort des hypothèses de non-changement et de changement. L'établissement de ces valeurs incombe aux décideurs (Hydro-Québec) et non au statisticien. C'est pourquoi, dans les analyses présentées dans ce qui suit, nous donnons les valeurs de R (calculées par l'éq. 4.7) seulement à titre indicatif.

4.2.2 Vérification de la stationnarité (série agrégée)

La figure 4.2 présente les résultats de l'application de la procédure bayésienne aux données d'apports énergétiques annuels agrégés. La figure 4.2(a) présente les apports énergétiques annuels agrégés ainsi que les probabilités a posteriori $f(\tau | x_1, \dots, x_n)$ associées à chacune des années (courbe en trait gras). La figure 4.2(b) donne la loi a posteriori inconditionnelle de l'amplitude du changement $f(\delta | x_1, \dots, x_n)$. La procédure bayésienne indique que la date la plus probable d'occurrence d'un changement de moyenne est l'année 1984 (cf. figure 4.2(a)). On peut observer aussi une faible probabilité que cette rupture survienne en 1963. Cela pourrait signifier la présence de deux changements de moyenne dans la série. On constate, d'autre part, que la distribution a posteriori de l'amplitude de la rupture (cf. figure 4.2(b)) est bimodale, ce qui peut constituer un argument supplémentaire en faveur de cette hypothèse de double rupture. Compte tenu que le modèle utilisé est un modèle à saut unique (eq. 4.1),

la méthode bayésienne a aussi été appliquée séparément aux sous-échantillons 1943-1983 et 1984-1993. Le tableau 4.1 donne les principaux résultats de l'analyse. On y présente, pour les trois périodes considérées (1943-1993, 1943-1983 et 1965-1993), l'année de changement la plus probable τ , sa probabilité a posteriori $f(\tau|x_1, \dots, x_n)$, la valeur de la statistique R , les moyennes avant et après τ , l'amplitude de la rupture δ et l'intervalle de crédibilité de niveau 95%.

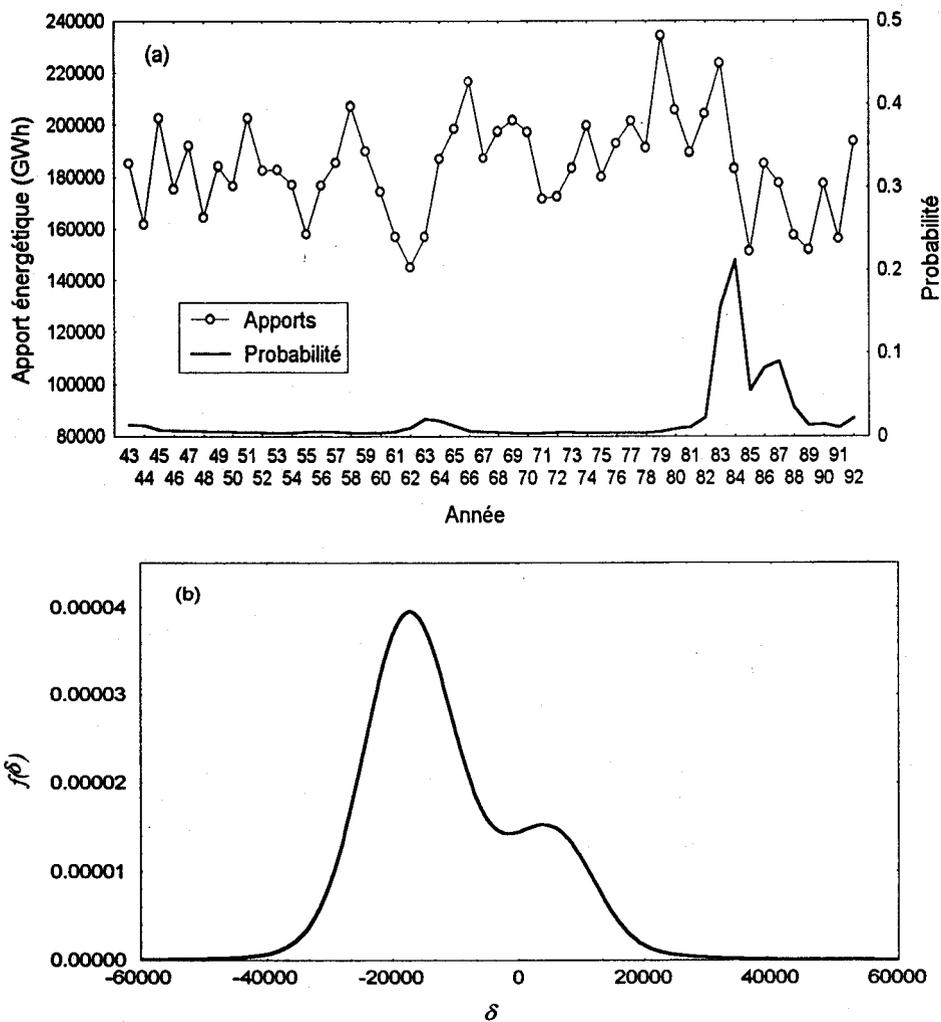


Figure 4.2. Résultats de la procédure bayésienne - Apports agrégés (1943-1993).

Tableau 4.1. Résultats de la procédure bayésienne pour la série agrégée

Période	τ	$f(\tau)$	R	\bar{x}_τ	$\bar{x}_{n-\tau}$	δ	Intervalle pour δ (95%)
1943-1993	1984	0.21	15.5	187 102	168 519	-18 583	[-31 424 ; 14 996]
1943-1983	1964	0.15	49.3	178 441	197 338	18 897	[7 156 ; 37 826]
1965-1993	1985	0.45	357.5	196 166	168 519	-27 647	[-39 280 ; -12 960]

L'examen du tableau 4.1 nous permet de constater, pour la série d'apports de 1943 à 1993, qu'une amplitude nulle ($\delta_0 = 0$) est tout à fait crédible. En effet, l'intervalle de crédibilité à 95% contient la valeur zéro. Le changement de moyenne détecté en 1984 n'est donc pas "significatif" au sens bayésien lorsqu'on considère toute la série. Toutefois, on constate que l'hypothèse de non-changement est peu crédible pour les sous-échantillons 1943-1983 et 1965-1993. La valeur de la statistique R calculée est d'ailleurs plus élevée pour ces sous-échantillons que pour la période entière d'observation. Selon ces résultats, la série d'apports énergétiques agrégés annuels pourrait être non stationnaire et constituée de trois sous-séries de moyennes différentes (1943-1964, 1965-1983, 1984-1993).

Ces résultats reposent sur une information de base restreinte (51 années d'observations ou moins). Il est alors difficile de conclure à un phénomène non stationnaire uniquement à partir de ces trois applications de la procédure bayésienne. En effet, ce type de déviation par rapport à la moyenne générale de la série peut très bien se produire alors que celle-ci est stationnaire (phénomène de persistance, cf. section 3.4.1). De plus, la présence de changements de moyennes a des implications majeures sur la prise de décision à Hydro-Québec. Aussi, faut-il pousser plus loin l'analyse en considérant d'autres séries reliées aux apports énergétiques annuels agrégés.

La série agrégée étant une agglomération spatiale de différentes séries, il apparaît logique d'appliquer d'abord la procédure aux apports énergétiques annuels de chacun des huit complexes. Les résultats sont présentés au tableau 4.2.

Tableau 4.2. Résultats de la procédure bayésienne pour chaque complexe (1943-1993).

Complexe	τ	$f(\tau)$	R	\bar{x}_τ	$\bar{x}_{n-\tau}$	δ	Intervalle pour δ
Bersimis	1991*	0.23	2.3	7 960	9 720	1 760	[-258 ; 2 662]
Church. Falls	1984	0.26	15.4	36 180	30 385	-5 795	[-10 494 ; 1 756]
La Grande	1984	0.09	8.0	78 539	70 506	-8 033	[-16 205 ; 9 745]
Manicouagan	1983	0.52	195.0	22 551	19 059	-3 492	[-5 265 ; -1 075]
Outardes	1983	0.26	12.0	10 198	8 965	-1 233	[-2 009 ; 1 501]
Outaouais	1946	0.05	4.8	8 905	10 051	1 146	[-2 321 ; 2 179]
St-Maurice	1950	0.13	13.4	8 150	9 502	1 352	[-750 ; 2 270]
St-Laurent	1970	0.31	230 731.0	11 632	13 434	1 802	[1 176 ; 2 356]

* Une probabilité non négligeable a aussi été détectée en 1972

Mentionnons d'abord que les intervalles de crédibilité obtenus sont bien représentatifs des lois de probabilité a posteriori inconditionnelles $f(\delta|x_1, \dots, x_n)$. Or, pour la plupart des complexes (à l'exception de Manicouagan et du St-Laurent), l'hypothèse de non-changement ($\delta_0 = 0$) est crédible. Toutefois, l'examen des résultats présentés au tableau 4.2, en particulier les dates τ obtenues, permet de constater une certaine cohérence régionale quant au moment de la possible rupture:

- l'année 1984 ou 1983 pour les complexes du nord du Québec (La Grande, Churchill Falls, Manicouagan et Outardes);
- l'année 1972 pour le complexe Bersimis;
- les années 1946 et 1950 respectivement pour les complexes Outaouais et St-Maurice (mentionnons toutefois que ces dates se situent au début de la série, ce qui peut indiquer un changement fortuit);
- l'année 1970 pour le complexe St-Laurent situé tout à fait au sud du Québec.

On remarque aussi, pour les complexes contribuant le plus aux apports énergétiques totaux (Churchill Falls, La Grande, Manicouagan et Outardes), que le changement de moyenne en 1984, quoique non crédible, correspond à une baisse d'hydraulicité ($\delta < 0$, cf. tableau 4.2). Cela est tout à fait en accord avec les résultats obtenus pour la série des apports énergétiques annuels (cf. figures 3.21, 3.33 et 4.2).

On constate enfin que le saut de moyenne observé en 1970 (année la plus probable, cf. tableau 4.2) pour le complexe St-Laurent est nettement plus significatif statistiquement que ceux des autres séries ($R = 230\ 731$). Une étude conjointe entre l'INRS-Eau et Hydro-Québec a d'ailleurs permis de détecter un changement de moyenne des apports naturels à cette période pour certains des Grands Lacs dont le comportement influence le système du St-Laurent (Mathier *et al.*, 1992). Toutefois, malgré l'intensité et la netteté de cet événement (cf. figures 4.3(a), 4.3(b)), ce saut de moyenne n'affecte pas la série agrégée pour laquelle la probabilité de changement est négligeable en 1970 (cf. figure 4.2(a)). La faible contribution des apports du St-Laurent aux apports totaux explique ce résultat.

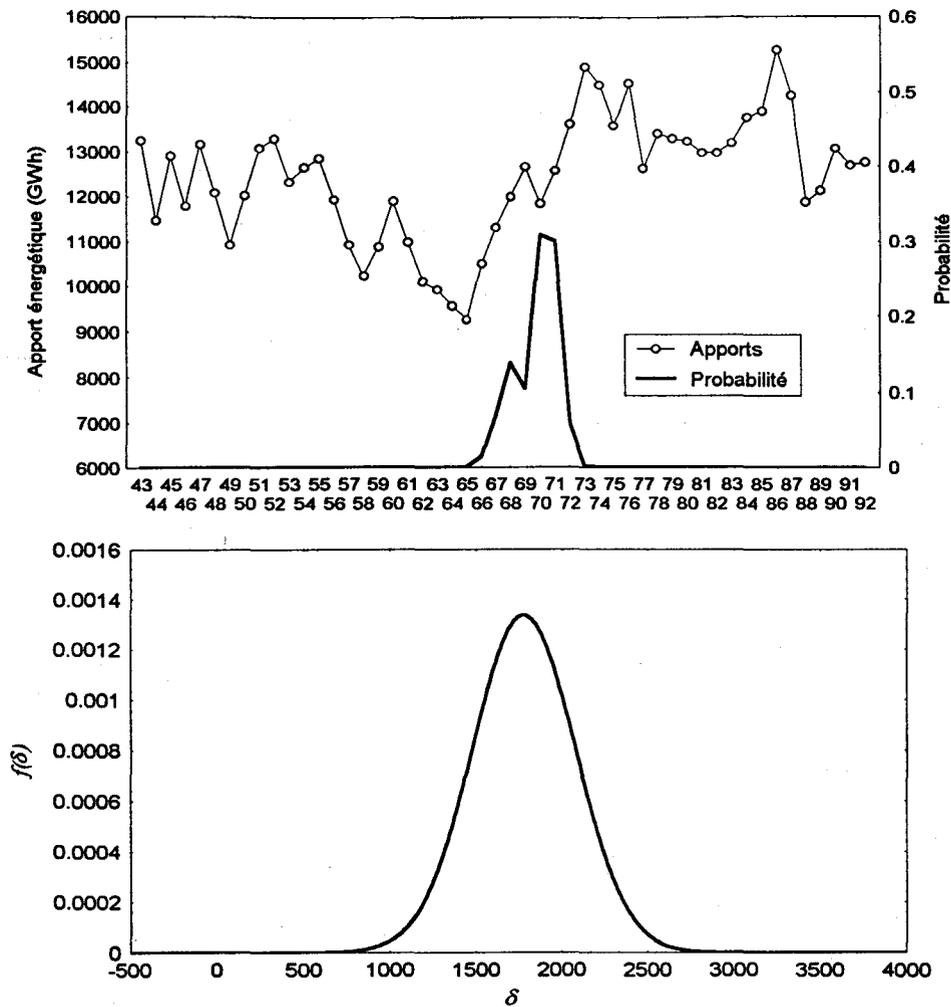


Figure 4.3. Résultats de la procédure bayésienne - Complexe St-Laurent.

Finalement, la procédure bayésienne a aussi été appliquée à des séries de précipitations annuelles. Les séries considérées représentent les précipitations de régions allant des plaines du sud des États-Unis aux provinces de l'Atlantique (tirées de Findlay *et al.* (1994) pour les séries canadiennes et Karl *et al.* (1994) pour les séries américaines). Des séries de précipitations sur les Grands Lacs ainsi que sur les bassins versants des rivières Richelieu et St-François ont aussi été considérées. La probabilité a posteriori $f(\tau|x_1, \dots, x_n)$ associée à l'année 1984 est négligeable pour l'ensemble des séries de précipitation contrairement à ce qui a été observé pour la série d'apports énergétiques agrégés (aussi pour les complexes Churchill Falls, La Grande, Manicouagan et Outardes). Toutefois, les résultats obtenus pour les précipitations montrent une probabilité de changement non négligeable entre 1963 et 1971 pour plusieurs de ces séries. Cela tend à montrer qu'un phénomène climatique serait effectivement survenu durant la décennie 1960-1970 et que cet événement aurait davantage

influencé les apports énergétiques du St-Laurent que ceux des autres systèmes hydriques (cf. tableau 4.2). Le tableau 4.3 donne les résultats obtenus pour l'ensemble des séries de précipitations considérées (en mm). On y présente l'année de changement la plus probable τ , sa probabilité a posteriori $f(\tau|x_1, \dots, x_n)$ et l'intervalle de crédibilité de niveau 90%. Mentionnons que certaines de ces séries représentent des précipitations régionales standardisées, donc adimensionnelles. Enfin, l'absence de longues séries de précipitations représentatives du nord du Québec pourrait expliquer le fait que l'année 1984 ne ressort pas de ces analyses. En effet, il est permis de supposer que les phénomènes météorologiques qui affectent le sud du Québec ainsi que les États-Unis ne sont pas les mêmes que ceux qui affectent le nord du Québec.

Tableau 4.3. Résultats de la procédure bayésienne pour les séries de précipitations (mm).

Série	Période d'observation	τ	$f(\tau)$	Intervalle pour δ (90%)
Atlantique/Canada*	1895-1992	1966	0.25	[0.47 ; 0.91]
St-Laurent/Grands Lacs*	1895-1992	1966	0.21	[0.25 ; 0.75]
Forêt nord-est*	1938-1992	1963	0.44	[0.09 ; 0.44]
Grands Lacs	1900-1991	1971	0.14	[37 ; 106]
Appalaches nord	1900-1991	1971	0.25	[-4 ; 100]
Piedmont nord	1900-1991	1971	0.13	[-60 ; 120]
Côtes nord-est	1900-1991	1970	0.09	[-22 ; 126]
NE US - SE Canada	1891-1990	1968	0.20	[82 ; 146]
Piedmont nord	1900-1991	1901	0.06	-
Appalaches sud	1900-1991	1901	0.10	-
Plaines du sud	1900-1991	1980	0.04	[-70 ; 135]
Plaines du nord	1900-1991	1906	0.10	[-106 ; 72]
Prairies de l'est	1900-1991	1988	0.06	-
Lac Ontario	1900-1994	1971	0.31	[53 ; 117]
Lac Érié	1900-1994	1971	0.12	[21 ; 109]
Lac Huron	1900-1994	1964	0.49	[74 ; 128]
Lac Michigan	1900-1994	1964	0.13	[24 ; 92]
Lac Ontario (NBS)	1900-1994	1971	0.35	[62 ; 121]
Lac Érié (NBS)	1900-1994	1971	0.23	[56 ; 128]
Lacs Michigan/Érié (NBS)	1900-1994	1966	0.07	[-28 ; 53]
Rivière Richelieu	1938-1990	1968	0.21	[50 ; 153]
Rivière St-François	1925-1990	1968	0.26	[9 ; 147]

* Précipitations régionales standardisées.

4.2.3 Conclusions et recommandations

(a) Moyenne à utiliser pour d'éventuelles simulations stochastiques

La détection d'un changement de moyenne dans une série d'observations à partir d'une procédure statistique est relativement simple à réaliser. Toutefois, compte tenu de l'information restreinte dont on dispose ici (51 données d'apports énergétiques), on peut difficilement distinguer entre la non-stationnarité de la moyenne et la présence de déviations soutenues engendrées par l'autocorrélation (persistance). Pour illustrer ce propos, la figure 4.4 présente une série de 50 données simulées à partir d'un modèle AR(1) (cf. section 4.3) avec les mêmes caractéristiques statistiques que celles de la série agrégée (moyenne, écart-type et autocorrélation). Ces observations, quoiqu'issues d'un modèle tout à fait stationnaire, présente des déviations persistantes d'amplitude similaire, et même plus grande, à celles observées dans la série d'apports énergétiques annuels agrégés (figure 4.2).

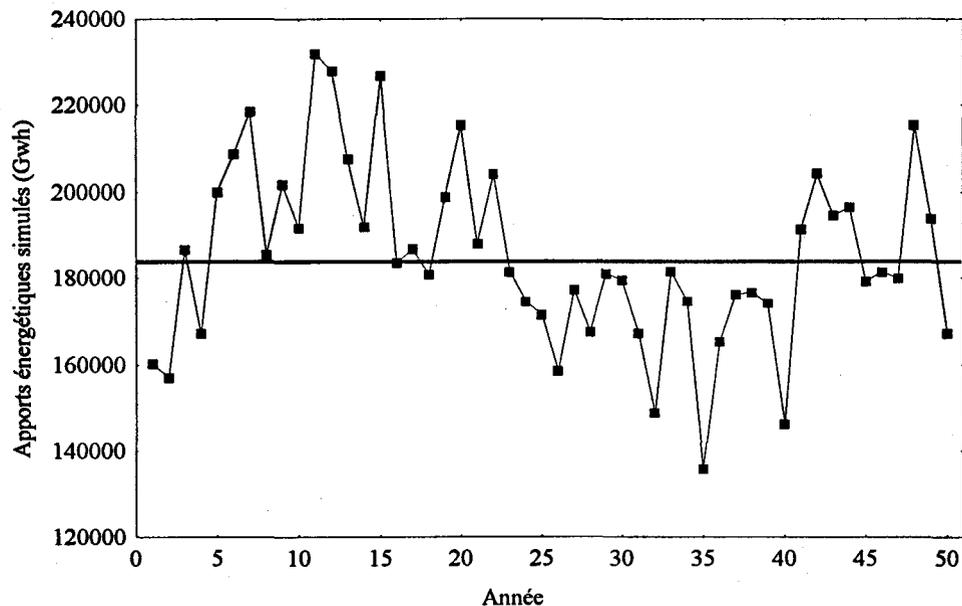


Figure 4.4. Série simulée à partir d'un modèle stationnaire AR(1)

Malgré les résultats obtenus à partir des séries de précipitations, permettant d'envisager la présence d'un changement de la moyenne des apports, il est difficile de conclure "hors de tout doute raisonnable" à la non-stationnarité de la moyenne dans la série des apports énergétiques annuels agrégés. C'est pourquoi, nous recommandons l'utilisation de la moyenne générale de la série comme intrant des modèles de simulations et de prévisions.

Nous considérons donc, dans la section 4.3, des modèles classiques qui reposent sur l'hypothèse de stationnarité.

(b) Nombre d'années nécessaires à la détection d'un éventuel changement de moyenne

Mentionnons d'abord que la seule comparaison des moyennes de deux sous-échantillons ne nous permet pas de conclure à la non-stationnarité, en particulier lorsque la date de rupture est inconnue a priori. L'adhésion, dans ce cas, à l'hypothèse de non-stationnarité d'un phénomène (due à un changement climatique par exemple) nécessite plus que la simple application d'un test statistique. Il faut identifier la source du changement de moyenne et en expliquer les causes. Ainsi, on ne peut pas, compte tenu de l'information actuellement disponible, déterminer explicitement le nombre d'années nécessaires à la détection d'une non-stationnarité dans la série des apports énergétiques annuels. Toutefois, lorsqu'on dispose d'information a priori sur la date d'un éventuel changement et qu'il est possible d'expliquer cet événement, l'application d'un test statistique classique est suffisant (par exemple, le test de Student). Dans ce cas, il est possible de déterminer le nombre d'observations permettant d'appliquer la procédure avec une précision désirée (Montgomery, 1984; Cochran, 1977).

La détermination du nombre d'observations nécessaires pour détecter un changement de moyenne est intimement liée à la puissance du test statistique. La puissance d'un test classique (non bayésien) est définie comme la probabilité de rejeter avec raison l'hypothèse d'égalité des moyennes et dépend à la fois de l'écart entre les moyennes et du nombre d'observations. Elle est aussi fonction du niveau de signification (probabilité de rejeter à tort l'hypothèse d'égalité des moyennes, souvent 5%). Généralement, étant donné un écart δ entre les moyennes, la puissance croît avec le nombre de données. Cela signifie qu'une différence fixée entre les moyennes théoriques est plus facile à détecter à l'aide d'un échantillon de plus grande taille. La fonction de puissance peut être évaluée pour la plupart des tests classiques (Student, Fisher, etc.) et permet de déterminer une taille d'échantillon optimale. Cet exercice nécessite, pour un niveau de signification donné, de choisir un écart δ critique et de fixer la puissance désirée. L'écart critique correspond habituellement à la valeur au delà de laquelle la différence des moyennes doit être détectée avec une forte probabilité. Le choix de ces paramètres (niveau de signification, puissance et écart) n'est généralement pas effectué par le statisticien seul mais nécessite la participation du planificateur.

(c) *Études concernant la non-stationnarité*

De nombreux chercheurs ont développé différentes méthodes de détection de changement de la moyenne de séries hydrologiques. La littérature sur le sujet est fort abondante. Un rapport de l'Organisation Mondiale de la Météorologie (OMM) fait une revue des diverses techniques statistiques employées en hydrologie (Cavadias, 1992).

Des travaux récents ont aussi été publiés concernant les changements climatiques qui sont susceptibles d'influencer le comportement des précipitations et par le fait même celui des apports. Citons en particulier Rodionov et Krovnin (1992), Rodionov (1994), Trenberth et Hurrell (1994), Hurrell (1995) et Hurrell et van Loon (1995). Ces auteurs ont utilisé différentes variables climatologiques (température, précipitation, pression et circulation atmosphériques) pour détecter des changements climatiques dans la région de l'Atlantique-Nord.

(d) *Présence d'un cycle*

Une série chronologique admet un cycle si les observations s'écartent de la moyenne générale selon un schéma régulier où les écarts se maintiennent sur une période donnée et changent ensuite de signe. Ce schéma est récurrent et peut par exemple être comparé à une onde sinusoïdale. Pour pouvoir affirmer statistiquement qu'un cycle est présent dans la série d'apports énergétiques annuels, il faudrait au moins observer quelques répétitions du phénomène. Or, en examinant la figure 4.2, il est difficile de conclure à la présence d'oscillations régulières (de durée quasi constante) des apports autour de la moyenne générale. De plus, même si la série admettait quelques cycles d'une quinzaine d'années (7 à 8 années d'écarts positifs suivis de 7 à 8 années d'écarts négatifs), le nombre d'événements actuellement observés ne serait pas suffisant pour conclure en toute confiance. L'identification d'un cycle, dans ce cas, nécessite une justification physique ainsi qu'un nombre élevé d'observations.

Étant donné la taille limitée de l'échantillon, on ne peut que la série d'apports énergétiques annuels agrégés présente un cycle; elle admet plutôt des fluctuations persistantes de durée aléatoire (cas fréquemment rencontré lorsque le phénomène causal est autocorrélé).

4.3 Modélisation et prévision des apports énergétiques

Hydro-Québec emploie actuellement un modèle AR(1) (autorégressif d'ordre 1) pour représenter la série des apports énergétiques annuels agrégés. Ce modèle est utilisé pour

simuler une longue série d'écarts (jusqu'à 10 000 répétitions) par rapport à la moyenne. Les écarts cumulés sur un horizon de 1 à 10 ans sont ensuite calculés à partir de cette série. Les probabilités d'occurrence des événements ainsi simulés sont enfin déterminées pour des fins de planification énergétique. La valeur de départ de la simulation est zéro (écart initial nul par rapport à la moyenne générale).

Les travaux prévus au sous-projet 2.4 concernant la modélisation des apports énergétiques annuels agrégés avaient pour objet l'étude de cette procédure de simulation et de certaines questions de recherche soumises par Hydro-Québec le 31 août 1995. En particulier :

- (a) La procédure de simulation répond-elle efficacement aux objectifs de planification ?
- (b) Le modèle AR(1) est-il adéquat ?
- (c) Quelle est la distribution des écarts cumulés sur différents horizons ?

Les travaux réalisés concernant la modélisation des apports énergétiques annuels agrégés sont présentés dans les sections qui suivent.

4.3.1 Généralités concernant les modèles de séries chronologiques

La modélisation de séries chronologiques à partir de modèles AR et ARMA (modèles autorégressifs et de moyenne mobile, cf. section 4.3.2) peut servir à des fins variées. Par exemple, l'estimation d'un modèle de ce type est un moyen efficace de résumer, à partir de quelques paramètres, les propriétés statistiques de base d'une série temporelle. En effet, la modélisation permet de mieux comprendre les caractéristiques statistiques clés d'un ensemble de données. Outre cette utilité, la modélisation d'une série chronologique est employée pour deux types d'applications: la prévision et la simulation.

La prévision est une extrapolation dans le temps et vise à fournir la meilleure estimation d'une réalisation spécifique du phénomène dans le futur. Dans le cas qui nous intéresse, on pourrait par exemple prédire l'apport énergétique annuel agrégé (ou de manière équivalente l'écart par rapport à la moyenne générale) de la prochaine année. La distribution des prévisions issues d'un modèle donné (le modèle AR(1) par exemple) est entièrement connue si les résidus satisfont aux hypothèses de base (normalité, indépendance temporelle, homogénéité de la variance). Il est donc possible de déterminer analytiquement les intervalles de confiance de ces estimations ainsi que certaines probabilités d'occurrence associées à ces prévisions (par exemple, la probabilité d'être en déficit l'année prochaine).

L'objectif général de la simulation est d'utiliser le modèle pour générer un ensemble de séries d'observations dont les propriétés statistiques reproduisent en moyenne celles de la série de données historiques. Ces séquences simulées sont aussi appelées séries synthétiques car elles ne constituent que des réalisations équiprobables pouvant éventuellement se réaliser dans le futur. La simulation est employée principalement dans trois cas:

- pour évaluer certaines caractéristiques associées aux prévisions. En effet, il est souvent difficile et même impossible de déterminer algébriquement certaines propriétés théoriques des prévisions (par exemple, la probabilité associée à une séquence de déficits consécutifs pour les prochaines années, etc.).
- pour comparer différents modèles de séries chronologiques. Dans ce cas, la simulation permet d'évaluer empiriquement les caractéristiques statistiques et ainsi d'identifier le modèle préservant le mieux, en moyenne, celles de la série originale.
- pour examiner la réaction d'un système face à une gamme de scénarios futurs possibles. On choisit généralement des scénarios simulés extrêmes. La simulation répond ici à des objectifs de planification généraux et les résultats ne peuvent être interprétés comme des indicateurs spécifiques de ce qui surviendra dans les prochaines années. Ce type d'application de la simulation est utilisée en particulier pour l'aménagement d'un nouveau bassin versant (Rasmussen *et al.*, 1996).

La prévision sert à estimer le comportement futur d'un phénomène et elle est essentielle à la planification énergétique. Elle permet d'évaluer la marge de manoeuvre à court terme qu'Hydro-Québec a pour réagir et garantir la production d'énergie avec une certaine fiabilité. La simulation est un outil complémentaire qui peut ajouter à l'information fournie par les prévisions.

4.3.2 Modèles considérés

Que l'on effectue des prévisions ou des simulations, le modèle utilisé est le même et il doit être représentatif de la série d'apports énergétiques annuels agrégés. Rappelons que les modèles considérés pour représenter cette série reposent tous sur l'hypothèse de stationnarité de la moyenne.

Compte tenu de la dépendance temporelle des apports énergétiques annuels agrégés (autocorrélation d'ordre 1 significative, cf. tableau 3.23), des modèles de série chronologique ont été considérés pour représenter cette série (Box *et al.*, 1994). Après quelques analyses

préliminaires, le modèle à 2 paramètres ARMA(1,1)² a été retenu pour examen. Le modèle ARMA(1,1) est un modèle mixte constitué de deux composantes: un terme autorégressif (AR) et un terme de moyenne mobile (MA). La partie autorégressive décrit comment une observation dépend directement de la donnée mesurée à la période précédente. Le terme de moyenne mobile traduit la relation existant entre une observation et l'erreur de prévision de la donnée précédente.

Pour une série chronologique $y_1, y_2, \dots, y_t, \dots, y_n$, le modèle ARMA(1,1) s'exprime de la manière suivante:

$$y_t = \mu + \phi \cdot (y_{t-1} - \mu) + \varepsilon_t - \theta \varepsilon_{t-1} \quad (4.8)$$

où μ est la moyenne générale théorique du phénomène, ϕ le paramètre autorégressif, θ le paramètre de moyenne mobile et ε_t le terme d'erreur au temps t distribué selon une loi normale $N(0, \sigma_\varepsilon^2)$. Le modèle AR(1), utilisé à Hydro-Québec, correspond au cas particulier ARMA(1,0) du modèle ARMA(1,1) où $\theta = 0$ et est donné par:

$$y_t = \mu + \phi \cdot (y_{t-1} - \mu) + \varepsilon_t \quad (4.9)$$

Pour appliquer un modèle ARMA(1,1), on doit estimer les paramètres ϕ et θ . Cela est effectué généralement à partir de la méthode classique des moindres carrés. Cette approche consiste à trouver les valeurs des paramètres qui minimisent la somme des carrés des erreurs de prévision ε_t . Dans cette étude, d'autres approches ont aussi été considérées. Neuf versions différentes du modèle ARMA(1,1), correspondant à différentes valeurs des paramètres ϕ et θ (éq. 4.8), ont été retenues pour représenter les apports énergétiques annuels agrégés. Le tableau 4.4 donne les valeurs des paramètres ϕ et θ correspondant à chacun des modèles.

Le modèle ARMA(0,0), qui s'exprime comme $y_t = \mu + \varepsilon_t$ (éq. 4.8), est une loi normale de moyenne μ . Ce modèle ne tient pas compte du caractère autorégressif des apports énergétiques. Il a été considéré dans l'étude comme modèle témoin et a été nommé modèle BLANC. La méthode d'estimation des moments modifiés ainsi que l'approche de O'Connell (cf. tableau 4.4) étant moins usuelles, elles sont décrites brièvement dans ce qui suit.

² Les modèles plus complexes (ARMA(2,1), etc.) ne s'ajustent pas significativement mieux aux données.

Tableau 4.4. Modèles considérés pour la simulation

Méthode d'estimation	Modèle	Nomenclature	Phi	Theta
Aucune	ARMA(0,0)	BLANC	0.000	0.000
Moments modifiés	AR(1)	AR	0.410	0.00
Moindres carrés	ARMA(1,1)	ARMA1	0.498	0.156
Approche de O'Connell	ARMA(1,1)	ARMA2	0.900	0.595
Approche de O'Connell	ARMA(1,1)	ARMA3	0.920	0.607
Approche de O'Connell	ARMA(1,1)	ARMA4	0.940	0.627
Approche de O'Connell	ARMA(1,1)	ARMA5	0.960	0.667
Approche de O'Connell	ARMA(1,1)	ARMA6	0.970	0.680
Approche de O'Connell	ARMA(1,1)	ARMA7	0.980	0.696

4.3.2.1 Méthode des moments modifiés (Modèle AR)

Pour le cas particulier du modèle AR(1), la valeur du paramètre ϕ déterminée par la méthode des moindres carrés correspond exactement au coefficient d'autocorrélation d'ordre 1 et donc à l'estimation issue de la méthode des moments (Box *et al.*, 1994). Cette valeur est ici de 0.356 pour la série d'apports énergétiques annuels agrégés (cf. tableau 3.23). Toutefois, il est connu (Quenouille, 1956) que la formule usuelle du calcul de l'autocorrélation d'ordre 1 $\hat{\rho}_1$ (éq. 3.1) est biaisée. Wallis et O'Connell (1972) ont proposé une correction de biais qui, après vérification par simulation, donne de bons résultats. L'expression corrigée développée par ces auteurs s'exprime de la manière suivante:

$$\tilde{\rho}_1 = \frac{\hat{\rho}_1 + (1/n)}{1 - (4/n)} \quad (4.10)$$

où n désigne la taille de l'échantillon. Wallis et O'Connell (1972) ont étudié le comportement de cet estimateur à l'aide de simulations intensives qui ont montré, pour diverses tailles d'échantillons, que la correction permettait d'obtenir un estimateur sans biais. De plus, l'augmentation de la variance induite par cette correction demeure assez faible, même pour de petites tailles d'échantillons.

La valeur corrigée du coefficient d'autocorrélation obtenue pour la série d'apports énergétiques est de 0.410 et est utilisée comme estimation de ϕ dans le modèle AR(1). Nous avons nommé cette approche "méthode des moments modifiés". Mentionnons enfin que cette correction de biais n'est valable que pour le modèle AR(1).

4.3.2.2 Approche de O'Connell (Modèles ARMA2 à ARMA7)

Le coefficient de Hurst (1951) est une statistique qui, comme l'autocorrélation, quantifie la persistance. Il est fonction du cumul des séquences de surplus et de déficits de la série (cf. Annexe C). Le coefficient de Hurst ne mesure toutefois pas la même chose que le coefficient d'autocorrélation et la communauté scientifique ne s'entend pas sur son interprétation. En effet, le phénomène de Hurst a suscité beaucoup de controverse (Klemeš, 1974; Hipel et McLeod, 1980). Nous avons tout de même convenu d'inclure dans l'étude des modèles qui reproduisent cette caractéristique qui, pour la série d'apports énergétiques annuels agrégés, vaut 0.81.

La méthode des moindres carrés ne permet pas d'obtenir un modèle ARMA(1,1) qui préserve le coefficient de Hurst. Cependant, O'Connell (1974) a construit, à partir de simulations, des tables permettant de déterminer des couples de paramètres ϕ et θ qui permettent de prendre en compte simultanément le coefficient de Hurst et le biais de l'autocorrélation. Les modèles ainsi obtenus préservent le coefficient de Hurst. Une description de cette approche est aussi présentée dans Bras et Rodriguez-Iturbe (1985).

Dans le cas qui nous intéresse, l'approche de O'Connell fournit de grandes valeurs des paramètres ϕ et θ . En effet, cela est nécessaire pour reproduire simultanément les coefficients de Hurst (0.81) et d'autocorrélation (0.41). En particulier, le paramètre autorégressif ϕ doit être proche de 1 (cf. tableau 4.4), ce qui suppose une très forte persistance dans les données. On peut montrer que ces valeurs de paramètres engendrent des coefficients d'autocorrélations d'ordre supérieurs à 1 beaucoup plus grands que ceux calculés à partir des données. L'emploi de tels modèles reposerait donc sur le fait que l'usager croit, en quelque sorte, que la persistance réelle du phénomène étudié est plus grande que celle observée à partir des observations.

Six modèles ARMA(1,1) de ce type ont été retenus (cf. tableau 4.4, ARMA2 à ARMA7).

4.3.3 Comparaison des modèles

Pour comparer les neuf modèles, nous avons d'abord analysé leurs résidus et ensuite procédé à une étude comparative par simulation.

4.3.3.1 Analyse des résidus

Le modèle de série chronologique ARMA(1,1) repose sur certaines hypothèses de base concernant les termes d'erreur ε_t . Ceux-ci doivent être indépendants (autocorrélation nulle) et identiquement distribués selon une loi normale de moyenne nulle et de variance constante (bruit blanc). Pour vérifier les hypothèses de base, il suffit d'étudier les résidus du modèle qui sont définis comme suit:

$$\hat{\varepsilon}_t = y_t - \hat{\mu} - \hat{\phi} \cdot (y_{t-1} - \hat{\mu}) + \hat{\theta} \hat{\varepsilon}_{t-1} \quad (4.11)$$

où $\hat{\mu}$, $\hat{\phi}$ et $\hat{\theta}$ sont les estimations des paramètres du modèles. La vérification des hypothèses de base garantissent l'efficacité des prévisions.

Le tableau 4.5 donne, pour chaque modèle considéré, quelques caractéristiques statistiques de leurs résidus ainsi que la valeur du critère d'information d'Akaïke (1974), noté *CIA*.

Tableau 4.5. Statistiques sur les résidus des modèles.

Modèle	Moyenne	Écart-type	Asymétrie	$\bar{\rho}_1$	CIA
BLANC	0.00	19 079	0.00	0.410	19.752
AR	-170.25	17 824	0.282	-0.063	19.655
ARMA1	-167.40	17 945	0.244	0.016	19.708
ARMA2	-502.11	18 603	0.122	0.122	19.780
ARMA3	-543.72	18 705	0.115	0.124	19.791
ARMA4	-600.02	18 818	0.106	0.135	19.803
ARMA5	-697.47	18 944	0.092	0.165	19.794
ARMA6	-741.96	19 014	0.086	0.173	19.802
ARMA7	-796.39	19 078	0.079	0.185	19.810

Le critère d'Akaïke est une mesure d'adéquation et est généralement utilisé pour comparer l'adéquation de plusieurs modèles possibles. La règle de décision consiste à choisir le modèle dont la valeur de *CIA* est minimum. Le critère d'Akaïke *CIA* est défini de la manière suivante:

$$CIA = \ln(\hat{\sigma}_\varepsilon^2) + \frac{2(p+1)}{n} \quad (4.12)$$

où $\hat{\sigma}_\varepsilon^2$ est la variance estimée des résidus du modèle, p le nombre de paramètres et n la taille d'échantillon. L'estimateur de la variance des résidus $\hat{\sigma}_\varepsilon^2$ utilisé dans la présente étude est donné par (Box *et al.*, 1994):

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{n-p} \sum_{i=1}^n \hat{\varepsilon}_i^2 \quad (4.13)$$

Le critère *CIA* a l'avantage de tenir compte de l'incertitude d'échantillonnage associé au nombre de paramètres estimés et donc du principe de parcimonie. En effet, on remarque qu'en augmentant la complexité du modèle (plus de paramètres), le second terme de l'équation (4.12) augmente, défavorisant ainsi les modèles à plusieurs paramètres et ce, même si le premier terme diminue à cause de l'amélioration de l'adéquation.

Les principales observations qui découlent de l'analyse des résidus des modèles considérés (tableau 4.5) sont les suivantes:

- Les résidus des modèles AR et ARMA1 possèdent des moyennes voisines de zéro que ceux des modèles ARMA2 à ARMA7. Ces valeurs ne sont toutefois pas significativement différentes de zéro à un niveau de signification de 5%.
- Les écart-types des résidus des modèles AR et ARMA1 sont légèrement plus faibles que ceux des modèles ARMA2 à ARMA7. Cela indique une meilleure adéquation des modèles classiques AR et ARMA1. Remarquons, en particulier, que les résidus des modèles ARMA6 et ARMA7 possèdent un écart-type pratiquement égal à celui des données originales.
- Les résidus des modèles ARMA2 à ARMA7 sont plus autocorrélés que ceux des modèles AR et ARMA1. Ces valeurs ne sont pas significatives à un niveau de signification de 5% mais elles indiquent toute de même qu'il reste une certaine persistance dans les résidus des modèles de O'Connell.
- La distribution des résidus des modèles ARMA2 à ARMA7 est plus symétrique que celle des modèles AR et ARMA1 indiquant ainsi que les modèles de O'Connell semblent plus compatibles avec l'hypothèse de normalité.
- D'après le critère d'Akaike, le modèle AR serait le plus adéquat pour représenter les apports énergétiques annuels agrégés. Remarquons aussi, selon ce critère, que le modèle BLANC serait même plus adéquat que les modèles de O'Connell. Toutefois, les résidus

de BLANC sont fortement autocorrélés indiquant ainsi la nécessité d'employer un modèle de type ARMA.

Les résidus des modèles AR et ARMA1 sont donc plus compatibles en général avec les hypothèses de base. Toutefois, les résidus des modèles de O'Connell semblent mieux distribués selon une loi normale (asymétrie proche de zéro, cf. tableau 4.5). Cela n'est pas étonnant puisque les poids θ affectés à la partie aléatoire des modèles ARMA(1,1) (terme de moyenne mobile) sont plus importants (cf. tableau 4.4).

4.3.3.2 *Étude comparative par simulation*

Nous avons déjà mentionné, à la section 4.3.1, que la simulation est un outil permettant de comparer des modèles. Une telle étude comparative a été effectuée afin de vérifier si les modèles préservent adéquatement certaines caractéristiques statistiques de la série originale. Les principales caractéristiques considérées sont la moyenne, l'écart-type, le coefficient d'autocorrélation d'ordre 1 et le coefficient de Hurst. Un ensemble de 1000 séries de 50 années de données issues de chacun des modèles ont donc été générées. Pour chacune des séries, les quatre caractéristiques statistiques ont été calculées et leur distribution a été étudiée.

Le tableau 4.6 donne pour chaque modèle la moyenne, l'écart-type, le maximum et le minimum calculés à partir des 1000 échantillons pour les quatre caractéristiques statistiques considérées (moyenne, écart-type, autocorrélation, coefficient de Hurst). Les figures 4.5 à 4.8 présentent des diagrammes en boîtes illustrant la distribution des caractéristiques statistiques pour les neuf modèles. Les principales conclusions sont les suivantes :

- les modèles AR et ARMA1 conduisent à des résultats très similaires sauf pour le coefficient d'autocorrélation qui est sous-estimé en moyenne par le modèle ARMA1 (cf. tableau 4.6 et figure 4.7). Cela s'explique par le fait que la méthode des moindres carrés, utilisée dans ce cas (cf. section 4.3.2), ne tient pas compte du biais de l'estimation de l'autocorrélation.
- les modèles AR et ARMA1 reproduisent en moyenne toutes les caractéristiques de la série d'apports énergétiques à l'exception du coefficient de Hurst (on obtient 0.72 en moyenne au lieu de 0.81) et du coefficient d'autocorrélation pour le modèle ARMA1.
- les modèles ARMA2 à ARMA7 donnent des résultats comparables entre eux et ils reproduisent en moyenne toutes les caractéristiques de la série d'apports énergétiques.

- les caractéristiques statistiques des modèles ARMA2 à ARMA7 sont systématiquement plus variables que celles des modèles classiques. En particulier, l'écart-type de la moyenne des modèles de O'Connell est au moins deux fois plus grand que celui des modèles AR et ARMA1. Les modèles ARMA4 à ARMA7 admettent même plus de 10% des séries d'apports énergétiques de moyenne, sur 50 années, supérieure à 200 TWh ou inférieure à 165 TWh (cf. tableau 4.6 et figure 4.5). Or, de tels scénarios impliquent l'existence de changements d'ordre climatologiques en réalité très peu probables.

Tableau 4.6. Principales statistiques des séries simulées pour chaque modèle.

	Moyenne				Écart-type			
	Moy	Éc-type	Min	Max	Moy	Éc-type	Min	Max
SÉRIE	183 822				19 079			
BLANC	183 654	2 603	175 251	192 424	19 013	1 854	13 581	25 056
AR	183 579	4 164	171 448	195 439	18 928	2 241	12 779	26 976
ARMA1	183 856	4 076	169 247	195 004	19 017	2 119	13 703	25 923
ARMA2	183 609	9 024	154 953	214 770	18 799	2 919	12 158	32 730
ARMA3	183 947	10 224	151 486	214 092	18 845	3 006	12 090	37 784
ARMA4	184 054	12 507	145 190	222 666	18 795	3 353	10 328	38 608
ARMA5	184 310	14 423	137 800	242 483	18 816	3 493	11 233	32 730
ARMA6	182 753	18 222	130 428	236 246	19 044	3 563	11 326	38 036
ARMA7	183 366	21 253	119 796	255 236	18 673	3 505	11 929	42 063
	Autocorrélation				Hurst			
	Moy	Éc-type	Min	Max	Moy	Éc-type	Min	Max
SÉRIE	0.41				0.81			
BLANC	0.000	0.156	-0.474	0.495	0.629	0.072	0.324	0.795
AR	0.415	0.143	-0.061	0.821	0.723	0.073	0.503	0.902
ARMA1	0.366	0.149	-0.143	0.744	0.721	0.071	0.515	0.897
ARMA2	0.398	0.199	-0.239	0.918	0.787	0.078	0.434	0.942
ARMA3	0.422	0.207	-0.289	1.000	0.798	0.081	0.516	0.957
ARMA4	0.437	0.214	-0.358	0.974	0.804	0.080	0.475	0.960
ARMA5	0.411	0.223	-0.236	0.925	0.803	0.083	0.458	0.952
ARMA6	0.428	0.223	-0.257	1.000	0.809	0.081	0.527	0.953
ARMA7	0.411	0.225	-0.375	0.963	0.809	0.081	0.528	0.956

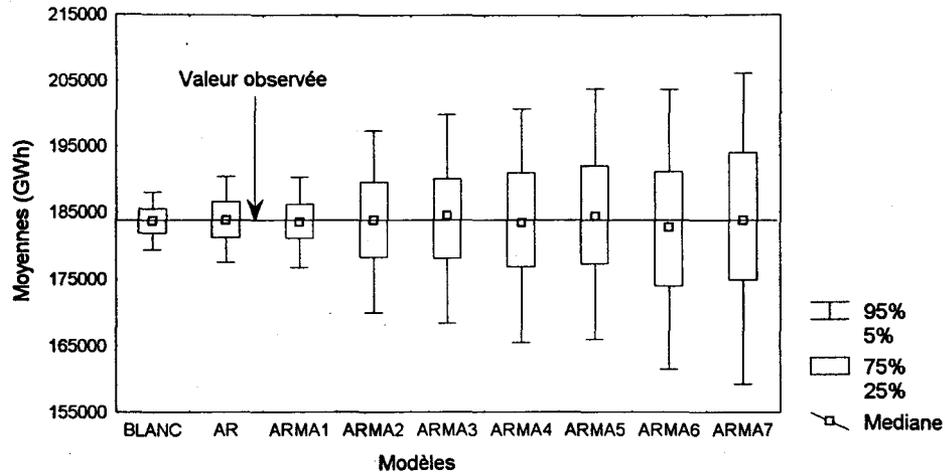


Figure 4.5. Distribution des moyennes des séries simulées.

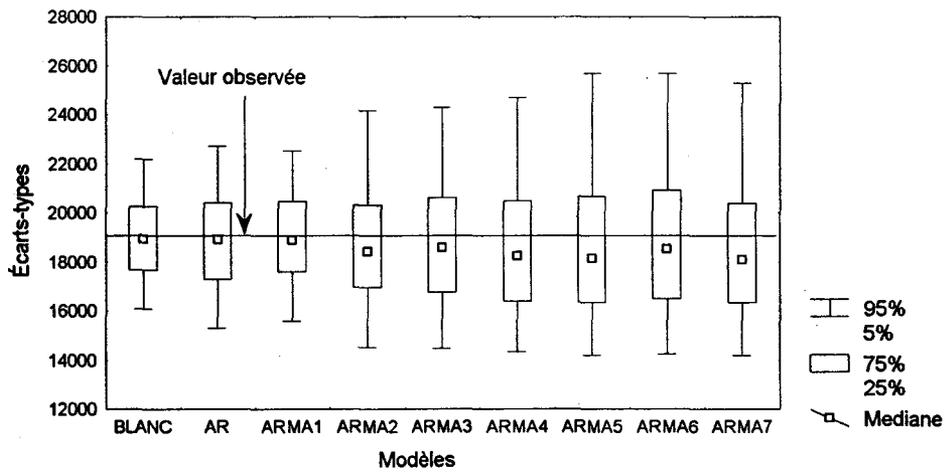


Figure 4.6. Distribution des écart-types des séries simulées.

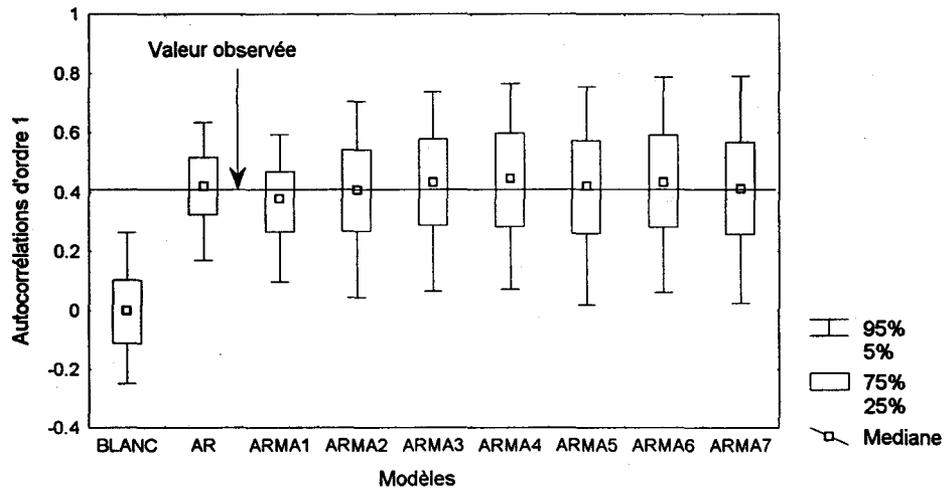


Figure 4.7. Distribution des coefficients d'autocorrélation $\tilde{\rho}_1$ des séries simulées.

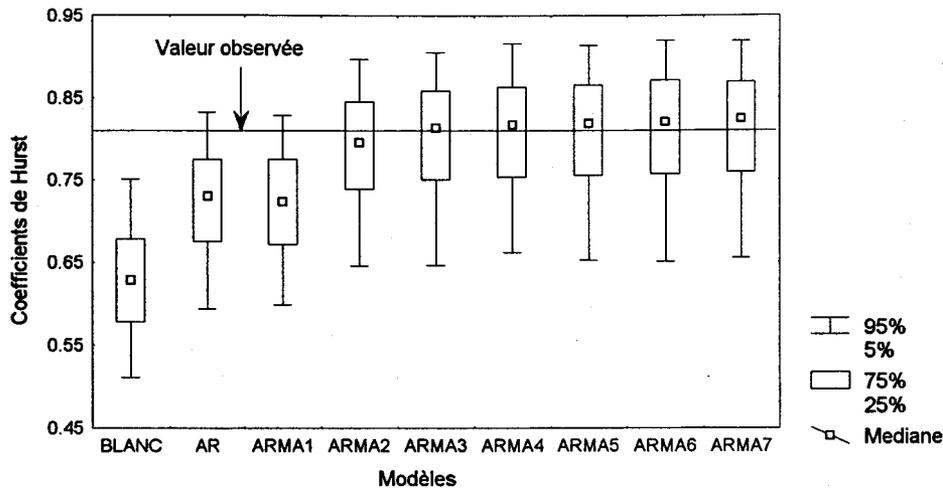


Figure 4.8. Distribution des coefficients de Hurst des séries simulées.

- les modèles ARMA2 à ARMA7 reproduisent mieux le coefficient de Hurst. Toutefois, certains de ces modèles (ARMA2 et ARMA4) admettent des valeurs minimums inférieures à celles des modèles classiques. Les valeurs maximums du coefficient de Hurst sont comparables.
- le modèle BLANC se comporte tout à fait différemment des autres modèles et ne préserve pas les coefficients d'autocorrélation et de Hurst.

D'autres caractéristiques statistiques concernant les séquences d'années consécutives d'apports énergétiques agrégés (runs) en déficits et en surplus, par rapport à la moyenne générale, ont aussi été considérées dans cette étude comparative des neuf modèles. Il s'agit des longueurs des séquences (notées respectivement par LSD et LSS pour les déficits et les surplus) et des apports cumulés durant ces événements consécutifs (notées respectivement par CSD et CSS pour les périodes déficitaires et excédentaires). Plus précisément, les modèles ont été comparés eu égard à la préservation des valeurs minimales, moyennes et maximales de ces caractéristiques statistiques.

La figure 4.9 présente les écarts énergétiques annuels agrégés par rapport à la moyenne pour la période 1943-1993. Cette figure illustre bien les séquences historiques déficitaires et excédentaires. On remarque, en particulier, la plus longue série d'années consécutives en déficit énergétique (cinq ans) de 1987 à 1991. Le tableau 4.7 donne les valeurs des caractéristiques des séquences observées pour cette série d'apports.

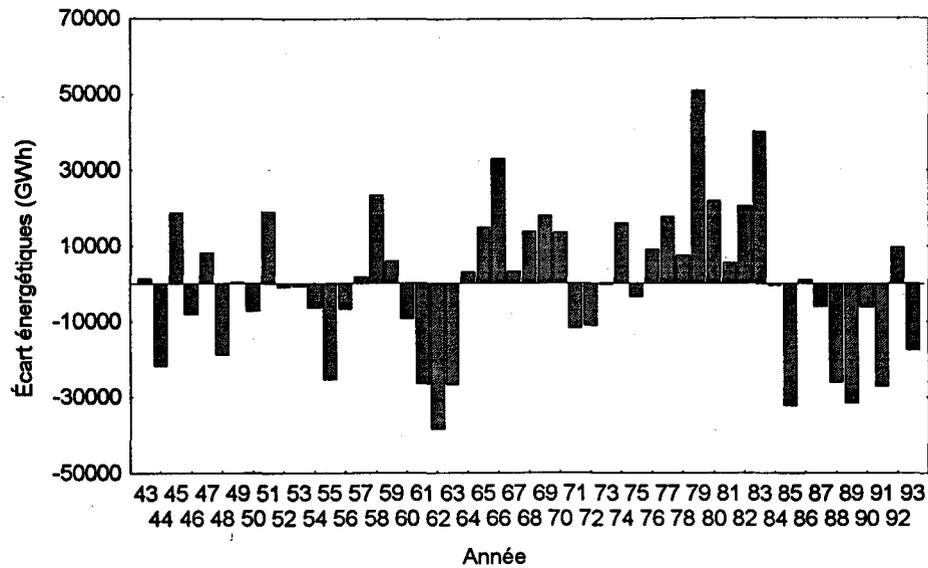


Figure 4.9. Écarts énergétiques annuels 1943-1993 (déficits et surplus).

Tableau 4.7. Caractéristiques de séquence d'apports agrégés déficitaires et excédentaires.

	LSD (années)	LSS (années)	CSD (GWh)	CSS (GWh)
MINIMUM	1.00	1.00	3 747	416
MOYEN	2.27	2.36	34 279	34 279
MAXIMUM	5.00	8.00	101 884	172 310

La distribution de ces caractéristiques statistiques est illustrée, pour les neuf modèles, par des diagrammes en boîte présentés à l'annexe D. On s'intéresse plus particulièrement aux caractéristiques des séquences déficitaires.

Exception faite de la longueur maximale des séquences excédentaires ($LSS = 8.00$, cf. tableau 4.7) que les modèles ARMA2 à ARMA7 reproduisent bien, les neuf modèles ne préservent aucune des caractéristiques statistiques de séquence. On constate, en particulier, que les modèles sur-estiment tous en moyenne la longueur maximale et le cumul maximal des séquences déficitaires ($LSD = 5$ et $CSD = 101\,883.89$, cf. tableau 4.7). Les neuf modèles produisent donc des scénarios de 50 années conservateurs par rapport à ces 2 critères. Enfin, la variabilité de la distribution des caractéristiques de séquence est systématiquement plus grande pour les modèles de O'Connell que pour les modèles classiques AR et ARMA1.

Dans certaines circonstances, par exemple pour l'aménagement d'un nouveau bassin (Rasmussen *et al.*, 1996), les caractéristiques statistiques des séquences sont très

importantes et il y a donc place à amélioration pour le modèle. Toutefois, nous émettons quelques réserves quant à la préservation de ces caractéristiques. En effet, il nous semble risqué, par exemple, de vouloir reproduire à tout prix la longueur maximale des séquences déficitaires (ici, $LSD = 5$) alors qu'il s'agit d'une seule réalisation de cet événement sur une série de taille restreinte.

4.3.3.3 Modèles retenus pour la prédiction

Dans une optique de prévision, les modèles classiques (AR et ARMA1) sont préférables, principalement à cause du comportement de leurs résidus. En effet, l'analyse effectuée à la section 4.3.3.2 a montré que les résidus des modèles AR et ARMA1 étaient plus compatibles avec les hypothèses de base garantissant l'efficacité des prévisions. Toutefois, certains modèles de O'Connell (ARMA2 et ARMA3), qui admettent une gamme de scénarios plus vaste, sont intéressants lorsqu'utilisés dans un but plus descriptif (simulation exploratoire). Comme cela a été vu à la section précédente, les caractéristiques de ces modèles sont plus variables et supposent des changements majeurs du climat. Ils permettent donc de générer des scénarios de 50 années d'apports plus extrêmes qui peuvent être ensuite soumis comme intrant à des modèles de planification.

Les modèles classiques AR et ARMA1 ont donc été retenus. Les modèles ARMA2 et ARMA3, qui préservent le coefficient de Hurst (cf. figure 4.8), ont aussi été considérés. Enfin, les modèles ARMA4 à ARMA7 admettent des scénarios d'apports énergétiques de moyenne et d'écart-type trop élevée ou trop faible sur 50 années (cf. figures 4.5 et 4.6). Ils ne répondent donc pas aux caractéristiques hydrologiques et climatiques du réseau actuel et ont été rejetés.

4.3.4 Prévision et distribution des écarts énergétiques annuels

Nous nous intéressons ici à la prévision ainsi qu'à la distribution des écarts énergétiques annuels agrégés par rapport à la moyenne générale (variable centrée, $z_t = y_t - \mu$). Dans la section précédente, quatre modèles de type ARMA(1,1) ont été retenus: les modèles AR, ARMA1, ARMA2 et ARMA3. L'expression mathématique du modèle ARMA(1,1) pour la variable centrée z_t est déduite de l'équation (4.8) et est donnée par:

$$z_t = \phi z_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1} \quad (4.14)$$

Les valeurs estimées des paramètres ϕ et θ pour les quatre modèles retenus sont données au tableau 4.4.

4.3.4.1 Prédiction des écarts énergétiques et probabilités d'occurrence

Considérons la prédiction $\hat{z}_t(l)$ de l'année $t + l$, $l \geq 1$, étant donné que nous sommes au temps t . La période t est l'origine de la prédiction. L'origine de toutes les prévisions présentées dans cette section est l'année 1993 qui représente la dernière année d'observation disponible (au moment où le projet a débuté). La prédiction $\hat{z}_t(l)$, issue d'un modèle ARMA(1,1) (éq. 4.14), est donnée par:

$$\begin{aligned}\hat{z}_t(1) &= \hat{\phi}z_t - \hat{\theta}\hat{\varepsilon}_t \\ \hat{z}_t(l) &= \hat{\phi}\hat{z}_t(l-1), \quad l \geq 2\end{aligned}\quad (4.15)$$

où $\hat{\phi}$ et $\hat{\theta}$ sont les estimations des paramètres ϕ et θ . Le terme $\hat{\varepsilon}_t$ correspond au résidu du modèle à la période d'origine t (éq. 4.11). La prédiction issue du modèle AR est déduite de l'équation (4.15) en posant $\hat{\theta} = 0$:

$$\begin{aligned}\hat{z}_t(1) &= \hat{\phi}z_t \\ \hat{z}_t(l) &= \hat{\phi}\hat{z}_t(l-1), \quad l \geq 2\end{aligned}\quad (4.16)$$

Les prévisions $\hat{z}_t(l)$ du modèle ARMA(1,1) tendent vers la moyenne générale de la série (ici zéro, puisque les données sont centrées) à mesure que l augmente. Leur distribution est entièrement connue si les résidus satisfont aux hypothèses de base (normalité, indépendance temporelle, homogénéité de la variance). Il s'agit d'une loi normale de moyenne $E\{\hat{z}_t(l)\} = \phi^l z_t$ et de variance donnée par (Box *et al.* 1994):

$$\text{Var}\{\hat{z}_t(l)\} = \sigma_\varepsilon^2 \left(1 + \sum_{j=1}^{l-1} \psi_j^2 \right) \quad (4.17)$$

où σ_ε^2 est la variance des résidus du modèle et où les ψ_j s'expriment, pour un modèle ARMA(1,1), de la manière suivante:

$$\psi_j = (\phi - \theta)\phi^{j-1}, \quad j = 1, 2, \dots \quad (4.18)$$

Pour la prédiction issue du modèle AR ($\theta = 0$), on peut déduire que l'expression (4.17) se réduit à:

$$\text{Var}\{\hat{z}_t(l)\} = \sigma_\varepsilon^2 \frac{1 - \phi^{2l}}{1 - \phi^2} \quad (4.19)$$

Ces variances peuvent être estimées en remplaçant, dans les expressions (4.17) et (4.19), ϕ , θ et σ_ε^2 par leurs estimations (tableau 4.4 pour les paramètres et tableau 4.5 pour la variance des résidus). Remarquons que la variance de la première prévision ($l = 1$) correspond à celle des résidus. La dispersion de la loi de probabilité des prévisions (loi normale) augmente toutefois à mesure que l'on s'éloigne de la période d'origine.

Connaissant la distribution des prévisions, un intervalle de confiance peut être construit. L'intervalle de niveau de signification $(1 - \alpha)100\%$ est donné par:

$$\hat{z}_t(l) \pm u_{\alpha/2} \hat{\sigma}_\varepsilon \left(1 + \sum_{j=1}^{l-1} \psi_j^2 \right)^{1/2} \quad (4.20)$$

Le tableau 4.8 présente la prévision de l'écart énergétique annuel ($z_t = y_t - \mu$) pour les années 1994 à 1998, son écart-type et l'intervalle de confiance associé de niveau 95%.

Puisque la loi de probabilité est entièrement connue, il est possible d'évaluer analytiquement la probabilité d'occurrence de certains types d'événements liés aux écarts énergétiques annuels. La probabilité d'observer un surplus (écart positif) ou un déficit (écart négatif) dans les prochaines années a été déterminée à titre d'exemple. Les résultats sont présentés au tableau 4.8. Enfin, le déficit correspondant à un risque donné (probabilité de non-dépassement dans ce cas) a aussi été évalué. Le tableau 4.9 donne, pour cinq probabilités au non-dépassement, les déficits prédits correspondants pour les années 1994 à 1998.

L'examen des tableaux 4.8 et 4.9 met particulièrement en lumière la grande variabilité des propriétés statistiques des modèles de O'Connell (cf. section 4.3.3.2) et le fait qu'ils admettent une persistance plus grande que les modèles classiques. Cette persistance est due aux valeurs des paramètres ϕ qui, déterminées à partir de l'approche de O'Connell, sont très proches de 1 (cf. tableau 4.4).

Plus précisément, les tableaux 4.8 et 4.9 permettent de constater que:

- tous les modèles retenus prédisent des déficits pour les années à venir;
- les écart énrgetiques prédits par les modèles ARMA2 et ARMA3 sont plus élevés en valeur absolue et tendent moins rapidement vers la moyenne générale de la série (persistance);
- les modèles AR et ARMA1 produisent des prévisions dont les variances sont inférieures à celles issues des modèles ARMA2 et ARMA3;

- la probabilité d'être en déficit, obtenue à partir des modèles ARMA2 et ARMA3, est systématiquement plus grande;
- les déficits prédits correspondant aux diverses probabilités au non-dépassement, issus des modèles ARMA2 et ARMA3, sont systématiquement plus sévères.

Tableau 4.8. Prévision (GWh), écart-type, intervalle de confiance et probabilité de surplus et de déficit pour les modèles retenus (1994-1998).

Modèle	Année	Prévision	Écart-type	Intervalle de confiance à 95%	Prob. de déficit	Prob. de Surplus
AR	1994	-7 279	17 824	[-42 214 27 656]	0.66	0.34
	1995	-2 984	19 264	[-40 742 34 773]	0.56	0.44
	1996	-1 223	19 496	[-39 435 36 988]	0.52	0.48
	1997	-502	19 534	[-38 789 37 785]	0.51	0.49
	1998	-206	19 541	[-38 505 38 094]	0.50	0.50
ARMA1	1994	-5 805	17 945	[-40 976 29 366]	0.63	0.37
	1995	-2 891	18 968	[-40 068 34 285]	0.56	0.44
	1996	-1 440	19 213	[-39 097 36 217]	0.53	0.47
	1997	-717	19 273	[-38 493 37 058]	0.52	0.48
	1998	-357	19 288	[-38 162 37 447]	0.51	0.49
ARMA2	1994	-8 967	18 604	[-45 429 27 494]	0.69	0.31
	1995	-8 077	19 449	[-46 190 30 049]	0.66	0.34
	1996	-7 263	20 108	[-46 675 32 148]	0.64	0.36
	1997	-6 537	20 627	[-46 965 33 891]	0.62	0.38
	1998	-5 884	21 037	[-47 116 35 349]	0.61	0.39
ARMA3	1994	-9 494	18 705	[-46 155 26 167]	0.69	0.31
	1995	-8 734	20 237	[-48 400 30 930]	0.67	0.33
	1996	-8 036	21 449	[-50 076 34 004]	0.65	0.35
	1997	-7 392	22 424	[-51 343 36 557]	0.63	0.37
	1998	-6 801	23 216	[-52 305 38 702]	0.62	0.38

Tableau 4.9. Déficit prédit correspondant à diverses probabilité au non-dépassement pour les modèles retenus (1994-1998).

Modèle	Année	Probabilité au non-dépassement				
		1%	2%	5%	10%	20%
AR	1994	-48 744	-43 885	-36 597	-30 121	-22 280
	1995	-47 799	-42 548	-34 671	-27 672	-19 197
	1996	-46 577	-41 262	-33 291	-26 208	-17 631
	1997	-45 945	-40 620	-32 632	-25 536	-16 942
	1998	-45 664	-40 337	-32 347	-25 248	-16 651
ARMA1	1994	-47 550	-42 658	-35 321	-28 801	-20 907
	1995	-47 016	-41 846	-34 090	-27 199	-18 855
	1996	-46 136	-40 899	-33 042	-26 062	-17 610
	1997	-45 554	-40 300	-32 419	-25 417	-16 938
	1998	-45 228	-39 970	-32 083	-25 076	-16 590
ARMA2	1994	-52 244	-47 173	-39 566	-32 808	-24 624
	1995	-53 315	-48 014	-40 061	-32 995	-24 439
	1996	-54 042	-48 560	-40 338	-33 033	-24 187
	1997	-54 522	-48 899	-40 465	-32 971	-23 897
	1998	-54 823	-49 088	-40 486	-32 843	-23 589
ARMA3	1994	-53 008	-47 909	-40 261	-33 465	-25 236
	1995	-55 813	-50 297	-42 022	-34 669	-25 766
	1996	-57 933	-52 086	-43 316	-35 523	-26 087
	1997	-59 558	-53 445	-44 276	-36 129	-26 265
	1998	-60 811	-54 482	-44 989	-36 554	-26 340

4.3.4.2 Prédiction des écarts énergétiques cumulés et probabilités d'occurrence

Considérons maintenant l'écart énergétique cumulé prédit $S_t(k)$ d'origine t et d'horizon k défini comme suit:

$$S_t(k) = \sum_{l=1}^k \hat{z}_t(l) \quad (4.21)$$

Puisque la loi des prévisions $\hat{z}_t(l)$ est connue (loi normale), il est possible de déterminer la distribution des écarts cumulés sur les k prochaines années. Cette distribution dépend non seulement des paramètres du modèle ARMA(1,1), mais aussi des corrélations entre les k prévisions $\hat{z}_t(l)$. En effet, la somme de ces k variables aléatoires normales corrélées entre

elles de moyenne et d'écart-type respectivement $\mu_l = \phi^l z_l$ et $\sigma_l^2 = \text{Var}\{\hat{z}_l(l)\}$, $l = 1, \dots, k$, est distribuée selon une loi normale de moyenne et de variance données par:

$$\mu_l(k) = \sum_{l=1}^k \mu_l \quad (4.22)$$

$$\sigma_l^2(k) = \sum_{l=1}^k \sigma_l^2 + 2 \sum_{l=1}^{k-1} \sum_{j=1}^{k-l} \sigma_{l,l+j} \quad (4.23)$$

où $\sigma_{l,l+j}$ désigne la covariance entre les prévisions $\hat{z}_l(l)$ et $\hat{z}_l(l+j)$. Cette covariance s'exprime (Box *et al.*, 1994, chapitre 5), pour un modèle de type ARMA, de la manière suivante:

$$\sigma_{l,l+j} = \sigma_\varepsilon^2 \sum_{i=0}^{l-1} \psi_i \psi_{j+i} \quad (4.24)$$

où les termes ψ_j sont donnés à l'équation (4.18).

Pour déterminer entièrement la distribution des écarts cumulés, il suffit donc d'évaluer les expressions (4.22) et (4.23) à partir des équations (4.17), (4.18) et (4.24). Après quelques manipulations algébriques (cf. annexe E), on peut montrer, pour le modèle AR, que la moyenne et la variance de l'écart énergétique cumulé prédit $S_l(k)$ sont respectivement:

$$\mu_l(k) = \frac{\phi(1-\phi^k)}{1-\phi} z_l \quad (4.25)$$

$$\sigma_l^2(k) = \frac{\sigma_\varepsilon^2}{(1+\phi)(1-\phi)^3} \left[k(1-\phi^2) - \phi(2+\phi) + 2\phi^{k+1}(1+\phi) - \phi^{2(k+1)} \right]$$

Le tableau 4.10 donne, pour les modèles AR, ARMA1, ARMA2 et ARMA3, les écarts cumulés prédits sur des horizons de 2 à 5 années. L'écart-type, l'intervalle de confiance à 95% et la probabilité d'occurrence de surplus et de déficits cumulés y sont aussi présentés. Le tableau 4.11 donne, pour cinq probabilités au non-dépassement, les déficits cumulés prédits.

On remarque, en particulier, la valeur élevée des écart-types des prévisions des écarts cumulés (cf. tableau 4.10). Cette variance devient très élevée lorsque l'horizon considéré est grand (nombre d'années k dans la somme $S_l(k)$, éq. 4.21). Cela est tout à fait normal

puisque le cumul fait intervenir plusieurs prévisions ayant chacune leur incertitude d'échantillonnage (cf. tableau 4.8). De plus, l'estimation des paramètres de ce type de modèle (ARMA(1,1)) nécessite un nombre assez grand d'observations (Box *et al.*, 1994).

Tableau 4.10. Prévision des écarts cumulés (GWh), écart-type, intervalle de confiance et probabilité de surplus et de déficit pour les modèles retenus (1994-1998).

Modèle	Années	Prévision	Écart-type	Intervalle de confiance à 95%	Prob. de déficit	Prob. de Surplus
AR	1994-95	-10 263	30 810	[-70 652 50 126]	0.630	0.370
	1994-96	-11 486	41 719	[-93 255 70 283]	0.608	0.392
	1994-97	-11 988	51 012	[-111 972 87 996]	0.593	0.407
	1994-98	-12 194	59 109	[-128 048 103 660]	0.582	0.418
ARMA1	1994-95	-8 696	30 032	[-67 560 50 168]	0.614	0.386
	1994-96	-10 136	40 478	[-89 473 69 201]	0.599	0.401
	1994-97	-10 853	49 597	[-108 063 86 357]	0.587	0.413
	1994-98	-11 210	57 665	[-124 235 101 815]	0.577	0.423
ARMA2	1994-95	-17 044	30 584	[-76 990 42 902]	0.711	0.289
	1994-96	-24 307	42 412	[-107 435 58 821]	0.716	0.284
	1994-97	-30 844	54 345	[-137 360 75672]	0.714	0.286
	1994-98	-36 728	66 379	[-166 831 93 375]	0.710	0.290
ARMA3	1994-95	-18 228	32 379	[-81 691 45 235]	0.713	0.287
	1994-96	-26 264	46 617	[-117 634 65 106]	0.713	0.287
	1994-97	-33 656	61 475	[-154 148 86 836]	0.708	0.292
	1994-98	-40 457	76 835	[-191 054 110 140]	0.701	0.299

Tableau 4.11. Déficit cumulé prédits correspondant à diverses probabilités au non-dépassement pour les modèles retenus (1994-1998).

Modèle	Année	Probabilité au non-dépassement				
		1%	2%	5%	10%	20%
AR	1994-95	-81 939	-73 540	-60 942	-49 748	-36 194
	1994-96	-108 539	-97 166	-80 107	-64 951	-46 597
	1994-97	-130 661	-116 755	-95 896	-77 363	-54 921
	1994-98	-149 703	-133 589	-109 420	-87 945	-61 941
ARMA1	1994-95	-78 562	-70 376	-58 095	-47 184	-33 972
	1994-96	-104 302	-93 267	-76 716	-62 010	-44 203
	1994-97	-126 233	-112 713	-92 433	-74 414	-52 595
	1994-98	-145 360	-129 641	-106 061	-85 111	-59 742
ARMA2	1994-95	-88 195	-79 857	-67 351	-56 240	-42 784
	1994-96	-122 973	-111 411	-94 069	-78 660	-60 002
	1994-97	-157 270	-142 455	-120 234	-100 490	-76 582
	1994-98	-191 149	-173 054	-145 912	-121 796	-92 594
ARMA3	1994-95	-93 553	-84 727	-71 487	-59 723	-45 479
	1994-96	-134 712	-122 004	-102 942	-86 006	-65 498
	1994-97	-176 669	-159 911	-134 774	-112 440	-85 395
	1994-98	-219 202	-198 257	-166 839	-138 925	-105 123

4.3.5 Conclusions et recommandations

(a) Procédure de simulation actuellement utilisée

La procédure utilisée actuellement à Hydro-Québec pour déterminer la distribution des écarts énergétiques repose entièrement sur une simulation dont la valeur de départ est un écart nul. Telle qu'appliquée, la simulation ne donne aucune information concernant les écarts pouvant survenir dans le futur à une date spécifique. D'une part, les données simulées ne tiennent pas compte du dernier événement de la série, malgré le fait qu'un écart énergétique est fortement influencé par l'événement de l'année précédente. En effet, la valeur initiale de la procédure est fixée à zéro. D'autre part, l'approche consiste à simuler une longue série que l'on partitionne ensuite pour déterminer les probabilités. Ceci peut induire un biais dans les résultats (Hipel et McLeod, 1994).

Nous recommandons donc l'emploi de prévisions qui permettent d'estimer l'écart énergétique susceptible de se produire, avec la plus grande probabilité, à une période donnée dans le futur. L'incertitude d'échantillonnage (variance et intervalle de confiance) peut être évaluée analytiquement. Cette approche permet aussi de calculer aisément la plupart des événements ou probabilités qui intéressent Hydro-Québec.

(b) Adéquation du modèle AR(1)

Compte-tenu des résultats présentés à la section 4.3.3, nous recommandons l'emploi du modèle stationnaire AR pour des fins de prévision et de simulation. Ce modèle, qui donne des résultats très comparables au modèle ARMA1, a l'avantage d'être très simple d'utilisation. Toutefois, nous ne rejetons pas les modèles ARMA2 et ARMA3 qui admettent une gamme de scénarios plus vaste. Comme il a été vu aux sections 4.3.3 et 4.3.4, les caractéristiques de ces modèles sont plus variables et supposent une plus grande persistance du phénomène. Ils permettent donc de générer des scénarios d'apports ou d'écarts énergétiques plus extrêmes qui peuvent être ensuite soumis comme intrants à des modèles de planification. Nous croyons, par contre, que l'emploi de tels modèles nécessitent une étude approfondie du phénomène de Hurst qui ne fait pas l'unanimité dans la communauté scientifique. Nous recommandons aussi d'utiliser la formule de correction du biais (cf. éq. 4.10) pour le calcul du coefficient d'autocorrélation et la détermination du paramètre du modèle AR (méthode des moments modifiés). La calibration du modèle doit être mise à jour à mesure qu'une nouvelle observation devient disponible.

Enfin, si les caractéristiques statistiques des séquences sont considérées très importantes, il y a place à amélioration pour le modèle. Toutefois, comme il a été mentionné à la section 4.3.3.2, nous émettons quelques réserves quant à la préservation, à tout prix, de ces caractéristiques.

(c) Distribution des écarts

Les propriétés théoriques des prévisions issues des modèles considérés sont bien connues (cf. section 4.3.4), ce qui permet de déterminer analytiquement les probabilités d'événements spécifiques plus complexes (cumul des écarts, par exemple). Toutefois, si on s'intéresse particulièrement aux durées des séquences consécutives de déficits ou de surplus, il n'existe pas à notre connaissance d'équations analytiques directes. On peut alors avoir recours à une simulation dont la valeur initiale doit correspondre à l'écart observé la dernière année.

4.4 Mise à jour: série 1943-1995

Comme convenu au début du projet C2, les travaux réalisés ont porté sur les séries d'apports énergétiques s'échelonnant de 1943 à 1993. Les apports des années 1994 et 1995 étant maintenant disponibles (182 822.4 GWh et 155 248 Gwh, respectivement), il est intéressant d'examiner l'influence qu'ont ces nouvelles observations sur le résultats des analyses statistiques. La présente section est donc consacrée à une brève mise à jour des analyses effectuées sur les apports énergétiques agrégés. Plus précisément, sont présentés dans ce qui suit:

- les principales caractéristiques statistiques de la série 1943-1995 (tableau 4.12);
- l'application de procédure bayésienne de sélection de non-stationnarité de la moyenne (figure 4.10 et tableau 4.13);
- les prévisions des écarts énergétiques et les probabilités d'occurrence de déficit et de surplus pour les années 1996-2000 (tableau 4.14);
- les déficits prédits de probabilité au non-dépassement de 1%, 2%, 5%, 10% et 20% (tableau 4.15);
- les prévisions des écarts énergétiques cumulés (horizons de 2 à 5 ans) et les probabilités d'occurrence de déficits et de surplus cumulés pour les années 1996-2000 (tableau 4.16);
- les déficits cumulés prédits de probabilité au non-dépassement de 1%, 2%, 5%, 10% et 20% (tableau 4.17);

Les caractéristiques statistiques de la série 1943-1995 diffèrent peu de celles de la série 1943-1993 (cf. tableaux 3.20 et 4.12). Toutefois, la moyenne générale a diminuée de 558 GWh et l'écart-type a augmenté quelque peu. Les deux nouvelles observations ne sont pas exceptionnelles puisque le minimum et le maximum demeurent inchangés. Le coefficient d'autocorrélation (estimateur biaisé, cf. éq. 3.1) est plus faible lorsqu'on ajoute les années 1994-1995 et il passe de 0.356 à 0.344. Statistiquement ces écarts sont non significatifs.

Tableau 4.12. Caractéristiques statistiques des apports énergétiques annuels agrégés (1943-1995).

MOYENNE	INT. CONF (95%)	MÉDIANE	ÉCART-TYPE	Cv	Cs	MINIMUM	MAXIMUM	TOTAL
183 264	[177 995 ; 188 533]	183 444	19 116	0.104	0.174	145 155	234 438	9 713 014

Les résultats de l'application de la procédure bayésienne de détection de non-stationnarité de la moyenne sont similaires à ceux obtenus sans les observations de 1994 et 1995 (cf. section 4.2.2). La figure 4.10(a) présente les apports énergétiques annuels agrégés ainsi que les

probabilités a posteriori $f(\tau|x_1, \dots, x_n)$ associées à chacune des années (courbe en trait gras). La figure 4.10(b) donne la loi a posteriori inconditionnelle de l'amplitude du changement $f(\delta|x_1, \dots, x_n)$. La procédure bayésienne indique, comme pour la série 1943-1993, que la date la plus probable d'occurrence d'un changement de moyenne est l'année 1984 (cf. figure 4.10(a)). On peut observer aussi une très faible probabilité que cette rupture survienne en 1963. Toutefois, le second mode de la distribution a posteriori de l'amplitude de la rupture (cf. figure 4.10(b)) est moins marqué que celui obtenu à partir de la série 1943-1993. Le changement de moyenne en 1984 serait donc plus important que celui de 1963 suite à l'ajout des deux nouvelles observations. Cela n'est pas surprenant puisqu'elles sont toutes les deux inférieures à la moyenne générale. Le tableau 4.13 donne les principaux résultats de l'analyse. On y présente l'année de changement la plus probable τ , sa probabilité a posteriori $f(\tau|x_1, \dots, x_n)$, la valeur de la statistique R , les moyennes avant et après τ , l'amplitude de la rupture δ et l'intervalle de crédibilité de niveau 95%. On constate, en particulier, qu'une amplitude de rupture nulle demeure crédible puisque la valeur 0 appartient à l'intervalle de crédibilité de niveau 95%, et ce même si la probabilité a posteriori du changement passe de 0.21 à 0.25.

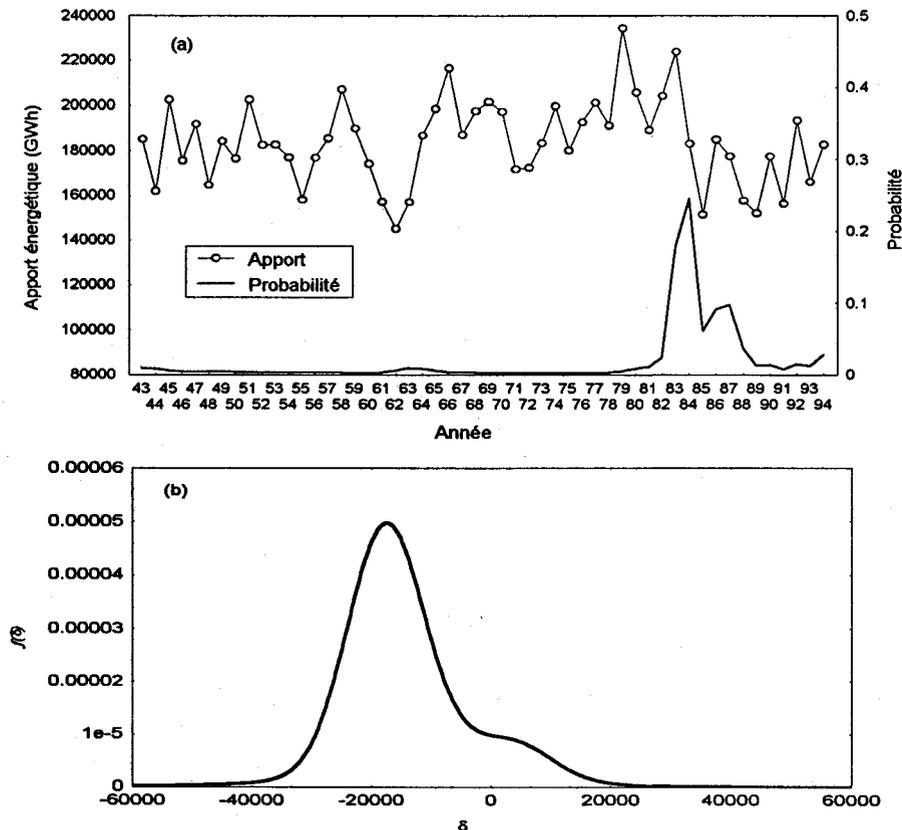


Figure 4.10. Résultats de la procédure bayésienne - Apports agrégés (1943-1995).

Tableau 4.13. Résultats de la procédure bayésienne pour la série agrégée (1943-1995)

Période	τ	$f(\tau)$	R	\bar{x}_τ	$\bar{x}_{n-\tau}$	δ	Intervalle pour δ
1943-1995	1984	0.25	16.1	187 102	168 613	-18 488	[-31 700 ; 11 200]

Enfin, l'analyse des résidus du modèle autorégressif d'ordre 1, AR, montre qu'il est toujours adéquat pour représenter les apports énergétiques annuels agrégés. Ce modèle a donc été utilisé pour mettre à jour la prévision des écarts énergétiques annuels et le calcul des probabilités d'occurrence. Les prévisions des écarts énergétiques et les probabilités d'occurrence issues des modèles AR avec et sans correction de biais sur le coefficient d'autocorrélation (éq. 4.10 et 3.1) sont présentées aux tableaux 4.14 à 4.17.

Les résultats permettent de constater que l'utilisation de la méthode des moments modifiés (décrite à la section 4.3.2.1 et recommandée à la section 4.3.5) produit des déficits prédits systématiquement plus importants. La différence entre les prévisions des deux modèles est généralement non négligeable du point de vue planification. En particulier, le modèle non biaisé prédit pour l'année 1996 un déficit de 10 982 GWh par rapport à 9 638 GWh pour le modèle sans correction de biais.

Tableau 4.14. Prévision (GWh), écart-type, intervalle de confiance et probabilité de surplus et de déficit pour les modèles retenus (1996-2000).

Modèle	Année	Prévision	Écart-type	Intervalle de confiance à 95%	Prob. de déficit	Prob. de Surplus
AR	1996	-9 638	17 636	[-44 203 24 928]	0.708	0.292
Biaisé	1997	-3 315	18 986	[-40 528 33 898]	0.569	0.431
$\hat{\phi} = 0.344$	1998	-1 140	19 140	[-38 654 36 373]	0.524	0.476
	1999	-392	19 158	[-37 942 37 157]	0.508	0.492
	2000	-135	19 160	[-37 688 37 418]	0.503	0.497
AR	1996	-10 982	17 642	[-45 561 23 597]	0.733	0.267
Non-biaisé	1997	-4 305	19 438	[-42 403 33 792]	0.588	0.412
$\hat{\phi} = 0.392$	1998	-1 688	19 699	[-40 299 36 922]	0.534	0.466
	1999	-662	19 739	[-39 350 38 027]	0.513	0.487
	2000	-259	19 745	[-38 960 38 441]	0.505	0.495

Tableau 4.15. Déficit prédit correspondant à diverses probabilité au non-dépassement pour les modèles retenus (1996-2000).

Modèle	Année	Probabilité au non-dépassement				
		1%	2%	5%	10%	20%
AR	1996	-50 664	-45 856	-38 645	-32 238	-24 480
Biaisé	1997	-47 484	-42 308	-34 545	-27 647	-19 294
$\hat{\phi} = 0.344$	1998	-45 666	-40 449	-32 622	-25 669	-17 249
	1999	-44 960	-39 738	-31 904	-24 944	-16 516
	2000	-44 708	-39 485	-31 650	-24 689	-16 260
AR	1996	-52 024	-47 215	-40 001	-33 592	-25 830
Non-biaisé	1997	-49 524	-44 225	-36 277	-29 215	-20 664
$\hat{\phi} = 0.392$	1998	-47 514	-42 144	-34 089	-26 933	-18 267
	1999	-46 581	-41 200	-33 129	-25 958	-17 274
	2000	-46 193	-40 811	-32 737	-25 564	-16 877

Tableau 4.16. Prévision des écarts cumulés (GWh), écart-type, intervalle de confiance et probabilité de surplus et de déficit pour les modèles retenus (1996-2000).

Modèle	Années	Prévision	Écart-type	Intervalle de confiance à 95%	Prob. de déficit	Prob. de Surplus
AR	1996-1997	-12 953	29 544	[-70 859 44 953]	0.669	0.331
Biaisé	1996-1998	-14 093	39 217	[-90 958 62 772]	0.640	0.360
$\hat{\phi} = 0.344$	1996-1999	-14 485	47 335	[-107 262 78 292]	0.620	0.380
	1996-2000	-14 620	54 373	[-121 191 91 951]	0.606	0.394
AR	1996-1997	-15 287	30 238	[-74 553 43 979]	0.693	0.307
Non-biaisé	1996-1998	-16 975	40 717	[-96 781 62 831]	0.662	0.338
$\hat{\phi} = 0.392$	1996-1999	-17 637	49 604	[-114 861 79 587]	0.639	0.361
	1996-2000	-17 896	57 332	[-130 267 94 475]	0.622	0.378

Tableau 4.17. Déficit cumulés prédit correspondant à diverses probabilité au non-dépassement pour les modèles retenus (1996-2000).

Modèle	Année	Probabilité au non-dépassement				
		1%	2%	5%	10%	20%
AR	1996-1997	-81 683	-73 629	-61 549	-50 815	-37 818
Biaisé	1996-1998	-105 325	-94 634	-78 599	-64 351	-47 099
$\hat{\phi} = 0.344$	1996-1999	-124 603	-111 699	-92 344	-75 147	-54 323
	1996-2000	-141 111	-126 289	-104 056	-84 302	-60 382
AR	1996-1997	-85 630	-77 388	-65 024	-54 039	-40 736
Non-biaisé	1996-1998	-111 698	-100 598	-83 949	-69 156	-51 244
$\hat{\phi} = 0.392$	1996-1999	-133 033	-119 511	-99 228	-81 207	-59 385
	1996-2000	-151 271	-135 642	-112 199	-91 370	-66 148

5 RÉFÉRENCES BIBLIOGRAPHIQUES

- Akaike, H., 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, AC-19: 716-723.
- Beard, L.R., 1971. HEC-4 Monthly Streamflow Simulation. The Hydrologic Engineering Center, Corps of Engineers, US Army, Davis, California 95616.
- Berger, J.O., 1985. *Statistical Decision Theory and Bayesian Analysis*. Berlin, Springer.
- Bernier, J., 1977. Étude de la stationnarité des séries hydrométéorologiques. *La Houille Blanche*, 4:313-319.
- Bernier, J., 1994. Statistical detection of changes in geophysical series *dans* *Engineering Risk in Natural Resources Management*, 159-176, L.Duckstein and E. Parent (eds).
- Bickel, P.J. and Doksum, K.A., 1977. *Mathematical Statistics*. Holden-Day, Oakland.
- Box, G.E.P., Jenkins, G.M. and Reinsel, G.C., 1994. *Time Series Analysis*. Prentice Hall, Englewood.
- Box, G.E.P. and Taio, G.C., 1973. *Bayesian Inference in Statistical Analysis*. Wiley, New York.
- Bras, R.L. and Rodriguez-Iturbe, 1985. *Random Functions and Hydrology*. Addison-Wesley, Reading, Massachusetts.
- Bruneau, P., 1984. Validation de longues séries de débits observés. Rapport interne, No 08741-RA-84/45, Hydro-Québec.
- Bruneau, P., 1986a. Étude sur la productibilité: Séries de 50 années de débits aux sites. Rapport interne, No 08741-RA-86/26, Hydro-Québec.
- Bruneau, P., 1986b. Validation de longues séries de débits observés: Phase II. Rapport interne, No 08741-RA-86/45, Hydro-Québec.
- Bruneau, P., 1986c. Validation de longues séries de débits observés: Phase II - La Grande. Rapport interne, No 08741-RA-86/45, Hydro-Québec.
- Bruneau, P., 1988. L'analyse bayésienne: une méthode simple pour identifier les dates probables de changement de paramètres. *Proceedings of the VIth IWRA Congress on Water Resources*, Ottawa. IWRA, Urbana, USA, 2, 171-189.
- Bruneau, P. et Rassam, J.-C., 1983. Application d'un modèle bayésien de détection de changements de moyennes dans une série. *Journal des Sciences Hydrologiques*, 28(3): 341-354.

- Cavadias, G., 1992. A survey of current approaches to modelling of hydrological time-series with respect to climate variability and change. World Meteorological Organization, WMO/TD-No. 534.
- Cochran, W.G., 1977. Sampling Techniques. Wiley, New York.
- Conover, W.J., Johnson, M.E. and Johnson, M.M., 1981. A comparative study of tests for homogeneity of variances with applications to the outer continental shelf bidding data. *Technometrics*, 23: 351-361.
- D'Agostino, R.B. and Stephens, M.A., 1986. Goodness-of-fit Techniques. Marcel-Dekker, New York.
- Draper, N.R. and Smith, H., 1966. Applied Regression Analysis. Wiley, New York.
- Feller, W. 1951. The asymptotic distribution of the range of sums of independent random variables, *Ann. Math. Stat.*, 22: 427-432
- Findlay, B.F., Gullett, D.W., Malone, L., Reycraft, J., Skinner, W.R., Vincent, L., and Whitewood, R., 1994. Canadian national and regional standardized annual precipitation departures, pp 800-828. In T.A. Boden, D.P. Kaiser, R.J Sepanski, and F.W. Stoss (eds.), *Trends '93: A Compendium of Data on Global Change*, ORNL/CDIAC-65. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tenn. U.S.A.
- Grygier, J.C. and Stedinger, J.R., 1990. SPIGOT: A synthetic streamflow generation software package. Technical description. Cornell University, Ithaca, New York.
- Haché, M., Perreault, L. et Bobée, B., 1996. Projet C1 de la chaire en hydrologie statistique: Validation des apports historiques. Rapport de recherche, INRS-Eau (en préparation).
- Hipel, K.W. and McLeod, A.I., 1980. Perspectives in stochastic hydrology présenté dans *Time Series*. Anderson ed., North-Holland: 73-102.
- Hipel, K.W. and McLeod, A.I., 1994. *Time Series Modelling of Water Resources and Environmental System*. Elsevier, Amsterdam.
- Hurrell, J.W., 1995. Decadal trends in the North Atlantic oscillation: regional temperatures and precipitation. *Science*, 269: 676-679.
- Hurrell, J.W. and van Loon, H., 1995. Analysis of low-frequency climate variations over the north atlantic using historical atmospheric data. Sixth Symposium on Global Change Studies, Dallas, Texas: 15-20.
- Hurst, H.E., 1951. Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, 116: 770-808.

- International St-Laurence Board of Control, 1963. Regulation of lake Ontario: plan 1958-D. Report to the International Joint Commission.
- Jambu, M., 1978. Classification Automatique pour l'Analyse des Données, Tome I: Méthodes et Algorithmes. Dunod, Paris.
- Jenkins, G.M. and Watts, D.G., 1968. Spectral Analysis and its Applications. Holden-Day, San-Francisco.
- Karl, T.R., Esterling, D.R. and Groisman, P.Y., 1994. United States historical climatology network, National and regional estimates of monthly and annual precipitation, pp.830-995. In T.A. Boden, D.P. Kaiser, R.J. Sepanski, and F.W. Stoss (eds.), Trends '93: A Compendium of Data on Global Change, ORNL/CDIAC-65. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tenn. U.S.A.
- Klemes, V., 1974. The Hurst phenomena: a puzzle ? Water Resources Research, 10: 675-688.
- Lebart, L., Morineau, A. et Tabard, N., 1977. Techniques de la Description Statistique. Dunod, Paris.
- Lee, A.F.S. and Heghinian, S.M., 1977. A shift of the mean level in a sequence of independent normal random variables: a Bayesian approach. Technometrics, 19(4): 503-506.
- Lehmann, E.L., 1975. Nonparametrics: Statistical Methods Based of Ranks. Holden-Day, Oakland.
- Levene, H., 1960. Robust tests for the equality of variances, in Contributions to Probability and Statistics, ed. I. Olkin. Palo Alto, Stanford University Press: 278-292.
- Mathier, L., Fagherazzi, L., Rassam, J.C. et Bobée, B., 1992. Great Lakes net basin supply simulation by a stochastic approach. INRS-Eau, rapport de recherche no 362.
- McLeod, A.I. and Hipel, K.W., 1978. Preservation of the rescaled adjusted range, 1, A reassessment of the Hurst phenomenon, Water Resour. Res., 14(3): 491-508.
- Montgomery, D.C., 1984. Design and Analysis of Experiments. Wiley, New York.
- Moran, P.A.P., 1959. The Theory of Storage, Methuen, London.
- Neter, J., Wasserman, W. and Kutner, M.H., 1985. Applied Linear Statistical Models. Irwin, Homewood, Illinois.
- Nguyen, V.T.V., 1993. Analyse statistique des apports énergétiques mensuels et annuels. Rapport de recherche No WRME93/2, Université McGill.

- O'Connell, P.E., 1974. Stochastic modelling of long-term persistence in stream flow sequences. Thèse de doctorat, Département de Génie Civil, Imperial College, Londres.
- Perreault, L., Bobée, B., Bruneau, P. et Mathier, L. 1994a. Rapport Général du Logiciel *RéMuS: Théorie*. Rapport de recherche No R-422, INRS-Eau.
- Perreault, L., Bobée, B. et Legendre, P. 1994b. Rapport Général du Logiciel *AJUSTE-II: Théorie et Application*. Rapport de recherche No R-421, INRS-Eau.
- Perron H., Perreault L. et Bobée, B. 1996. Logiciel *RéMuS: Guide de l'Usager*. Rapport de recherche (en préparation).
- Quenouille, M.H., 1956. Notes in bias estimation. *Biometrika*, 43: 353-360.
- Rasmussen, P.F., Fagherazzi, L., Duquette, R., Legendre, P. et Perreault, L., 1996. Détermination de séries mensuelles types pour le modèle MINERVE. INRS-Eau, rapport de recherche en préparation.
- Rodionov, S.N. et Krovvin, A.S., 1992. The 1980s in the context of climatic changes in the North Atlantic region. *ICES Sciences Symposium*, 195: 93-102.
- Rodionov, S.N., 1994. Association between winter precipitation and water level fluctuation in the Great Lakes and atmospheric circulation patterns. *American Meteorological Society*, 7: 1693-1706.
- Roy, R., Mathier, L. et Bobée, B. (1995). Analyse probabiliste des conditions de faible hydraulité dans une perspective de production hydro-électrique.
- Salas, J.D., Delleur, J.W., Yevjevich, V. and Lane, W.L., 1980. *Applied Modeling of Hydrological Time Series*. Water Resources Publications, Chelsea, Michigan.
- Shapiro, S.S. and Wilk, M.B., 1965. An analysis of variance test for normality (complete samples). *Biometrika*, 52: 591-611.
- Srivastava, M.S. and Carter, E.M., 1983. *An Introduction to Applied Multivariate Statistics*. North-Holland, New York.
- Technik-Eaucan, 1995. Simulation du parc d'équipement d'Hydro-Québec avec les modèles Énergie III et SIMHYD (2.67). Apports naturels intermédiaires aux aménagements Vol. I et II, Rapports internes TEC6103-1 et TEC6103-2.
- Trenberth, K.E. and Hurrell, J.W., 1994. Decadal atmosphere-ocean variations in the Pacific. *Climate Dynamics*, 9: 303-319.
- UNESCO, 1981. *Methods of Computation of the Water Balance of Large Lakes and Reservoirs. Volume I - Methodology*. Edited by H.L. Ferguson and V.A. Znamensky.

Wallis, J.R. and O'Connell, P.E., 1972. Small sample estimation of ρ_1 . *Water Resources Research*, 8: 707-712.

Wilcoxon, F., 1945. Individual comparison by ranking methods. *Biometrics*, 1: 80-83.



ANNEXE A : DIAGRAMME EN BOÎTE

Le diagramme en boîte (boxplot) est une représentation schématique de la forme de la distribution des observations. Cette méthode graphique, au même titre qu'un histogramme, est très utile pour comparer visuellement les caractéristiques statistiques de plusieurs séries. Il existe plusieurs types de diagrammes en boîte. Ceux qui sont présentés dans le présent rapport sont construits de la manière suivante (cf. figure A.1):

- le point central représente la médiane des moyennes obtenues;
- les extrémités inférieure et supérieure de la boîte représentent respectivement le 25ème et 75ème quantiles des moyennes obtenues (valeurs correspondant à 25% et 75% de probabilité au non-dépassement). Ainsi, 50% des moyennes obtenues se situent à l'intérieur de cette boîte;
- les extrémités des tirets représentent les quantiles de 5% et 95% de probabilité au non-dépassement.

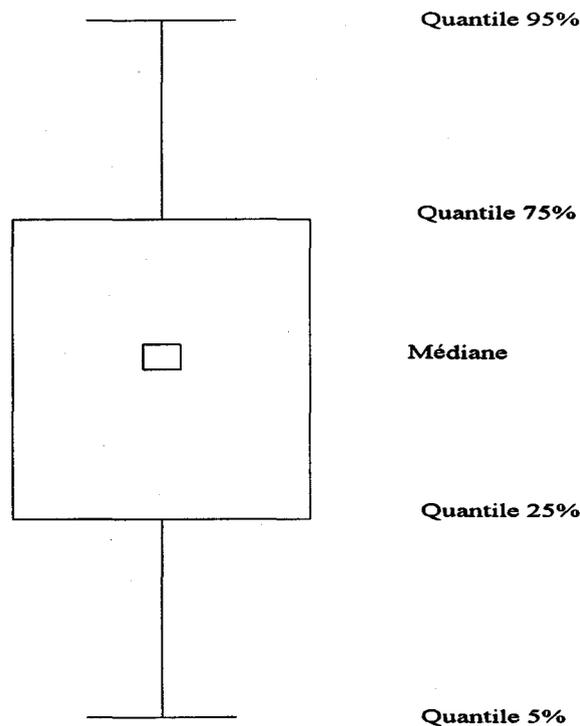


Figure A.1. Diagramme en boîte.



ANNEXE B : AUTOCORRÉLATION DU CUMUL

Soit le vecteur aléatoire $\mathbf{x} = (X_1, X_2, \dots, X_7)$ distribué selon une loi normale multidimensionnelle $N_7(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, où $\boldsymbol{\mu}$ désigne le vecteur des moyennes et $\boldsymbol{\Sigma}$ la matrice des variances et covariances. Considérons, de plus, une variable aléatoire Y représentée par un modèle autorégressif d'ordre 1 qui relie ses réalisations Y_t successives dans le temps:

$$Y_t = \mu_Y + \rho_Y(Y_{t-1} - \mu_Y) + \varepsilon_t \quad (\text{B.1})$$

où μ_Y désigne la moyenne de la série, ε_t un terme d'erreur aléatoire distribué selon une loi normale de moyenne nulle et de variance σ_ε^2 ($N(0, \sigma_\varepsilon^2)$), et ρ_Y l'autocorrélation d'ordre 1.

Le cumul des réalisations de ces huit variables aléatoires Z_t pour l'année t s'écrit donc comme suit:

$$\begin{aligned} Z_t &= \sum_{i=1}^7 X_{i,t} + Y_t \\ &= \sum_{i=1}^7 X_{i,t} + \mu_Y + \rho_Y(Y_{t-1} - \mu_Y) + \varepsilon_t \end{aligned} \quad (\text{B.2})$$

Or, si $\mathbf{x} \approx N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ alors, pour une matrice \mathbf{C} de dimension $(r \times p)$, $\mathbf{C}\mathbf{x} \approx N_r(\mathbf{C}\boldsymbol{\mu}, \mathbf{C}\boldsymbol{\Sigma}\mathbf{C}')$ (cf. Srivastava et Carter, 1983). En posant $\mathbf{C} = (1, 1, \dots, 1)$, on en déduit que la somme des 7 variables aléatoires X_1, X_2, \dots, X_7 est distribuée selon une loi normale unidimensionnelle de moyenne et de variance données respectivement par:

$$\mu_X = \sum_{i=1}^7 \mu_i \quad (\text{B.3})$$

$$\sigma_X^2 = \sum_{i=1}^7 \sigma_i^2 + 2 \sum_{i \neq j}^7 \sigma_i \sigma_j \rho_{ij} \quad (\text{B.4})$$

où $\mu_i, i = 1, \dots, 7$, et $\sigma_i^2, i = 1, \dots, 7$, sont respectivement les moyennes et les variances de X_1, X_2, \dots, X_7 et où ρ_{ij} désigne la corrélation entre les variables X_i et X_j .

Pour simplifier les calculs algébriques, certains termes de l'équation (B.2) sont regroupés pour obtenir l'expression équivalente:

$$Z_t - \mu_Z = \rho_Y(Y_{t-1} - \mu_Y) + \alpha_t \quad (\text{B.5})$$

où $\mu_z = \mu_x + \mu_y$ est la moyenne de Z_t et $\alpha_t \approx N(0, \sigma_\varepsilon^2 + \sigma_x^2)$.

Par définition, l'autocorrélation d'ordre 1 ρ_z de la série des Z_t est donnée par:

$$\rho_z = \frac{E\{(Z_t - \mu_z)(Z_{t-1} - \mu_z)\}}{E\{(Z_t - \mu_z)^2\}} \quad (\text{B.6})$$

En utilisant les équations (B.5) et (B.6), on a que:

$$\begin{aligned} E\{(Z_t - \mu_z)(Z_{t-1} - \mu_z)\} &= \rho_Y^2 E\{(Y_{t-1} - \mu_Y)(Y_{t-2} - \mu_Y)\} + \rho_Y E\{(Y_{t-1} - \mu_Y)\alpha_{t-1}\} + \\ &\quad \rho_Y E\{(Y_{t-2} - \mu_Y)\alpha_{t-1}\} + E\{\alpha_t \alpha_{t-1}\} \end{aligned} \quad (\text{B.7})$$

Or, en supposant que les variables aléatoires X_t sont non corrélées avec Y et que l'autocorrélation des X_t est nulle, on obtient:

$$E\{(Z_t - \mu_z)(Z_{t-1} - \mu_z)\} = \rho_Y^3 \sigma_Y^2 \quad (\text{B.8})$$

De la même manière, on déduit que le dénominateur de l'expression (B.6) est donné par:

$$\begin{aligned} E\{(Z_t - \mu_z)^2\} &= \rho_Y^2 E\{(Y_{t-1} - \mu_Y)^2\} + 2\rho_Y E\{(Y_{t-1} - \mu_Y)\alpha_{t-1}\} + E\{\alpha_t^2\} \\ &= \rho_Y^2 \sigma_Y^2 + (1 - \rho_Y^2) \sigma_Y^2 + \sigma_x^2 \end{aligned} \quad (\text{B.9})$$

car, pour le modèle AR(1) de l'équation (B.1) (Box *et al.*, 1994), $\sigma_\varepsilon^2 = (1 - \rho_Y^2) \sigma_Y^2$.

Enfin, après quelques manipulations algébriques, on obtient le résultat recherché:

$$\rho_z = \frac{\rho_Y^3 \sigma_Y^2}{\rho_Y^2 \sigma_Y^2 + (1 - \rho_Y^2) \sigma_Y^2 + \sigma_x^2} = \left(\frac{\sigma_Y^2}{\sigma_Y^2 + \sigma_x^2} \right) \rho_Y^3 \quad (\text{B.10})$$

■

ANNEXE C : COEFFICIENT DE HURST

C.1 Quelques définitions

Sommes partielles

Considérons une série chronologique, y_1, y_2, \dots, y_n . La séquence des sommes partielles ajustées est définie de la manière suivante:

$$S_k^* = \sum_{i=1}^k y_i - k\bar{y}_n \quad k = 1, 2, \dots, n$$

où $\bar{y}_n = \frac{1}{n} \sum_{i=1}^n y_i$ est la moyenne de la série. En particulier, on a $S_n^* = 0$.

Pour caractériser une série donnée, on peut considérer le maximum et le minimum des S_k^* :

$$M_n^* = \max(0, S_1^*, S_2^*, \dots, S_n^*)$$

$$m_n^* = \min(0, S_1^*, S_2^*, \dots, S_n^*)$$

Domaine de variation

Le domaine de variation des sommes partielles ajustées est donné par:

$$R_n^* = \max_{k \in [0, n]} S_k^* - \min_{k \in [0, n]} S_k^* = M_n^* - m_n^*$$

Cette caractéristique statistique peut être interprétée de façon physique. Par exemple, R_n^* peut représenter le stockage nécessaire d'un réservoir si l'on veut exploiter le volume total d'eau à un taux constant. La figure A.1 illustre les définitions de la somme partielle ajustée et de son domaine de variation.

Domaine de variation standardisé

Le domaine de variation standardisé est défini de la façon suivante:

$$\bar{R}_n^* = \frac{R_n^*}{\hat{\sigma}_n^2}$$

où $\hat{\sigma}_n^2 = n^{-1} \sum_{i=1}^n (X_i - \bar{X}_i)^2$ est l'estimation de la variance de la série. Les aspects théoriques de cette caractéristique statistique ont été largement étudiés en hydrologie (McLeod et Hipel, 1978; Salas et al., 1980).

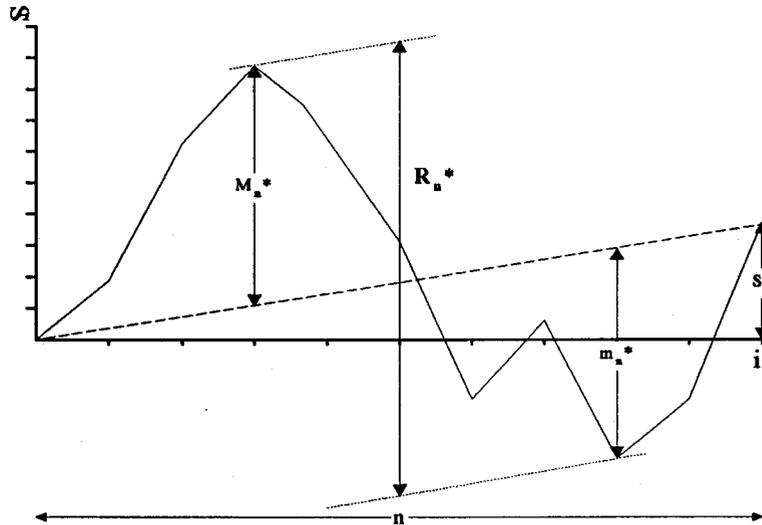


Figure C.1. Définitions des sommes partielles ajustées et du domaine de variation ajusté

C.2 Coefficient de Hurst

Suite à des études expérimentales intensives, Hurst (1951) a identifié une relation non-linéaire approximative entre \bar{R}_n^* et la taille d'échantillon n :

$$\bar{R}_n^* = \frac{R_n^*}{\sigma_n^2} = k n^H$$

où k et h sont des constantes. On peut montrer que pour $n=2$, $\bar{R}_n^* = 1$. Pour cette raison, Hurst a simplifié son expression de la manière suivante:

$$\bar{R}_n^* = (n/2)^H$$

où H est le coefficient de Hurst. Il a trouvé que pour une série d'éléments indépendants et pour n grand, l'espérance du coefficient H est égale à 0.5. Cependant, pour les séries réelles examinées par Hurst, la moyenne obtenue du coefficient H était de 0.73. Cette observation a été appelée le phénomène de Hurst. Feller (1951) a suggéré qu'un modèle markovien pourrait expliquer le phénomène, mais Moran (1959) a pu démontrer que même les processus autocorrélés ont asymptotiquement une valeur de H égale à 0.5. En raison du lien entre H et le domaine de variation des sommes ajustées et le besoin de stockage, on a traditionnellement accordé beaucoup d'importance au coefficient de Hurst en hydrologie. En pratique, on estime H comme suit:

$$H = \frac{\log \bar{R}_n^*}{\log(n/2)}$$

ANNEXE D : DISTRIBUTION DES SÉQUENCES

Les diagrammes en boîte des diverses caractéristiques statistiques de séquences obtenues par simulation sont présentées, pour les neuf modèles, dans ce qui suit (cf. section 4.3.3.2).

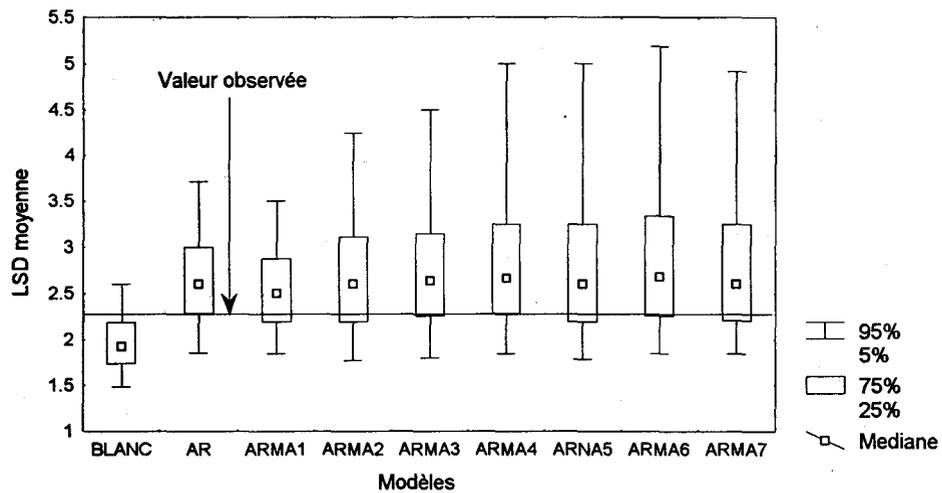


Figure D.1. Distribution des longueurs moyennes des séquences déficitaires.

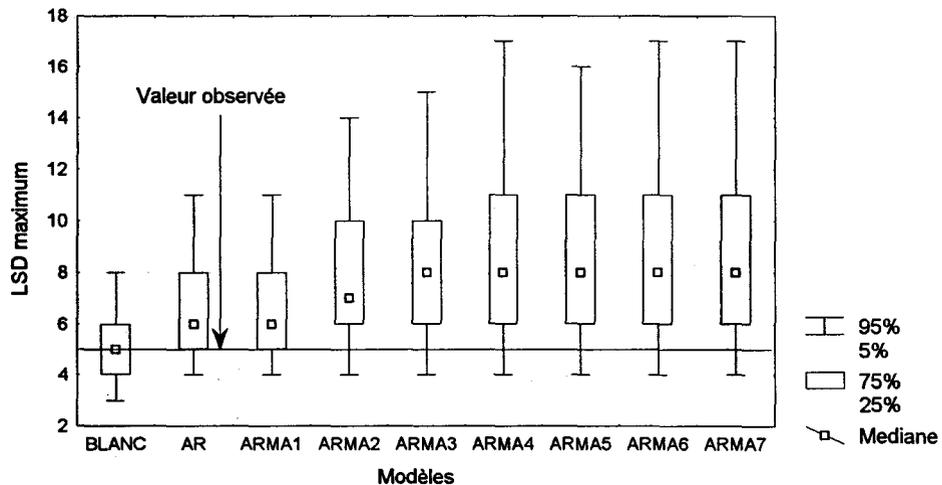


Figure D.2. Distribution des longueurs maximums des séquences déficitaires.

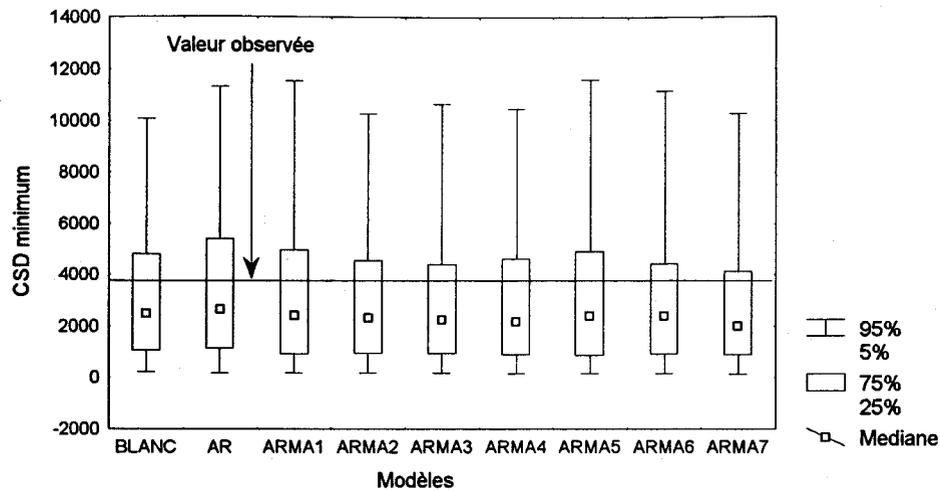


Figure D.3. Distribution des cumuls minimums des séquences déficitaires.

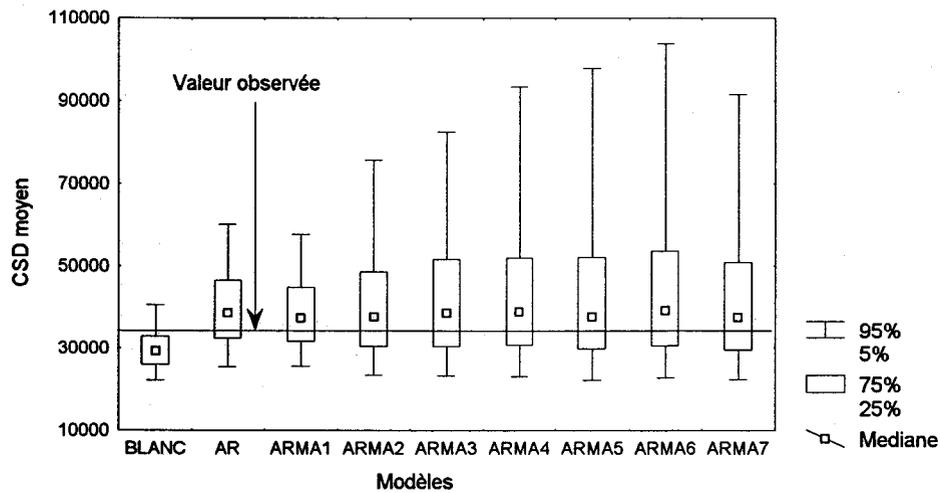


Figure D.4. Distribution des cumuls moyens des séquences déficitaires.

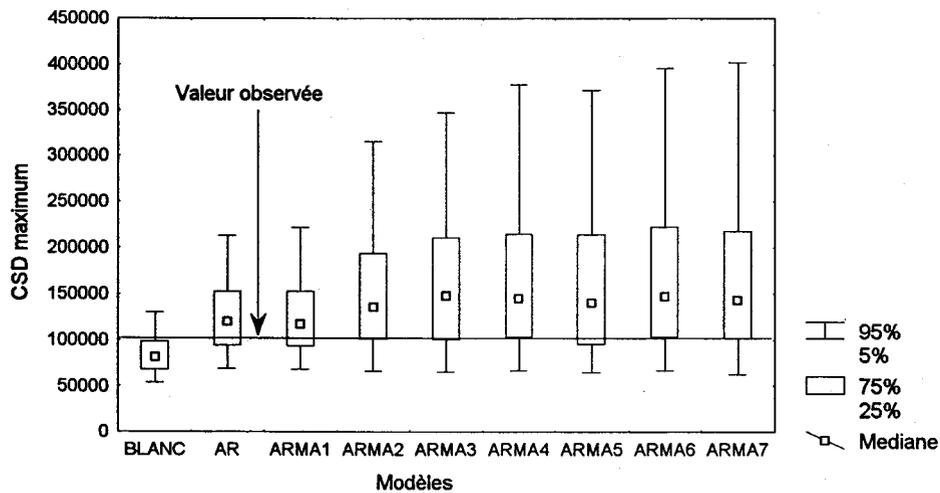


Figure D.5. Distribution des cumuls maximums des séquences déficitaires.

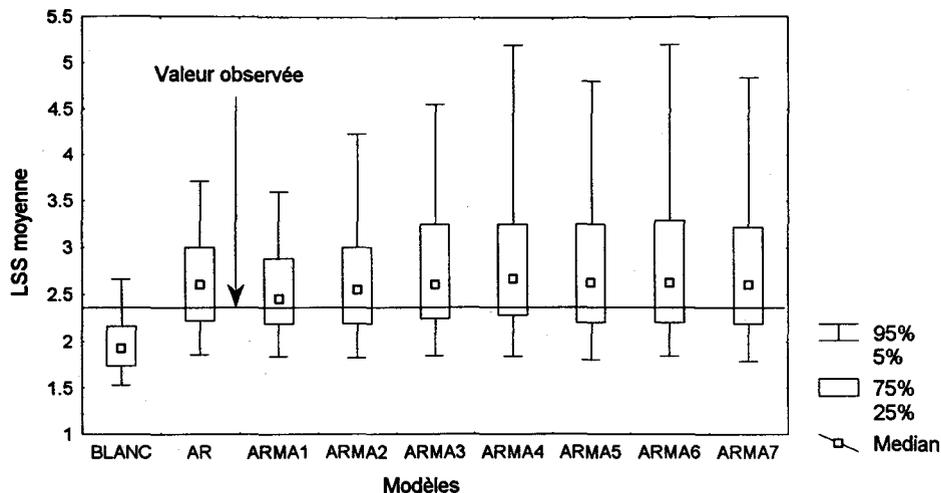


Figure D.6. Distribution des longueurs moyennes des séquences excédentaires.

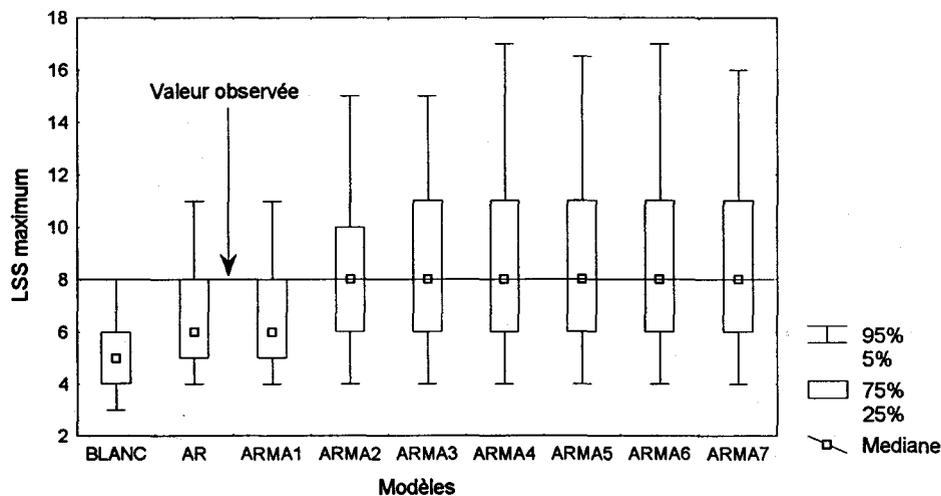


Figure D.7. Distribution des longueurs maximums des séquences excédentaires.

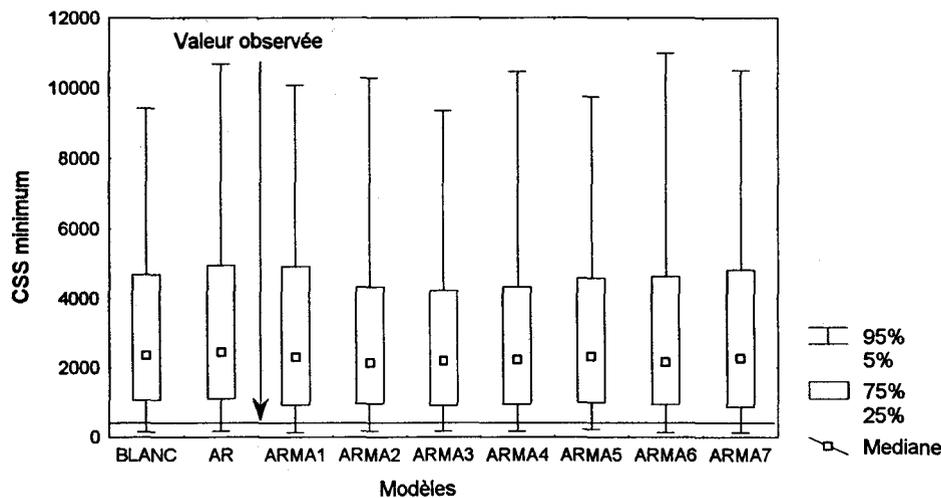


Figure D.8. Distribution des cumuls minimums des séquences excédentaires.

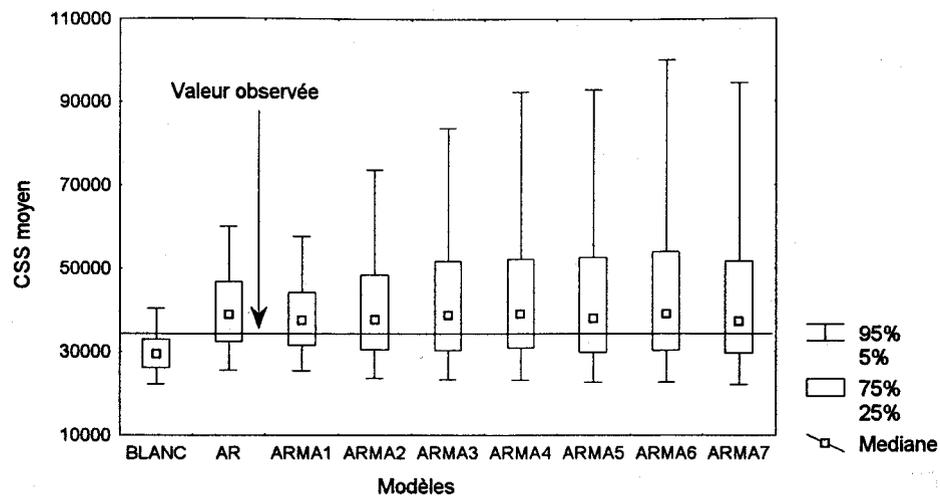


Figure D.9. Distribution des cumuls moyens des séquences excédentaires.

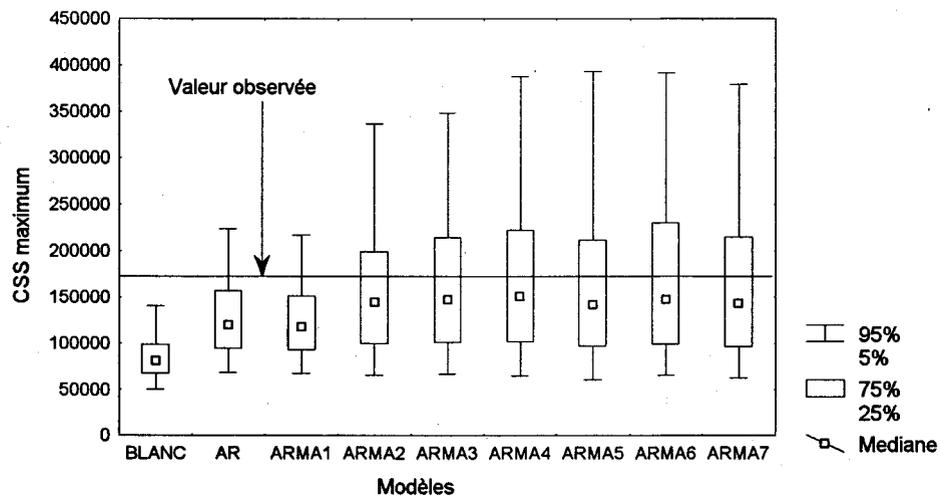


Figure D.10. Distribution des cumuls maximums des séquences excédentaires.

ANNEXE E : DISTRIBUTION DES CUMULS (AR)

Considérons l'écart énergétique cumulé prédit $S_t(k)$ d'origine t et d'horizon k défini comme suit:

$$S_t(k) = \sum_{l=1}^k \hat{z}_t(l) \quad (\text{E.1})$$

Chacune des prévisions $\hat{z}_t(l)$, $l = 1, \dots, k$, est distribuée selon une loi normale de moyenne μ_l et de variance σ_l^2 . Les k prévisions $\hat{z}_t(l)$ sont corrélées et donc conjointement distribuées selon une loi normale multidimensionnelle. On déduit des propriétés de cette loi de probabilité (cf. Srivastava et Carter, 1983, théorème 2.3.2), que la somme $S_t(k)$ de ces variables aléatoires est alors distribuée selon une loi normale unidimensionnelle de moyenne et de variance données respectivement par:

$$\mu_t(k) = \sum_{l=1}^k \mu_l \quad (\text{E.2})$$

$$\sigma_t^2(k) = \sum_{l=1}^k \sigma_l^2 + 2 \sum_{l=1}^{k-1} \sum_{j=l+1}^k \sigma_{l,j} \quad (\text{E.3})$$

où $\sigma_{l,l+j}$ désigne la covariance entre les prévisions $\hat{z}_t(l)$ et $\hat{z}_t(l+j)$.

Moyenne de $S_t(k)$ (modèle AR)

Sachant, pour un modèle autorégressif d'ordre 1 (AR), que $\mu_l = E\{\hat{z}_t(l)\} = \phi^l z_t$, on déduit des expressions (E.1) et (E.2) que:

$$\mu_t(k) = \sum_{l=1}^k \phi^l z_t = z_t \sum_{l=1}^k \phi^l = \frac{\phi(1-\phi^k)}{1-\phi} z_t$$

car, pour $0 < r < 1$, on a que:

$$\sum_{l=1}^k r^l = \frac{r(1-r^k)}{1-r} \quad (\text{E.4})$$



Variance de $S_t(k)$ (modèle AR)

La variance du cumul d'horizon k est composée de deux termes (éq. E.3). L'expression (4.19) nous permet de déduire le premier terme de $\sigma_t^2(k)$:

$$\sum_{i=1}^k \sigma_i^2 = \frac{\sigma_\varepsilon^2}{1-\phi^2} \sum_{i=1}^k (1-\phi^{2i}) \quad (\text{E.5})$$

En utilisant le résultat (E.4) pour évaluer la somme et après quelques manipulations algébriques nous pouvons écrire:

$$\begin{aligned} \sum_{i=1}^k \sigma_i^2 &= \frac{\sigma_\varepsilon^2}{(1-\phi^2)(1-\phi)} \left[k(1-\phi) - \frac{\phi^2(1-\phi)(1-\phi^{2k})}{1-\phi^2} \right] \\ &= \frac{\sigma_\varepsilon^2}{(1-\phi^2)^2(1-\phi)} \left[k(1-\phi^2)(1-\phi) - \phi^2(1-\phi)(1-\phi^{2k}) \right] \end{aligned} \quad (\text{E.6})$$

Le second terme fait intervenir la covariance $\sigma_{l,l+j}$ entre les prévisions $\hat{z}_t(l)$ et $\hat{z}_t(l+j)$. Or, cette covariance peut être déduite des équations (4.18) et (4.24):

$$\sigma_{l,l+j} = \sigma_\varepsilon^2 \phi^j \left(\frac{1-\phi^{2l}}{1-\phi^2} \right) \quad (\text{E.7})$$

On peut donc écrire:

$$2 \sum_{l=1}^{k-1} \sum_{j=1}^{k-l} \sigma_{l,l+j} = \frac{2\sigma_\varepsilon^2}{1-\phi^2} \sum_{l=1}^{k-1} (1-\phi^{2l}) \sum_{j=1}^{k-l} \phi^j \quad (\text{E.8})$$

Si on emploie l'expression (E.4) ainsi que le résultat suivant:

$$\sum_{l=1}^k q^{l-1} = \frac{q^k - 1}{q - 1}, \quad \forall q \neq 1 \quad (\text{E.9})$$

l'équation (E.8) peut s'exprimer de la manière suivante:

$$\frac{\sigma_\varepsilon^2}{(1-\phi^2)^2(1-\phi)} \left[2\phi(1-\phi^2)(k-1) + 2(\phi^{k+2} - \phi^2)(1-\phi^{k-1})(1+\phi) - 2\phi^3(1-\phi^{2(k-1)}) \right] \quad (\text{E.10})$$

Enfin, la variance peut être obtenue en sommant les deux termes, soit les expressions (E.6) et (E.10). Après quelques manipulations algébriques et une factorisation adéquate utilisant le fait que $(1 - \phi^2) = (1 - \phi)(1 + \phi)$, on obtient le résultat recherché:

$$\sigma_i^2(k) = \frac{\sigma_\varepsilon^2}{(1 + \phi)(1 - \phi)^3} \left[k(1 - \phi^2) - \phi(2 + \phi) + 2\phi^{k+1}(1 + \phi) - \phi^{2(k+1)} \right] \quad (\text{E.11})$$

■